Information Acquisition and Expected Returns: Evidence from EDGAR Search Traffic*

Frank Weikai Li[†] Chengzhu Sun[‡]

This Draft: December 2017

Abstract

This paper examines expected return information embedded in investors' information acquisition activity. Using a novel dataset containing investors' access of company filings through the SEC's EDGAR system, we reverse engineer investors' expectations of future payoffs and show that the abnormal number of IPs searching for firms' financial statements strongly predicts future returns. The return predictability stems from investors allocating more effort to firms with improving fundamentals and following exogenous shock to underpricing. A long-short portfolio based on our measure of information acquisition activity generates a monthly abnormal return of 80 basis points that is not reversed in the long-run. The return predictability is stronger for firms with larger and lengthier financial filings that are more costly to process. Collectively, these findings support theories of endogenous information acquisition that costly information acquisition reveals the value of information.

JEL classification: G12, G14

Keywords: Information Acquisition, EDGAR Search, Return Predictability, Market Efficiency

*We are grateful to Utpal Bhattacharya, Ekkehart Boehmer, Fangjian Fu, Claire Yurong Hong, Dashan Huang, Hongqi Liu, Roger Loh, Wenlan Qian, Rik Sen, Tao Shen, Melvyn Teo, Baolian Wang, Rong Wang, Chishen Wei, K.C. John Wei, Jialin Yu and Huai Zhang (Discussant) and seminar participants at the Singapore Scholars Symposium, Peking University, Tsinghua SEM and CUHK Shenzhen for helpful comments. All errors are our own.

[†]Singapore Management University, Lee Kong Chian School of Business (Email: wkli@smu.edu.sg)

[‡]Hong Kong University of Science and Technology (Email: csunab@connect.ust.hk)

1 Introduction

Information acquisition and dissemination is key to understanding asset price movements and market efficiency. When information is costly to acquire and price is only partially revealing, economic agents will expend resources and effort to become informed (Grossman and Stiglitz (1980); Verrecchia (1982)), and in doing so, will move prices closer to the fundamental value. A central prediction from theories of costly information acquisition is that more investors will choose to become informed when they perceive greater benefits from doing so, holding the cost of information acquisition constant. Although theories offer clear and rich predictions, empirical evidence of the relation between information acquisition behavior and the value of information is sparse in financial markets, potentially due to the difficulty of directly measuring the information acquisition activities of investors.

In this paper, we take advantage of a novel dataset containing investors' access of regulatory filings through the Security and Exchange Commission (SEC)'s EDGAR (Electronic Data Gathering, Analysis, and Retrieval) system to study the implications of information acquisition activities on firm value. Because the EDGAR system is the main source of firms' regulatory filings, and the SEC maintains a log file of all activities performed by users on EDGAR, we are able to directly observe investors' information acquisition activity for a broad cross-section of firms over a sample period of more than 10 years.

Our research objectives in this paper are twofold. First, we examine the determinants of investors' information acquisition through the EDGAR website. Motivated by theories of information acquisition[1], we posit that information acquisition activities should be negatively related to the cost of gathering information and positively related to the value of information. To test this, we use the number of unique IP addresses searching for SEC filings through EDGAR as a proxy for investors' information acquisition. We then run cross-sectional regressions of our information acquisition proxy on several firm characteristics associated with the cost of information acquisition. Specifically, we hypothesize that firms with higher investor visibility and attention, and a better information environment,

[1]There is a large body of theoretical literature on information acquisition, e.g., Grossman and Stiglitz (1980), Diamond and Verrecchia (1981), Verrecchia (1982), Hellwig (1980), Admati (1985), and Mele and Sangiorgi (2015).

will attract more information acquisition, as these stocks are more accessible in investors' minds and less costly to analyze. We also expect investors to have stronger incentives to acquire information about firms with higher valuation uncertainty. Using firm size as a proxy for investor visibility, trading volume as a proxy for investor attention (Gervais, Kaniel, and Mingelgrin (2001); Barber and Odean (2007)), analyst coverage as a proxy for information environment (Hong, Lim, and Stein (2000)), and idiosyncratic volatility as a proxy for valuation uncertainty (Zhang (2006)), we find evidence consistent with the theories. These four firm characteristics explain 55% of cross-sectional variation of information acquisition across firms. Further tests show that information acquisition through EDGAR also increases following negative return performance and among firms with lower institutional ownership, but these additional characteristics do not significantly improve the explanatory power of our baseline model.

After implementing a simple characteristic-based model of expected information acquisition, we proceed to examine our second research question, that an abnormal level of information acquisition reflects the expected benefits of trading on information. This prediction is based on the simple premise that when resource-constrained investors decide how to allocate their time and effort, they will have a strong preference for firms with the largest price appreciation or depreciation potential. In reality, due to short-sale constraints, investors will more likely engage in costly information acquisition when the expected return of a stock is positive.[2]

To test this hypothesis, we extract the number of IPs unexplained by firm characteristics to reverse engineer investors' expectations of future payoffs. Consistent with the idea that information acquisition embeds the value of information, we show that an abnormal number of IPs (denoted as AIP) requesting EDGAR filings strongly predicts subsequent stock returns. An equal-weighted, monthly rebalanced, long-short strategy that buys stocks in the highest decile of AIP and sells stocks in the lowest decile of AIP generates 52 to 82 basis points per month after adjustment for the Carhart (1997) four factors and is highly significant. Adjusting for the recently proposed factor models – the Fama and French (2016) five-factor

[2]Since the EDGAR log file contains millions of unique IPs, most EDGAR users must be retail investors who face even higher short-sale constraints than institutional investors.

model, the Hou, Xue, and Zhang (2015) q-factor model, and the Stambaugh and Yuan (2016) mispricing factor model – does not affect the return spread of the long/short portfolio much. The abnormal return of AIP strategy is much weaker for value-weighted portfolios. The high-minus-low AIP strategy generates approximately 30 basis points per month, which is mostly insignificant. This is expected given that short-sale constraints are less binding for big stocks, so the direction of the information contained in AIP is more ambiguous for big stocks. Using several proxies of short-sale constraints, we confirm that the positive expected return information embedded in information acquisition is more pronounced for stocks that are more difficult to short ex-ante.

The return predictability associated with the abnormal number of IPs persists for two quarters, and is not reversed in the subsequent months. This persistence in return predictability alleviates concerns that our findings is the result of temporary price pressure caused by noise traders, which is reversed over the long-run (Da, Engelberg, and Gao (2011)).

With a Fama-MacBeth regression setting, we confirm that AIP has additional explanatory power for future stock returns when we control for the standard cross-sectional return predictors, such as firm size, book-to-market ratio, momentum, short-term reversal, idiosyncratic volatility, turnover, and institutional ownership. The return predictability of AIP is also **not** affected by alternative explanations such as post-earnings announcement drift, earnings announcement premium, and investor disagreement. Looking into different types of EDGAR filings, we find that the return predictability of AIP comes mainly from those searching for firms' annual accounting reports 10-Ks (AIP_10K). As gathering and analyzing 10-Ks is more costly than other SEC filings and those searching activities are more indicative of deliberate information acquisition, the stronger predictability of AIP for 10-Ks is consistent with theories of costly information acquisition. To further substantiate our argument, we use the file size and word count of 10-Ks as proxies for the complexity of financial disclosure (Loughran and McDonald (2014)), and find that the return predictability of AIP is indeed stronger among firms with larger and lengthier 10-Ks.

Having established the robustness of the return predictability of the abnormal number of IPs, we test the sources of return predictability. The underlying assumption in this paper is that under short-sale constraints, investors rationally allocate more effort and resources to underpriced stocks with high expected returns. As mispricing implies the separation of stock price from the fundamental value of a firm, we conjecture two non-mutually exclusive channels through which investors can identify mispricing. The first channel is investors' information acquisition activity revealing their favorable expectation of the fundamental performance of firms that are yet to be priced in the market. Consistent with the first channel, we find that AIP strongly predicts the future changes in firms' quarterly Return-on-Assets and revisions in analyst consensus forecast, even after controlling for past profitability and other determinants of firms' fundamental performances. The second channel of investors identifying mispricing is that investors could observe changes in stock prices due to exogeneous reasons. Supporting the second channel, we show that the abnormal number of IPs searching for EDGAR filings increases significantly for firms experiencing mutual fund outflow-induced selling pressure. Taken together, our evidence suggests that investors expend greater resources and effort on undervalued stocks and these findings are much more difficult to reconcile with alternative explanations such as omitted risk factor or changes in investor visibility (Merton (1987))[3].

Finally, we provide some suggestive evidence of the types of investors conducting informed searches of firms' fundamentals through EDGAR. We show that the abnormal number of IPs positively predicts net purchases by hedge funds in the following quarter. In contrast, AIP does not have predictability for net purchases by mutual fund managers. These results are consistent with the idea that investors searching for financial filings through EDGAR are more sophisticated than those searching through the Google search engine, and the information contained in EDGAR searches overlaps with the information of most sophisticated institutional investors like hedge funds.[4]

This paper contributes to several strands of the existing literature. First, our results offer strong empirical evidence supporting theories of costly information acquisition that costly

[3]Alternative explanations based on omitted risk factor or changes in investor base all work through the discount-rate channel, while the return predictability of AIP operates (partially) through the cashflow channel.

[4]Drake, Quinn, and Thornock (2017) report that EDGAR users tend to have higher education levels and are more likely to work in major cities with more accounting and finance jobs.

information acquisition is positively related to the expected benefits of information. Using the novel EDGAR log file dataset, we construct a direct measure of investors' information acquisition activity, and show its strong predictability for firms' future returns and fundamentals. Du (2015) shows that the number of web visits to SEC filings of insider trades predicts post-filing stock return in the short-run. Although similar in spirit, our paper differs from his paper as we study a much broader sample of SEC regulatory filings and longer horizon returns. We also test the channels underlying the return predictability results. Using EDGAR search data, Chen, Cohen, Gurun, Lou, and Malloy (2017) find that mutual funds tend to track a particular set of firms and insiders, and that their tracked trades generate abnormal performance. Lee and So (2017) study the information content of analysts' coverage decisions and show that an abnormal amount of analyst coverage positively predicts future firm performance. By extracting the information acquisition activities of all internet users through the EDGAR site, our measure captures the expected return information embedded in the collective behavior of a much larger set of market participants, i.e., millions of unique users. In addition, analysts' incentives have been found to be distorted by generating trading commissions for their brokerage houses or currying favor with firm managers (Ke and Yu (2006)); such distortions are less likely among EDGAR users. Empirically, we show that the return predictability of AIP is not affected after controlling for analyst coverage proxies.

This paper also contributes to the growing literature on the effect of investor attention and information acquisition on asset prices and capital market efficiency. Da, Engelberg, and Gao (2011) show that the attention of retail investors, as captured by Google search volume, causes transitory price pressures on attention-grabbing stocks. Using news-searching activity via the Bloomberg terminal as a proxy for institutional investors' attention, Ben-Rephael, Da, and Israelsen (2017) find that institutional attention facilitates the timely incorporation of fundamental information into asset prices. More pertinent to this study, Drake, Roulstone, and Thornock (2015) show that EDGAR-based information acquisition affects the pricing of earnings-related news. However, the aforementioned papers mainly examine the effect of information acquisition on the pricing of *publicly* announced news, while this paper directly infers investors' *private* expectations of firm value through their collective actions.

Finally, our work contributes to the emerging literature on extracting intelligence latent in the collective "wisdom of crowds". Chen, De, Hu, and Hwang (2014) find evidence that investors' social media posts help predict stock return. Lee, Ma, and Wang (2015) show that investors' co-search patterns via the EDGAR website could help identify peer firms better than traditional industry benchmarks. Huang (2016) finds that consumer opinions of firms' products on Amazon.com contain value-relevant information about firm fundamentals and stock prices. Similarly, Green, Huang, Wen, and Zhou (2017) document that employer reviews on Glassdoor reveal valuable information about employers' fundamentals. This paper complements the above studies as we infer agents' expectations not from what they "say", but from what they actually "do".

Our finding that information acquisition activity predicts future returns does not necessarily imply that the market is inefficient. As pointed out by Grossman and Stiglitz (1980), a fully efficient market where prices instantaneously reflect all available information cannot sustain an equilibrium when information is costly to acquire and analyze. Rather, our evidence is mostly consistent with the idea of "efficiently inefficient markets" (Pedersen (2015)), where competition among investors makes the market almost efficient, but the market also remains inefficient enough that these investors are compensated for their costs of acquiring and analyzing information.

The remainder of this paper is organized as follows. Section 2 describes the data, presents summary statistics, and examines the determinants of information acquisition. Section 3 shows that the abnormal level of information acquisition reveals investors' expectations of future returns. Section 4 tests the channels underlying the return predictability results. Finally, Section 5 concludes.

2 Data and Methodology

2.1 Sample Construction

Our IP search volume data comes from the Security and Exchange Commission's (SEC) EDGAR log file database, which has recorded all website search traffic for SEC filings since

2003.[5] Each search record contains information about the user's unique Internet Protocol (IP) address (anonymized)[6], timestamp, searched company (identified by the Central Index Key (CIK)) and searched specific filing (identified by the unique SEC accession number).[7] Following Lee, Ma, and Wang (2015) and Ryans (2017), we first filter the raw log data to eliminate the requests made by robots or automated webcrawlers, since such numerous and indiscriminate requests are uninformativeness for our research question.[8] Next, we match the CIK in the EDGAR log filings to that in COMPUSTAT to identify public companies, and retrieve the filing type and filing date for each requested file by linking the accession number to the Master Index files maintained by the SEC.[9] We classify these filings into six groups: 10-K, 10-Q, 8-K, insider, registration, and proxy.[10] Finally, we calculate the monthly IP search volume for each filing category at firm level by counting the total number of unique IP addresses that searched one category of SEC filings of a specific company within a one-month window. We define IP_total as the total number of unique IP addresses searching all six types of EDGAR filings. Drake, Roulstone, and Thornock (2015) report that periodic accounting reports are the type of SEC filings most frequently requested by

[5]The data is available for download at https://www.sec.gov/data/edgar-log-file-data-set.html.

[6]The EDGAR log file dataset provides the first three octets of the IP address with the fourth octet obfuscated with a three character string that preserves the uniqueness of the last octet without revealing the full identity of the IP.

[7]The detailed log file record elements are described at https://www.sec.gov/files /EDGAR_variables_FINAL.pdf

[8]First, following Lee, Ma, and Wang (2015), we exclude the searching records of those users who download more than 50 unique firms' filings in one day. The user is identified by their unique IP address. Second, following Ryans (2017) and Drake, Roulstone, and Thornock (2015), we remove log records that reference an index (idx=1), as index pages only provide the links to filings rather than the filings themselves. Third, following Ryans (2017), we keep the request records with successful document delivery (code=200). We then further exclude the search records of users who make more than 25 filing requests per minute or more than 500 requests per day, or with more than three unique CIKs searching per minute. Finally, we only keep one search record for a specific filing (unique accession number) to each user in a given day. This step is to avoid duplicated records due to users viewing the same document multiple times, a particular concern after the adoption of XBRL filing in 2009. For users who view the financial reports of XBRL-adopted firms in interactive data format, every click on a different footnote will generate a new search record, although it references the same document.

[9]Further details of the EDGAR index files can be found at https://www.sec.gov/edgar/searchedgar/accessing-edgar-data.htm

[10]We define the 10-K category as the filing type in "10-K", "10-K/A", "10-K405", "10-K405/A", "10-KSB", "10KSB", "10-KSB-A", "10KSB/A", "10-KT", "NT 10-K", and "10-KSB40"; the 10-Q category as the filing type in "10-Q", "10-Q/A", "10QSB", "10-QSB", "10QSB-A", and "NT 10-Q"; the 8-K category as the filing type in "8-K" and "8-K/A"; the insider category as the filing type in "SC 13G", "SC-13D", "SC 13G/A", "SC 13D/A", "3", "4", and "5"; the registration category as the filing type in "S-1", "S-1/A", "S-3", "S-3/A", "S-3ASR", "424B5", "424B4", "424B3", "424B2", and "FWP"; and the proxy category as the filing type in "DEF 14A", "DEF 14C", "DEFA14A", "DEFM14A", "DEFR14A", and "DEFM14C".

investors through the EDGAR website. We therefore also compute two additional measures of information acquisition specifically targeting firms' periodic accounting reports. IP_funtl (IP_10K) is the total number of unique IP addresses searching 10-K, 10-Q, and 8-K files (10-K files). Our sample runs from January 2003 to December 2014.[11]

It is important to note that there are other ways for investors to access financial filings, such as a firm's investor relations website and Yahoo! Finance. Data vendors such as Bloomberg and FactSet also provide investors with access to these financial statements. As a result, our analysis of the EDGAR server log cannot capture all the views/downloads that the entire universe of investors are performing on company filings. However, the EDGAR server still possesses several advantages over other information sources. First, it is questionable that investors primarily use the company website to retrieve SEC filings. As an example, Monga and Chasan (2015) quote General Electric (GE) CFO Jeffrey Bornstein, who noted that GE's 2013 annual report was downloaded from their investor relations website just 800 times.[12] However, for the same annual report, the EDGAR logs record 21,987 (4,325) downloads in the year (two months) following its filing. Second, some firms, such as Google (Alphabet, Inc) and ExxonMobil, refer investors directly to the EDGAR website to obtain their SEC filings. For such cases in which the investor relations department links the investors to the EDGAR site, these views/downloads will be captured in the SEC server. Third, other sources of company information often condense income-statement and balance-sheet information into pre-specified bins. As a result, some critical components of firms' financial information may be misrepresented. Finally, investors could better assess a firm's future prospects by reading the qualitative information contained in 10-K filings, which is not available in these data consolidators (Loughran and McDonald (2011)).

We obtain monthly stock returns from the Center for Research in Security Prices (CRSP), and annual accounting data from Compustat. Our sample of stocks starts with all common stocks traded on the NYSE, Amex, and NASDAQ. We adjust the stock returns by delisting. If a delisting return is missing and the delisting is performance-related, we set the delisting

[11]There are significant gaps in the data prior to March 2003 and between September 2005 and May 2006, due to lost or corrupt log file. As a result, we exclude these months from our sample in our analysis.

[12]https://www.wsj.com/articles/the-109-894-word-annual-report-1433203762.

return at -30% (Shumway (1997)).

We use standard control variables in our empirical analysis. *Size* (LnME) is defined as the natural logarithm of market capitalization at the end of June in each year. *Book-to-market ratio* (LnBM) is the most recent fiscal year-end report of book value divided by the market capitalization at the end of calendar year t-1. Book value equals the value of common stockholders' equity, plus deferred taxes and investment tax credits, and minus the book value of preferred stock. *Momentum* (Mom) is defined as the cumulative holding-period return from month t-12 and t-2. We follow the literature by skipping the most recent month's return when constructing the *Momentum* variable. The *short term reversal measure* (REV) is the prior month's return. *Turnover12* is the monthly trading volume over shares outstanding, averaged from the past 12 months. Since the dealer nature of the NASDAQ market makes its turnover difficult to compare with the turnover observed on NYSE and AMEX, we follow Gao and Ritter (2010) by adjusting the trading volume for NASDAQ stocks.[13] *Institutional ownership* (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by total shares outstanding. *Idiosyncratic volatility* (IVOL) is the standard deviation of the residuals from the regression of daily stock excess returns on the Fama and French (1993) three-factor returns within a month (Ang, Hodrick, Xing, and Zhang (2006)). Institutional ownership data of stocks are available from the Thomson Reuters (formerly CDA/Spectrum) Institutional Holdings database (13F). Coverage is the log one plus the number of analysts following a firm from I/B/E/S. We download the file size and number of words of the 10-Ks for all publicly-traded firms from WRDS SEC Analytics.

We also construct measures for trading activities by hedge funds. Using the list of hedge funds provided by Jiang (2014), we retrieve their quarterly holdings from the Thomson Reuters CDA/Spectrum Institutional (13F) database. We define net purchases of stock i by hedge funds in quarter q as follows:

$$NetBuy_{i,q} = \frac{Shrown_{i,q}}{Shrout_{i,q}} - \frac{Shrown_{i,q-1}}{Shrout_{i,q-1}} \tag{1}$$

[13]Specifically, we divide NASDAQ volume by 2.0, 1.8, 1.6, and 1.0 for the periods before February 2001, between February 2001 and December 2001, between January 2002 and December 2003, and after January 2004, respectively.

where $Shrown_{i,q}$ is the number of shares of firm i held by hedge funds in quarter q, and $Shrout_{i,q}$ is the number of firm i's shares outstanding in quarter q. To provide a basis for comparison, we construct a similar measure for mutual fund investors.

2.2 Summary Statistics

Panel A of Table 1 displays the time-series average of the cross-sectional means and standard deviations of the variables for the full sample. The average number of unique IPs searching for all six types of EDGAR filings of a firm is 155 in a month. The cross-sectional standard deviation is 317, indicating a large cross-sectional variation among firms. Consistent with Drake, Roulstone, and Thornock (2015), the annual report 10-Ks is the most frequently searched type of SEC filings, with an average of 60 IPs requesting it in a month. IPs searching for 10-Q and 8-K files are relatively less frequent. The average institutional ownership in our sample is 55%, reflecting the rapid growth of assets managed by institutional investors during our sample period. The remaining summary statistics are well known and do not require additional discussion.

Panel B reports the pairwise rank correlation among our variables. As we can see, the three IP variables are highly correlated. This is expected as periodic accounting reports consist of the largest fraction of EDGAR search requests. The number of IPs is also highly correlated with firm size, analyst coverage, and turnover, suggesting that firms with high investor visibility and attention have more EDGAR users. The number of IPs is negatively correlated with stock idiosyncratic volatility. However, this is mainly due to the size effect: small firms with high return volatility attract less EDGAR searching. As will be explained later, once we control for firm size, the number of IPs becomes positively correlated with idiosyncratic volatility, potentially because the incentives of acquiring information are greater when firm valuation is more uncertain (Grossman and Stiglitz (1980)).

2.3 Cross-sectional Determinants of Number of IPs

Theories of endogenous information acquisition suggest that information acquisition activity is a function of both the cost of acquiring information and the benefits of trading on acquired information. In order to isolate investors' expected payoff from information acquisition, we need a model of expected information acquisition activities. To this end, we develop and implement a simple characteristics-based model of expected information acquisition, and identify the discrepancies between the realized and expected level of information acquisition. Calculating these discrepancies requires proxies for information acquisition and firm characteristics useful in estimating the expected level of information acquisition activities.

Our proxy for information acquisition activity is the number of unique IP addresses searching for EDGAR filings for each firm in a given month. To mitigate data mining concerns, we use three measures capturing information acquisition activities for different types of EDGAR filings. IP_total is the total number of unique IPs searching for all types of EDGAR filings, and IP_funtl (IP_10K) is the total number of unique IPs searching for 10-K, 10-Q and 8-K files (10-K files). Our choice of firm characteristics is guided by information acquisition theories. Specifically, we hypothesize that firms with higher visibility and investor attention would attract more information acquisition, as these firms are more accessible in investors' minds. We also conjecture that the strength of firms' information environments would affect information acquisition, although the direction of the effect is not clear. On one hand, firms with abundant public information will be less costly to analyze, so we expect information acquisition to increase with the quality of a firm's information environment. On the other hand, a better information environment also means that the stock is less likely to be mispriced ex-ante, so investors' incentives to acquire private information will be reduced. Finally, we expect investors to have stronger incentives to acquire information about firms with higher valuation uncertainty. Following prior literature, we use firm size as a proxy for investor visibility, trading volume as a proxy for investor attention (Gervais, Kaniel, and Mingelgrin (2001); Barber and Odean (2007)), analyst coverage as a proxy for information environment[14] (Hong, Lim, and Stein (2000)), and idiosyncratic volatility as a proxy for valuation uncertainty (Zhang (2006)).

[14]Another motivation for including analyst coverage is that according to Lee and So (2017), analyst coverage contains information about future stock return. By including analyst coverage as a regressor, any expected return information embedded in the number of IPs will be incremental to that contained in analyst coverage proxies.

We calculate the abnormal number of IPs by fitting monthly cross-sectional regressions of the raw number of IPs to isolate the components of the number of IPs not attributable to firms' size, turnover, analyst coverage, and idiosyncratic volatility. To mitigate the effect of outliers, we use the log of one plus the number of IPs when estimating the abnormal number of IPs for firms. Specifically, we calculate the abnormal number of IPs for firm i in month t by estimating the following regression:

$$Log(1 + IP_{i,t}) = \beta_0 + \beta_1 LnME_{i,t} + \beta_2 Coverage_{i,t} + \beta_3 Turnover12_{i,t} + \beta_4 IVOL_{i,t} + \epsilon_{i,t} \quad (2)$$

where LnME is the log of market capitalization, Coverage is the log of one plus analyst coverage, Turnover12 is the monthly turnover averaged over the past 12 months, and IVOL is the daily idiosyncratic volatility calculated following Ang, Hodrick, Xing, and Zhang (2006). We define the abnormal number of IPs for each firm-month as the regression residuals from equation (2). We use the notation AIP to refer to the abnormal number of IPs, where higher values correspond to firms that have greater number of IPs searching for their EDGAR filings given their size, trading volume, analyst coverage, and volatility.

Table 2 reports the time-series average coefficients from estimating equation (2). The three panels correspond to three different measures of IPs as dependent variables. To see the improvement of R^2, we add the explanatory variables one by one from Column (1) to Column (7). Consistent with our hypothesis, information acquisition activities increase with firm size (t-stat=69.44), as larger firms are more visible to investors. Size alone explains 40% of the cross-sectional variation of the number of IPs. Columns (2) and (3) show that information acquisition increases with the strength of firms' information environments and investor attention, proxied by analyst coverage and turnover, respectively. Column (4) further shows that the number of IPs increases with return volatility after controlling for firm size. This finding suggests that investors' demand for information is larger for firms with more uncertain value. Column (4) also shows that these four firm characteristics explain 55% of the cross-sectional variation of the number of IPs on average. The results are similar in Panels B and C, where the dependent variables are IP_fundl and IP_10K, respectively.

The four firm characteristics used in equation (2) were selected based on theories and parsimony, and may therefore omit other firm characteristics that drive variation in the expected level of information acquisition activity. For example, investors may be attracted to firms with extreme past performance and glamour characteristics (Barber and Odean (2007)). To examine the explanatory power of other firm characteristics, we add the stock's past 12-month return, book-to-market ratio, and institutional ownership iteratively from Column (5) to Column (7). The results suggest that more investors search for EDGAR filings when the firm has performed poorly in the past year, and behaves like value stocks. However, adding these additional three characteristics improves the average R^2 of equation (1) by only 0.5 percentage points, suggesting the limited incremental explanatory power of past return performance, book-to-market ratio and institutional ownership. In the robustness test below, we show that the inclusion of other firm characteristics does not significantly affect the return predictability of AIP.

As there might be a nonlinear relationship between the abnormal number of IPs and firm characteristics, we further look at average stock characteristics across decile portfolios sorted by abnormal number of IPs searching for 10-K files (AIP_10K). Higher (lower) deciles correspond to firms with abnormally high (low) number of IPs. Panel C of Table 1 reports the time-series average of the cross-sectional mean values of each variable for each decile. First, the observation counts show that there are about 330 firms in each decile, suggesting that our measure of information acquisition is available for a broad cross-sectional sample of 3300 firms per month. Second, the table shows that AIP is positively correlated with the raw number of IPs searching for EDGAR filings. Third, AIP is, by construction, uncorrelated with firm size, analyst coverage, turnover, and volatility, although middle portfolios are slightly larger in terms of size and turnover. Finally, the panel shows that firms in the extreme deciles have lower institutional ownership and are more likely to be value stocks.

3 Information Acquisition and Future Stock Returns

When investors expend effort and time to acquire firms' fundamental information, they must perceive some benefits of utilizing such information. Hence a key hypothesis in this paper is that costly information acquisition activities reveal investors' perceptions of expected payoffs. Although in theory, the direction of the information content could be either positive or negative, in reality we expect firms with larger abnormal numbers of IPs searching for their EDGAR filings to have better future performance due to short-sale constraints. In addition, the positive predictive power of AIP should be stronger for small firms with severe short-selling constraints. In this section, we test the relation between the abnormal number of IPs and future returns using both portfolio sorts and the Fama-MacBeth regression.

3.1 Portfolio Sorts

In this section, we show that stocks sorted based on their abnormal numbers of IPs generate significant return spreads. We conduct the decile portfolio sorts as follows. At the end of each month, we sort stocks into deciles by their AIP. We then compute the average return of each decile portfolio over the next month, which provide a time series of monthly returns for each decile. We use these time series to compute the average excess return of each decile over the entire sample. As we are most interested in the return spread between the two extreme portfolios, we also report the return to a long–short portfolio (i.e., a zero-investment portfolio that longs the stocks in the highest AIP decile and shorts the stocks in the lowest decile). Our sample is from January 2003 to December 2014.

Table 3 reports the average monthly excess return of each decile portfolio. Panel A reports the equal-weighted portfolio return, and Panel B reports the value-weighted return. The three columns in each panel correspond to sorting based on the abnormal number of IPs searching for three different types of EDGAR filings. Panel A shows a strong positive relation between AIP and future returns, regardless of which IP variables are used. For sorts based on AIP_total, firms in the highest decile of AIP outperform the firms in the lowest decile by 71 basis points per month on an equal-weighted basis (t-stat=3.18). The results are stronger when we do the portfolio sorts based on AIP_funtl and AIP_10K. Specifically, the high-minus-low monthly return spread is 100 basis points (t-stat=4.70) based on AIP_10K, which corresponds to an annualized return of 12%. The evidence shows that aggregate information acquisition activities across EDGAR users reveal an economically large source of predictable return across firms. The economic magnitude is quite impressive given the fact that many other well-documented asset pricing anomalies are no longer profitable in our sample period (Chordia, Subrahmanyam, and Tong (2014); Green, Hand, and Zhang (2017)).

The larger return spread based on IPs searching for 10K compared with IPs searching for other types of EDGAR filings is consistent with information acquisition theories. A firm's annual report is among the lengthiest and most difficult-to-read SEC filings. Annual reports contain detailed annual operating and financial performance and metrics, suggesting that digesting these reports requires a large amount of effort and attention from investors. Compared with 10-Ks, 10-Q and 8-K files are usually shorter and easier to digest, and investors driven to these types of filings are more likely to respond to current news events, and less likely to reflect a deliberate information acquisition choice. Given the substantially higher cost of acquiring and analyzing 10-K files, the expected benefits perceived by investors should also be larger, which is consistent with our results.

The return spread of the high-minus-low-AIP portfolio is considerably smaller and less significant when returns are value weighted. The high-minus-low return is only about 30 basis points per month, and mostly not significant. This is consistent with our prior that when short-sale constraints are less binding for big firms, the information content embedded in EDGAR searching could be either positive or negative. Investors could take unconstrained short positions on big stocks to benefit from the negative information they obtained through EDGAR filings. This implies that, ex-ante, we do not have a clear **directional** prediction of a relatioship between the abnormal number of IPs and future returns. In other words, firms in the top decile of AIP are a mixture of firms with high and low expected returns, and in aggregate they cancel each other out.

Table 4 examines the relation between the abnormal number of IPs and firms' future return after controlling for the portfolios' exposure to standard asset-pricing factors. The table reports the monthly Carhart (1997) four factor alpha for decile portfolios sorted on AIP, as well as the long/short hedge portfolio. The four factor alpha is the intercept from a regression of the portfolio's excess return on the contemporaneous excess market return (MKTRF), the size factor (SMB), the value factor (HML), and the momentum factor (UMD). Panel A shows that AIP continuously predicts a strong positive return spread cross-sectionally for equal-weighted portfolios. The four-factor alphas of the long/short portfolio range from 52 to 82 basis points per month and are highly significant. Moreover, in the case of AIP_10K, the alphas are fairly symmetric across deciles. The lowest AIP decile portfolio generates a four factor alpha of about -34 basis points (t-stat=-2.84), and the highest AIP decile generates a positive alpha of 48 basis points (t-stat=3.30). The evidence suggests that when short-sale constraints are binding, investors will rationally allocate less effort towards firms with negative expected returns. Panel B of Table 4 shows the portfolio alpha for value-weighted returns. Again, we find the results are generally weaker, both economically and statistically. The four-factor alpha of the long/short portfolio ranges from 14 to 42 basis points, which are either insignificant or only marginally significant.

To emphasize the importance of measuring the abnormal level of information acquisition activity when uncovering expected return information, we conduct a parallel portfolio test when ranking firms into deciles based on the raw number of IPs searching for EDGAR filings, as shown in Table 5. Panel A reports the equal-weighted excess return and Panel B reports the equal-weighted four-factor alpha. The results show that the raw number of IPs is not significantly correlated with firms' future returns, regardless of which IP variable we use. The monthly four-factor alpha of the long-short portfolio based on the raw number of IPs ranges from -20 to 9 basis points, which are never significant. The lack of significant predictive power of the raw IP suggests that it is important to control for the expected level of information acquisition activities when uncovering investors' expected payoffs.

To obtain a better sense of how the AIP strategy performs if an investor could get access to the monthly EDGAR log file data in real-time, we plot the cumulative returns to the low and high AIP_10K decile portfolio, as well as the long-short hedge portfolio, in Figure 1. The blue line shows that one dollar invested in the lowest AIP_10K decile portfolio at the beginning of 2003 would have grown to two at the end of 2014. One dollar invested in the highest AIP_10K decile portfolio would have grown to 7.5. The grey line shows that one dollar would have grown to almost four dollars if invested in the long-short hedged portfolio, with a smooth return path.

3.2 Robustness

In Table A1, we examine the robustness of our portfolio sorts. For brevity, we focus on the sorts based on AIP_10K. The first row shows the return spread when returns are weighted by past month gross return, as suggested by Asparouhova, Bessembinder, and Kalcheva (2013). The gross-return-weighted return spread is 1.1% (t=5.16). Rows (2) and (3) show that our results barely change when we subtract the characteristic-matched portfolio (Daniel, Grinblatt, Titman, and Wermers (1997)) or industry-level return from stock return. This suggests that the nature of information contained in costly information acquisition behavior is firm-specific. In the fourth row, we augment the Carhart (1997) four-factors with the Pástor and Stambaugh (2003) liquidity factor. The Pástor and Stambaugh (2003) five-factor adjusted alpha is 0.80% (t=4.23) for the equal-weighted portfolio and 0.35% (t=1.78) for the value-weighted portfolio. The fifth row shows that our results hold when we use the Fama and French (2016) five factors to calculate alphas, with a monthly return spread of 0.69% (t=3.36) for the equal-weighted portfolio. This suggests that our long-short portfolio is not merely loading on the profitability and investment factor as proposed by Fama and French (2016). The sixth row shows that our results still hold when we use the Stambaugh and Yuan (2016) mispricing factor model to compute alpha. The portfolio generates an equal-weighted alpha of 0.89%(t=4.42) and value-weighted alpha of 0.27%(t=1.35). Using Hou, Xue, and Zhang (2015)'s Q-factor model also does not change our results, as shown in the seventh row. The eighth row of Table A1 shows that our results survive when we exclude stocks whose market capitalizations are in the bottom quintile of the NYSE size distribution. Again, the strategy based on AIP generates a monthly four-factor alpha of 0.52% (t=2.58) and 0.28% (t=1.35) when returns are equal-weighted and value-weighted, respectively. The ninth row reports the long-short alphas if we implement a six-months interval between when we sort stocks and when we measure strategy returns. The purpose of this test is to mimic the profits an investor would generate in reality since SEC delays the release of EDGAR log file data by six months. The equal-weighted alpha is quite substantially reduced in this case, but nonetheless still significant, with an equal-weighted four-factor alpha of 0.53% (t=2.23). The tenth and eleventh rows show that the long-short portfolio generates significant alpha in two subperiods: one from 2003 to 2008 and another from 2009 to 2014.

Our results are not sensitive to the specific model of calculating the abnormal number of IPs, as shown in Table A2. The first row shows that the long-short portfolio based on AIP_10K calculated using model (7) of equation (2) generates a four-factor alpha of 0.66% (t=3.95). In the second row, we show that a positive relation between AIP and returns holds for change-based specifications, which mitigates concerns that the return predictability of AIP is driven by an omitted firm-fixed effect not controlled for in our model of AIP or multivariate regressions. The long-short portfolio based on the change of AIP_10K relative to its 12-month moving average generates an equal-weighted four-factor alpha of 0.88% (t=4.82). In the third row, we include the square terms of the four firm characteristics when calculating AIP to account for the nonlinear relation between number of IPs and firm characteristics. The four-factor alpha is 0.689% and 0.552% for the equal- and value-weighted portfolio, respectively. In the fourth row, we control for the lagged number of IPs when calculating AIP, and the alpha is still significant.

3.3 The Role of Firm Size and Arbitrage Frictions

Our previous results show that the long/short portfolio alpha is only significant for equal-weighted returns, but not value-weighted returns. To take a closer look at the role of firm size, we report the portfolio sorting results based on AIP by size quintiles in Table 6. For each month, we group all stocks into size quintiles based on the NYSE size breakpoints. We then independently sort stocks into quintiles based on AIP_10K. The table reports the Carhart (1997) four-factor alpha for the 25 portfolios: equal-weighted returns in Panel A and value-weighted returns in Panel B. We also report the alpha for each size quintile of the high-minus-low-AIP portfolios. The result shows that the return predictability of AIP is strongest among microcap stocks, but is not limited to only the smallest size quintile. The high-minus-low AIP portfolio generates a significant four-factor alpha of approximately 0.4% among the three middle-sized quintiles, both equal-weighted and value-weighted. The alpha is insignificant in the largest size quintile.

The findings in Table 6 show that the return predictability of AIP is more pronounced for small firms than for large firms, which could be driven by two non-mutually exclusive channels. The first is that the latent information embedded in the number of IPs searching EDGAR files could be either positive or negative when short-sale constraints are not binding. Given that large firms have fewer short-sale impediments, the direction of return predictability for large firms is more ambiguous. An independent channel that could reinforce the weak return predictability among these stocks is that whatever information is contained in the EDGAR searches, they are arbitraged away quickly due to less arbitrage friction (e.g., short-sale costs, liquidity, non-fundamental volatility) among large firms. We now explore the return predictability of AIP with other measures of limits to arbitrage.

Following the literature, we investigate the roles of three limits-to-arbitrage measures: idiosyncratic volatility (Stambaugh, Yu, and Yuan (2015); Pontiff (2006)), residual institutional ownership (Nagel (2005)), and residual analyst coverage (Hong, Lim, and Stein (2000)). At the end of each month, we sort all stocks into terciles based on each limits-to-arbitrage variable X. We then independently sort stocks into quintiles based on the abnormal number of IPs searching for 10-Ks. Table 7 displays the equal-weighted four-factor alphas of the lowest and highest AIP portfolios in the lowest and highest X groups. Consistent with the limits to arbitrage predictions, the alpha of the high-minus-low AIP_10K portfolio is more pronounced among stocks with higher idiosyncratic volatility, lower institutional ownership, and less analyst coverage. For example, the high-minus-low AIP_10K portfolio generates 1.24% (t=4.44) monthly alpha for high-volatility stocks, and only 0.23% (t=1.76) for low-volatility stocks.

3.4 Variation in the Complexity of Financial Filings

Our maintained hypothesis in this paper is that investors' costly information acquisition activity should be positively related to the expected payoff from using the information. If this is true, we would expect the payoff to be larger when the information acquisition/processing cost is higher. To test this prediction, we use the complexity of a firm's financial filings as a proxy for the cost of information acquisition/processing. The idea is intuitive, as more complex filings require more effort and time for investors to process and digest. Following the recent literature (Loughran and McDonald (2014); You and Zhang (2009)), we use the natural log of the gross 10-K file size (complete submission text file) and the number of words contained in the filing as a proxy for filing complexity.[15]

To this end, we first obtain the filing size and number of words contained in firms' most recent 10-K reports. However, as big firms have more business lines and more diverse sets of operations, they would naturally have lengthier and larger 10-K filings.[16] To remove the confounding effect of firm size, we regress the log of filing size and number of words on the log of firms' market capitalizations, and use the regression residual as our proxy of filing complexity. At the end of each month, we sort all stocks into terciles based on either the residual file size or the residual word count. We then independently sort stocks into quintiles based on AIP_10K. Table 8 shows the equal-weighted four-factor alpha of the lowest and highest AIP_10K portfolios in the highest and lowest groups of filing complexity. Consistent with theories of endogenous information acquisition, the alpha of the high-minus-low portfolio is indeed larger and more significant for firms with more complex financial filings. For example, the high-minus-low AIP_10K portfolio generates 0.92% (t=4.46) monthly alpha among firms with the largest 10-K sizes, and 0.65% (t=3.51) among firms with the smallest 10-K sizes. The result is similar when we use the word count in 10-K as a proxy for filing complexity. Overall, the evidence strongly supports to our hypothesis that the more costly information acquistion/processing is, the larger the expected payoff revealed by information acquisition activity.

3.5 Fama-MacBeth Regression

We now test our main hypothesis using the Fama and MacBeth (1973) regression methodology. One advantage of this methodology is that it allows us to examine the predictive power of AIP while controlling for other known predictors of cross-sectional stock returns. This is important because, as shown in Table 1, AIP is correlated with some of these predictors.

[15]Loughran and McDonald (2014) report that the 10-K file size is positively associated with high return volatility in a one-month period following 10-K filings, supporting the use of file size as a proxy for the linguistic complexity of 10-K disclosure. You and Zhang (2009) find that investors' underreaction to information contained in 10-Ks is stronger for 10-Ks with larger numbers of words.

[16]The rank correlation is 0.34 between 10-K file size and firm size, and 0.40 between word count and firm size.

We conduct the Fama-MacBeth regressions in the usual way. For each month, starting in February 2003 and ending with December 2014, we run the following cross-sectional regression:

$$Ret_{i,t+1} = \beta_0 + \beta_1 AIP_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \tag{3}$$

where $Ret_{i,t+1}$ is the return of stock i in month $t+1$, $AIP_{i,t}$ is the abnormal number of IPs searching for firm i's EDGAR filings in month t, and X is a set of control variables known to predict returns, including the natural logarithm of the book-to-market ratio (LnBM), the natural logarithm of the market value of equity (LnME), returns from the prior month (Rev), returns from the prior 12-month period excluding month t-1 (Mom), institutional ownership (IO), and idiosyncratic volatility (IVOL) and past 12-month turnover (Turnover12).

Table 9 reports the time-series averages of the coefficients of the independent variables, and the t-statistics are Newey-West adjusted with four lags to control for heteroskedasticity and autocorrelation. We report the results for AIP_total in Columns (1) to (3), AIP_fundl in Columns (4) to (6) and AIP_10K in Columns (7) to (9). Columns (1), (4), and (7) show the coefficient of AIP without any other return predictors. The coefficients of all three AIP variables are positive and significant at 1% level. This is consistent with our portfolio sorting results in which stocks with abnormally large numbers of IPs searching for their EDGAR filings have higher expected returns. In Columns (2), (5), and (8), we add the usual controls including size, book-to-market ratio, past 1-month returns, and past 12-month returns. The coefficients of AIP barely change, and retain their strong predictive power. In Columns (3), (6), and (9), we further add institutional ownership, turnover, and idiosyncratic volatility to the regression, and AIP still positively predicts future returns. The economic magnitude is also quite large. The difference in AIP_10K between the lowest decile portfolio and highest decile portfolio is 2.39, which implies a monthly return spread of 105 basis points between these two extreme deciles. The magnitude estimated from the Fama-MacBeth regression is in line with our portfolio sorting results. For the control variables, the signs of the coefficients are consistent with those reported in the previous literature, except for momentum, which attracts a negative coefficient.[17] Due to the short and recent sample period, however, the

[17] This is due to the 2009 momentum crash (see Daniel and Moskowitz (2016)). The coefficient of momentum becomes positive once we exclude the year 2009 from our sample.

coefficients of most control variables are not significantly different from zero.

3.6 Alternative Explanations

3.6.1 Earnings Announcement

EDGAR searching activity is positively related to scheduled firm events such as earnings announcements (Drake, Roulstone, and Thornock (2015)). Since an earnings surprise leads to post-earnings announcement drift (Bernard and Thomas (1989)) and announcement months are generally associated with positive stock returns (Lamont and Frazzini (2007)), the return predictability of AIP may be driven by these earnings-related return predictability effects. As a robustness check, we add standardized unexpected earnings (SUE) and an earnings-announcement month dummy (EAM) to the Fama-MacBeth regression. Columns (1) to (3) of Table A4 show that the coefficients of AIP become stronger after controlling for earnings-related variables, suggesting that the information contained in AIP is not driven by earnings-related return predictability effects.

3.6.2 Breadth of Ownership

Chen, Hong, and Stein (2002) show that reduction of the breadth of institutional ownership is a proxy for overvaluation when short-sale constraints are binding for some investors. To the extent that breadth of ownership is positively correlated with the number of IPs searching for EDGAR filings, our result may simply be a rediscovery of their findings. However, Columns (4) to (6) of Table A4 show that this is not the case. The coefficients of AIP barely change after controlling for change of breadth of ownership (dBreadth). The coefficient of change of breadth of ownership is positive but insignificant, probably due to the short sample period.

3.6.3 Attention-Driven Price Pressure

We also examine the persistence of the return predictability of AIP. This test could help rule out another alternative explanation, namely that the short-run predictability is due to temporary price pressure driven by investors' demand for attention-grabbing stocks. Da, Engelberg, and Gao (2011) show that an increase in Google Search Volume for a stock predicts higher stock prices in the short-run that are eventually reversed within a year. As we hypothesize that AIP contains expected return information driven by firms' fundamental changes, the return predictability of AIP should not be reversed in the long-run. To test this, we run Fama-MacBeth regression of cumulative returns from month $t + j$ to $t + k$ on the abnormal number of IPs searching for 10-Ks in the EDGAR database (AIP_10K) in month t. The result is reported in Table 10. We separately show the return predictability of AIP_10K for the next quarter return skipping the immediate month in Column (1), the second quarter return in Column (2), the second half-year return in Column (3), and the second year return in Column (4). The table shows that the lagged value of AIP significantly predicts returns for up to two quarters, and eventually levels off for longer horizons. The coefficient of AIP is always positive and never reversed, mitigating concerns that the predictive power of AIP comes from transitory price pressure that is subsequently reversed. Investors searching firm fundamentals through the EDGAR system appear to be more sophisticated than those searching through Google Search Engine, and their aggregate information acquisition activities contain value-relevant information about firms that slowly diffuses into stock prices.

3.6.4 Investor Recognition

The positive return predictability of AIP could potentially be explained by Merton (1987)'s investor recognition hypothesis. In his model, equilibrium stock return is affected by investors' recognition of a stock because investors are not aware of all securities. Stocks with lower investor recognition have higher expected returns to compensate investors who hold the stock for insufficient diversification. An increase in investor recognition (proxied by abnormal number of IPs) of a stock will reduce its expected return going forward and lead to a contemparenous increases in stock price. This could explain why AIP predicts short-run increase in stock returns. However, other evidence is not consistent with this alternative explanation. First, a stock experiencing an increase in investor recognition should have **lower** expected returns going forward, which is inconsistent with the fact that AIP also positively predicts long-horizon returns, as presented in the last subsection. Second, the investor recognition hypothesis implies that the return predictability of AIP comes only from the reduction in discount rate, which has no implication for firms' future cash flows and profitability. However, in the next section, we show that part of the return predictability of AIP comes from its predictability for a firm's fundamental performance, because investors allocate greater effort to firms with improving fundamentals that are not fully priced in the market. The predictability of AIP for future earnings is more difficult to explain with the investor recognition hypothesis, but is more consistent with the costly information acquisition explanation.

3.6.5 Omitted Risk Factors

Last but not least, there is always the possibility that AIP captures some omitted risk factors, despite our best efforts to control for it using an extensive list of asset-pricing models. First, to the extent that omitted risk factors are persistent at firm level, a within-firm change of AIP should be less able to predict returns if the return predictability of AIP is purely driven by risk factors. However, row (2) of Table A2 shows a similarly strong return predictability of the within-firm change of AIP. Second, in the next section, we show that the return predictability of AIP partially comes from its predictability for firms' future fundamental performance. We also show that more IPs begin to search a firm through EDGAR when the firm experiences underpricing due to exogeneous reasons. Overall, the omitted risk factor explanation is difficult to square with these additional evidences.

3.7 Which Types of EDGAR Filings Matter Most?

Given the high correlation between the three types of IP measures, as shown in Table 1, we next examine whether the expected return information embedded in the three AIP variables are incremental to each other. To test this, we run a horse race by including all three AIP variables in the Fama-MacBeth regression. The result is reported in Table 11. Column (1) shows the result without other controls, and Column (2) includes all the usual return predictors. The results clearly show that the return predictability of AIP comes mainly from those searching for firms' annual reports. While AIP_10K retains its strong predictive power, the coefficient of AIP_total and AIP_fundl becomes insignificant. Acquiring and analyzing 10-Ks is more costly than for other SEC filings and more indicative of deliberate information acquisition behavior. The result is thus consistent with our hypothesis that costly information acquisition activity contains expected benefits from utilizing that information.

3.8 IPs or Searches?

Our measure of information acquisition activity essentially equal weights each investor searching through EDGAR regardless of the number of searches they requested through the EDGAR system during a one-month window. An alternative measure of information acquisition activity is the total number of searches for a firm requested by investors through the EDGAR system. This measure is problematic because, as documented by Drake, Roulstone, and Thornock (2015), the number of requests through EDGAR is dominated by a small fraction of investors who access EDGAR very frequently, and their activities are over-represented in this alternative measure.[18] Under the assumption that information is dispersed among a large group of economic agents (Hayek (1945)), we believe that our measure of the abnormal number of IPs should be more powerful in terms of inferring the latent information embedded in "the wisdom of crowd". Nevertheless, to test which measure of information acquisition activity has the stronger return predictability, we conduct a horse race between the abnormal number of searches (Asearch) and abnormal number of IPs (AIP) using the Fama-MacBeth regression approach. Using the same decomposition method, we extract the abnormal number of searches for each firm as the residual from a monthly regression of log one plus the raw number of EDGAR requests for SEC filings on the same set of firm characteristics used in equation (2).

The result is reported in Table 12. Searches/IPs for all types of EDGAR files are shown in Columns (1) and (2), searches/IPs for 10K, 10Q and 8K in Columns (3) and (4), and searches/IPs for annual reports only in Columns (5) and (6). Columns (1), (3), and (5) show that the return predictability of Asearch is generally positive but weaker than that

[18] Drake, Roulstone, and Thornock (2015) report that 86% of the users accessing EDGAR do so infrequently and only about 2% of the users access EDGAR actively during a given quarter.

of AIP. Columns (2), (4), and (6) show that once we control for AIP, the coefficient of Asearch is no longer significant and even changes sign. Importantly, the coefficients of AIP are still positive and highly significant. The result supports our use of the number of IPs as a cleaner measure of information acquisition activity, and indirectly supports the underlying assumption that private information is dispersed among market participants.

3.9 Historical vs. Current Filings

We also examine the return predictability results for IPs searching for current versus historical 10-K filings. This test could further distinguish our information acquisition hypothesis from the news-announcement explanations. On one hand, if the return predictability of AIP is entirely driven by news announcements, the result should be stronger when most IPs are searching for current financial filings as investors rush to understand the implications of current news on firm value. On the other hand, although historical filings are unlikely to provide any news to investors, they still make up an important component of the information mosaic assembled by investors, and thus should be valuable to acquire (Drake, Roulstone, and Thornock (2016)). To test this, for each stock in each month, we compute the number of days between the searching and filing dates of 10-Ks averaged across all IPs searching this stock. We then sort all stocks into two groups based on whether the average time lag between the searching and filing date of the 10-K is less or more than one year. We independently sort stocks into quintiles based on the abnormal number of IPs searching for 10-K files (AIP_10K). Interestingly, Table A3 shows that the return predictability of AIP is actually stronger when most IPs are searching for 10-Ks filed more than one year ago. Specifically, the alpha of the high-minus-low portfolio generates 0.61% (t-stat=3.08) monthly alpha when the average IP is searching for 10-Ks filed within one year, while that figure is 1% (t-stat=5.28) when most IPs are searching for historical 10-Ks.

4 Channels

The key hypothesis of this paper is that information acquisition activity embeds expected return information because investors rationally allocate greater effort to analyzing firms that are underpriced with large price appreciation potential. As mispricing implies the separation of stock prices from firms' fundamental value, there are two non-mutually exclusive channels through which investors can identify mispricing. The first channel is investors' costly information acquisition containing their favorable expectations of firms' fundamental performances that are not fully priced in by the market. The second channel is investors identifying mispricing by observing changes in stock prices that are not attributable to firms' fundamental changes. In this section, we test both channels.

4.1 Predicting Fundamental Performance

We first test whether information acquisition via EDGAR reveals novel information about firms' fundamental performance changes. We use two measures of a firm's fundamental performance. The first is the change in quarterly Return-on-Assets (dROA) from four quarters ago, which takes into account of the seasonality of firms' operating performances. The second measure is the monthly forecast revision of analysts' consensus Earnings-per-Share (EPS) forecast (FREV) scaled by stock prices 12 months ago, which is a higher-frequency measure of firms' fundamental performances. We run a panel regression of dROA and FREV on lagged AIP, controlling for other firm characteristics that are associated with firms' fundamental performances, including size, book-to-market, past 12-month returns, analyst coverage, turnover, institutional ownership, idiosyncratic volatility, and lagged quarterly ROA. Since quarterly ROA is measured at quarterly frequency, we calculate the AIP at quarterly frequency as the monthly AIP averaged within a quarter. We also control for time-fixed effects, and standard errors are double clustered by firm and time following Petersen (2009). If the return predictability of AIP is partially driven by its predictive power for firm fundamentals, the coefficient of AIP should be significantly positive.

Table 13 reports the results of predicting fundamental performance based on AIP. The dependent variable is the change in quarterly ROA from Columns (1) to (3), and analyst forecast revision from Columns (4) to (6). We show the predictability result for all three AIP measures. The coefficients of AIP are significantly positive for both measures of fundamental performance, regardless of which AIP measures we use. The economic magnitude is non-trivial. For example, Column (3) shows that an interquartile increase in AIP_10K is associated with an increase of 0.22 percentage points in dROA, which is about 17% of the interquartile range of quarterly change in ROA. This finding suggests that information acquisition via EDGAR contains investors' expectations of firms' future operating performances and even leads analysts' revisions of their forecasts of firms' fundamentals. It is worth noting that the predictability of AIP is obtained after controlling for other determinants of firms' fundamental performances. For example, the past 12-month returns strongly and positively predict changes in ROA and analyst forecast revision, while turnover and idiosyncratic volatility negatively predict fundamental performance. Overall, the test supports our hypothesis that the source of return predictability comes from investors allocating greater effort to firms with improving fundamentals.

4.2 Underpricing Driven by Outflow-induced Fire Sale

A second channel through which mispricing could occur is exogeneous shock to stock prices that is not attributable to fundamentals. One such example would be index addition events. Shleifer (1986), Wurgler and Zhuravskaya (2002), and Chang, Hong, and Liskovich (2014) show that forced buying from index-tracking institutional investors around such events could lead to large price pressure on affected stocks. However, index-addition events are rare, which limits their applicability in our setting. In this paper, we use mutual fund outflow-induced fire sale as an exogeneous shock to stock prices. Coval and Stafford (2007), Khan, Kogan, and Serafeim (2012) and Edmans, Goldstein, and Jiang (2012) find that mutual funds sell a firm's shares roughly in proportion to its portfolio weights when the funds are facing severe outflows. The forced selling behavior results in significant downward price pressure that persists for more than a year. This is a relatively exogenous and clean measure of underpricing as it is associated with who is selling–funds facing large investor redemptions–rather than what is being sold, and so is unlikely to be driven by (unobserved) changes in firms' fundamental performances.

To test whether investors expend more effort on firms experiencing fire-sale induced underpricing, we examine the change in the abnormal number of IPs following flow-induced fire sale. Specifically, we run the following Fama-MacBeth regression:

$$dAIP_{i,q+1} = \beta_0 + \beta_1 Outflows_{i,q} + \beta_2 X_{i,q} + \epsilon_{i,q+1} \tag{4}$$

where $Outflow_{i,q}$ is the flow-induced fire sale measure calculated in accordance with Edmans, Goldstein, and Jiang (2012), which reflects fund outflow expressed as a percentage of firms' shares outstanding. Our dependent variable $dAIP_{i,q+1}$ is the within-firm change of AIP in quarter $q+1$ following mutual fund outflows. X is a set of firm characteristics that may affect the change of AIP.

Table 14 reports the result. Again we report the results for all three AIP measures. Columns (1), (3), and (5) show that the coefficients of "Outflows" are significantly negative without other controls, for all three IP measures. The negative coefficient means that more investors are searching for the EDGAR filings of firms that are underpriced due to exogenous reasons. Columns (2), (4), and (6) show that the negative relation between outflow-induced selling pressure and change in AIP is robust after controlling for firms' size, book-to-market ratio, analyst coverage, idiosyncratic volatility, turnover, institutional ownership, and past returns, suggesting that our findings are likely driven by variation in underpricing.

In sum, by using mutual funds outflow-induced selling pressure to identify stock-level underpricing, our test also supports the second channel that part of the return predictability we document is attributable to investors allocating more attention and resources to firms experiencing exogenous sources of undervaluation that deviate from firm fundamentals.

4.3 Information Acquisition and Institutional Trading

As investors' information acquisition through EDGAR contains value-relevant information about stocks, a natural question that emerges is who these sophisticated investors are? Although we do not have the identify of those searching through the EDGAR system, hedge fund managers appear to fit the profile of informed investors in the equity market. A growing literature shows that hedge funds possess stock picking abilities and are able to identify stock-level mispricing (Brunnermeier and Nagel (2004); Jiao, Massa, and Zhang (2016); Agarwal, Jiang, Tang, and Yang (2013)). If hedge fund managers rationally allocate more resources and effort to undervalued firms with improving fundamentals, they should also trade in the direction of the latent information indicated by EDGAR search traffic.

To examine whether the abnormal number of IPs predict hedge fund trades, we run a Fama-MacBeth regression of net purchases by hedge funds and mutual funds in quarter q on lagged AIP, and control for other stock characteristics that might influence fund trading decisions. Net purchase is measured as the quarterly change in hedge fund holding on a stock, with holding expressed as a fraction of a firm's shares outstanding. Since hedge fund trades are inferred from quarterly holding reports, we calculate the AIP at quarterly frequency as the monthly AIP averaged within a quarter. Specifically, in each quarter, we run the following cross-sectional regression:

$$NetBuy_{i,q} = \beta_0 + \beta_1 AIP_{i.q-1} + \beta_2 Hold_{i,q-1} + \gamma X_{i,q-1} + \epsilon_{i,q} \tag{5}$$

where $NetBuy_{i,q}$ is either the net purchases by hedge funds or those by mutual funds in quarter q, $AIP_{i,q-1}$ is the abnormal number of IPs searching for firm i's EDGAR filings in quarter $q-1$, $Hold_{i,q-1}$ is either the hedge fund ownership or mutual fund ownership at the end of quarter $q-1$, and $X_{i,q-1}$ is a vector of firm characteristics at the end of quarter $q-1$, including firm size, book-to-market, analyst coverage, volatility, turnover, institutional ownership, and momentum. If hedge funds contribute to informed searches through EDGAR, the coefficient on AIP should be significant and positive.

Table 15 reports the time-series averages of the cross-sectional regression coefficients. The dependent variable from Column (1) to Column (3) is net buying by hedge funds. The coefficient of AIP in the regression of hedge funds' net purchases is positive and significant for all three measures of AIP. In terms of economic magnitude, one interquartile increase in AIP_10K is associated with an increase of 0.26 percentage points in hedge funds' net purchases, which is about 25% of the interquartile range of net buying by hedge funds. The economic magnitude is reasonable given that not all hedge funds are fundamental investors, and that they also have other information sources to aid their investment decisions. In contrast, Columns (4) to (6) show that an abnormal number of IPs searching for EDGAR filings does not significantly predict mutual funds' net purchases.

Overall, the evidence suggests that either hedge funds are part of these sophisticated investors making informed searches through EDGAR system, or that the information sources of hedge funds are consistent with the latent information embedded in "the wisdom of crowds". Either interpretation would support our premise that investors' information acquisition activity reveals their expected benefits of trading on such information.

5 Conclusion

In this paper, we examine the expected return information contained in investors' costly information acquisition activities. Specifically, we use a novel dataset of investors' requests for company filings through the EDGAR system to back out their expectations of future payoffs. To this end, we develop and implement a simple characteristic-based model to decompose the total number of IPs searching for EDGAR filings into abnormal and expected components, and show that the abnormal number of IPs searching for firms' financial reports positively predicts subsequent stock returns. A long-short portfolio that buys stocks with an abnormal number of IPs in the top decile and sells stocks in the bottom decile generates an equal-weighted monthly four-factor alpha of up to 82 basis points that is not reversed in the long run. We also find that the abnormal number of IPs predicts firms' ascending fundamental performances, and that it also increases following exogenous underpricing, suggesting that investors rationally allocate greater resources and effort to firms with large price appreciation potential. Finally, information acquisition via EDGAR also predicts subsequent purchases by hedge fund managers, suggesting that sophisticated investors are making informed searches for firms with the highest potential payoffs.

Taken together, our findings provide empirical support to theoretical models of endogenous information acquisition that costly information acquisition activity is positively associated with the value of information (Grossman and Stiglitz (1980)). Our research also highlights the promise of using the collective wisdom of investors–extracted from their EDGAR search behavior–to study expected returns and other important economic outcomes.

References

Admati, A. R., 1985, “A noisy rational expectations equilibrium for multi-asset securities markets,” *Econometrica: Journal of the Econometric Society*, pp. 629–657.

Agarwal, V., W. Jiang, Y. Tang, and B. Yang, 2013, “Uncovering hedge fund skill from the portfolio holdings they hide,” *The Journal of Finance*, 68(2), 739–783.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang, 2006, “The cross-section of volatility and expected returns,” *The Journal of Finance*, 61(1), 259–299.

Asparouhova, E., H. Bessembinder, and I. Kalcheva, 2013, “Noisy prices and inference regarding returns,” *The Journal of Finance*, 68(2), 665–714.

Barber, B. M., and T. Odean, 2007, “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors,” *The Review of Financial Studies*, 21(2), 785–818.

Ben-Rephael, A., Z. Da, and R. D. Israelsen, 2017, “It Depends on Where You Search: Institutional Investor Attention and Underreaction to News,” *The Review of Financial Studies*, p. hhx031.

Bernard, V. L., and J. K. Thomas, 1989, “Post-earnings-announcement drift: delayed price response or risk premium?,” *Journal of Accounting research*, pp. 1–36.

Brunnermeier, M. K., and S. Nagel, 2004, “Hedge Funds and the Technology Bubble,” *Journal of Finance*, pp. 2013–2040.

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

Chang, Y.-C., H. Hong, and I. Liskovich, 2014, “Regression discontinuity and the price effects of stock market indexing,” *The Review of Financial Studies*, 28(1), 212–246.

Chen, H., L. Cohen, U. G. Gurun, D. Lou, and C. J. Malloy, 2017, “IQ from IP: Simplifying Search in Portfolio Choice,” .

Chen, H., P. De, Y. Hu, and B.-H. Hwang, 2014, “Wisdom of crowds: The value of stock opinions transmitted through social media,” *The Review of Financial Studies*, 27(5), 1367–1403.

Chen, J., H. Hong, and J. C. Stein, 2002, "Breadth of ownership and stock returns," *Journal of financial Economics*, 66(2), 171–205.

Chordia, T., A. Subrahmanyam, and Q. Tong, 2014, "Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?," *Journal of Accounting and Economics*, 58(1), 41–58.

Coval, J., and E. Stafford, 2007, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 86(2), 479–512.

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

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

Daniel, K., and T. J. Moskowitz, 2016, "Momentum crashes," *Journal of Financial Economics*, 122(2), 221–247.

Diamond, D. W., and R. E. Verrecchia, 1981, "Information aggregation in a noisy rational expectations economy," *Journal of Financial Economics*, 9(3), 221–235.

Drake, M. S., P. J. Quinn, and J. R. Thornock, 2017, "Who Uses Financial Statements? A Demographic Analysis of Financial Statement Downloads from EDGAR," *Accounting Horizons*.

Drake, M. S., D. T. Roulstone, and J. R. Thornock, 2015, "The determinants and consequences of information acquisition via EDGAR," *Contemporary Accounting Research*, 32(3), 1128–1161.

——— , 2016, "The usefulness of historical accounting reports," *Journal of Accounting and Economics*, 61(2), 448–464.

Du, Z., 2015, "Endogenous Information Acquisition: Evidence from Web Visits to SEC Filings of Insider Trdes," working paper, Working paper, Kellogg School of Management.

Edmans, A., I. Goldstein, and W. Jiang, 2012, "The real effects of financial markets: The impact of prices on takeovers," *The Journal of Finance*, 67(3), 933–971.

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

——— , 2016, "Dissecting anomalies with a five-factor model," *Review of Financial Studies*, 29(1), 69–103.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, return, and equilibrium: Empirical tests," *The Journal of Political Economy*, pp. 607–636.

Gao, X., and J. R. Ritter, 2010, "The marketing of seasoned equity offerings," *Journal of Financial Economics*, 97(1), 33–52.

Gervais, S., R. Kaniel, and D. H. Mingelgrin, 2001, "The high-volume return premium," *The Journal of Finance*, 56(3), 877–919.

Green, J., J. R. Hand, and X. F. Zhang, 2017, "The characteristics that provide independent information about average us monthly stock returns," *The Review of Financial Studies*, p. hhx019.

Green, T. C., R. Huang, Q. Wen, and D. Zhou, 2017, "Wisdom of the Employee Crowd: Employer Reviews and Stock Returns," .

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

Hayek, F. A., 1945, "The use of knowledge in society," *The American economic review*, pp. 519–530.

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

Hong, H., T. Lim, and J. C. Stein, 2000, "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *The Journal of Finance*, 55(1), 265–295.

Hou, K., C. Xue, and L. Zhang, 2015, "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28(3), 650–705.

Huang, J., 2016, "The customer knows best: The investment value of consumer opinions," *Browser Download This Paper*.

Jiang, W., 2014, "Leveraged speculators and asset prices," .

Jiao, Y., M. Massa, and H. Zhang, 2016, "Short selling meets hedge fund 13F: An anatomy of informed demand," *Journal of Financial Economics*, 122(3), 544–567.

Ke, B., and Y. Yu, 2006, "The effect of issuing biased earnings forecasts on analysts' access to management and survival," *Journal of Accounting Research*, 44(5), 965–999.

Khan, M., L. Kogan, and G. Serafeim, 2012, "Mutual fund trading pressure: Firm-level stock price impact and timing of SEOs," *The Journal of Finance*, 67(4), 1371–1395.

Lamont, O., and A. Frazzini, 2007, "The earnings announcement premium and trading volume," working paper, National Bureau of Economic Research.

Lee, C. M., P. Ma, and C. C. Wang, 2015, "Search-based peer firms: Aggregating investor perceptions through internet co-searches," *Journal of Financial Economics*, 116(2), 410–431.

Lee, C. M., and E. C. So, 2017, "Uncovering expected returns: Information in analyst coverage proxies," *Journal of Financial Economics*, 124(2), 331–348.

Loughran, T., and B. McDonald, 2011, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of Finance*, 66(1), 35–65.

——— , 2014, "Measuring readability in financial disclosures," *The Journal of Finance*, 69(4), 1643–1671.

Mele, A., and F. Sangiorgi, 2015, "Uncertainty, information acquisition, and price swings in asset markets," *The Review of Economic Studies*, 82(4), 1533–1567.

Merton, R. C., 1987, "A simple model of capital market equilibrium with incomplete information," *The journal of finance*, 42(3), 483–510.

Monga, V., and E. Chasan, 2015, "The 109,894-Word Annual Report: As Regulators Require More Disclosures, 10-Ks Reach Epic Lengths; How Much Is Too Much?," *Wall Street Journal*.

Nagel, S., 2005, "Short sales, institutional investors and the cross-section of stock returns," *Journal of Financial Economics*, 78(2), 277–309.

Pástor, L., and R. F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111(3), 642–685.

Pedersen, L. H., 2015, *Efficiently inefficient: how smart money invests and market prices* *are determined*. Princeton University Press.

Petersen, M. A., 2009, "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies*, 22(1), 435–480.

Pontiff, J., 2006, "Costly arbitrage and the myth of idiosyncratic risk," *Journal of Accounting and Economics*, 42(1), 35–52.

Ryans, J. P., 2017, "Using the EDGAR Log File Data Set," .

Shleifer, A., 1986, "Do demand curves for stocks slope down?," *The Journal of Finance*, 41(3), 579–590.

Shumway, T., 1997, "The delisting bias in CRSP data," *The Journal of Finance*, 52(1), 327–340.

Stambaugh, R. F., J. Yu, and Y. Yuan, 2015, "Arbitrage asymmetry and the idiosyncratic volatility puzzle," *The Journal of Finance*.

Stambaugh, R. F., and Y. Yuan, 2016, "Mispricing factors," *The Review of Financial Studies*, 30(4), 1270–1315.

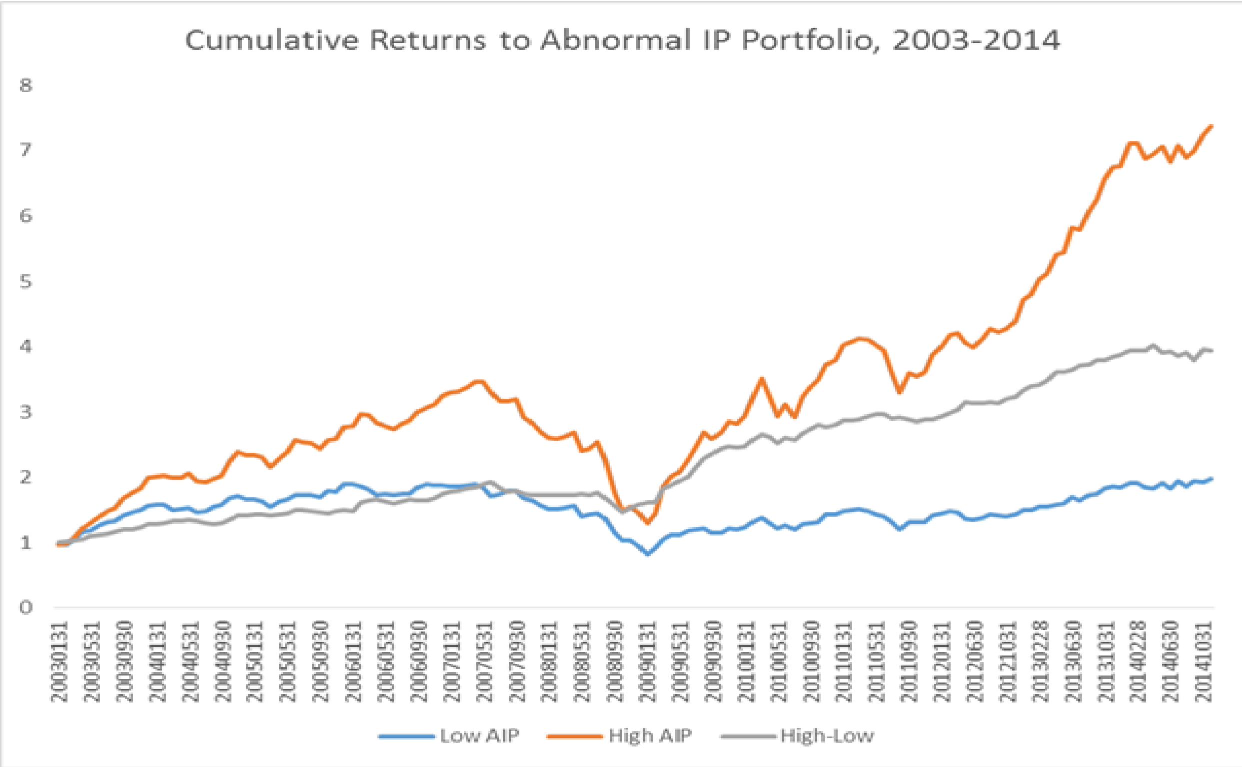
Verrecchia, R. E., 1982, "Information acquisition in a noisy rational expectations economy," *Econometrica: Journal of the Econometric Society*, pp. 1415–1430.

Wurgler, J., and E. Zhuravskaya, 2002, "Does arbitrage flatten demand curves for stocks?," *The Journal of Business*, 75(4), 583–608.

You, H., and X.-j. Zhang, 2009, "Financial reporting complexity and investor underreaction to 10-K information," *Review of Accounting Studies*, 14(4), 559–586.

Zhang, X., 2006, "Information uncertainty and stock returns," *The Journal of Finance*, 61(1), 105–137.

Figure 1: **Cumulative Returns to AIP strategy**



This figure shows the cumulative equal-weighted returns to the lowest and highest decile portfolios, sorted by the abnormal number of IPs searching for 10-K files in the EDGAR system (AIP_10K). The grey line represents the cumulative returns to the top-minus-bottom portfolio formed on AIP_10K. The sample runs from January 2003 to December 2014.

Table 1: **Stock-Level Descriptive Statistics**

This table presents the descriptive statistics of our variables. Panel A reports the summary statistics for the full sample. Panel B reports the pairwise rank correlation between our variables where they overlap. Panel C reports the characteristics of portfolios sorted by the abnormal number of IPs searching for 10-K files in the SEC's EDGAR database (AIP_10K). IP_total is the total number of unique IP addresses searching for all six types of EDGAR filings. IP_funtl is the total number of unique IP addresses searching for 10-K, 10-Q, and 8-K files. AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database. For each month, we sort all stocks into deciles based on their AIP_10K. We first calculate the mean of each variable for each decile in each month, and then calculate the time-series average of cross-sectional means. LnME is the natural log of a firm's market capitalization at the end of June of each year in millions of US dollars. Coverage is log one plus analyst coverage. Turnover12 is the monthly turnover ratio averaged over the past 12 months. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. The overall sample period is from January 2003 to December 2014.

Panel A: Summary Statistics					
Variable	Mean	Median	STD	P25	P75
		Number of IPs searching for EDGAR filings			
IP_total	155	94	317	56	159
IP_funtl	107	64	213	37	111
IP_10K	60	32	135	17	60
IP_10Q	37	24	61	13	42
IP_8K	33	19	79	10	36
		Firm-level characteristics			
LnME	6.16	6.08	1.98	4.74	7.47
LnBM	-0.66	-0.56	0.84	-1.11	-0.12
Mom	16.67%	7.64%	57.57%	-12.06%	31.78%
Coverage	1.49	1.59	1.01	0.59	2.30
IVOL	0.02	0.02	0.02	0.01	0.03
Turnover12	0.17	0.12	0.19	0.05	0.21
IO	55.30%	59.15%	31.41%	28.92%	80.58%
		Firm fundamentals and institutional trades			
dROA (%)	0.032	-0.018	4.844	-0.684	0.599
FREV (%)	-0.106	-0.001	22.185	-0.070	0.052
Net buying by HFs (%)	0.102	-0.002	2.106	-0.475	0.582
Net buying by MFs (%)	0.266	0.093	2.649	-0.640	1.204

Table 1 Continued

Panel B: Rank Correlations										
	IP_total	IP_funtl	IP_10K	LnME	Cov	Turnover12	Ivol	LnBM	Mom	IO
IP_total	1.000									
IP_funtl	0.918	1.000								
IP_10K	0.812	0.897	1.000							
LnME	0.671	0.664	0.672	1.000						
Cov	0.594	0.605	0.603	0.832	1.000					
Turnover12	0.588	0.579	0.539	0.544	0.621	1.000				
Ivol	-0.134	-0.149	-0.212	-0.523	-0.360	-0.016	1.000			
LnBM	-0.239	-0.229	-0.224	-0.319	-0.326	-0.303	0.051	1.000		
Mom	0.031	0.023	0.044	0.112	0.051	0.049	-0.117	0.008	1.000	
IO	0.469	0.494	0.514	0.650	0.647	0.615	-0.306	-0.193	0.095	1.000

Table 1 Continued

Panel C: Descriptive statistics by AIP_10K deciles												
	Obs	AIP_10K	IP_total	IP_funtl	IP_10K	LnME	Cov	Turnover12	Ivol	LnBM	Mom	IO
1(Low)	330	-1.25	59	35	12	5.977	1.369	0.154	0.025	-0.590	0.150	45.53%
2	330	-0.60	76	51	22	6.074	1.513	0.163	0.024	-0.719	0.164	53.38%
3	330	-0.38	91	63	30	6.166	1.573	0.166	0.024	-0.742	0.163	57.21%
4	330	-0.21	104	72	36	6.248	1.611	0.170	0.024	-0.741	0.172	59.23%
5	330	-0.07	116	82	42	6.270	1.623	0.171	0.024	-0.711	0.176	60.20%
6	330	0.07	128	91	48	6.284	1.634	0.170	0.024	-0.700	0.173	60.79%
7	330	0.22	141	101	55	6.218	1.594	0.165	0.024	-0.662	0.174	60.19%
8	330	0.39	160	116	66	6.118	1.526	0.164	0.024	-0.623	0.164	58.91%
9	330	0.62	201	147	87	6.032	1.454	0.158	0.025	-0.563	0.162	56.09%
10(High)	330	1.14	464	342	226	6.257	1.483	0.163	0.025	-0.537	0.168	53.28%

Table 2: **Cross-Sectional Determinants of Number of IPs Searching EDGAR Filings**

This table presents the Fama-MacBeth regression of log number of IPs searching for SEC EDGAR files. In Panel A, the dependent variable is log one plus the number of unique IP addresses searching for EDGAR filings in a month. In Panel B, the dependent variable is log one plus the number of unique IP addresses searching for 10-K, 10-Q and 8-K files in a month. In Panel C, the dependent variable is log one plus the number of unique IP addresses searching for 10-K files in a month. LnME is the natural log of a firm's market capitalization at the end of June of each year in millions of US dollars. Coverage is log one plus analyst coverage. Turnover12 is the average monthly turnover ratio over the past 12 months. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. The overall sample period is from January 2003 to December 2014.

Panel A: Dependent Variable is log(1+# of unique IP addresses searching for all EDGAR filings)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LnME	0.2713***	0.2356***	0.2475***	0.2943***	0.2992***	0.3015***	0.3026***
	(69.44)	(71.54)	(73.46)	(75.60)	(76.94)	(77.29)	(77.58)
Coverage		0.1310***	0.0422***	0.0382***	0.0321***	0.0332***	0.0360***
		(32.65)	(14.39)	(14.36)	(12.17)	(12.56)	(14.17)
Turnover12			1.0083***	0.7934***	0.7862***	0.7912***	0.7877***
			(30.21)	(29.08)	(30.04)	(29.75)	(30.52)
Ivol				9.1266***	9.0159***	9.0510***	9.0215***
				(34.65)	(33.38)	(33.16)	(32.36)
Mom					-0.0518***	-0.0529***	-0.0507***
					(-6.00)	(-6.19)	(-5.99)
LnBM						0.0171***	0.0158***
						(8.19)	(7.25)
IO							-0.0299**
							(-1.99)
Constant	2.5352***	2.6342***	2.5357***	2.0730***	2.0483***	2.0408***	2.0449***
	(39.20)	(40.68)	(40.19)	(33.45)	(33.37)	(33.32)	(32.62)
Ave.R-sq	0.404	0.483	0.520	0.554	0.558	0.559	0.563
N.of Obs.	610651	488129	488129	488123	488123	488123	484835

Table 2 Continued

Panel B: Dependent Variable is log(1+# of unique IP addresses searching for 10-K, 10-Q, and 8-K files)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LnME	0.2723***	0.2355***	0.2468***	0.2931***	0.2984***	0.3015***	0.3005***
	(64.05)	(61.21)	(60.97)	(65.17)	(66.87)	(67.27)	(67.10)
Coverage		0.1405***	0.0530***	0.0492***	0.0421***	0.0436***	0.0369***
		(35.59)	(15.60)	(15.87)	(13.86)	(14.48)	(15.57)
Turnover12			0.9833***	0.7702***	0.7708***	0.7787***	0.7560***
			(29.18)	(26.62)	(27.23)	(26.95)	(27.32)
Ivol				9.0866***	8.9652***	9.0334***	9.0934***
				(36.40)	(34.66)	(34.19)	(33.54)
Mom					-0.0684***	-0.0698***	-0.0685***
					(-7.72)	(-8.00)	(-7.95)
LnBM						0.0251***	0.0223***
						(10.23)	(9.01)
IO							0.0411***
							(2.76)
Constant	2.2017***	2.2804***	2.1866***	1.7281***	1.7033***	1.6943***	1.6868***
	(34.86)	(36.21)	(35.81)	(28.85)	(28.72)	(28.66)	(27.97)
Ave.R-sq	0.386	0.458	0.491	0.522	0.526	0.527	0.533
N.of Obs.	610651	488129	488129	488123	488123	488123	484835

Table 2 Continued

Panel C: Dependent Variable is log(1+# of unique IP addresses searching for 10-K files)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LnME	0.2979***	0.2674***	0.2765***	0.3120***	0.3169***	0.3201***	0.3155***
	(61.33)	(60.72)	(59.37)	(61.48)	(62.64)	(63.24)	(62.55)
Coverage		0.1453***	0.0729***	0.0698***	0.0637***	0.0649***	0.0431***
		(35.85)	(23.41)	(23.28)	(21.42)	(21.49)	(16.48)
Turnover12			0.8122***	0.6461***	0.6415***	0.6522***	0.5924***
			(30.68)	(28.59)	(28.59)	(28.74)	(28.38)
Ivol				6.9981***	6.9145***	7.0130***	7.2542***
				(30.56)	(29.46)	(28.94)	(29.41)
Mom					-0.0484***	-0.0510***	-0.0521***
					(-5.54)	(-5.93)	(-6.09)
LnBM						0.0267***	0.0213***
						(9.03)	(7.48)
IO							0.1600***
							(10.36)
Constant	1.3873***	1.4159***	1.3396***	0.9886***	0.9639***	0.9554***	0.9267***
	(25.17)	(25.47)	(24.67)	(18.65)	(18.51)	(18.43)	(17.62)
Ave.R-sq	0.388	0.467	0.486	0.501	0.504	0.506	0.511
N.of Obs.	610651	488129	488129	488123	488123	488123	484835

Table 3: **Portfolio Excess Returns Sorted by Abnormal Number of IPs**

This table reports the monthly average excess returns for each of the decile portfolios, as well as the long-short portfolio (High-Low). AIP_total is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all types of filings in the EDGAR database on a set of firm characteristics. Similarly, AIP_funtl (AIP_10K) is constructed using the number of IPs searching for 10-K, 10-Q, and 8-K files (10-K) in the EDGAR database. In the end of each month, all stocks are sorted into deciles based on their abnormal numbers of IPs, and a long-short portfolio is formed by buying the highest decile and shorting the lowest decile portfolio. Portfolio returns are computed over the next month. Panel A reports the results for equally weighted portfolios and Panel B shows the results for value-weighted portfolios. The sample runs from January 2003 to December 2014.

Panel A: Equal-weighted Decile Portfolio Excess Return						
	AIP_total	t-stat	AIP_funtl	t-stat	AIP_10K	t-stat
Low	0.46	1.20	0.50	1.29	0.47	1.22
2	0.78	1.78	0.76	1.73	0.63	1.40
3	0.81	1.83	0.80	1.79	0.75	1.68
4	1.08	2.33	1.04	2.27	0.85	1.81
5	1.00	2.15	1.00	2.13	0.93	1.99
6	1.07	2.24	0.99	2.07	1.02	2.11
7	1.19	2.40	1.14	2.34	1.11	2.28
8	1.12	2.19	1.06	2.05	1.26	2.51
9	1.14	2.21	1.24	2.35	1.32	2.54
High	1.18	2.29	1.29	2.55	1.48	2.98
High - Low	**0.71**	3.18	**0.79**	3.61	**1.00**	4.70
Panel B: Value-weighted Decile Portfolio Excess Return						
	AIP_total	t-stat	AIP_funtl	t-stat	AIP_10K	t-stat
Low	0.40	1.01	0.57	1.60	0.48	1.42
2	0.80	2.01	0.72	1.72	0.59	1.39
3	0.76	1.93	0.86	2.15	0.68	1.61
4	1.04	2.58	0.97	2.35	0.83	2.03
5	0.85	2.09	0.92	2.20	0.99	2.54
6	0.80	2.03	0.89	2.23	0.75	1.83
7	1.00	2.62	0.90	2.38	0.88	2.18
8	0.89	2.26	0.84	2.13	1.01	2.70
9	0.94	2.60	0.87	2.43	0.74	2.04
High	0.71	2.13	0.66	2.01	0.75	2.28
High - Low	0.31	1.23	0.09	0.44	0.26	1.32

Table 4: **Factor-adjusted Alphas of Portfolios Sorted by Abnormal Number of IPs**

This table reports the monthly Carhart (1997) four factor alphas for each of the 10 decile portfolios, as well as the long-short portfolio (High-Low). AIP_total is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all type of files in the EDGAR database on a set of firm characteristics. Similarly, AIP_funtl (AIP_10K) is constructed using the number of IPs searching for 10-K, 10-Q, and 8-K filings (10-K) in the EDGAR database. In the end of each month, all stocks are sorted into deciles based on their abnormal numbers of IPs, and a long-short portfolio is formed by buying the highest decile and shorting the lowest decile portfolio. Portfolio returns are computed over the next month. Panel A reports the results for equally weighted portfolios and Panel B shows the results for value-weighted portfolios. The sample runs from January 2003 to December 2014.

Panel A: Equal-weighted Decile Portfolio 4-factor alpha						
	AIP_total	t-stat	AIP_funtl	t-stat	AIP_10K	t-stat
Low	-0.36	-3.40	-0.32	-2.98	-0.34	-2.84
2	-0.16	-1.77	-0.19	-2.13	-0.33	-3.68
3	-0.15	-2.02	-0.16	-1.79	-0.22	-2.33
4	0.07	0.74	0.04	0.44	-0.18	-2.13
5	-0.01	-0.15	-0.02	-0.20	-0.09	-1.05
6	0.04	0.46	-0.04	-0.50	-0.02	-0.21
7	0.14	1.33	0.11	0.89	0.08	0.63
8	0.06	0.47	-0.01	-0.10	0.20	1.80
9	0.08	0.56	0.16	1.20	0.27	1.85
High	0.16	0.94	0.29	1.87	0.48	3.30
High - Low	**0.52**	2.74	**0.62**	3.33	**0.82**	4.35

Panel B: Value-weighted Decile Portfolio 4-factor alpha						
	AIP_total	t-stat	AIP_funtl	t-stat	AIP_10K	t-stat
Low	-0.40	-2.15	-0.17	-1.05	-0.22	-1.38
2	-0.07	-0.52	-0.20	-1.57	-0.34	-2.69
3	-0.11	-0.96	-0.02	-0.16	-0.25	-2.10
4	0.14	1.27	0.05	0.44	-0.08	-0.72
5	-0.06	-0.52	0.01	0.07	0.13	1.21
6	-0.07	-0.68	0.00	0.03	-0.16	-1.57
7	0.16	1.71	0.07	0.70	-0.01	-0.12
8	0.05	0.43	-0.02	-0.14	0.20	2.40
9	0.15	1.69	0.10	1.19	-0.04	-0.41
High	0.02	0.20	-0.03	-0.32	0.05	0.57
High - Low	**0.42**	1.79	0.14	0.68	0.27	1.38

Table 5: **Returns and Alphas of Portfolios Sorted by Raw Number of IPs**

This table reports the monthly excess returns and Carhart (1997) four-factor alphas for decile portfolios sorted by the raw number of IPs searching for EDGAR files. In the end of each month, all stocks are sorted into deciles based on their raw numbers of IPs, and a long-short portfolio is formed by buying the highest decile and shorting the lowest decile portfolio. Portfolio returns are computed over the next month. Panel A reports the results for equally weighted excess return and Panel B shows the results Carhart (1997) four-factor alphas. The sample runs from January 2003 to December 2014.

Panel A: Equal-weighted Decile Portfolio Excess Return						
	IP_total	t-stat	IP_funtl	t-stat	IP_10K	t-stat
Low	0.73	2.17	0.87	2.62	0.73	2.04
2	0.92	2.17	0.80	1.90	0.80	1.87
3	1.01	2.19	0.91	1.89	0.63	1.32
4	1.12	2.22	1.12	2.28	0.95	1.86
5	1.12	2.19	0.89	1.73	1.05	2.01
6	1.07	2.08	1.17	2.23	1.12	2.10
7	1.01	1.92	1.12	2.11	1.12	2.07
8	1.14	2.06	1.05	1.91	1.22	2.25
9	0.99	1.84	1.04	1.96	1.19	2.26
High	0.98	1.99	1.09	2.20	1.10	2.31
High - Low	0.26	1.19	0.22	0.68	0.37	1.58
Panel B: Equal-weighted Decile Portfolio 4-factor alpha						
	IP_total	t-stat	IP_funtl	t-stat	IP_10K	t-stat
Low	0.05	0.30	0.18	1.18	0.04	0.23
2	0.06	0.44	-0.05	-0.39	-0.12	-0.78
3	0.08	0.68	-0.07	-0.54	-0.26	-1.96
4	0.00	0.00	0.05	0.35	-0.08	-0.59
5	0.01	0.08	-0.11	-0.94	-0.08	-0.70
6	-0.09	-0.83	-0.02	-0.17	-0.01	-0.12
7	-0.09	-0.94	-0.08	-0.96	0.01	0.15
8	-0.11	-1.17	-0.08	-0.73	0.06	0.77
9	-0.13	-1.33	-0.05	-0.49	0.05	0.49
High	-0.05	-0.50	-0.02	-0.20	0.13	1.49
High - Low	-0.09	-0.56	-0.20	-1.15	0.09	0.47

Table 6: **Two-way sorts by Firm Size and Abnormal Number of IPs**

This table reports the monthly Carhart (1997) four-factor alphas (in percentages) sorted by stock's market capitalization and the abnormal number of IPs searching 10-K files (AIP_10K). AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. In the end of each month, all the stocks are sorted into quintiles based on NYSE size breakpoints. We then independently sort the stocks into quintiles based on their AIP_10K. We also report, for each size quintile, the high-AIP minus low-AIP portfolio alpha. Panel A reports the resuls on an equal-weighted basis and Panel B the results on a value-weighted basis. The sample runs from January 2003 to December 2014.

Panel A: Equal-weighted 4 factor alpha					
	Small firms	2	3	4	Large firms
Low AIP	-0.51	-0.14	-0.27	-0.17	-0.19
2	-0.19	-0.17	-0.22	-0.04	-0.23
3	-0.13	0.09	-0.04	0.01	-0.02
4	0.17	0.11	0.10	0.16	0.20
High AIP	0.64	0.22	0.16	0.20	-0.26
High-Low	**1.14**	**0.36**	**0.43**	**0.37**	-0.07
t-stat	5.38	1.72	2.01	1.68	-0.26
Panel B: Value-weighted 4 factor alpha					
	Small firms	2	3	4	Large firms
Low AIP	-0.57	-0.20	-0.27	-0.19	-0.20
2	-0.28	-0.17	-0.21	-0.04	-0.19
3	-0.15	-0.04	-0.02	-0.01	-0.02
4	-0.03	0.09	0.11	0.15	0.23
High AIP	0.41	0.02	0.19	0.21	-0.30
High-Low	**0.98**	0.22	**0.46**	**0.40**	-0.10
t-stat	4.80	0.97	2.18	1.78	-0.37

Table 7: **Limits to Arbitrage**

This table reports the results for limits to arbitrage. We sort stocks into terciles based on each limits-to-arbitrage variable X, including idiosyncratic volatility (IVOL), institutional ownership (IO) and analyst coverage (Coverage). We then independently sort stocks into quintiles based on the abnormal number of IPs searching for 10-K files (AIP_10K). AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in EDGAR database on a set of firm characteristics. We report the Carhart (1997) four-factor alpha of the lowest and highest AIP portfolios in the lowest and highest X groups. The "High-Low" column reports the Carhart (1997) four-factor alpha of the high-AIP minus low-AIP portfolios. The sample runs from January 2003 to December 2014.

	Low AIP_10K	High AIP_10K	High-Low
High IVOL	-0.76	0.48	1.24
	(-3.27)	(1.95)	(4.44)
Low IVOL	0.03	0.27	0.23
	(0.30)	(3.34)	(1.76)
High IO	-0.17	0.23	0.40
	(-1.61)	(1.75)	(2.36)
Low IO	-0.56	0.48	1.03
	(-3.53)	(1.91)	(4.41)
High Coverage	-0.33	0.18	0.51
	(-3.08)	(1.54)	(3.07)
Low Coverage	-0.41	0.68	1.10
	(-2.59)	(3.23)	(5.77)

Table 8: **Complexity of Financial Filings**

This table reports the return predictability results for variation in the complexity of financial filings. For each month, we run cross-sectional regression of the log of filing size and number of words on the log of a firm's market capitalization, and use the regression residual as our proxy for filing complexity. We sort stocks into terciles based on the residual size or residual number of words of the most recent 10-K filing. We then independently sort stocks into quintiles based on the abnormal number of IPs searching for 10-K files (AIP_10K). AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. We report the Carhart (1997) four-factor alpha of the lowest and highest AIP portfolios in the lowest and highest information cost groups. The "High-Low" column reports the Carhart (1997) four-factor alpha of the high-AIP minus low-AIP portfolios. The sample runs from January 2003 to December 2014.

	Low AIP_10K	High AIP_10K	High-Low
Large File Size	-0.48	0.44	0.92
	(-3.98)	(2.86)	(4.46)
Small File Size	-0.29	0.36	0.65
	(-2.13)	(2.64)	(3.51)
More word count	-0.48	0.49	0.97
	(-4.08)	(3.18)	(5.06)
Lesser word count	-0.36	0.32	0.68
	(-3.05)	(2.48)	(4.21)

Table 9: **Fama-MacBeth Regression: Baseline**

This table reports the results of the Fama and MacBeth (1973) regression of monthly stock returns on the abnormal number of IPs searching for EDGAR files (AIP). AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all types of files in the EDGAR database on a set of firm characteristics. Columns (1) to (3) show the results for IPs searchings for all types of EDGAR filings. Columns (4) to (6) show the results for IPs searching for 10-K, 10-Q, and 8-K files. Columns (7) to (9) show the results for IPs searching for 10-K files. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	All EDGAR Filings			10-K, 10-Q and 8K			10-K		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
AIP	0.0060***	0.0053***	0.0050***	0.0047***	0.0041***	0.0042***	0.0051***	0.0046***	0.0044***
	(2.68)	(2.64)	(2.88)	(2.70)	(2.78)	(2.94)	(3.73)	(3.81)	(3.74)
Rev		-0.0247***	-0.0283***		-0.0245***	-0.0281***		-0.0247***	-0.0284***
		(-3.18)	(-3.74)		(-3.16)	(-3.72)		(-3.19)	(-3.75)
LnME		-0.0006	-0.0014**		-0.0006	-0.0014**		-0.0006	-0.0014**
		(-0.89)	(-2.59)		(-0.92)	(-2.60)		(-0.93)	(-2.58)
LnBM		0.0019	0.0014		0.0019	0.0013		0.0019	0.0013
		(1.64)	(1.29)		(1.59)	(1.24)		(1.58)	(1.24)
Mom		-0.0058	-0.0048		-0.0057	-0.0047		-0.0058	-0.0048
		(-0.95)	(-0.88)		(-0.94)	(-0.86)		(-0.94)	(-0.86)
Ivol			-0.0015			-0.0025			-0.0007
			(-0.02)			(-0.04)			(-0.01)
Turnover12			-0.0094			-0.0091			-0.0089
			(-1.37)			(-1.32)			(-1.28)
IO			0.0122***			0.0119***			0.0114***
			(4.00)			(3.94)			(3.86)
Constant	0.0123**	0.0122	0.0119**	0.0122**	0.0122*	0.0120**	0.0122**	0.0123*	0.0119**
	(2.18)	(1.65)	(2.33)	(2.18)	(1.66)	(2.36)	(2.18)	(1.67)	(2.35)
Ave.R-sq	0.003	0.030	0.046	0.003	0.030	0.046	0.003	0.030	0.046
N.of Obs.	483667	483667	480793	483667	483667	480793	483667	483667	480793

Table 10: **Predicting Long-horizon Returns**

This table reports the results from the Fama and MacBeth (1973) regression of cumulative returns from month $t + j$ to $t + k$ on the abnormal number of IPs searching for 10-K files in the EDGAR database (AIP_10K) in month t. AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	Ret(2,4)	Ret(5,7)	Ret(8,13)	Ret(14,25)
AIP_10K	0.0102***	0.0068**	0.0150	0.0175
	(2.95)	(2.05)	(1.57)	(0.64)
Rev	-0.0072	0.0037	0.0033	-0.0451
	(-0.53)	(0.21)	(0.11)	(-0.93)
LnME	-0.0023	-0.0013	-0.0015	-0.0048
	(-1.64)	(-1.03)	(-0.61)	(-1.11)
LnBM	0.0046*	0.0041	0.0118**	0.0197*
	(1.72)	(1.57)	(2.36)	(1.79)
Mom	-0.0193	-0.0117	-0.0300*	-0.0421
	(-1.24)	(-0.88)	(-1.75)	(-1.26)
Ivol	0.0407	-0.0184	0.2652	0.5759
	(0.20)	(-0.10)	(0.73)	(0.84)
Turnover12	-0.0165	-0.0312*	-0.0451	-0.0488
	(-0.92)	(-1.95)	(-1.53)	(-1.08)
IO	0.0116	0.0152**	0.0414**	0.0956**
	(1.63)	(2.18)	(2.42)	(2.47)
Constant	0.0370**	0.0281*	0.0451	0.0947
	(2.41)	(1.72)	(1.53)	(1.51)
Ave.R-sq	0.051	0.044	0.036	0.035
N.of Obs.	469185	456068	425505	360584

Table 11: **Which Types of EDGAR Files?**

This table reports the results of the Fama and MacBeth (1973) regression of monthly stock returns on the abnormal number of IPs searching for EDGAR filings (AIP). AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all types of files in the EDGAR database on a set of firm characteristics. AIP_total shows the results for IPs searchings for all types of EDGAR files. AIP_fundl shows the results for IPs searching for 10-K, 10-Q, and 8-K files. AIP_10K shows the results for IPs searching for 10-K files. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	(1)	(2)
AIP_total	-0.0014	-0.0003
	(-0.63)	(-0.17)
AIP_fundl	0.0022	0.0012
	(1.11)	(0.70)
AIP_10K	0.0049***	0.0043***
	(3.96)	(4.02)
Rev		-0.0287***
		(-3.80)
LnME		-0.0014**
		(-2.52)
LnBM		0.0013
		(1.24)
Mom		-0.0048
		(-0.88)
Ivol		-0.0027
		(-0.04)
Turnover12		-0.0088
		(-1.27)
IO		0.0112***
		(3.84)
Constant	0.0122**	0.0120**
	(2.18)	(2.34)
Ave.R-sq	0.005	0.048
N.of Obs.	483667	480793

Table 12: **Number of IPs or Number of Searches?**

This table reports the results of the Fama and MacBeth (1973) regression. Asearch is the residual from a monthly regression of log one plus the total number of EDGAR requests for SEC filings. AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for EDGAR files on a set of firm characteristics. Columns (1) and (2) show the results for searching for all types of EDGAR files. Columns (3) and (4) show the results for searching activities for 10-K, 10-Q, and 8-K files. Columns (5) and (6) show the results for searching activities for 10-K files. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	All EDGAR Files		10K, 10Q, 8K		10K	
Asearch	0.0014	-0.0004	0.0020*	-0.0024	0.0033***	-0.0039
	(1.54)	(-0.42)	(1.90)	(-1.49)	(3.93)	(-1.57)
AIP		0.0055**		0.0062***		0.0084***
		(2.45)		(2.83)		(2.90)
Rev	-0.0283***	-0.0284***	-0.0283***	-0.0284***	-0.0284***	-0.0289***
	(-3.73)	(-3.76)	(-3.74)	(-3.77)	(-3.75)	(-3.75)
LnME	-0.0014**	-0.0014***	-0.0014**	-0.0014**	-0.0014***	-0.0013***
	(-2.59)	(-2.63)	(-2.61)	(-2.52)	(-2.64)	(-3.11)
LnBM	0.0013	0.0014	0.0014	0.0014	0.0012	0.0015*
	(1.26)	(1.31)	(1.34)	(1.36)	(1.13)	(1.71)
Mom	-0.0049	-0.0048	-0.0048	-0.0049	-0.0048	-0.0049
	(-0.89)	(-0.88)	(-0.87)	(-0.89)	(-0.86)	(-1.15)
Ivol	0.0048	-0.0014	0.0065	-0.0033	0.0039	-0.0021
	(0.07)	(-0.02)	(0.09)	(-0.05)	(0.05)	(-0.03)
Turnover12	-0.0100	-0.0096	-0.0095	-0.0091	-0.0095	-0.0088
	(-1.46)	(-1.39)	(-1.38)	(-1.33)	(-1.37)	(-1.33)
IO	0.0127***	0.0123***	0.0122***	0.0115***	0.0120***	0.0109***
	(4.10)	(4.04)	(4.06)	(3.86)	(4.03)	(3.57)
Constant	0.0115**	0.0120**	0.0116**	0.0119**	0.0117**	0.0120***
	(2.26)	(2.35)	(2.29)	(2.33)	(2.32)	(3.19)
Ave.R-sq	0.046	0.047	0.046	0.048	0.046	0.049
N.of Obs.	480793	480793	480793	480793	480793	480793

Table 13: **Predicting Fundamental Performance**

This table reports the results of the panel regression of future fundamental performance measure on the abnormal number of IPs searching for 10-K files in the EDGAR database in month *t*. AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for EDGAR filings on a set of firm characteristics. The dependent variable in Columns (1) to (3) is the change of quarterly Return-on-Assets from four quarters ago. The dependent variable in Columns (4) to (6) is the monthly revision of analysts consensus annual EPS forecast. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. Coverage is log one plus analyst coverage. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). We control for the year-quarter fixed effects in Columns (1) to (3) and the year-month fixed effects in Columns (4) to (6). Turnover12 is the monthly turnover ratio averaged over the past 12 months. Standard errors are double clustered at both firm and time level. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	Change of ROA			Forecast Revision		
	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K
AIP	0.0017*	0.0026**	0.0028***	0.0007***	0.0016***	0.0019***
	(1.96)	(2.51)	(2.92)	(2.78)	(6.19)	(5.28)
LROA	-0.3425***	-0.3428***	-0.3430***			
	(-4.71)	(-4.73)	(-4.74)			
LnME	0.0008	0.0008	0.0008	-0.0005	-0.0005	-0.0005
	(1.27)	(1.31)	(1.33)	(-1.51)	(-1.55)	(-1.63)
LnBM	-0.0013	-0.0012	-0.0012	-0.0008**	-0.0008**	-0.0009**
	(-0.87)	(-0.84)	(-0.83)	(-2.37)	(-2.39)	(-2.47)
Mom	0.0100***	0.0099***	0.0100***	0.0025***	0.0025***	0.0025***
	(3.55)	(3.56)	(3.57)	(5.20)	(5.14)	(5.18)
Coverage	0.0004	0.0004	0.0005	0.0021***	0.0021***	0.0021***
	(0.29)	(0.31)	(0.32)	(3.30)	(3.29)	(3.28)
Turnover12	-0.0118**	-0.0117**	-0.0117**	-0.0082***	-0.0082***	-0.0082***
	(-2.43)	(-2.42)	(-2.43)	(-3.15)	(-3.16)	(-3.17)
IO	-0.0010	-0.0011	-0.0013	0.0049***	0.0051***	0.0052***
	(-0.48)	(-0.51)	(-0.61)	(5.47)	(5.61)	(5.73)
Ivol	-0.0777	-0.0773	-0.0775	-0.1111**	-0.1115**	-0.1138**
	(-1.41)	(-1.41)	(-1.42)	(-2.35)	(-2.37)	(-2.41)
Time FE	yes	yes	yes	yes	yes	yes
Adj.R-sq	0.056	0.056	0.056	0.002	0.002	0.002
N.of Obs.	128504	128504	128504	348130	348130	348130

Table 14: **Mutual Fund Outflows Induced Mispricing and Abnormal Number of IPs**

This table reports the results of the Fama and MacBeth (1973) regression of the quarterly change in the abnormal number of IPs searching for EDGAR files on quarterly mutual fund outflows. Outflows is calculated following Edmans, Goldstein, and Jiang (2012). AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for EDGAR filings on a set of firm characteristics. dAIP equals the within-firm change in AIP in the quarter in which mutual fund outflows occur. LnME is the natural log of a firm's market capitalization at the end of June of each year in millions of US dollars. Coverage is log one plus analyst coverage. Turnover12 is the monthly turnover ratio averaged over the past 12 months. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	dAIP_total		dAIP_funtl		dAIP_10K	
	(1)	(2)	(3)	(4)	(5)	(6)
Outflows	-2.4242***	-1.7256***	-1.9145***	-1.3527***	-1.9303**	-1.5459**
	(-4.02)	(-4.92)	(-3.36)	(-3.27)	(-2.06)	(-2.31)
LnME		-0.0091***		-0.0094***		-0.0093***
		(-6.03)		(-5.68)		(-5.81)
LnBM		0.0013		-0.0014		-0.0017
		(0.56)		(-0.57)		(-0.75)
Coverage		0.0080***		0.0076***		0.0087***
		(4.50)		(4.28)		(3.70)
Ivol		-1.8233***		-1.9963***		-1.8354***
		(-6.48)		(-7.68)		(-6.19)
Turnover12		-0.0015		0.0158		0.0203
		(-0.09)		(1.13)		(1.56)
IO		-0.0023		-0.0141**		-0.0143**
		(-0.36)		(-2.54)		(-2.28)
Mom		-0.0336***		-0.0370***		-0.0398***
		(-5.17)		(-5.70)		(-7.68)
Constant	0.0007	0.0901***	0.0050**	0.1036***	0.0049**	0.0967***
	(0.29)	(7.79)	(2.09)	(8.54)	(2.06)	(6.54)
Ave.R-sq	0.001	0.031	0.001	0.034	0.001	0.026
N.of Obs.	131863	131041	131863	131041	131863	131041

Table 15: **Information Acquisition and Institutional Trading**

This table reports Fama and MacBeth (1973) regression of institutional trading in quarter q on the abnormal number of IPs searching for EDGAR filings (AIP) in quarter $q-1$. The dependent variable in Columns (1) to (3) is the net buying by hedge fund in that quarter. The dependent variable in Columns (4) to (6) is the net buying by mutual funds in that quarter. AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all types of files in the EDGAR database on a set of firm characteristics. Columns (1) and (4) show the results for IPs searchings for all types of EDGAR filings. Columns (2) and (5) show the results for IPs searching for 10-K, 10-Q, and 8-K files. Columns (3) and (6) show the results for IPs searching for 10-K files. Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	Net Buying by Hedge Funds			Net Buying by Mutual Funds		
	AIP_total	AIP_funtl	AIP_10K	AIP_total	AIP_funtl	AIP_10K
AIP	0.0053**	0.0043**	0.0034**	0.0030	0.0046	0.0065
	(2.45)	(2.43)	(2.27)	(0.52)	(0.77)	(1.00)
Lagged Holding	-0.1245***	-0.1234***	-0.1218***	-0.1557***	-0.1569***	-0.1573***
	(-7.62)	(-7.72)	(-9.24)	(-3.91)	(-3.83)	(-3.77)
LnME	-0.0011***	-0.0011***	-0.0011***	-0.0003	-0.0003	-0.0003
	(-3.37)	(-3.39)	(-4.12)	(-1.07)	(-1.08)	(-1.06)
LnBM	-0.0002*	-0.0002*	-0.0002	0.0003	0.0004	0.0003
	(-1.88)	(-1.74)	(-1.64)	(0.32)	(0.38)	(0.34)
Cov	-0.0006	-0.0006	-0.0005	-0.0008	-0.0007	-0.0005
	(-1.40)	(-1.33)	(-1.51)	(-0.30)	(-0.26)	(-0.22)
Ivol	0.0145	0.0153	0.0164	-0.1965*	-0.1975*	-0.2012*
	(0.76)	(0.80)	(0.97)	(-1.96)	(-1.94)	(-1.95)
Turnover12	0.0041*	0.0042*	0.0044**	-0.0069**	-0.0068**	-0.0064**
	(1.76)	(1.78)	(2.23)	(-2.04)	(-2.22)	(-2.38)
IO	0.0150***	0.0145***	0.0140***	0.0739***	0.0736***	0.0728***
	(3.09)	(3.14)	(3.87)	(2.86)	(2.90)	(2.92)
Mom	0.0003	0.0004	0.0002	0.0066***	0.0067***	0.0065***
	(1.14)	(1.22)	(0.97)	(6.44)	(6.29)	(7.74)
Constant	0.0070***	0.0070***	0.0069***	0.0048*	0.0050*	0.0053*
	(4.06)	(4.01)	(4.68)	(1.95)	(1.91)	(1.96)
Ave.R-sq	0.091	0.091	0.091	0.113	0.113	0.113
N.of Obs.	131795	131795	131795	131795	131795	131795

Appendices

Table A1: **Robustness of Decile Portfolio Sorts**

This table reports the results of several robustness tests for a long/short portfolio based on the abnormal number of IPs searching for 10-K files in the EDGAR database (AIP_10K). AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. For the first robustness test, we report the gross return-weighted portfolio returns, for which the weights are 1 + the stock's lagged monthly return, following Asparouhova, Bessembinder, and Kalcheva (2013). The second robustness test shows the portfolio returns adjusted using the DGTW method. The third set of robustness tests shows the Fama-French 48 industry-adjusted excess return. The fourth row shows the alpha using the Pástor and Stambaugh (2003) liquidity factor augmented with the Fama-French factors and the momentum factor. For the fifth set of tests, we report the alphas using the Fama and French (2016) Five Factor model. For the sixth and seventh sets of tests, we report the alphas using the Stambaugh and Yuan (2016) Mispricing Factors model and the Hou, Xue, and Zhang (2015) Q-factor model. For the eighth set of analyses, we exclude stocks whose market capitalizations are in the bottom quintile based on NYSE size breakpoints. In the ninth row, we skip six months between the moment an abnormal IP is constructed and the moment at which we start measuring returns. In the tenth and eleventh rows, we report the four-factor alpha for two sub-sample periods, one from 2003 to 2008 and the another from 2009 to 2014.

	EW	VW
Gross return-weighed portfolio	1.096	NA
	(5.16)	
DGTW adjusted	0.910	0.410
	(4.51)	(2.22)
FF48 Industry-adjusted	0.739	0.155
	(3.26)	(1.16)
FF + Cahart + PS Factor	0.800	0.348
	(4.23)	(1.78)
FF five factor (2015)	0.685	0.248
	(3.36)	(1.19)
Mispricing factors (Stambaugh and Yuan 2017)	0.892	0.276
	(4.42)	(1.35)
Q-factor (Hou, Xue and Zhang 2015)	0.897	0.183
	(4.66)	(0.87)
Remove microcap stocks	0.518	0.276
	(2.58)	(1.35)
Skip six months	0.532	0.266
	(2.23)	(1.28)
2003-2008	0.620	0.261
	(2.41)	(0.89)
2009-2014	1.073	0.121
	(3.74)	(0.45)

Table A2: **Alternative Implementations of AIP**

This table reports several alternative implementations of AIP_10K when calculating the long/short portfolio Carhart (1997) four-factor alpha. AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. In the first row, we calculate AIP_10K using model (7) of equation 2. In the second row, we sort portfolios based on changes in AIP_10K relative to its 12-month moving average. In the third row, we also include the square term of the four firm characteristics when calculating AIP. In the fourth row, we include the lagged number of IPs in the expected IP regression. Column (1) reports the results for the equal-weighted portfolio, and Column (2) reports for the value-weighted portfolio. The sample runs from January 2003 to December 2014.

	EW	VW
Model (7) of Expected IP Regression	0.658	0.156
	(3.95)	(0.82)
Change in AIP relative to 12-months average	0.883	0.388
	(4.82)	(1.44)
Nonlinear functional form of Expected IP Regression	0.689	0.552
	(4.30)	(2.39)
Control for lagged # of IPs in Expected IP Regression	0.698	0.508
	(5.44)	(2.03)

Table A3: **Current or Historical Accounting Report?**

This table reports the return predictability results for IPs searching for current versus historical 10-K reports. For each month and each stock, we compute the number of days between the searching and filing dates of 10-Ks averaged across all IPs searching for this stock. We then sort all stocks into two groups based on whether the average time lag between the searching and filing dates of 10-K is less or more than one year. We independently sort stocks into quintiles based on the abnormal number of IPs searching for 10-K files (AIP_10K). AIP_10K is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for 10-K files in the EDGAR database on a set of firm characteristics. We report the Carhart (1997) four-factor alpha of the lowest and highest AIP portfolios. The "High-Low" column reports the Carhart (1997) four-factor alpha of the high-AIP minus low-AIP portfolios. The sample runs from January 2003 to December 2014.

	Low AIP_10K	High AIP_10K	High-Low
less than 1 year	-0.41	0.21	0.61
	(-2.29)	(1.29)	(3.08)
more than 1 year	-0.45	0.55	1.00
	(-4.54)	(3.63)	(5.28)

Table A4: **Fama-MacBeth Regression: Controlling for Earnings Surprise, Earnings Announcement Premium and Change of Breadth of Ownership**

This table reports the results of the Fama and MacBeth (1973) regression of monthly stock returns on the abnormal number of IPs searching for EDGAR filings (AIP). AIP is the residual from a monthly regression of log one plus the total number of unique IP addresses searching for all types of files in the EDGAR site on a set of firm characteristics. Columns (1) and (4) show the results for IPs searching for all types of EDGAR filings. Columns (2) and (5) show the results for IPs searching for 10-K, 10-Q, and 8-K files. Columns (3) and (6) show the results for IPs searching for 10-K files. SUE is a firm's standardized unexplained earnings, defined as the realized earnings per share (EPS) minus EPS from four quarters prior, divided by the standard deviation of this difference over the prior eight quarters. EAM is a dummy variable that equals one when a given firm announces earnings in the month. dBreadth is the percentage change of breadth of 13F institutional ownership, following Chen, Hong, and Stein (2002). Size (LnME) is the natural log of a firm's market capitalization at the end of June of each year. Book-to-market (LnBM) is the natural log of the book-to-market ratio. The cases with negative book value are deleted. Momentum (MOM) is defined as the cumulative returns from month t-12 to t-2. The short term reversal measure (REV) is the lagged monthly return. Institutional ownership (IO) is the sum of shares held by institutions from 13F filings in each quarter divided by the total shares outstanding. IVOL is the idiosyncratic volatility, calculated following Ang, Hodrick, Xing, and Zhang (2006). Turnover12 is the monthly turnover ratio averaged over the past 12 months. All t-statistics are Newey-West adjusted to control for heteroskedasticity and autocorrelation. ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K
AIP	0.0057***	0.0047***	0.0044***	0.0056***	0.0046***	0.0044***
	(3.31)	(3.32)	(4.02)	(3.29)	(3.27)	(3.97)
Rev	-0.0300***	-0.0298***	-0.0300***	-0.0299***	-0.0297***	-0.0299***
	(-4.08)	(-4.06)	(-4.09)	(-4.10)	(-4.08)	(-4.10)
LnME	-0.0017***	-0.0017***	-0.0016***	-0.0017***	-0.0017***	-0.0016***
	(-3.13)	(-3.16)	(-3.14)	(-3.13)	(-3.16)	(-3.15)
LnBM	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
	(1.56)	(1.51)	(1.53)	(1.57)	(1.51)	(1.54)
Mom	-0.0065	-0.0064	-0.0064	-0.0068	-0.0067	-0.0067
	(-1.13)	(-1.12)	(-1.11)	(-1.18)	(-1.17)	(-1.17)
Ivol	0.0116	0.0104	0.0121	0.0077	0.0063	0.0083
	(0.16)	(0.15)	(0.17)	(0.11)	(0.09)	(0.12)
Turnover12	-0.0097	-0.0093	-0.0092	-0.0093	-0.0089	-0.0088
	(-1.38)	(-1.32)	(-1.29)	(-1.33)	(-1.27)	(-1.24)
IO	0.0118***	0.0114***	0.0109***	0.0118***	0.0114***	0.0109***
	(3.42)	(3.35)	(3.26)	(3.43)	(3.36)	(3.27)
SUE	0.0028***	0.0028***	0.0027***	0.0027***	0.0027***	0.0027***
	(8.36)	(8.39)	(8.46)	(8.29)	(8.31)	(8.37)
EAM	0.0032**	0.0034**	0.0027**	0.0032**	0.0033**	0.0027**
	(2.51)	(2.59)	(2.23)	(2.48)	(2.56)	(2.19)
dBreadth				0.0710	0.0789	0.0842
				(0.95)	(1.05)	(1.13)
Constant	0.0121**	0.0122**	0.0122**	0.0121**	0.0123**	0.0123**
	(2.46)	(2.49)	(2.50)	(2.42)	(2.46)	(2.47)
Ave.R-sq	0.051	0.051	0.051	0.052	0.052	0.052
N.of Obs.	443261	443261	443261	442794	442794	442794