Demand for Information and Stock Returns: Evidence from EDGAR

Pingle Wang*

January 10, 2020

Abstract

This paper empirically shows that information acquisition affects stock returns by reducing firm-level information asymmetry. When firms disclose material information that was known by insiders, demand for information transforms private information into public information, drives up the contemporaneous price, and reduces the risk premium in the future. The supply of information has no direct effect on information asymmetry, but acts as a catalyst, which magnifies the effect of the demand for information. Moreover, demand for information has stronger effects when investors are geographically close to firm headquarters or have prior experience in collecting firm-specific information, suggesting that the cost of information acquisition affects information asymmetry.

*Simon Business School, University of Rochester. E-mail: pingle.wang@simon.rochester.edu. I am extremely grateful to Ron Kaniel for his relentless support and guidance. I am also grateful to my dissertation committee members: Alan Moreira and Jerold Warner for their invaluable support. I would like to thank Shuaiyu Chen, Ramona Dagostino, Michael Gofman, Xiaoxia Lou, Robert Novy-Marx, Christian Opp, Dmitry Orlov, Mykola Pinchuk, Bill Schwert, Clifford Smith, Giulio Trigilia, Pavel Zryumov, and seminar participants at Simon Business School for their helpful discussions and comments. All errors are my own.

1 Introduction

Information plays a central role in determining asset prices. From the supply side, a large and extensive empirical literature has investigated how information disclosure affects stock returns through changing the information asymmetry between insiders and investors, but papers have found mixed results under different settings.[1] One potential explanation for the inconclusive evidence is that the supplied information may not be fully absorbed by investors (Huberman and Regev (2001)), and these papers do not directly incorporate the information acquisition. Therefore, it is crucially important to study the demand side of information as well. The empirical literature regarding the effect of information acquisition on stock returns is growing but still very small due to data limitations. Empirical studies have used direct and indirect measures of information acquisition, all of which predict positive abnormal returns in the future.[2] Researchers explain the empirical findings with an attention-based behavioral channel: Given the limited attention and the short-selling constraint, investors are more likely to buy rather than sell attention-grabbing stocks. Therefore, information acquisition reflects investors' increased attention and recognition of the firm, and it predicts positive abnormal returns in the future.

While the empirical findings suggest a positive correlation between information acquisition and subsequent stock returns, it is puzzling from a theoretical standpoint. Theoretical papers show that acquiring material information reduces the payoff uncertainty and the asymmetric information between insiders and investors (*i.e., information asymmetry channel*).[3] Therefore, stocks become less risky to hold, and information acquisition should predict a lower risk premium in the future, which is in the opposite direction of the existing empiri-

[1] For example, Lee, Mucklow, and Ready (1993), and Coller and Yohn (1997) argue that earnings announcement and management forecasts increase the firm-level information asymmetry, which predicts increased risk premiums. Bushee, Core, Guay, and Hamm (2010), and Kelly and Ljungqvist (2012) argue that press release and analyst coverage reduce the firm-level information asymmetry and future risk premiums.

[2] Gervais, Kaniel, and Mingelgrin (2001) and Barber and Odean (2007) use abnormal trading volume or extreme stock returns to capture investor attention indirectly. Da, Engelberg, and Gao (2011) and Ben-Rephael, Carlin, Da, and Israelsen (2017) use Google Trends and Bloomberg Search Index to directly measure information acquisition of retail and institutional investors.

[3] See Grossman and Stiglitz (1980),Verrecchia (1982),Wang (1993),Easley and O'Hara (2004).

cal findings. To the best of my knowledge, no empirical evidence has directly supported the information asymmetry channel. Therefore, my paper tries to examine *whether information acquisition reduces information asymmetry and lowers the stock return.*

The existing proxy for information acquisition is limited to aggregated measures, such as Google Trends or Bloomberg Search Index. These measures capture the search intensity but fail to distinguish the types of information collected by investors. On the one hand, if investors acquire *material* information, it can potentially reduce the information asymmetry and lower the risk premium. On the other hand, if investors are attracted by attention-grabbing events and collect *stale* information to put events into context, it will lead to a positive return in the future. Given the opposite predictions of returns under each channel, it is critical to disentangle the types of information acquisition.

My paper proposes a novel empirical setting through which stale and material information acquisition can be cleanly distinguished, which allows me to test the theoretical predictions of information acquisition under the information asymmetry channel. Based on the filing downloads from the EDGAR log data between 2003 and 2016, I show that acquiring historical filings that contain *stale* information predicts positive abnormal returns that gradually decay over time, consistent with the attention-based behavioral channel documented in the literature. However, when investors collect unanticipated *material* information that was privately known by insiders, the contemporaneous price increases, followed by a persistent reduction of risk premium in the future. To directly test the underlying mechanism, I show that acquiring material information reduces the firm-level information asymmetry proxied by the price impact. Furthermore, to alleviate the concern of endogenous attention allocation and information asymmetry, I use the Northeast Blackout event as an exogenous shock to investors' information acquisition, which establishes a causal link between information acquisition and the firm-level information asymmetry. My finding provides the first empirical evidence that supports the information asymmetry channel and highlights the important distinction between *material* and *stale* information acquisition.

It is challenging to test whether information acquisition reduces the information asymmetry, as firms may have unobserved incentives to disclose information, which can correlate with both subsequent information acquisition and the information asymmetry. To alleviate the concern from the supply side of information, I focus on the mandatory disclosure, Form 8-K, which is required by the SEC and is filed by firms within four business days after a triggering event. The filings typically contain event-specific material information that was privately known by firm insiders. Compared with other types of information disclosure, such as management forecasts or annual reports, Form 8-K leaves firms with little room to manipulate regarding *when* and *what* to disclose. Accordingly, I use the number of 8-K downloads to proxy for material information acquisition through the information asymmetry channel.

I find that the effect of material information acquisition on stock returns is consistent with the information asymmetry channel, which not only implies a lower return in the subsequent period but also has the predictability over the long-term. Specifically, demand for 8-K predicts a significant and negative abnormal return of -12 basis points (bps) for the following month using Fama-Macbeth regression and controlling for media coverage and a set of risk factors. Moreover, the long/short portfolio sorted on 8-K downloads yields a positive return in the formation week and a persistent and negative return over the long-term.[4] Such a long-term predictability lends further support for the information asymmetry channel. As suggested by the theoretical work, acquiring material information reduces the risk of holding the stock, which leads to an increase in current stock price, followed by a lower and persistent risk premium in the long-term.

To further sharpen my hypothesis that the information asymmetry channel rests on acquiring unanticipated information, I explore the variation of information content with two analyses. In the first analysis, I decompose demand for 8-K into two parts: demand for *unscheduled* and *scheduled* 8-K. Scheduled 8-K filings contain little new information, whereas

[4]Throughout the paper, I use Fama-French five factors and UMD factor as the testing model unless specified otherwise. All results are robust under other specifications, such as CAPM model or Fama-French three/four-factor models. When conducting portfolio sort analysis, I use the size-adjusted information acquisition, because larger firms typically attract more downloads in EDGAR.

unscheduled 8-Ks disclose material information only known by insiders.[5] I show that demand for unscheduled 8-K predicts lower subsequent returns, whereas demand for scheduled 8-K has no predictability.

In the second analysis, I distinguish between *good* and *bad* news to alleviate the following concern: information contents could drive both returns and the 8-K downloads, which leads to the spurious findings in the main analysis. For example, investors may overreact to good news so that future returns are lower. Alternatively, bad news can attract more downloads than good news, and firms with bad (good) news disclosure are likely to release bad (good) news in the near future, which can also predict subsequent negative returns. To test whether the result is driven by information content, I calculate cumulative abnormal returns (CAR) around event dates from the 8-K filings. The more positive (negative) the CAR is, the better (worse) the news is perceived by the market. The CAR has a mean of 0.4% and a standard deviation of 12%. The histogram of the CAR suggests that good and bad news are equally likely among 8-K filings. I then sort 8-K filings by CAR into quintiles and show that 8-K downloads of either good or bad news predict a lower return in the future, whereas 8-K downloads of neutral news have no predictability. Only when there is a lot of market reaction around the event date due to unexpected information release, can information acquisition greatly reduce the information asymmetry between insiders and investors.

Having demonstrated the effect of material information acquisition on returns, I then directly test its effect on information asymmetry proxied by the price impact measure.[6] Contrary to the existing literature, I show that 8-K disclosure itself has no direct effect on firm-level information asymmetry, but acts as a catalyst, which magnifies the effect of 8-K downloads in reducing information asymmetry. In terms of the economic magnitude, a 10%

[5]Around 10% of the total 8-K filings are scheduled filings. The SEC typically categorizes such filings under Item 2.02. A typical example is the announcement of the recent conference calls to discuss its earnings. All the non-public information is transmitted to the public during the conference calls, leaving the scheduled 8-K filing with no additional information to be learned by investors.

[6]The results are robust to other measures, such as bid-ask spread, PINs, and Amihud Illiquidity. These measures typically have a large liquidity component, whereas the price impact measure captures the information component of the trade.

increase in 8-K downloads reduces a firm's price impact measure by 8 bps in the subsequent week, corresponding to a 20% reduction of the price impact measure around the mean. My result highlights the interaction between the supply and demand sides of information, which can potentially explain the inconclusive results in the previous studies, as they focused almost exclusively on the relation between information supply and information asymmetry.

Since information acquisition and information asymmetry can be endogenously determined, it poses an empirical challenge for causal interpretation based on the OLS analysis. An ideal setup would be to have an exogenous shock to information acquisition. Accordingly, I use the Northeast blackout of 2003 as a natural experiment and identify the causal effect of information acquisition on firm-level information asymmetry in a difference-in-differences setting. On August 14, 2003, there was a widespread unanticipated power outage throughout parts of the Northeastern and Midwestern United States and the Canadian province of Ontario, beginning just after 4:10 p.m. Although Wall Street got its power back at 6 a.m. on the 15th, most traders were unable to commute due to the train system shutdown, and a lot of infrastructure was suffered from power outage, which limits investors from the affected regions to get access to firm filings during the blackout.[7] Even though the geographical distribution of information acquisition before the shock is endogenous, the power outage will affect firms differently. Ex-ante, we expect firms with a larger fraction of historical downloads from the affected regions to incur a larger reduction in downloads during the blackout. Using a difference-in-differences estimation on a sample of firms with material information disclosure one day before the blackout, I show that firms with an additional 1% increase in historical 8-K downloads from the affected regions before the shock suffer a 3% increase in information asymmetry due to the blackout. Moreover, I show that such an effect is not driven by either different trading volumes between two types of firms, or different sophistication levels of investors due to the blackout.

To lend further support to the theoretical predictions under the information asymmetry

[7] "Nobody is here. It's dead. It's a complete ghost town. It really is like coming in on the weekend," said Richard Koss, bond portfolio manager at Brown Brothers Harriman. From CNN news.

channel, I explore the heterogeneity in investor and firm types, and examine their differential effects of material information acquisition on stock returns. First, I find that the effect of material information acquisition is stronger for firms with a higher fraction of *local* investors who are geographically close to firm headquarters, and with a higher fraction of *recurring* investors who have prior experience in collecting information. These investors typically have a lower cost when acquiring and processing information, and the effect of their information acquisition should be stronger (Verrecchia (1982)). Second, I show that the effect of material information acquisition is stronger for firms with higher ex-ante information asymmetry, such as small firms, illiquid firms, and firms with large dispersion in analyst forecasting errors. In other words, information acquisition has more potential to reduce information asymmetry when insiders typically possess more material information. For example, the low media and analyst coverage makes it more difficult for small firms to disseminate information, resulting in a high cost of external financing. Therefore, small firms and their investors rely more heavily on the EDGAR platform, and timely information acquisition has larger effects on stock returns.

Information acquisition occurs not only when investors trade on time-sensitive material information, but also when investors are attracted by attention-grabbing events. When the latter case happens, investors are more likely to buy rather than sell attention-grabbing stocks, given the short-selling constraint faced by certain investors. As a result, information acquisition captures increased attention and predicts positive abnormal returns in the future. I use the number of 10-K downloads as a proxy for *stale* information acquisition under the behavioral channel for the following reasons: First, Form 10-K provides investors with comprehensive financial and operational statements, which are useful for fundamental investment.[8] Second, unlike 8-K filings, 10-K filings have a significant reporting lag between fiscal-end date and disclosure date.[9] The long reporting lag of 10-Ks discourages investors

[8]For example, Deaves, Dine, and Horton (2006) surveys 1,600 retail investors and finds that the majority of shareholders read and use financial statements to make investment decisions.

[9]Starting in 2003, the filing deadline for 10-Ks is 75 days for accelerated filers and 90 days for non-accelerated filers. The average report lag in my sample is 81 days.

who trade on time-sensitive information since most information in 10-Ks is released to the market either through earnings announcements or previous 8-K filings. I also check that the 10-K downloads are spread out throughout the year after its release, thereby validating the argument that 10-K downloads capture stale information acquisition. As a result, demand for 10-K is expected to affect stock returns through the behavioral channel, rather than the information asymmetry channel that requires timely processing of material information.

I show that, in stark contrast to material information acquisition, stale information acquisition predicts higher stock returns, consistent with the behavioral channel. Specifically, the long/short portfolio sorted on 10-K downloads earns an alpha of 40 bps in the first holding month but reverses to zero over the next year. The alpha decay pattern aligns with previous findings in the literature. For example, Gervais et al. (2001) shows that stocks with high abnormal trading volumes have high visibility and outperform over the next 100 days. Chen, Hong, and Stein (2002) shows that there is a strong autocorrelation in investor recognition, which is measured by the breadth of institutional ownership. Stocks with increasing investor recognition outperform in the first twelve months after portfolio formation. Gargano and Rossi (2018) uses individual brokerage account data and shows that investors buy attention-grabbing stocks whose positive performance persists for six months.

I further show that the distinctive predictions between 8-K downloads and 10-K downloads on stock returns are driven by the timeliness of the information. Given the low disclosure frequency of 10-K filing, the majority of 10-K downloads happened far away from the releasing date. As a result, demand for 10-K is dominated by the behavioral channel most of the time. When I limit the sample to firms with the 10-K release, demand for newly filed 10-K in the given week predicts a lower return going forward, similar to the prediction of 8-K downloads. In other words, the effect of 10-K downloads flips from the behavioral channel to the information asymmetry channel if information acquisition happens during a time window close to disclosure.

My paper contributes to the literature in several ways. First of all, it is the first paper to empirically test the role of information acquisition in reducing information asymmetry, examine its effect on stock returns, and explore its heterogeneous effect through the cost of information acquisition channel, which complements a stream of theoretical work starting from Grossman and Stiglitz (1976). Prior literature tests whether information asymmetry affects stock returns (Easley, Hvidkjaer, and O'Hara (2002), Kelly and Ljungqvist (2012)). My paper takes a step further by directly testing how information acquisition affects returns through the information asymmetry channel.

Second, my paper informs the debate on whether the supply of information generally reduces information asymmetry. For example, Bushee et al. (2010) finds that press coverage reduces firm-level information asymmetry. Amiram, Owens, and Rozenbaum (2016) finds that analyst forecast disclosure reduces information asymmetry during the announcement period. In contrast, Coller and Yohn (1997) shows that earnings announcements and management forecasts increase information asymmetry. I show that the supply of information has no direct effect on information asymmetry, but instead acts as a catalyst through which demand for information reduces information asymmetry. Therefore, the seemingly contradicting evidence regarding the effect of information supply on information asymmetry is likely driven by the omitted demand for information.

Lastly, my paper contributes to the recently growing literature on investor attention by highlighting the different aspects of information acquisition on stock returns. Prior literature has focused on the behavioral channel by examining Google Trends (Da et al. (2011)) or Bloomberg Search Index (Ben-Rephael et al. (2017)). These aggregated measures cannot differentiate sources of information so that the behavioral channel overcasts the information asymmetry channel. I show that acquiring unanticipated material information reduces information asymmetry and lowers stock returns, whereas collecting stale information reflects investors' increased attention and predicts higher abnormal returns. Only by taking different types of information acquisition into account can we distinguish between the behavioral channel and the information asymmetry channel.

The paper proceeds as follows. Section 2 discusses the sample selection and data collection. Section 3 jointly tests the information asymmetry channel and the behavioral channel. Section 4 discusses the mechanisms of material information acquisition. Section 5 shows the differential effects of material information acquisition conditional on ex-ante information asymmetry, cost of information acquisition, and information content. Section 6 discusses how stale information acquisition affects returns. Section 7 concludes.

2 Data and Sample Selection

The paper uses data from several sources. I use the CRSP, Compustat, I/B/E/S, and TAQ databases to obtain stock related information, the Thompson Reuters database to obtain institutional ownership data, the EDGAR server log to obtain daily log of page requests for SEC filings[10], and the EDGAR Master File to obtain filing type and date. To control for media coverage, I use Ravenpack news data. Ravenpack news data provide news coverage for a large sample of public companies[11]. I also control for Google Trends and Bloomberg News Heat Index. Google Trends data provide within-firm daily Google search volume index and are often used to capture retail investors' attention. Bloomberg News Heat Index captures the news search volume by Bloomberg users and is used to capture institutional investors' attention, which is available starting from 2010/02/17.

The sample starts in 2003 when the EDGAR log data became available and ends in 2016. I use all domestic equity stocks with share code 10 or 11. I require stocks with a valid market value at month-end in the CRSP, valid financial statement data in Compustat, and valid earning announcement data in I/B/E/S. I also require that stocks in the CRSP have matched identifiers in the SEC EDGAR database. The matched sample has 5,989 unique stocks. After merging with Ravenpack and Google Trends data, the sample reduces to 4,106 unique stocks, where most of the sample loss occurs for microcap stocks. For the main

[10]I use the link file provided by WRDS to link stock identifiers "permno" in CRSP and "cik" in SEC.
[11]I match Ravenpack data with the CRSP data using 8-digit CUSIP, ticker symbol, and company names.

analysis, I use the full sample. All my results are robust when using the smaller sample.

2.1 The EDGAR Server Log

The EDGAR log is publicly available and can be obtained from its website. The data contain daily log files from 2003 forward. The log file contains the timestamps of page requests, the firm identifier, the filing accession number, the IP address of the request[12], the index page flag[13], server status code[14], the crawler flag, and so on. Log files between September 24, 2005, and May 10, 2006, were labeled by the SEC as "lost or damaged", and are excluded from the empirical analysis. Some users employ automated programs to crawl SEC filings, but not all crawling activities are flagged by the EDGAR. Following Lee, Ma, and Wang (2015), I label an IP address as a crawler if it is associated with more than 50 daily requests.

The sample starts with over 21.89 billion records. Following Lee et al. (2015), I first reduce the sample to 9.84 billion records by excluding requests with the index page flag or failed connection, since these requests do no capture any information acquisition activities. I then link the Central Key Index (CIK) provided by EDGAR with the stock identifier in CRSP. After the merge, the sample reduces to 3.36 billion records. I further reduce the sample by focusing on filings of the following three types, Forms 10-K, 10-Q, and 8-K, which leaves me with 1.36 billion records. Finally, I get the physical locations and service providers of IP addresses in the record.

2.2 Overview of EDGAR Downloads

Figure 1 shows the monthly aggregated downloads in my sample. I separate crawling activities ("robots") from human viewing activities ("human"). Figure 1a shows the plot for all filing types. There has been an increasing trend for viewing activities on EDGAR. The

[12]Only the first three octets of the IP address are available, and the last octet is replaced with random characters, so that the IP address is uniquely identifiable.

[13]There is an index page containing all documents for a filing. The index page flag indicates that the user simply visits the index page without downloading any documents.

[14]The server status code indicates whether the request is successful, which is typically below 300.

number of human downloads starts at 0.25 million in 2003 and ends at 1.5 million in 2016. The number of crawling requests is about 15 times greater than the number of human downloads. Figures 1b to 1d show the monthly aggregated plot by file types. There is strong seasonality in 10-K and 10-Q downloads driven by the filing cycles. On the contrary, 8-K downloads exhibit weak seasonality, because 8-K filings are triggered by unanticipated material events. In terms of the aggregated magnitude, 10-K downloads account for around a half of all downloads, with the remaining half split by 8-K and 10-Q downloads.

3 The Two Channels of Information Acquisition

In this section, I use three approaches to test the effects of information acquisition on returns through the behavioral channel and the information asymmetry channel. First, I run monthly Fama-Macbeth (1973) cross-sectional regression of returns on heterogeneous measures of information acquisition. Second, I use a non-parametric approach by forming long/short portfolios and regressing portfolio returns on factors. Third, I use weekly level data to provide additional robustness check for the results.

3.1 Fama-Macbeth (1973) Approach

I first study the relation between future stock returns and the overall demand for information. I run Fama-Macbeth (1973) regression of monthly individual stock returns from month $t+1$ on information acquisition variables from month t.

All regressions control for the following characteristics. For firms' fundamental variables, I include *Asset Growth*, *log(BM)*, *log(ME)*, and *Operating Profit*. *Asset Growth* is the annual growth rate of assets; *log(BM)* is the natural logarithm of the book-to-market ratio; *log(ME)* is the natural logarithm of the firm market capitalization; *Operating Profit* is the ratio of operating profits to book equity. I include the current month stock return $r_{1,0}$ and the past-12 month stock return $r_{12,2}$ to control for firms' past performance, which may drive both investor demand and future returns. Gervais et al. (2001) and Barber and Odean (2007) document that abnormal trading volume increases a firm's visibility, which could affect both demand and future stock returns. Therefore, I include *Abnormal Trading Volume*, which is the difference between monthly trading volume and past 12-month average trading volume, scaled by the standard deviation of the past 12-month trading volume. Since many of my information acquisition variables capture information acquisition of firms' annual and quarterly filings, I include earnings surprise and earnings drift from the most recent earnings announcement. *SUE* is the unexpected quarterly earnings scaled by market cap; *Earning Drift* is the sum of daily returns in three days around earning announcement. Lastly, I control for firm disclosure. *file 8K*, *file 10K*, and *file 10Q* are the numbers of Form 8-K, 10-K, and 10-Q filed on the EDGAR in the given month, respectively. Column (1) of Table 1 shows the baseline result. Consistent with the previous literature, asset growth, firm size, operating profit, unexpected earnings, abnormal trading volume, and abnormal earnings announcement returns can explain the cross-section of stock returns.

Column (2) of Table 1 shows the effect of aggregated information acquisition on subsequent stock returns. $log\ views_{all}$ is the natural logarithm of all filing downloads of the firm in the current month. The estimate of $log\ views_{all}$ is positive and significant. Firms with high filing downloads earn a premium of roughly 18.3 basis points per month (2.2% per year), which is consistent with prior studies that use direct measures of information acquisition, such as Google Trends, or indirect measures, such as abnormal trading volume and extreme stock returns.

To disentangle the behavioral channel and the information asymmetry channel, I split the aggregated downloads by filing types. Demand for 10-K and 10-Q is more likely to capture the general demand for assets, as forms 10-K and 10-Q provide investors with a comprehensive overview of the firm. Information on the firm's balance sheet is also widely used to make fundamental investing decisions. Moreover, 10-K/Qs are often filed with significant delays, so that the demand for these filings responds to information that is not time-sensitive. Forms

8-K, on the other hand, are generally filed unanticipated and contain information that is privately known by insiders. Under the SEC disclosure regulation, the material information needs to be disclosed within four business days. Therefore, demand for 8-K is more likely to transform the disclosed information into public information and reduces the information asymmetry between insiders and investors.

Column (3) of Table 1 shows the effects of information acquisition through two channels. $log\ views_j$ is the monthly natural logarithm of Form j's downloads. On the one hand, the coefficient estimate of $log\ views_{8K}$ is significantly negative (-12 bps per month), consistent with the prediction of the information asymmetry channel. Demand for 8-K reduces the information asymmetry between insiders and investors, decreases the payoff uncertainty, and leads to a lower risk premium.[15] On the other hand, the coefficient estimate of $log\ views_{10K}$ is significantly positive (39 bps per month), which is consistent with the prediction of the behavioral channel. Moreover, the effect of 8-K demand on stock returns is empirically smaller in magnitude than the effect of 10-K demand. The result explains why using aggregated measures of information acquisition only finds evidence of the behavioral channel, but overlooks the information asymmetry channel. In fact, what we have seen in the previous literature is the overall effect of information acquisition on stock returns, and my paper is the first to document each channel separately. Lastly, the coefficient of $log\ views_{10Q}$ is insignificant, and its magnitude is relatively small. There are two reasons for the result. The correlation between 10-K downloads and 10-Q downloads is 0.91 over the full panel. As a result, 10-Q downloads do not provide additional variation beyond 10-K downloads in explaining stock returns. Moreover, the substance and quality of forms 10-K and 10-Q differ. Forms 10-K are required to be audited, whereas Forms 10-Q are not. In addition, the MD&A section in Forms 10-K are much more detailed than in Forms 10-Q.[16] As a result, Form 10-K is a more reliable source for investment reference than Form 10-Q.

[15] I postpone the analysis for the direct effect of 8-K demand on information asymmetry in Section 4. I mainly focus on the effect of information acquisition on stock returns in this section.

[16] For example, MD&A section of IBM Form 10-K spans 50 pages in 2018 and only 20 pages in 2019Q1.

To further refine the identification of acquiring unanticipated and material information through the information asymmetry channel, I decompose demand for 8-K into two parts: demand for scheduled and unscheduled 8-K. The scheduled 8-K filing includes pre-scheduled event, such as earnings announcements and annual shareholder meetings. These scheduled reports typically contain information that is well anticipated by the market. Therefore, demand for scheduled 8-K should have no predictability under the information asymmetry channel. Unscheduled filings disclose material information that was only known by insiders. It requires investors to timely collect, process, and incorporate the information into the market. Therefore, acquiring unanticipated and material information can reduce the information asymmetry between insiders and investors, leading to a lower risk premium. In Column (4), I show that the return predictability of demand for 8-K is entirely driven by unscheduled 8-K downloads, which further supports the information asymmetry channel.

In Column (5), I control for the change in Google Trends and media coverage. The result is robust, but the sample is smaller than the ones in previous columns.

3.2 Portfolio Sort Approach

The previous section demonstrated the effects of information acquisition through the behavioral channel and the information asymmetry channel using Fama-Macbeth regressions. In this section, I provide additional supporting evidence using a portfolio sort approach. This approach not only is less parametric than the Fama-Macbeth regression but also has the flexibility to show the effects over a longer period. I show that demand for 8-K and demand for 10-K not only predict opposite returns over the subsequent month, but also yield different return paths over the long term.

Since large firms naturally receive more downloads than small firms, it is important to control for firm size when sorting on downloads, especially for the 10-K filings.[17] At each month, I sort stocks into quintiles using NYSE breakpoints. Conditional on each NYSE

[17]The correlation between the log of 10-K downloads and firm size is 0.56, and the correlation between the log of 8-K downloads and firm size is 0.34.

quintile, I then sort stocks by the number of downloads into quintiles.[18] Finally, I form a long/short portfolio by buying the top quintile stocks and selling the bottom quintile stocks and regress the monthly portfolio returns on benchmark factors. The factor models include CAPM, Fama-French three-factor (FF3), Fama-French-Carhart (FFC), Fama-French five-factor plus momentum (FF5+UMD), and an eight-factor model by including betting-against-beta and liquidity factors.

Table 2 shows the univariate portfolio sort result for 10-K and 8-K downloads. The results are consistent with the ones using the Fama-Macbeth regressions. In the short term, demand for 10-K predicts positive abnormal returns, whereas demand for 8-K predicts negative abnormal returns. In particular, the long/short 10-K portfolio yields a monthly return of 0.81%. After controlling for common pricing factors, the average alpha is around 0.42% per month. Moreover, the effect of 10-K demand is short-lived, which can be seen from the insignificant alphas with three or twelve holding months. The long/short 8-K portfolio, on the contrary, earns a monthly alpha of -0.48%, consistent with the short-term prediction of the information asymmetry channel. Moreover, the effect of 8-K demand is long-lasting, averaging -0.5% per month for the next 12 months.[19]

Since I use size-adjusted downloads as the sorting variables, it is interesting to see how 10-K and 8-K portfolios perform under different size groups. At each month, I first sort stocks by their previous month market capitalization into quintiles. Conditional on each size quintile, I then sort stocks by the 10-K (8-K) downloads into quintiles and form the long/short portfolio by buying the top quintile stocks and selling the bottom quintile stocks.

Figure 3 shows the result. Portfolios sorted on size-adjusted 10-K downloads yield positive

[18] The results are robust using other approaches to control for firm size. In the previous version, I first run a cross-sectional regression of the natural logarithm of downloads on the natural logarithm of lag firm size and extract regression residuals as the size-adjusted demand for information. I then sort stocks into quintiles by the size-adjusted demand. Such procedure is also used in Nagel (2005). Using NYSE breakpoints is less-parametric.

[19] Table A2 separate the robot crawling activity into retail and institutional crawling activities. Institutional crawling for unscheduled 8-K filings predicts lower subsequent risk premiums, whereas retail crawling has no predictability. The result is consistent with the increased popularity of algorithm trading employed by institutional investors.

and significant alphas across all size quintiles. The result is the strongest in the small size quintile, yielding 1.2% alpha per month. The magnitude of the alpha decreases with firm size. Both the liquidity and the short-selling constraint contribute to the result. Small firms are more illiquid than large firms. When facing a demand shock, small stocks face a larger trading pressure than large stocks. Moreover, small stocks have a tighter short-selling constraint than large stocks, which limits the potential arbitrage opportunities and results in a large price increase.

Portfolios sorted on size-adjusted 8-K downloads yield negative alphas across all size quintiles, but alphas are significant for the bottom three size quintiles and insignificant for large stocks. For example, the 8-K portfolio yields an average alpha of -40 bps per month for the bottom three size quintiles, and -10 bps for the top two quintiles. This result is consistent with the information asymmetry hypothesis. Small firms have less media/analyst coverage and institutional holding than large firms. Investors of small firms face a higher degree of information asymmetry and rely more on themselves in processing and incorporating the disclosed information. Moreover, for large firms, there are more channels to disseminate information, so that the downloading activities on the EDGAR may not necessarily represent a significant portion of information acquisition. Therefore, demand for 8-K filings has a stronger effect in small firms than in large ones.

3.3 Weekly Frequency Result

To zoom in the supply and demand side of information, I conduct the analyses at a weekly frequency. Forms 10-K are filed once a year in general. Forms 8-K are filed irregularly, but once a month on average. Therefore, the weekly level analysis allows me to study the interaction between the supply and the demand for information and their effects on prices.

I aggregate the daily stock returns and daily downloads to weekly frequency (Friday close to Friday close). My main variable of interest is $\log(views_t^k)$, which is the natural logarithm of total downloads of filing type k in week t. I then create a set of dummies to capture the information supply. The dummy variable $Filing\ k_t$ is equal to one if the firm has issued filings of type k in week t. The dummy variable $News_t$ is equal to one if the firm appears in the Ravenpack news database in week t. The dummy variable $Earnings\ Release_t$ is equal to one if the firm releases its earnings in week t. For a subset of the analysis[20], I also control for the Bloomberg search index and Google Trends, which capture the institutional and retail demand studied in the previous literature (Ben-Rephael et al. (2017), Da et al. (2011)). The dummy variable AIA_t is equal to one if the Bloomberg News Heat daily index has a maximum of 3 or above in week t. The dummy variable $DADSVI_t$ is equal to one if the Google Trends daily index in any day of the week is above its 90 percentile in the past month.

Table 3 shows the weekly regression result. I regress the weekly stock returns on demand for filings, controlling for the supply of firm filings, media coverage, earnings announcements, firm characteristics, lag returns, and time fixed effects. To capture the interaction between supply and demand of information under the information asymmetry channel, I add an interaction term between the supply and demand of 8-K filings. Columns (1) and (3) study the contemporaneous relation between stock returns and demand for information, where the dependent variable is the current week stock returns. The dependent variables in Columns (2) and (4) are stock returns in the subsequent week.

As shown in Columns (1) and (2) of Table 3, the coefficient estimates of $\log(views_t^{10K})$ are all positive and significant, consistent with the behavioral channel well documented in the existing literature. The coefficient estimate of the interaction term between 8-K supply and demand is positive and significant in Column (1), and is negative and significant in the Column (2). The results have not been documented empirically, and they support the information asymmetry channel. When firms release new information through 8-K filings, acquiring information reduces the information asymmetry between insiders and investors, and stocks become less risky to hold. Therefore, the contemporaneous price increases, and

[20]Bloomberg News Heat index is only available after 2010/02/17.

future risk premium decreases. Both the behavioral channel and the information asymmetry channel are robust after controlling for the Bloomberg and Google Trends search indexes, which are shown in Columns (3) and (4).

The two channels of information acquisition not only have opposite predictions in the short term, but also suggest distinct patterns in the long term. For the behavioral channel, we should see the strongest evidence of positive contemporaneous return spread, followed by an alpha decay pattern. The speed of the alpha decay process relies on the liquidity of the underlying asset and the time lag between information acquisition and investment decision. For the information asymmetry channel, although the contemporaneous price also increases, the underlying mechanism is completely different. The reduced risk premium in the future drives up the contemporaneous price, followed by a negative and persistent return spread in the future.

Figure 4 shows long-term return patterns of both channels. At the end of each week, I first sort stocks by size into five groups using NYSE breakpoints. Conditional on each NYSE size-group, I sort stocks by the weekly 10-K and 8-K downloads into quintiles and form long/short portfolios. Portfolios are held throughout the next 24 weeks. The alphas of portfolios at each holding week are plotted. For 8-K portfolios, I limit the set of stocks that filed 8-K filings in the week, as the evidence suggested in Table 3 shows that the effect of 8-K demand is stronger, conditional on the supply of information.

The results do not imply that acquiring information from 10-K does not affect information asymmetry. It merely states that the behavioral channel dominates for the annual filing 10-K, and it is hard to empirically disentangle the two channels because of the low disclosing frequency. To test whether demand for newly disclosed 10-K filings reduces information asymmetry, I limit my sample to a set of stocks that just disclosed 10-Ks in a week. I then sort these stocks on the size-adjusted downloads of the newly issued 10-K filings into quintiles. The result is plotted in Figure 5. Conditional on firms just issued 10-Ks in week 0, firms with higher downloads of newly released 10-Ks yield higher returns in the contemporaneous week, and lower returns in the upcoming weeks. The effect of 10-K demand flips to the information asymmetry channel in this small subset. The long-term return pattern is comparable to the one found in 8-K filings. However, the alpha is noisily estimated, since only a small portion of firms file 10-K in a given week.

4 Mechanisms of 8-K Demand on Stock Returns

Easley and O'Hara (2004) documents that investors demand higher returns for stocks with more private information. Boot and Thakor (2001) suggests that disclosing information that is only known to informed investors decreases the information advantage informed investors have over the uninformed. However, little research has shown the effect of information acquisition on information asymmetry, as past literature almost exclusively focuses on the supply side. In this section, I show that investors' 8-K demand decreases the firm-level information asymmetry. As a result, the stock becomes less risky for uninformed investors to hold, and the risk premium decreases.

4.1 Panel Regressions of Information Asymmetry

To test whether demand for information reduces the information asymmetry, I run a weekly panel regression of future information asymmetry on information acquisition, controlling for firm characteristics and information disclosure. The result is shown in Table 4. I use the price impact measure estimated following Holden and Jacobsen (2014) as a proxy for information asymmetry of the firm.[21] For a given stock, the price impact on the k^{th} trade is defined as

$$Price\ Impact_k = \frac{2D_k(M_{k+5} - M_k)}{M_k}, \tag{1}$$

[21] I have also used alternative information asymmetry measures, such as Amihud illiquidity measure, bid-ask spread, or PINs. The results are similar and available upon request.

where M_{k+5} is the midpoint five minutes after the midpoint M_k, and D_k is the buy-sell indicator of the trade. The price impact measure captures the permanent component of the effective spread and is widely used in the microstructure literature to proxy the firm-level information asymmetry. In column (1), the coefficient estimate of $\textit{log views}_{8K}$ is negative and significant, suggesting that higher 8-K downloads are associated with lower information asymmetry in the next week. The supply of 8-K also reduces the information asymmetry, as can be seen by the negative coefficient estimate of *Filing 8K*. However, once we interact the demand and supply of 8-K filing, the supply of 8-K does not have any significance, which is shown in column (2). The interaction term between 8-K demand and supply is negative and significant, showing that the demand for 8-K has a stronger effect on reducing information asymmetry, conditional on new information arrivals. The economic magnitude is also large. A 10% increase in 8-K downloads leads to a nine bps reduction in the price impact measure, which has an average of 45 bps.

Consistent with the mechanism, the effect of 8-K downloads on stock returns should be larger when the ex-ante information asymmetry is higher. The previous section (Figure 3) shows that the effect is indeed larger for small stocks than large stocks, as the market capitalization is one of the most important proxies for the information asymmetry. I also use the Amihud illiquidity measure and previous quarter analyst forecast dispersion to proxy for ex-ante information asymmetry for additional robustness checks. I first sort stocks by these measures into terciles. Conditional on each tercile, I sort stocks by 8-K downloads within the NYSE size-group into quintiles. Table 5 shows the portfolio double-sort results for 8-K downloads and the ex-ante information asymmetry. When Amihud measure is low, the alpha of long/short 8-K downloads portfolio is -11 bps per month. When Amihud measure is high, the magnitude of alpha increases to -53 bps per month. The results are similar using the analyst forecast dispersion measure.

4.2 Natural Experiment: Northeast Blackout of 2003

Given the endogenous nature of information acquisition and information asymmetry, there could be potential alternative explanations for the mechanism shown in the previous section. For example, it could be the reverse causality explanation. Investors may be aware of firms with high information asymmetry. Upon releasing of new information, firms with higher information asymmetry attracts more downloads, and the supply of information reduces the information asymmetry. Then we may observe a spurious relation between information acquisition and information asymmetry. To rule out such endogeneity concern, we need a shock to information acquisition and the shock has to be orthogonal to the supply of information. Therefore, in this section, I test how demand for 8-K filings reduces information asymmetry through a natural experiment setting, using the Northeast blackout of 2003. On August 14, 2003 (Thursday), there was a widespread power outage throughout parts of the Northeastern and Midwestern United States, beginning just after 4:10 p.m. EDT. For the next 30 minutes, outages were reported in parts of Ohio, New York, and New Jersey. Major cities include New York City, Toronto, Baltimore, and Detroit. Manhattan, including Wall Street, was completely shut down. Although Wall Street got its power back at 6 a.m. on the 15th, most traders were not able to commute due to the shut down of train system, and lots of infrastructure was suffered from the power outage. Shares were lightly traded on Friday, and the NYSE ended up with just under 624 million shares traded. Figure 6 shows the EDGAR 8-K hourly downloading traffic by investors from the affected and non-affected regions. The number of hourly downloads on August 15 fell a lot, especially for the affected regions.

The Northeast blackout provides an opportunity for a natural experiment. Since demand for 8-K is an endogenous choice, the ideal experiment would be to have an exogenous shock to the endogenous choice. In a simplistic world, suppose we have only two types of firms that just disclosed unanticipated information. Before the shock arrives, type A firms have a higher proportion of 8-K downloads from the affected regions than type B firms. The geographical distribution of attention allocation before the shock can be completely endogenous. When the shock arrives, it (ideally would) shuts down all information acquisition from the affected regions, and investors from non-affected regions are free to acquire information. Such a shock impacts the information acquisition of investors for two firms differently. Investors who would have acquired information on type A firms during the blackout cannot do so and stay uninformed during the blackout, which leaves the firm-level information asymmetry at a relatively high level.

The difference-in-differences estimation is a natural fit for this problem. In this setting, I have two periods: August 14, 2003, and August 15, 2003. I first calculate the fraction of historical 8-K downloads from the affected regions, denoted as *frac*, for each firm before August 13, 2003. I then limit my sample to firms that just disclosed material information on August 13, so that information acquisition is crucially important for both types of firms. I also limit my sample to firms with headquarters outside the affected regions, which alleviates the concern that firms from the affected regions may suffer additional economic disruption than firms outside the affected regions. I define a firm as treated if the fraction *frac* is above the median. The regression specifications are the following:

$$
\begin{aligned}
\textit{Price Impact}_{i,t} &= \beta_0 + \beta_1 \textit{treated}_i + \beta_2 \textit{post}_t + \beta_3 \textit{treated}_i \times \textit{post}_t + \epsilon_{i,t}, \\
\textit{Price Impact}_{i,t} &= \beta_0 + \beta_1 \textit{frac}_i + \beta_2 \textit{post}_t + \beta_3 \textit{frac}_i \times \textit{post}_t + \epsilon_{i,t},
\end{aligned}
$$

where *Post* is equal to one if it is the August 15, 2003, and zero otherwise. The first regression specification considers the binary case, whereas the second specification considers the continuous treatment effect. The dependent variable is the daily *Price Impact* measure. The regression results are presented in Table 6. The coefficient on the interaction term is positive and significant, suggesting that the reduction of information acquisition increases the firm-level information asymmetry. In terms of the magnitude, firms with 1% more downloads from the affected regions before the shock suffer a 3% increase in information asymmetry after the blackout. To alleviate the concern that headquarter locations may not perfectly control for business activities, I get the frequency of state name appearances from the firm 10-K filings as a proxy for state-level operation intensity, and exclude firms with top state name appearances in the affected regions. The results do not change in the smaller sample.

I also test and rule out the following alternative explanations. First, it could be that traders of firms with a higher fraction of historical downloads from affected regions cannot trade as freely as traders of firms with a lower fraction. Second, it could be that the blackout creates a selection among traders, so that the level of sophistication between two types of firms has changed. In table 7, I use trading volumes normalized by the number of shares outstanding to rule out the first explanation, since there is no differential effect in trading volumes among firms due to the blackout. I also use the number of trades after the market close to capture trader's sophistication level. There is no evidence to suggest that the composition of sophisticated traders has changed.

5 Heterogeneous Effect of 8-K Demand

This section shows heterogeneous effects of information acquisition through the information asymmetry channel. Specifically, I examine how the effect varies with the cost of information acquisition and the information content.

5.1 Cost of Information Acquisition

The cost of information acquisition plays an important role in reducing information asymmetry. In Verrecchia (1982) Corollary 4, the informativeness of price increases when information acquisition cost is reduced. Although I do not directly observe the cost of information acquisition for each investor, an investor's past information acquisition history and his/her geographical location are observed in the data. Moreover, institutional investors tend to have lower cost of information acquisition than retail investors. I use the firm-level fraction of local demand, the fraction of recurring viewers, and abnormal institutional investors' attention to capture the cost of information acquisition.

Local investors have an information advantage over non-local investors in collecting and processing information. Therefore, holding the level of information acquisition fixed, firms with more local demand for information have a lower cost of information acquisition. Moreover, I make the explicit assumption that the cost of information acquisition is lower for an investor who acquired information on the firm in the past quarter than one who did not. Therefore, recurring visitor ratio defined in Figure A4 can be used as a proxy for the cost of information acquisition. The higher the recurring visitor ratio is, the lower the cost of information acquisition.

I also take advantage of the names of IP service providers (ISP) and classify investors' downloads by whether the ISP is from institutions, such as banks and investment companies. I then calculate the fraction of 8-K downloads from institutions. To capture the spike in institutional attention, I create an abnormal institutional attention variable, ab_inst^{8K}, which is the fraction of institutional 8-K downloads normalized by the past 12-month average and standard deviation. The higher the abnormal institutional attention measure is, the lower the cost of information acquisition.

Panel A of Table 8 shows the portfolio double-sort results by 8-K demand and the average distance of viewer location to firms' headquarters. For each stock in each month, I calculate the average distance between IP addresses and firm headquarters for each filing type. I then double sort stocks by the average distance into terciles and by the downloads within the NYSE size-group into quintiles. The effect of 8-K demand is mainly concentrated in the low (-61 bps/month) and medium (-41 bps/month) distance tercile. Moreover, the difference between high and low terciles is statistically significant.

Panel B of Table 8 studies the effect of 8-K demand on prices, conditional on visitors' past visiting patterns. For each firm-month, I calculate the proportion of recurring visitors. I then double sort stocks by the frequency ratio into terciles and by 8-K downloads within the NYSE size-group into quintiles. Portfolios sorted by 8-K demand show significant and negative alphas when downloads are from recurring visitors (-55 bps/month). When the recurring ratio is low, however, the 8-K portfolio yields an insignificant alpha. These results suggest that the effect of information acquisition on stock returns is magnified when investors have a low cost of information acquisition.

Panel C of Table 8 studies the effect of 8-K demand on prices, conditional on the abnormal institution attention. Portfolios sorted by 8-K demand show more negative alphas when there is a spike in institutional investors' attention. For example, conditional on high abnormal institutional attention, the 8-K portfolio yields an alpha of -70 bps in the subsequent month, compared to -49 bps when the abnormal institutional attention is low. The difference is significant at 10% level.

5.2 Information Content

Moreover, the effect of 8-K demand should be a function of the information content provided in the filings. The demand for information only reduces the information asymmetry if the information provided by the firm was previously private. Some filings, such as reports about the pre-scheduled meetings, do not convey any private information. Others, such as material agreement and change of officers, require investors' attention to interpret the information. Therefore, it is important to see how the effect of information acquisition interacts with the information content provided in the filings.

I extract the "event date" and "post date" for each filing and calculate the three-day market excess abnormal return of the firm around both dates[22]. Two measures are then used to quantify the importance of each filing. The first measure is simply the maximum of absolute abnormal returns around event and post dates. This measure captures the market response to the information provided in the filing. If the new information is good (bad) news, the measure is high (low). If the information conveyed in the filing is already anticipated or even well understood by the market, the measure should be small in absolute terms. In my

[22]Starting 2004, the SEC requires firms to disclose any material information within four days of the event. In practice, however, the lag can be more than four days as firms can ask for some additional grace periods.

sample, the measure has a mean of 0.4% and standard deviation of 12%.

The second measure is constructed using textual analysis and machine learning. For each filing i, I build a document classifier based on the past one-year 8-K filings of all firms in my sample. I then compute the document similarity vector between the filing i and all past year filings. The similarity vector represents how similar the pair of documents is. I calculate the expected market response to the filing i as the weighted average of three-day abnormal returns of filings in the past year, with the weight determined by the similarity vector. The expected market response captures what the abnormal return level should be, given the similarity of information content between the filing i and past filings. Lastly, I calculate the difference between the realized market response and the expected market response, and use this "unexpected market response" as a proxy for information importance. The measure has a mean of 0.1% and standard deviation of 10%. The difference between the two measures is that the second measure more clearly captures the shock in information content beyond the part expected by the market.

To see how the effect of 8-K demand varies with the importance of information content of the filing, I double sort stocks by the 8-K downloads and the above two measures. The result is shown in Table 9. In Panel A, the information importance measure is the raw abnormal cumulative return around the event. In Panel B, the information importance measure is the unexpected abnormal return. Both panels yield a similar result. The relation between the effect of 8-K downloads on returns and cumulative abnormal return around event date exhibits a "V-shape". The effect of 8-K demand is concentrated in the low and high abnormal return terciles, and relatively weak in the middle tercile, where the average abnormal return is around zero. When abnormal returns are high (low), firms are likely to have disclosed good (bad) private information. The demand for 8-K filings then plays an important role in interpreting the piece of information and reduce information asymmetry, which leads to a negative spread in future returns, regardless of whether the information itself is good or bad. However, when there is little abnormal return around event/post date, it is likely that the market has already taken into account the information content, which leaves investors with little to learn. As a result, the spread in 8-K demand does not predict future returns.

6 Mechanism of 10-K Demand on Stock Returns

Barber and Odean (2007) documents that attention is a scarce resource, and demand for assets is rooted in the stocks that grab investor attention. When investors make purchasing decisions for a stock, 10-K filings provide the most comprehensive coverage of the operational and financial details of a firm. Therefore, the effect of demand for 10-Ks on stock prices is a byproduct of demand shocks to assets. That is, the demand for asset drives up the demand for 10-K filings and stock prices. As a result, we should expect the effect of 10-K demand on stock prices to be higher for attention-grabbing stocks, where the demand shock is potentially greater.

I use stocks with high abnormal trading volume and high daily absolute returns to proxy for attention-grabbing stocks. Abnormal trading volume and absolute daily returns are constructed as follows,

$$abvol_{i,t} = \frac{vol_{i,t} - \bar{vol}_{i,t-1,t-12}}{std\_vol_{i,t-1,t-12}}, \tag{2}$$

$$max\_dret_{i,t} = \max_{d \in t} \mid ret_{i,t,d} \mid \tag{3}$$

where $\bar{vol}_{i,t-1,t-12}$ and $std\_vol_{i,t-1,t-12}$ are the mean and standard deviation of monthly trading volume during the past 12 month, respectively. $ret_{i,t,d}$ is the daily return of stock i on month t and day d. Gervais et al. (2001) first documents that stocks with abnormally high trading volume earn a return premium in the future. The argument is that shocks to the trading volume of a stock increase its visibility, which draws investor attention and drives up stock prices. Barber and Odean (2007) uses abnormal trading volume and maximum daily return to proxy for attention-grabbing. I first sort stocks by abnormal trading volume (absolute daily returns) into terciles. Conditional on each tercile, I then sort stocks by the size-adjusted 10-K demand into quintiles.

Panel A of Table 10 show the alphas of double-sorted portfolios for abnormal trading volume and 10-K downloads. The last column shows the alphas of long/short 10-K demand portfolios conditional on abnormal volume terciles. For low abnormal trading volume tercile, the spread in alpha is 31 bps per month. The spread in alpha increases to 87 bps per month for stocks in high abnormal trading volume tercile. Panel B shows a similar result using the maximum daily return as a proxy for attention-grabbing. The results suggest that the variation in information acquisition of 10-K better captures the variation in asset demand when stocks are more likely to draw investors' attention.

7 Conclusion

In this paper, I empirically test two channels where the demand for information affects asset prices. On the one hand, acquiring unanticipated material information reduces information asymmetry, which in turn drives up contemporaneous prices, followed by a persistent decline in risk premiums. On the other hand, investors acquire stale information due to attention-grabbing events. Such information acquisition reflects investors' increased attention and predicts positive abnormal returns that decay fast over time. Two channels of information acquisition have opposite predictions of subsequent stock returns, yet the empirical literature so far only provided evidence for the latter channel. My findings point out that it crucially depends on what information is acquired in order to disentangle the two channels.

Empirically documenting how information acquisition reduces information asymmetry has important implications. Firstly, this is the first paper to empirically show that the demand side of information reduces firm-level information asymmetry. The previous literature has focused solely on the supply side of information and find inconclusive evidence. My result informs the debate and suggests that the omitted demand side of information can reconcile the seemingly contradictive findings.

Secondly, information acquisition has a larger effect through the information asymmetry channel for small firms, which typically have lower institutional holdings, analyst coverage, and media exposure than large firms. Therefore, communication through EDGAR plays a more important role for management teams of small firms to deliver their messages to investors. Investors of small firms also rely more heavily on EDGAR to gain insights into the firm's operations. Timely processing the disclosed information reduces the information asymmetry between insiders and investors, leading to a reduction in the cost of capital. Such a reduction is valuable for small firms, as they are the ones facing high financing costs.

Thirdly, my paper has important implications for firms' information dissemination. I find that the effect of acquiring unanticipated information is higher when the cost of information acquisition is lower. The cost of information acquisition is not just about collecting information, but also about processing and interpreting the information. Local investors and recurring investors have such information advantages, and their information acquisition has a higher effect on stock returns than non-local and inexperienced investors. Although firms cannot choose the composition of their investors, they have control over how to efficiently disclose the information and lower the cost of information acquisition by their investors.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of financial markets* 5, 31–56.

Amiram, Dan, Edward Owens, and Oded Rozenbaum, 2016, Do information releases increase or decrease information asymmetry? new evidence from analyst forecast announcements, *Journal of Accounting and Economics* 62, 121–138.

Barber, Brad M, and Terrance 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, 785–818.

Barlevy, Gadi, and Pietro Veronesi, 2000, Information acquisition in financial markets, *The Review of Economic Studies* 67, 79–90.

Barlevy, Gadi, and Pietro Veronesi, 2007, Information acquisition in financial markets: a correction, Technical report.

Bauguess, Scott W, John Cooney, and Kathleen Weiss Hanley, 2018, Investor demand for information in newly issued securities, *Available at SSRN 2379056* .

Ben-Rephael, Azi, Bruce Carlin, Zhi Da, and Ryan Israelsen, 2017, Demand for Information and Asset Pricing, Technical Report w23274, National Bureau of Economic Research, Cambridge, MA.

Bernile, Gennaro, Alok Kumar, and Johan Sulaeman, 2015, Home away from Home: Geography of Information and Local Investors, *The Review of Financial Studies* 28, 2009–2049.

Boot, Arnoud WA, and Anjan V Thakor, 2001, The many faces of information disclosure, *The Review of Financial Studies* 14, 1021–1057.

Brennan, Michael J, and Avanidhar Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of financial economics* 38, 361–381.

Brown, Stephen, and Stephen A Hillegeist, 2007, How disclosure quality affects the level of information asymmetry, *Review of Accounting Studies* 12, 443–477.

Brown, Stephen, Stephen A Hillegeist, and Kin Lo, 2004, Conference calls and information asymmetry, *Journal of Accounting and Economics* 37, 343–366.

Bushee, Brian J, John E Core, Wayne Guay, and Sophia JW Hamm, 2010, The role of the business press as an information intermediary, *Journal of Accounting Research* 48, 1–19.

Chen, Huaizhi, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher Malloy, 2018, IQ from IP: Simplifying Search in Portfolio Choice, Working Paper 24801, National Bureau of Economic Research.

Chen, Joseph, Harrison Hong, and Jeremy C Stein, 2002, Breadth of ownership and stock returns, *Journal of financial Economics* 66, 171–205.

Coller, Maribeth, and Teri Lombardi Yohn, 1997, Management forecasts and information asymmetry: An examination of bid-ask spreads, *Journal of accounting research* 35, 181–191.

Corwin, Shane A, and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *The Journal of Finance* 67, 719–760.

Coval, Joshua D., and Tobias J. Moskowitz, 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *The Journal of Finance* 54, 2045–2073.

Coval, Joshua D., and Tobias J. Moskowitz, 2001, The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy* 109, 811–841.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In Search of Attention, *The Journal of Finance* 66, 1461–1499.

Deaves, Richard, Catherine Dine, and William Horton, 2006, How are investment decisions made, *Task Force to Modernize Securities Legislation in Canada* .

Duarte, Jefferson, and Lance Young, 2009, Why is pin priced?, *Journal of Financial Economics* 91, 119–138.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *The journal of finance* 57, 2185–2221.

Easley, David, Nicholas M Kiefer, and Maureen O'Hara, 1997, One day in the life of a very common stock, *The Review of Financial Studies* 10, 805–835.

Easley, David, and Maureen O'Hara, 2004, Information and the Cost of Capital, *The Journal of Finance* 59, 1553–1583.

Epstein, Larry G., and Martin Schneider, 2008, Ambiguity, Information Quality, and Asset Pricing, *The Journal of Finance* 63, 197–228.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *The Journal of Finance* 64, 2023–2052.

Gargano, Antonio, and Alberto G Rossi, 2018, Does it pay to pay attention?, *The Review of Financial Studies* 31, 4595–4649.

Gervais, Simon, Ron Kaniel, and Dan H Mingelgrin, 2001, The high-volume return premium, *The Journal of Finance* 56, 877–919.

Gibbons, Brian, Peter Iliev, and Jonathan Kalodimos, 2019, Analyst information acquisition via edgar, *Available at SSRN 3112761* .

Grossman, Sanford J, and Joseph E Stiglitz, 1976, Information and competitive price systems, *The American Economic Review* 246–253.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American economic review* 70, 393–408.

Hellwig, Martin F, 1980, On the aggregation of information in competitive markets, *Journal of economic theory* 22, 477–498.

Holden, Craig W, and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *The Journal of Finance* 69, 1747–1785.

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of finance* 54, 2143–2184.

Huberman, Gur, and Tomer Regev, 2001, Contagious speculation and a cure for cancer: A nonevent that made stock prices soar, *The Journal of Finance* 56, 387–396.

Hwang, Byoung-Hyoun, and Hugh Hoikwang Kim, 2017, It pays to write well, *Journal of Financial Economics* 124, 373–394.

Juergens, Jennifer L, and Laura Lindsey, 2009, Getting out early: an analysis of market making activity at the recommending analyst's firm, *The Journal of Finance* 64, 2327–2359.

Kelly, Bryan, and Alexander Ljungqvist, 2012, Testing asymmetric-information asset pricing models, *The Review of Financial Studies* 25, 1366–1413.

Kim, Oliver, and Robert E Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of accounting and economics* 17, 41–67.

Lawrence, Alastair, 2013, Individual investors and financial disclosure, *Journal of Accounting and Economics* 56, 130–147.

Lee, Charles MC, Paul Ma, and Charles CY Wang, 2015, Search-based peer firms: Aggregating investor perceptions through internet co-searches, *Journal of Financial Economics* 116, 410–431.

Lee, Charles MC, Belinda Mucklow, and Mark J Ready, 1993, Spreads, depths, and the impact of earnings information: An intraday analysis, *The Review of Financial Studies* 6, 345–374.

Lee, Charles MC, and Mark J Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Loughran, Tim, and Bill McDonald, 2014, Information decay and financial disclosures, Technical report, working paper, University of Notre Dame.

Loughran, Tim, and Bill McDonald, 2017, The Use of EDGAR Filings by Investors, *Journal of Behavioral Finance* 18, 231–248.

Lucca, David O, and Emanuel Moench, 2015, The pre-fomc announcement drift, *The Journal of Finance* 70, 329–371.

Mele, Antonio, and Francesco Sangiorgi, 2015, Uncertainty, information acquisition, and price swings in asset markets, *The Review of Economic Studies* 82, 1533–1567.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of financial economics* 78, 277–309.

Nieuwerburgh, Stijn Van, and Laura Veldkamp, 2009, Information Immobility and the Home Bias Puzzle, *The Journal of Finance* 64, 1187–1215.

Savor, Pavel, and Mungo Wilson, 2013, How Much Do Investors Care About Macroeconomic Risk? Evidence from Scheduled Economic Announcements, *Journal of Financial and Quantitative Analysis* 48, 343–375.

Savor, Pavel, and Mungo Wilson, 2014, Asset pricing: A tale of two days, *Journal of Financial Economics* 113, 171–201.

Veldkamp, Laura L., 2006, Information Markets and the Comovement of Asset Prices, *The Review of Economic Studies* 73, 823–845.

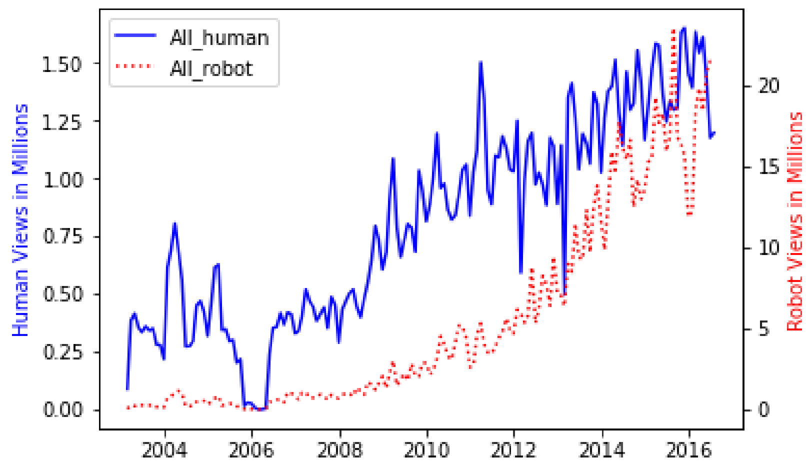
Verrecchia, Robert E, 1982, Information acquisition in a noisy rational expectations economy, *Econometrica: Journal of the Econometric Society* 1415–1430.

Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *The Review of Economic Studies* 60, 249–282.

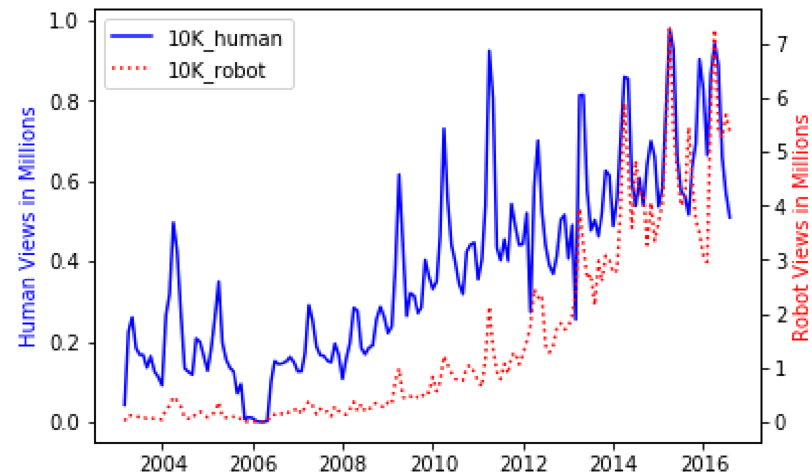
Figure 1

Time-series EDGAR Viewing Activity

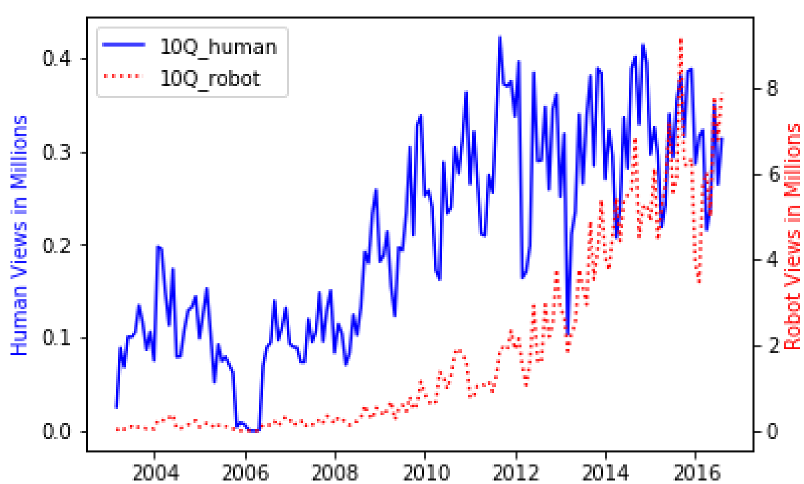
The figure shows the monthly aggregated number of downloads on EDGAR Log system. Following Lee et al. (2015), I separate crawling activities ("robot") from human viewing activities ("human"). Figures (b) to (d) show the number of downloads for 10-K, 10-Q, and 8-K filings, respectively.



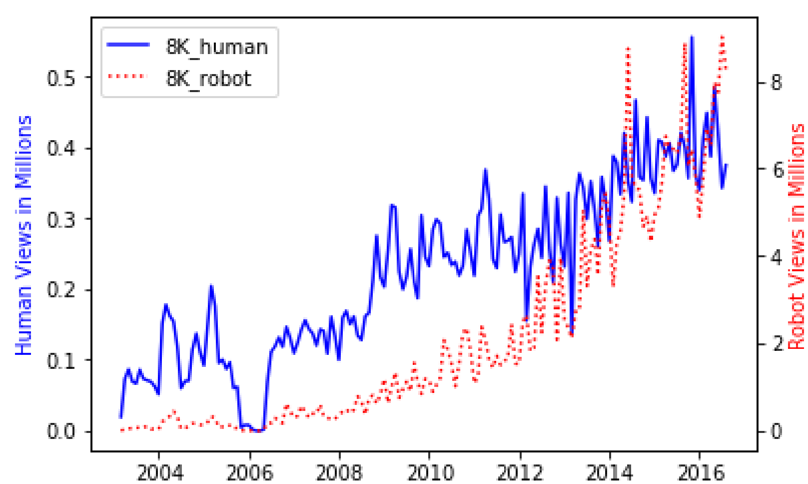
(a) All Filings



(b) 10-K Filings



(c) 10-Q Filings



(d) 8-K Filings

Figure 2

Investors' Demand for Filings Histogram by Firm Sizes

The figure shows the histogram of investor demand for filings on EDGAR, grouped by firm sizes. The horizontal axis is the natural logarithm of monthly filing downloads of a firm. A small firm is defined with a firm market cap below 20% NYSE percentile. A large firm is defined with a firm market cap above 80% NYSE percentile. A medium-size firm is defined with a firm market cap between 20% and 80% NYSE percentile.

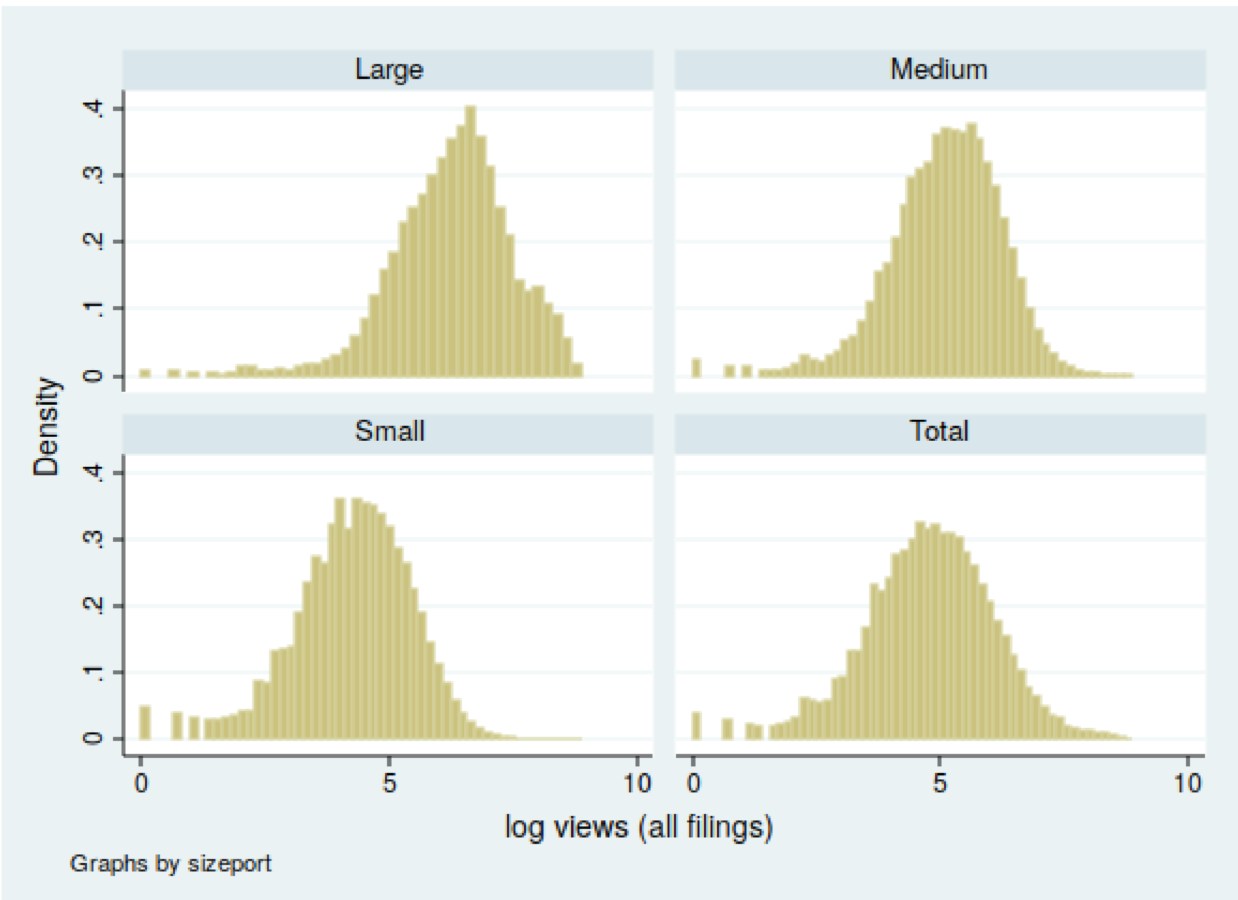


Figure 3
10-K (8-K) Portfolio Alpha and Firm Size

The figure shows the monthly 10-K (8-K) portfolio alpha, conditional on size quintiles. Stocks are sorted by the 10-K (8-K) downloads and the lag firm size into quintiles. Conditional on each size quintile, I form long/short portfolios and regress portfolio return on Fama-French five-factor and UMD. I then plot the average alpha of long/short portfolio for each size quintile, with t-statistics in parenthesis.

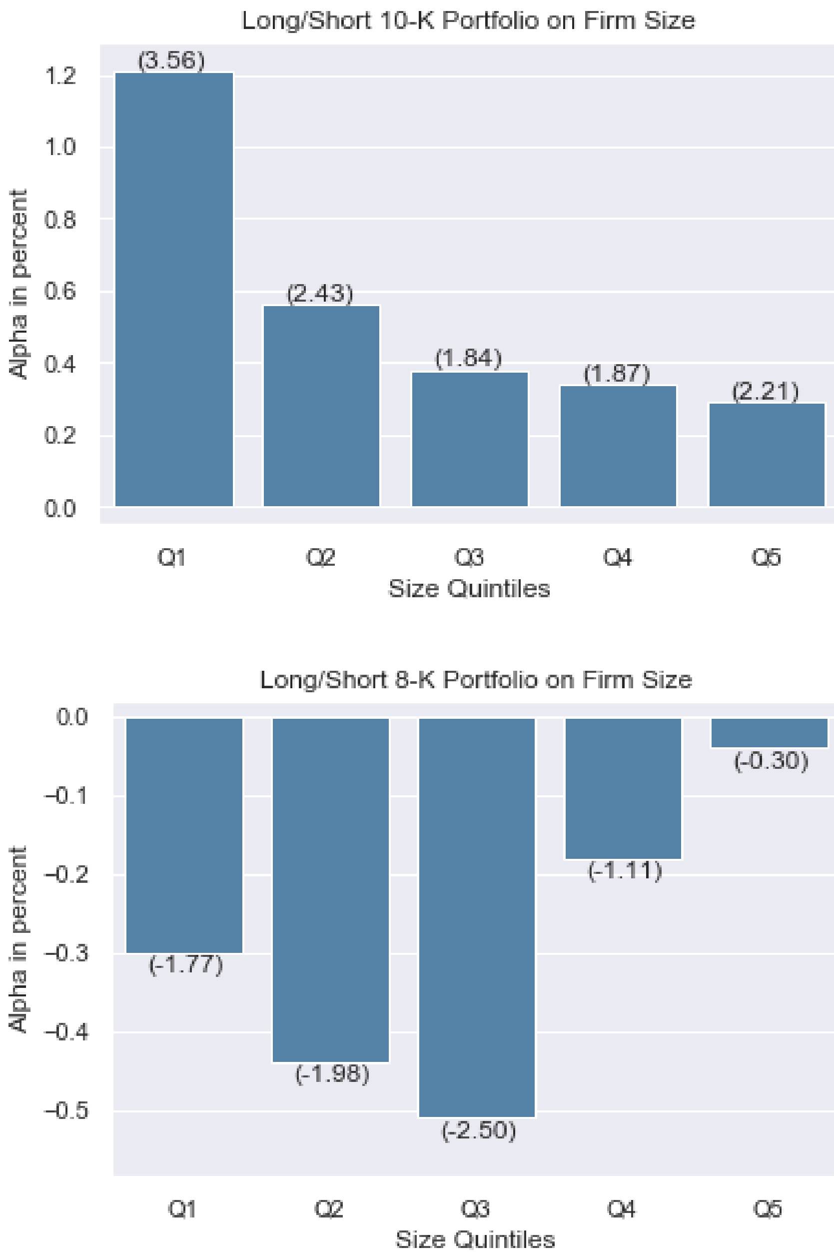
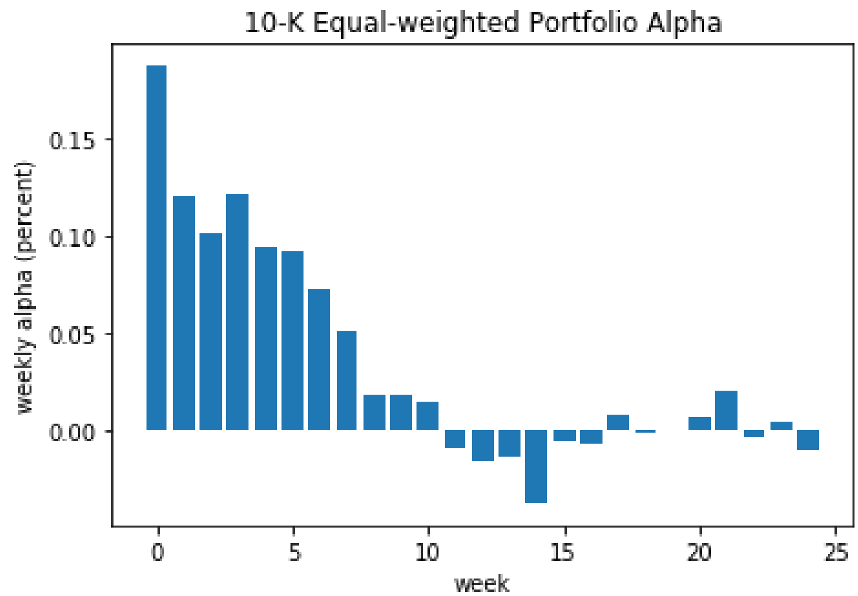


Figure 4
Long/Short 10-K (8-K) Demand Portfolio - Weekly Returns
The figure shows the weekly Fama French 5-factor alphas of 10-K and 8-K portfolios. Stocks are sorted by the weekly downloads at the end of Friday within the NYSE size-group. Long/short portfolios are held throughout the next 24 weeks.



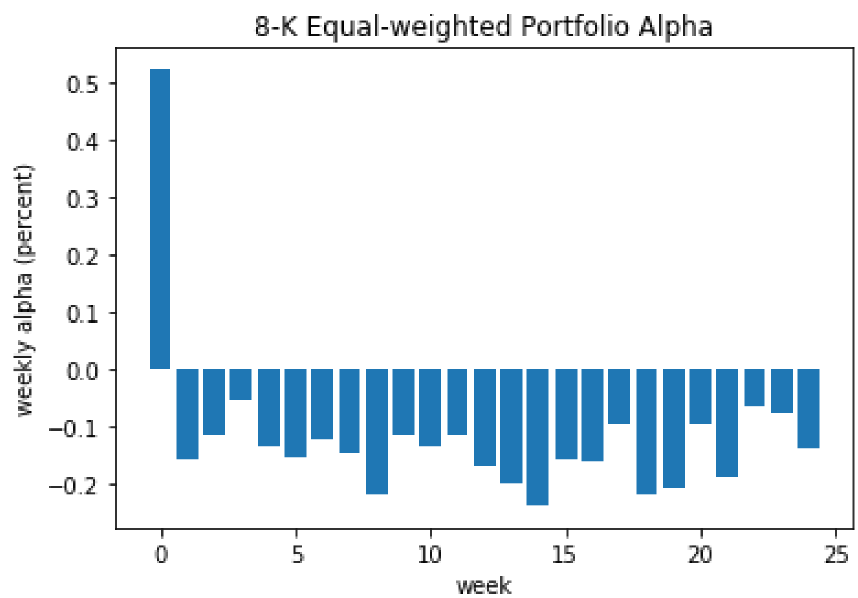


Figure 5

Newly Disclosed 10-K Portfolios

The figure shows the weekly Fama French 5-factor alphas of 10-K portfolios, conditional on a set of firms just disclosed 10-K in a week. At each week, I limit the sample to firms just disclosed 10-K in the week. Stocks are then sorted within the NYSE size-groups by the downloads of the newly disclosed 10-K filing into quintiles. Long/short portfolios are held throughout the next 24 weeks.

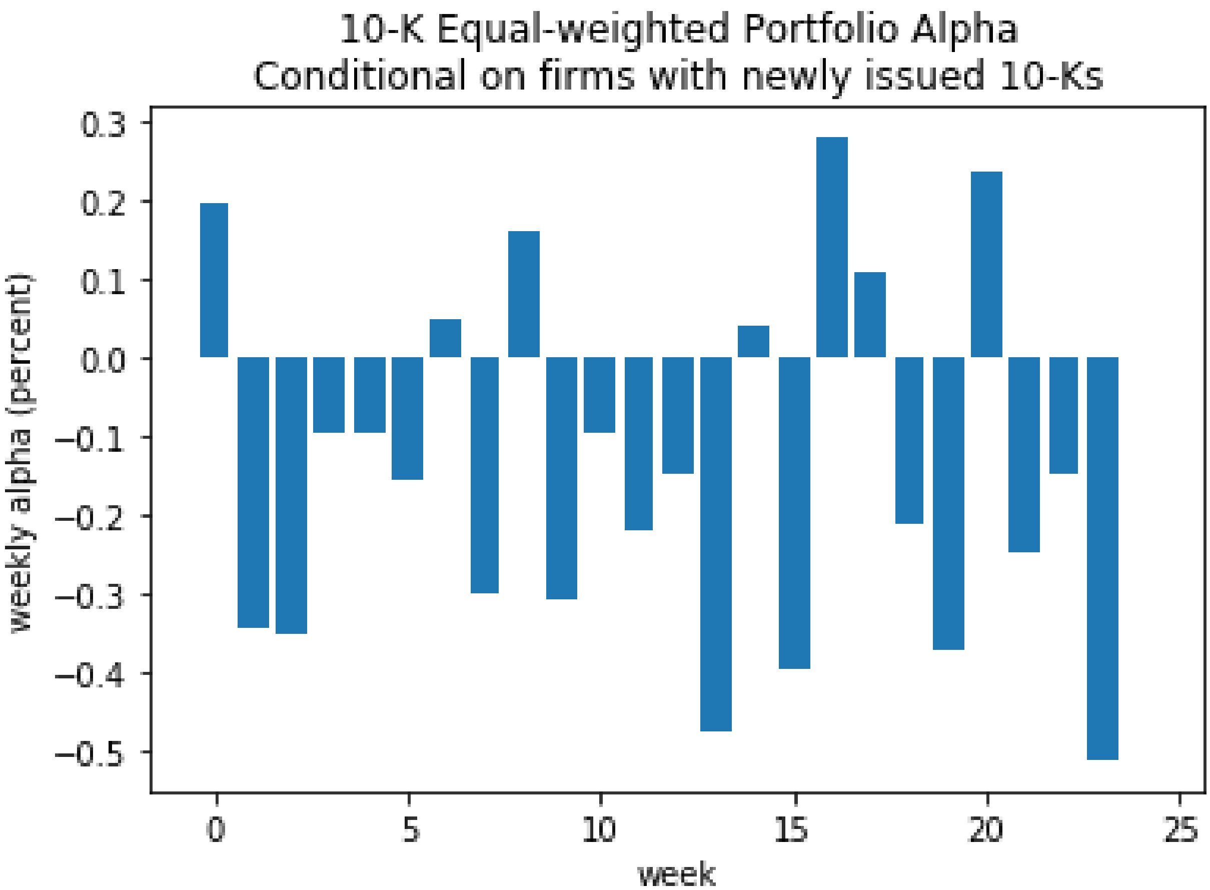
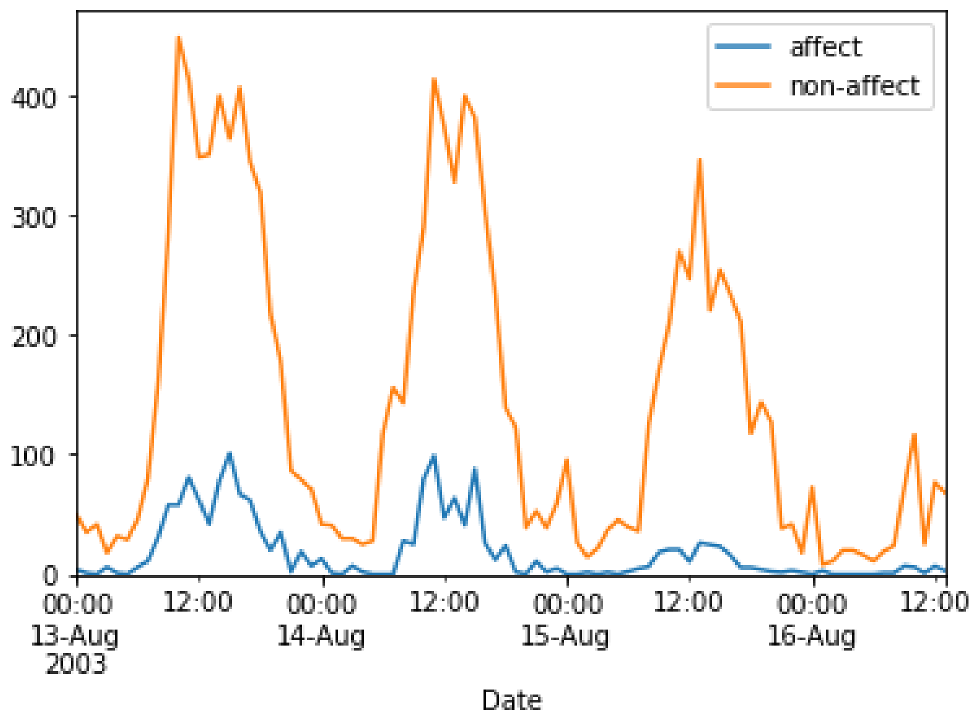


Figure 6
Viewing Activity during Northeast Blackout of 2003
The figure shows the hourly 8-K download pattern by investors in affected and non-affected regions during the blackout event.



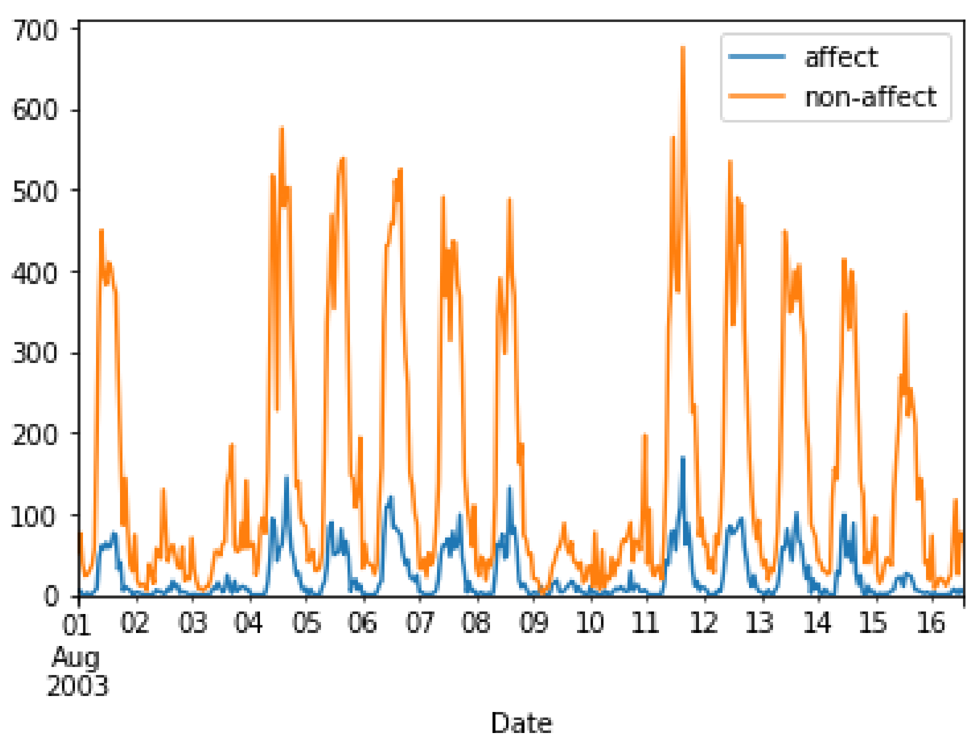


Figure 7

Demand for 8-K and Abnormal Return around Events

The figure studies the long/short portfolio of 8-K demand and abnormal returns around 8-K filing and event date. For each unscheduled 8-K filings, I calculate the cumulative abnormal return relative to the market around event and filing date. I then double sort stocks by the 8-K demand within the NYSE size-groups and the cumulative abnormal return into 5-by-5 blocks. Conditional on each abnormal return quintile, I regress the long/short 8-K portfolio return on Fama French five factors and momentum factor, and plot the alphas and 95% confidence intervals. For stocks with multiple unscheduled filings in a month, I choose the one with the highest absolute abnormal return.

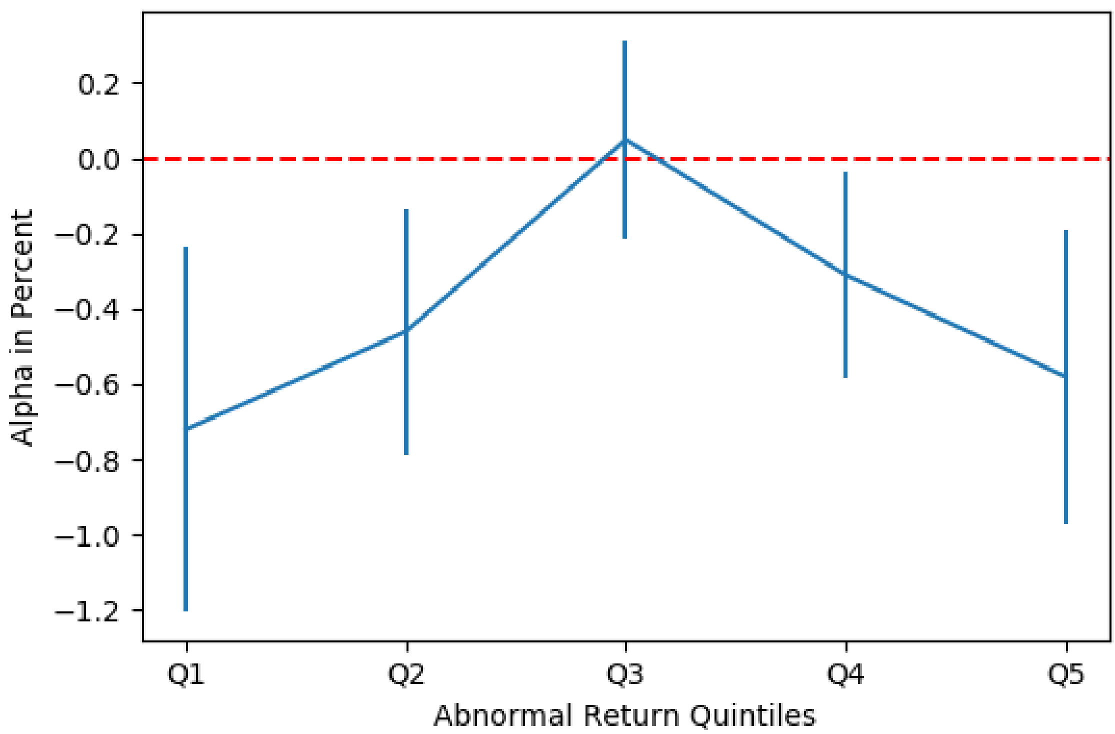


Table 1
Fama-Macbeth Regression on EDGAR Demand for Filings

The table shows results from Fama-Macbeth regressions of monthly individual stock returns on EDGAR downloads. The variable *log views*$_k$ is the natural logarithm of human downloads of the firm for filing type k. Regressions include controls for other variables that are known to predict cross-section variation in returns. Independent variables are winsorized at one and 99% levels. The sample covers from 2003 to 2016, with the dates determined by the availability of EDGAR Log data. Asset Growth is the annual percentage change in total assets. log(BM) is the natural logarithm of the book-to-market ratio. log(ME) is the natural logarithm of market capitalization. Operating Profit is the revenue minus cost of goods sold, SG&A expenses, and interest expense, divided by lagged common shareholders' equity. Abnormal Trading Volume is the difference between trading volume and previous 12-month average trading volume, scaled by the standard deviation of previous 12-month trading volume. SUE is the unexpected quarterly earnings (adjusted by median forecast earnings) divided by fiscal-quarter-end market capitalization. Earning Drift is the sum of daily returns in three days around the earnings announcement. Media Coverage is the total number of news covered by Ravenpack. Count Variables file 10K/10Q/8K are the number of 10-K/10-Q/8-K filings in the month.

	(1)	(2)	(3)	(4)	(5)
	Ret$_{t+1}$	Ret$_{t+1}$	Ret$_{t+1}$	Ret$_{t+1}$	Ret$_{t+1}$
log views$_{all}$		0.183*			
		(1.75)			
log views$_{10K}$			0.390***	0.388***	0.347***
			(7.42)	(7.35)	(5.89)
log views$_{10Q}$			-0.0691	-0.0697	-0.0531
			(-1.15)	(-1.15)	(-1.09)
log views$_{8K}$			-0.120**		
			(-2.23)		
log views$_{8K}^{unscheduled}$				-0.117**	-0.174***
				(-2.32)	(-3.10)
log views$_{8K}^{scheduled}$				0.0237	0.0128
				(0.70)	(0.30)
file 10K	0.222*	0.149	-0.0693	-0.0729	-0.0797
	(1.87)	(1.22)	(-0.55)	(-0.58)	(-0.52)
file 8K	-0.0760***	-0.120***	-0.0584	-0.0585	-0.0417
	(-2.99)	(-4.79)	(-1.13)	(-1.20)	(-1.37)
file 10Q	-0.0725*	-0.0873*	-0.0531	-0.0441	-0.0471
	(-1.75)	(-1.90)	(-1.44)	(-1.49)	(-1.55)
Asset Growth	-0.723***	-0.680***	-0.622***	-0.625***	-0.527***
	(-4.76)	(-4.70)	(-4.30)	(-4.32)	(-3.56)
log(BM)	0.134	0.108	0.0946	0.0937	0.0503
	(0.87)	(0.70)	(0.62)	(0.61)	(0.32)
log(ME)	-0.0696	-0.133*	-0.178**	-0.178**	-0.0658
	(-1.39)	(-1.81)	(-2.48)	(-2.49)	(-0.90)
Operating Profit	0.0834**	0.0654*	0.0482	0.0477	0.0475
	(2.31)	(1.74)	(1.30)	(1.29)	(1.39)
$r_{1,0}$	-2.319***	-2.397***	-2.398***	-2.394***	-2.224***
	(-3.49)	(-3.71)	(-3.72)	(-3.71)	(-3.07)
$r_{12,2}$	-0.608	-0.498	-0.496	-0.495	-0.415
	(-1.41)	(-1.30)	(-1.30)	(-1.29)	(-1.04)
Abnormal Trading Volume	0.141***	0.132***	0.135***	0.136***	0.128***
	(4.08)	(3.90)	(4.02)	(4.02)	(3.35)
SUE	3.930***	3.900***	3.861***	3.860***	3.945***
	(4.88)	(4.90)	(4.86)	(4.86)	(4.24)
Earning Drift	1.250***	1.261***	1.244***	1.240***	1.116***
	(3.30)	(3.38)	(3.34)	(3.33)	(2.73)
Change in Google Trend					-0.108
					(-0.82)
Media Coverage					0.00314
					(0.69)
Constant	1.719**	1.801**	2.399***	2.395***	1.052
	(2.06)	(2.12)	(2.63)	(2.62)	(1.14)
N	502662	502662	502662	502662	347381
Average R^2	0.0351	0.0385	0.0402	0.0403	0.0431
F	11.87	11.55	14.10	13.11	8.328

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2
Long/Short Portfolio by 10-K and 8-K Demand

The table shows monthly alphas and factor loadings of portfolios sorted by the 10-K/8-K viewing activity. To control for firm sizes, I first sort stocks by size into five groups using NYSE breakpoints. Conditional on each NYSE size-group, I then sort stocks by 10-K (8-K) downloads into quintiles and form equal-weighted portfolios. Panel A and B show the long/short portfolio returns and alphas with one, three, and twelve holding months for the 10-K and 8-K portfolios. Panel C and D show the factor loadings of 10-K and 8-K portfolios. For 8-K downloads, I only focus on the downloads on unscheduled filings.

Panel A: Size-adjusted 10-K Views Equal Weighted L/S Alpha

Holding Months	Raw Return	$Alpha^{CAPM}$	$Alpha^{FF3}$	$Alpha^{FFC}$	$Alpha^{FF5+UMD}$	$Alpha^{8-factor}$
1	0.81***	0.39*	0.39**	0.47***	0.37**	0.46***
	(3.05)	(1.89)	(1.99)	(3.31)	(2.6)	(3.14)
3	0.51**	0.11	0.11	0.19	0.07	0.16
	(2.09)	(0.57)	(0.62)	(1.48)	(0.55)	(1.22)
12	0.34	-0.03	-0.05	0.01	-0.12	-0.02
	(1.57)	(-0.22)	(-0.33)	(0.12)	(-1.03)	(-0.21)

Panel B: Size-adjusted 8-K Views Equal Weighted L/S Alpha

Holding Months	Raw Return	$Alpha^{CAPM}$	$Alpha^{FF3}$	$Alpha^{FFC}$	$Alpha^{FF5+UMD}$	$Alpha^{8-factor}$
1	-0.12	-0.54***	-0.55***	-0.47***	-0.43***	-0.41***
	(-0.46)	(-2.82)	(-2.96)	(-3.52)	(-3.11)	(-2.94)
3	-0.14	-0.57***	-0.58***	-0.49***	-0.45***	-0.45***
	(-0.57)	(-3.13)	(-3.29)	(-3.98)	(-3.5)	(-3.41)
12	-0.13	-0.55***	-0.56***	-0.49***	-0.44***	-0.46***
	(-0.56)	(-3.34)	(-3.65)	(-4.17)	(-3.65)	(-3.69)

Panel C: Factor Loadings of 10-K Portfolio

level	alpha	mktrf	smb	hml	umd	rmw	cma
Low	0.08	0.779***	0.525***	0.148***	-0.029	-0.325***	-0.173**
	(0.85)	(28.23)	(11.53)	(3.3)	(-1.34)	(-5.34)	(-2.32)
2	0.07	0.897***	0.697***	0.088**	-0.067***	-0.328***	-0.116*
	(0.9)	(38.33)	(18.05)	(2.3)	(-3.6)	(-6.35)	(-1.85)
3	0.22***	0.994***	0.733***	0.022	-0.129***	-0.292***	-0.001
	(2.71)	(41.34)	(18.49)	(0.57)	(-6.72)	(-5.5)	(-0.02)
4	0.32***	1.018***	0.796***	0.003	-0.212***	-0.204***	0.029
	(3.38)	(36.56)	(17.32)	(0.07)	(-9.54)	(-3.34)	(0.38)
High	0.46***	1.097***	0.887***	0.107*	-0.441***	-0.117	0.106
	(3.42)	(27.91)	(13.68)	(1.68)	(-14.05)	(-1.35)	(1.01)
H-L	0.37**	0.32***	0.366***	-0.042	-0.41***	0.21**	0.277**
	(2.6)	(7.62)	(5.28)	(-0.62)	(-12.25)	(2.27)	(2.45)

Panel D: Factor Loadings of 8-K Portfolio

level	alpha	mktrf	smb	hml	umd	rmw	cma
Low	0.31***	0.8***	0.554***	0.106***	-0.039**	-0.252***	-0.0
	(4.2)	(37.69)	(15.83)	(3.09)	(-2.3)	(-5.38)	(-0.01)
2	0.19***	0.899***	0.656***	0.091***	-0.106***	-0.165***	-0.143**
	(2.67)	(42.27)	(18.71)	(2.63)	(-6.23)	(-3.51)	(-2.5)
3	0.28***	0.954***	0.734***	0.038	-0.159***	-0.234***	-0.051
	(2.96)	(34.91)	(16.29)	(0.85)	(-7.27)	(-3.88)	(-0.69)
4	0.13*	1.019***	0.761***	0.072*	-0.166***	-0.282***	0.01
	(1.87)	(40.01)	(18.1)	(1.74)	(-8.18)	(-5.03)	(0.14)
High	-0.11	1.097***	0.854***	0.056	-0.423***	-0.354***	-0.051
	(-0.77)	(25.98)	(12.25)	(0.82)	(-12.56)	(-3.81)	(-0.45)
H-L	-0.43***	0.3***	0.304***	-0.053	-0.383***	-0.101	-0.053
	(-3.11)	(7.52)	(4.62)	(-0.82)	(-12.02)	(-1.15)	(-0.5)

t statistics in parentheses

Table 3

Weekly Regression of Stock Returns on EDGAR Demand for Filings

The table shows results from regressions of weekly individual stock returns on EDGAR downloads. The dependent variable in columns (1) and (2) is the current week stock returns in basis points. The dependent variable in columns (3) and (4) is the next week stock returns in basis points. $views_t^k$ is the cumulative downloads of filing type k at week t. *Filing* k_t is a dummy variable, which is equal to one if the firm issued any filings with type k at week t. $News_t$ is a dummy variable, which is equal to one if there is any news coverage of the firm in Ravenpack at week t. *Earnings Release*$_t$ is a dummy variable, which is equal to one if the firm releases its earnings at week t. AIA_t is a dummy variable, which is equal to one if the Bloomberg News Heat daily index has a maximum of 3 or above in week t. $DADSVI_t$ is a dummy variable, which is equal to one if the Google Trends daily index in any day of the week is above its 90 percentile in the past month. Firm controls include the log of firm market capitalization, and the book-to-market ratio. Time fixed effects are included, and standard errors are clustered by week.

	(1)	(2)	(3)	(4)
	ret_t	ret_{t+1}	ret_t	ret_{t+1}
$\log(views_t^{10K})$	14.11***	11.68***	6.809***	4.523***
	(8.58)	(7.69)	(4.27)	(2.95)
$\log(views_t^{8K})$	-0.405	-2.046	-1.955	-1.691
	(-0.26)	(-1.53)	(-0.88)	(-1.42)
$Filing\ 10K_t \times \log(views_t^{10K})$	-1.504	5.867	-2.450	-0.0753
	(-0.23)	(1.01)	(-0.39)	(-0.01)
$Filing\ 8K_t \times \log(views_t^{8K})$	11.54***	-3.262*	16.09***	-4.804*
	(4.79)	(-1.82)	(3.81)	(-1.79)
$Filing\ 10K_t$	-28.96	-25.01	0.292	0.375
	(-1.18)	(-1.27)	(0.01)	(0.01)
$Filing\ 8K_t$	-5.198	8.463*	-32.67**	11.28
	(-0.86)	(1.86)	(-2.51)	(1.16)
$Media\ Coverage_t$	34.38***	5.645***	24.83***	2.975
	(16.84)	(3.29)	(7.33)	(0.92)
$Earning\ Release_t$	36.70***	14.36***	1.390	17.51***
	(6.20)	(3.39)	(0.17)	(3.18)
AIA_t			51.67***	0.999
			(11.19)	(0.34)
$DADSVI_t$			21.89***	2.301
			(11.05)	(1.32)
lag returns	Yes	Yes	Yes	Yes
firm controls	Yes	Yes	Yes	Yes
week fe	Yes	Yes	Yes	Yes
N	2308554	2305463	529874	528806
Adjusted R^2	0.115	0.113	0.154	0.172
F	63.31	21.04	27.70	3.784

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4
Panel Regression of Information Asymmetry
The table shows the weekly panel regression of next-month information asymmetry proxy on current month investor demand for filings. The dependent variable is the price impact measure estimated following Holden and Jacobsen (2014). $views_t^k$ is the cumulative downloads of filing type k at week t. $Filing\ k_t$ is a dummy variable, which is equal to one if the firm issued any filings with type k at week t. $News_t$ is a dummy variable, which is equal to one if there is any news coverage of the firm in Ravenpack at week t. $Earnings\ Release_t$ is a dummy variable, which is equal to one if the firm releases its earnings at week t. Firm controls include the log of firm market capitalization, and the book-to-market ratio. Time and firm fixed effects are included. Standard errors are two-way clustered by time and firm.

	(1) $Price\ Impact_{t+1}$	(2) $Price\ Impact_{t+1}$
$\log(views_t^{10K})$	-0.00120*	-0.00109*
	(-1.83)	(-1.74)
$\log(views_t^{8K})$	-0.00352***	-0.00324***
	(-3.41)	(-3.19)
$Filing\ 10K_t$	0.00625	0.0373**
	(0.63)	(1.99)
$Filing\ 8K_t$	-0.00284*	0.00217
	(-1.79)	(0.68)
$Filing\ 10K_t \times \log(views_t^{10K})$		-0.00331*
		(-1.79)
$Filing\ 8K_t \times \log(views_t^{8K})$		-0.00508**
		(-1.97)
$Media\ Coverage_t$	0.00605**	0.00598**
	(2.50)	(2.46)
$Earnings\ Release_t$	-0.00957**	-0.00983**
	(-1.98)	(-2.04)
Firm and Time FEs	Yes	Yes
firm controls	Yes	Yes
N	2022081	2022081
Adjusted R^2	0.485	0.485
F	234.5	188.0

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5
8-K Demand and Information Asymmetry
The table shows monthly alphas of portfolios sorted by 8-K downloads and information asymmetry. I use Amihud illiquidity measure and previous quarter earning forecast dispersion to measure ex-ante information asymmetry. For each portfolio, I regress portfolio return on Fama French five factors and momentum factor. For 8-K downloads, I only focus on the downloads on unscheduled filings.

Panel A: Double Sort by Size-adjusted 8-K Views and Amihud

Amihud/Views	Low	2	3	4	High	H-L
Low	0.12	0.06	0.01	0.07	0.02	-0.11
	(1.58)	(0.84)	(0.21)	(0.82)	(0.17)	(-0.93)
2	0.20**	0.13	0.26**	0.13	-0.12	-0.33*
	(2.09)	(1.57)	(2.4)	(1.22)	(-0.74)	(-1.67)
High	0.45***	0.48**	0.54**	0.57**	-0.10	-0.53**
	(2.75)	(2.55)	(2.14)	(2.32)	(-0.27)	(-2.27)
H-L	0.33*	0.42**	0.53**	0.51*	-0.11	-0.42*
	(1.81)	(2.04)	(2.02)	(1.97)	(-0.32)	(-1.91)

Panel B: Double Sort by Size-adjusted 8-K Views and Past Forecast Dispersion

Forecast Dispersion/Views	Low	2	3	4	High	H-L
Low	0.28***	0.12	0.15*	0.20**	0.33***	0.05
	(3.56)	(1.58)	(1.82)	(2.13)	(2.7)	(0.34)
2	0.17**	-0.03	0.26***	0.13	-0.05	-0.22
	(2.09)	(-0.37)	(2.68)	(1.29)	(-0.34)	(-1.34)
High	0.17	0.02	-0.05	-0.15	-0.38**	-0.56***
	(1.19)	(0.13)	(-0.33)	(-1.16)	(-2.03)	(-2.85)
H-L	-0.11	-0.11	-0.20	-0.36**	-0.71***	-0.61**
	(-0.65)	(-0.64)	(-1.05)	(-2.09)	(-3.24)	(-2.25)

t statistics in parentheses

Table 6
Diff-in-diff Estimation on Blackout Event

The table shows the estimation result of diff-in-diff regressions. The dependent variable is the daily price impact measure to proxy for firm-level information asymmetry. The sample includes two days of data, from August 14 to August 15, 2003. Firms in the sample just disclosed material information on August 13 and have headquarters outside the affected regions. The variable *frac* is the fraction of historical 8-K download from the affected regions prior to the shock. The variable *treated* is equal to one if the firms have above median *frac*. Post is equal to one if it is the second day of the sample. The last two columns further exclude firms with top business activity from affected regions, where the level of business activity is measured by the frequency of state name appeared in 10-K filings.

	(1) Price Impact	(2) Price Impact	(3) Price Impact	(4) Price Impact
post	-0.000611**	-0.000499	-0.000574*	-0.000445
	(-2.00)	(-1.33)	(-1.83)	(-1.14)
treated	-0.000395		-0.000212	
	(-1.25)		(-0.63)	
treated $\times$ post	0.00167**		0.00169*	
	(2.00)		(1.81)	
frac		-0.00409*		-0.00284
		(-1.89)		(-1.21)
frac $\times$ post		0.00855**		0.00834**
		(2.34)		(2.02)
Constant	0.00234***	0.00246***	0.00228***	0.00239***
	(10.10)	(9.60)	(9.67)	(9.05)
Exclude Plans from Affected Regions	No	No	Yes	Yes
N	640	640	580	580
Adjusted R^2	0.00968	0.00427	0.0112	0.00409
F	3.960	3.995	3.816	3.487

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7

Alternative Explanations on Blackout Event

The table shows the estimation result of diff-in-diff regressions. The dependent variable in columns (1) and (2) is the daily trading volume normalized by the number of shares outstanding. The dependent variable in columns (3) and (4) is the daily number of trades after the market close. The sample includes two days of data, from August 14 to August 15, 2003. Firms in the sample just disclosed material information on August 13. The variable *frac* is the fraction of historical 8-K download from the affected regions prior to the shock. The variable *treated* is equal to one if the firms have above median *frac*. Post is equal to one if it is the second day of the sample.

	(1) Volume	(2) Volume	(3) N_Trades$^{after\ market}$	(4) N_Trades$^{after\ market}$
post	-0.00395***	-0.00347***	0.101	0.227
	(-3.67)	(-3.03)	(0.14)	(0.29)
treated	0.000266		1.534*	
	(0.23)		(1.90)	
treated × post	-0.000137		-0.167	
	(-0.08)		(-0.15)	
frac		0.0118		14.55***
		(1.45)		(2.60)
frac × post		-0.00764		-2.815
		(-0.66)		(-0.36)
Constant	0.00696***	0.00625***	5.548***	5.180***
	(9.16)	(7.72)	(10.61)	(9.32)
N	580	580	580	580
Adjusted R^2	0.0401	0.0439	0.0112	0.0190
F	8.017	8.817	2.170	3.715

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8

8-K Demand and the Cost of Information Acquisition

The table shows monthly alphas of equal-weighted portfolios sorted by 8-K downloads, conditional on geographical distance distribution to headquarters and recurring viewer ratios. Geographical distance is the value-weighted distance between the location of viewing IP and the firm headquarter. I classify a view as a recurring view if the IP address visited any firm filings in the past three months. Recurring visitor ratio is the ratio between the numbers of recurring and non-recurring downloads. Abnormal institutional attention, ab_inst^{8K}, is the fraction of institutional 8-K downloads normalized by the past 12-month average and standard deviation. For each stock at each month, I first sort stocks by geographical distance (recurring visitor ratio, abnormal institutional attention) into terciles. Conditional on each tercile, I then sort stocks by 8-K downloads within the NYSE size-group into quintiles. For each portfolio, I regress portfolio return on Fama French five factors and momentum factor, and report the alphas.

Panel A: Double Sort by Size-adjusted 8-K Views and Distance

distance/8-K views	Low	2	3	4	High	H-L
Low	0.38***	0.16*	0.28**	0.19**	-0.21	-0.61***
	(4.67)	(1.73)	(2.61)	(2.02)	(-1.35)	(-3.94)
2	0.23**	0.31***	0.22**	0.09	-0.19	-0.41**
	(2.32)	(3.45)	(2.01)	(0.8)	(-1.15)	(-2.43)
High	0.28**	0.12	0.28**	0.33**	0.05	-0.23
	(2.53)	(1.18)	(2.22)	(2.54)	(0.23)	(-1.13)
H-L	-0.12	-0.05	0.01	0.14	0.26	0.38**
	(-1.09)	(-0.4)	(0.07)	(1.07)	(1.29)	(2.31)

Panel B: Double Sort by Size-adjusted 8-K Views and 8-K Recurring Ratio

$freq^{8K}$/8-K views	Low	2	3	4	High	H-L
Low	0.29***	0.20**	0.27**	0.36***	0.16	-0.15
	(3.0)	(2.0)	(2.38)	(2.74)	(0.95)	(-0.81)
2	0.26***	0.26***	0.31***	0.23**	-0.15	-0.41**
	(2.62)	(3.1)	(2.8)	(2.15)	(-0.86)	(-2.25)
High	0.25***	0.11	0.20*	0.05	-0.30	-0.55***
	(2.81)	(1.14)	(1.74)	(0.45)	(-1.62)	(-2.95)
H-L	-0.02	-0.05	-0.03	-0.30**	-0.46**	-0.41**
	(-0.16)	(-0.47)	(-0.29)	(-2.1)	(-2.54)	(-2.24)

Panel C: Double Sort by Size-adjusted 8-K Views and Abnormal Institution Attention

ab_inst^{8K}/8-K views	Low	2	3	4	High	H-L
Low	0.35**	0.44***	0.11	0.21	-0.14	-0.49**
	(2.35)	(2.72)	(0.69)	(1.04)	(-0.76)	(-2.51)
2	0.37**	0.13	0.22	0.26	-0.1	-0.47*
	(2.07)	(0.81)	(1.47)	(1.37)	(-0.47)	(-1.76)
High	0.57***	0.37**	0.29*	0.06	-0.13	-0.7***
	(3.82)	(2.15)	(1.7)	(0.31)	(-0.65)	(-3.01)
H-L	0.22	-0.07	0.18	-0.15	0.01	-0.21*
	(1.57)	(-0.36)	(0.91)	(-0.74)	(0.03)	(-1.74)

t statistics in parentheses

Table 9
8-K Demand and Information Content
The table shows monthly alphas of portfolios sorted by 8-K downloads and cumulative abnormal returns around filing and event date of unscheduled 8-K filings. For each unscheduled 8-K filings, I calculate the cumulative abnormal return relative to the market around event and filing date. I then double sort stocks by the 8-K downloads within each NYSE size-group and the cumulative abnormal return into 5-by-3 blocks. Conditional on each abnormal return tercile, I regress the long/short 8-K portfolio return on Fama French five factors and momentum factor. For stocks with multiple unscheduled filings in a month, I choose the one with the highest absolute abnormal return.

Panel A: Double Sort by Size-adjusted 8-K Views and Abnormal Returns

abret/views	Low	2	3	4	High	H-L
Low	0.42*	0.08	0.33**	-0.05	-0.31	-0.75**
	(1.7)	(0.49)	(2.01)	(-0.29)	(-1.59)	(-2.51)
2	0.30**	0.24**	0.09	0.19*	0.16	-0.14
	(2.19)	(2.21)	(0.8)	(1.87)	(1.23)	(-0.72)
High	0.33*	0.45***	0.50***	0.34**	-0.14	-0.48**
	(1.71)	(2.72)	(2.87)	(2.35)	(-0.74)	(-1.99)
H-L	-0.09	0.36*	0.16	0.38**	0.17	0.27
	(-0.28)	(1.71)	(0.76)	(2.15)	(0.84)	(0.65)

Panel B: Double Sort by Size-adjusted 8-K Views and Unexpected Abnormal Returns

unexpected abret/views	Low	2	3	4	High	H-L
Low	0.42*	0.13	0.33**	-0.02	-0.33*	-0.77**
	(1.68)	(0.77)	(2.0)	(-0.11)	(-1.68)	(-2.56)
2	0.30**	0.17	0.11	0.17	0.25*	-0.05
	(2.22)	(1.57)	(1.01)	(1.63)	(1.75)	(-0.27)
High	0.32*	0.49***	0.48***	0.33**	-0.18	-0.51*
	(1.69)	(3.05)	(2.79)	(2.28)	(-0.94)	(-1.91)
H-L	-0.10	0.36*	0.14	0.35*	0.15	0.26
	(-0.33)	(1.68)	(0.66)	(1.94)	(0.78)	(0.49)

t statistics in parentheses

Table 10
10-K Demand and Attention-Grabbing

The table shows monthly alphas of portfolios sorted by 10-K downloads and attention-grabbing measure. I use abnormal trading volume and maximum daily absolute return to measure attention-grabbing. Abnormal trading volume is the difference between monthly trading volume and past 12-month average trading volume, scaled by the standard deviation of past 12-month trading volume. For each portfolio, I regress portfolio return on Fama French five factors and momentum factor.

Panel A: Double Sort by Size-adjusted 10-K Views and Maximum Return

Max Return/Views	Low	2	3	4	High	H-L
Low	0.29***	0.31***	0.38***	0.35***	0.60***	0.31**
	(2.9)	(3.84)	(4.58)	(4.09)	(5.86)	(2.31)
2	-0.03	-0.07	0.16*	0.33***	0.52***	0.55***
	(-0.26)	(-0.83)	(1.91)	(3.55)	(3.12)	(2.84)
High	-0.36**	-0.32**	-0.02	0.45**	0.51	0.87***
	(-2.17)	(-2.22)	(-0.11)	(2.16)	(1.63)	(3.14)
H-L	-0.65***	-0.63***	-0.40**	0.10	-0.09	0.56**
	(-3.3)	(-3.41)	(-2.18)	(0.49)	(-0.31)	(1.99)

Panel B: Double Sort by Size-adjusted 10-K Views and Abnormal Trading Volume

Abnormal Trading Volume/Views	Low	2	3	4	High	H-L
Low	-0.31***	-0.23**	-0.12	-0.19	0.06	0.38*
	(-2.7)	(-2.29)	(-1.05)	(-1.33)	(0.31)	(1.87)
2	0.01	0.03	0.27***	0.46***	0.63***	0.62***
	(0.07)	(0.35)	(2.76)	(4.04)	(3.34)	(3.31)
High	0.21*	0.09	0.31***	0.72***	1.16***	0.95***
	(1.85)	(0.9)	(2.91)	(5.5)	(5.49)	(4.16)
H-L	0.52***	0.32**	0.43***	0.91***	1.10***	0.57**
	(3.41)	(2.22)	(2.91)	(5.57)	(5.3)	(2.32)

t statistics in parentheses

Figure A1

Long-term Performance of 8-K portfolio: Pre-estimate loadings

The figure shows the long-term alphas of the long/short portfolio sorted by 8-K downloads. At each given date and within each NYSE size quintile, I sort stocks with unscheduled 8-K disclosure by the filing downloads into quintiles. I then estimate the stock-level factor loadings using one-year daily return before the disclosure and calculate the estimated alpha for the next 25 weeks. I then plot the average difference in alpha between the top quintile stocks and the bottom quintile stocks.

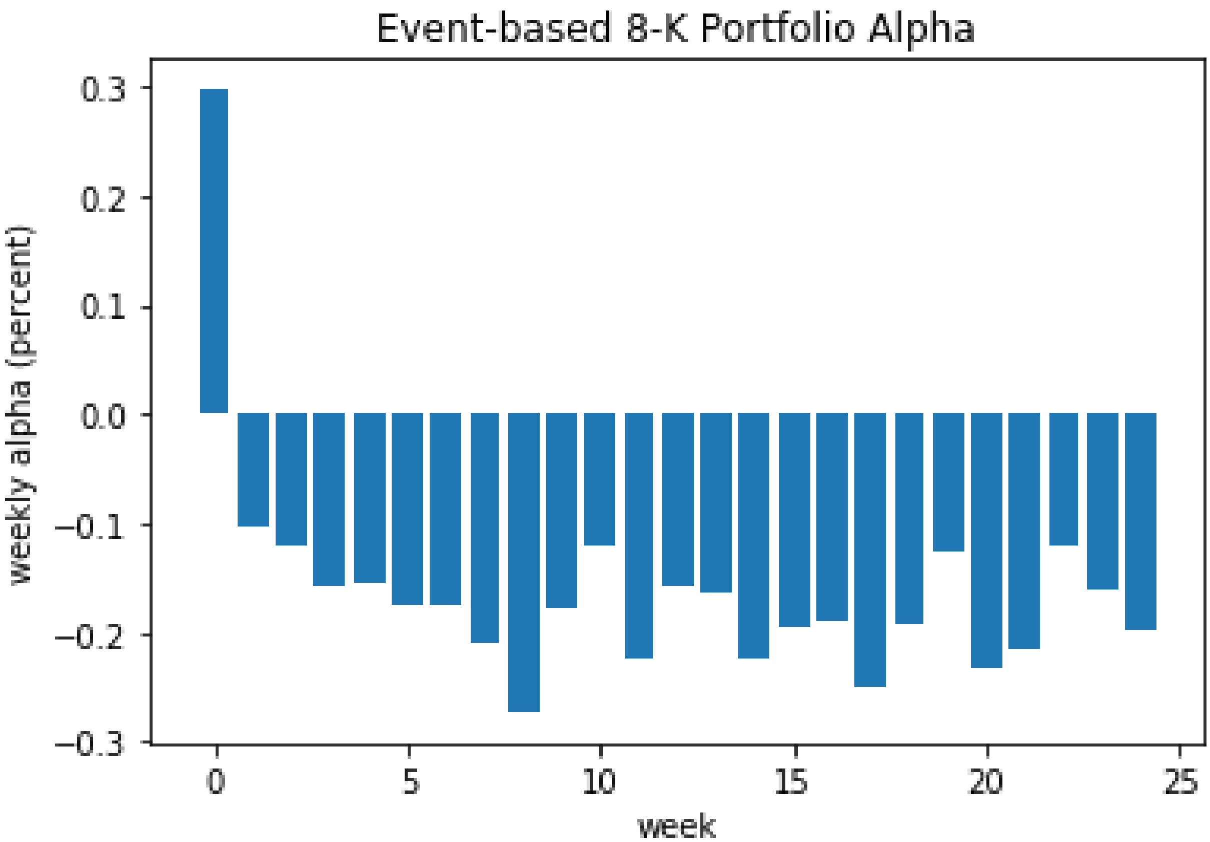


Figure A2

10-K downloads Conditional on 8-K downloads

The figure shows the time-series of 10-K viewing activity, conditional on whether the visitor also viewed any 8-K filings of the firm in the past three months. $views_{10K}^{only}$ is the total number of 10-K downloads by visitors who have not downloaded any 8-K filings of the firm. $views_{10K}^{both}$ is the total number of 10-K downloads by visitors who have downloaded one or more 8-K filings of the firm.

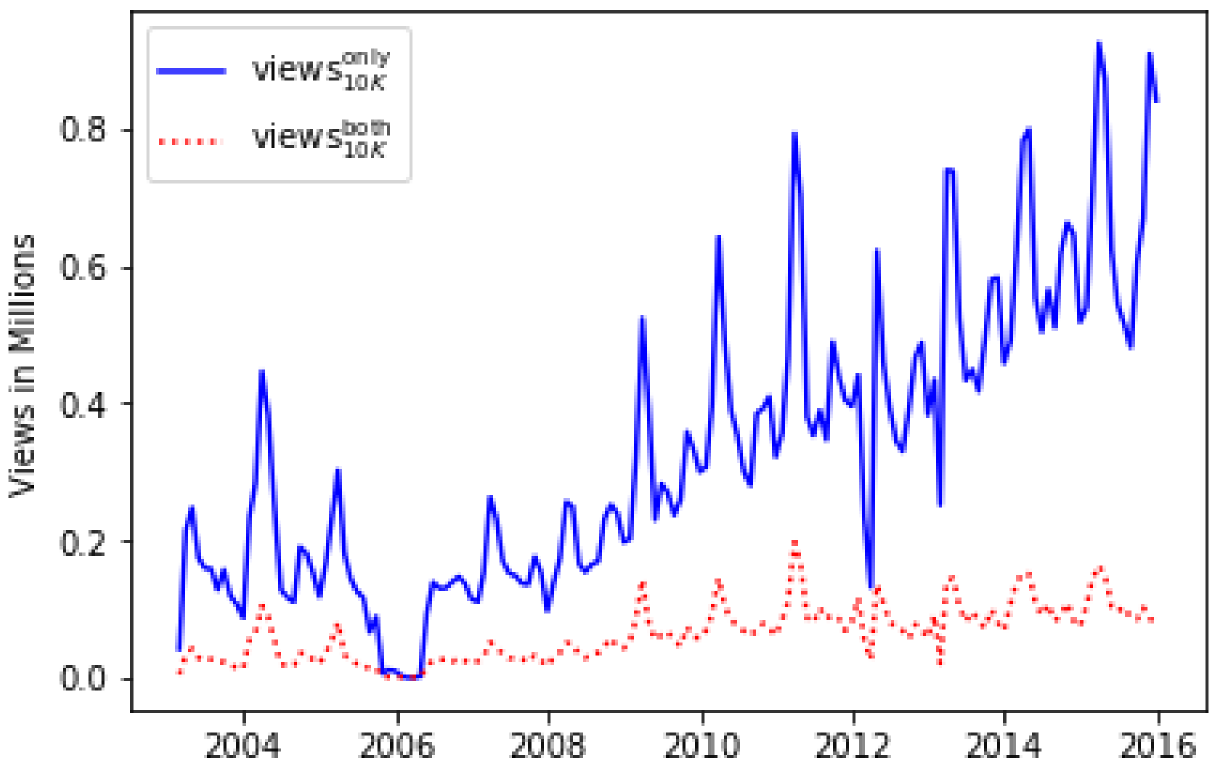
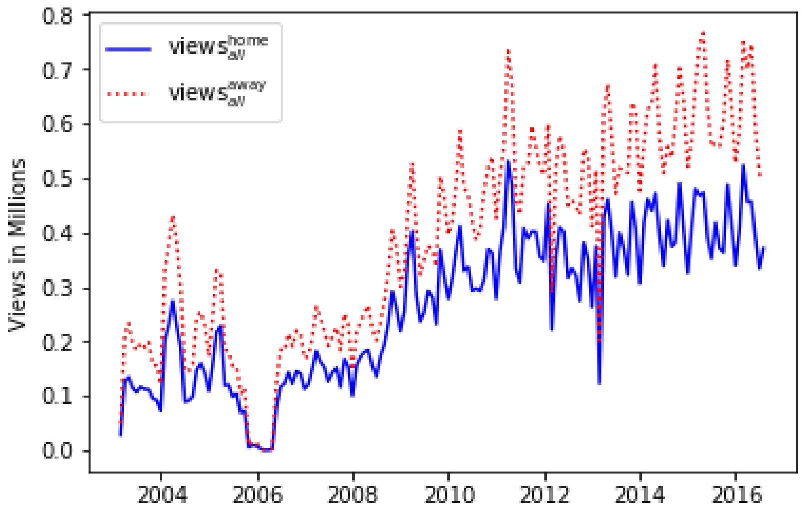


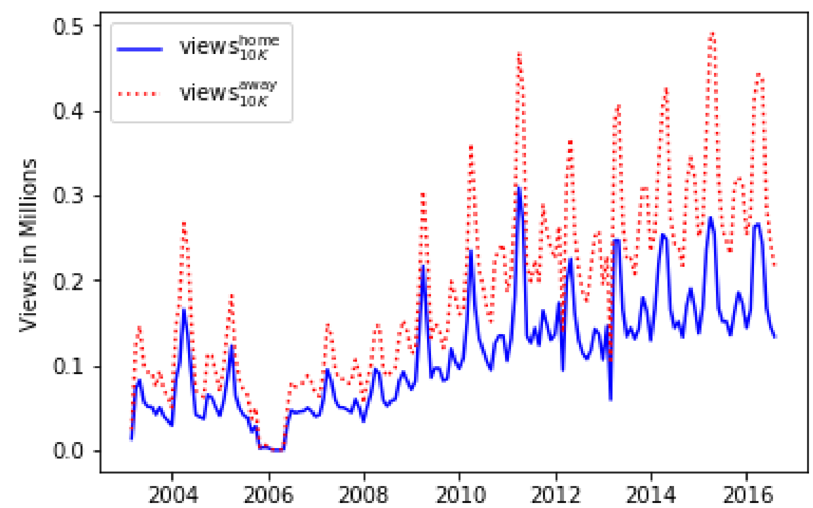
Figure A3

Viewing Activities by Geographical Distance

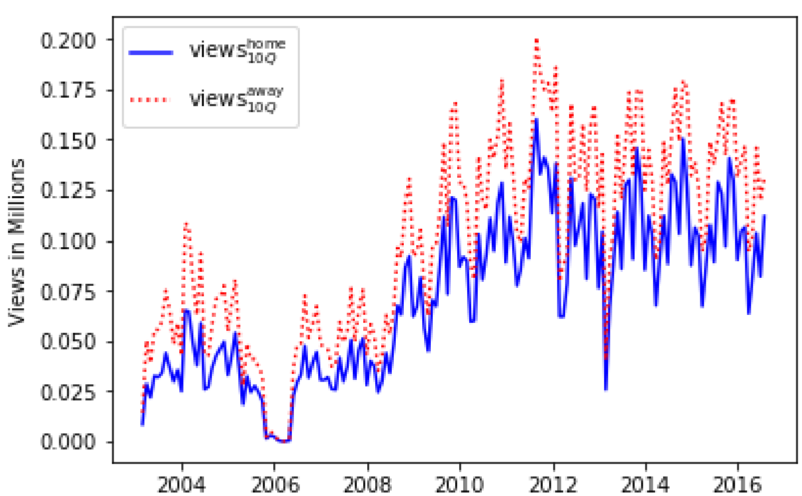
The figure shows the number of downloads by geographical distance. I classify a filing view as home if the distance between the locations of viewing IP and headquarter is less than 400 miles.



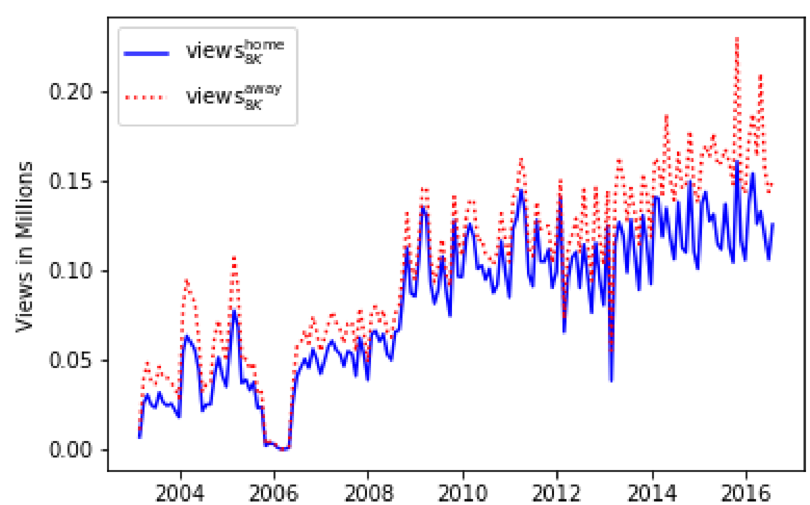
(a) All Filings



(b) 10-K Filings



(c) 10-Q Filings



(d) 8-K Filings

Figure A4

Time-series Recurring Visitor Ratios

The figure shows the time-series plot of recurring visitor ratios by 10-K and 8-K visitors. For each firm and IP address, I classify a filing view as recurring if the IP address submitted requests to view the company filings during the past three months. At each month, I then calculate the cross-section average of recurring ratios by 10-K and 8-K filings.

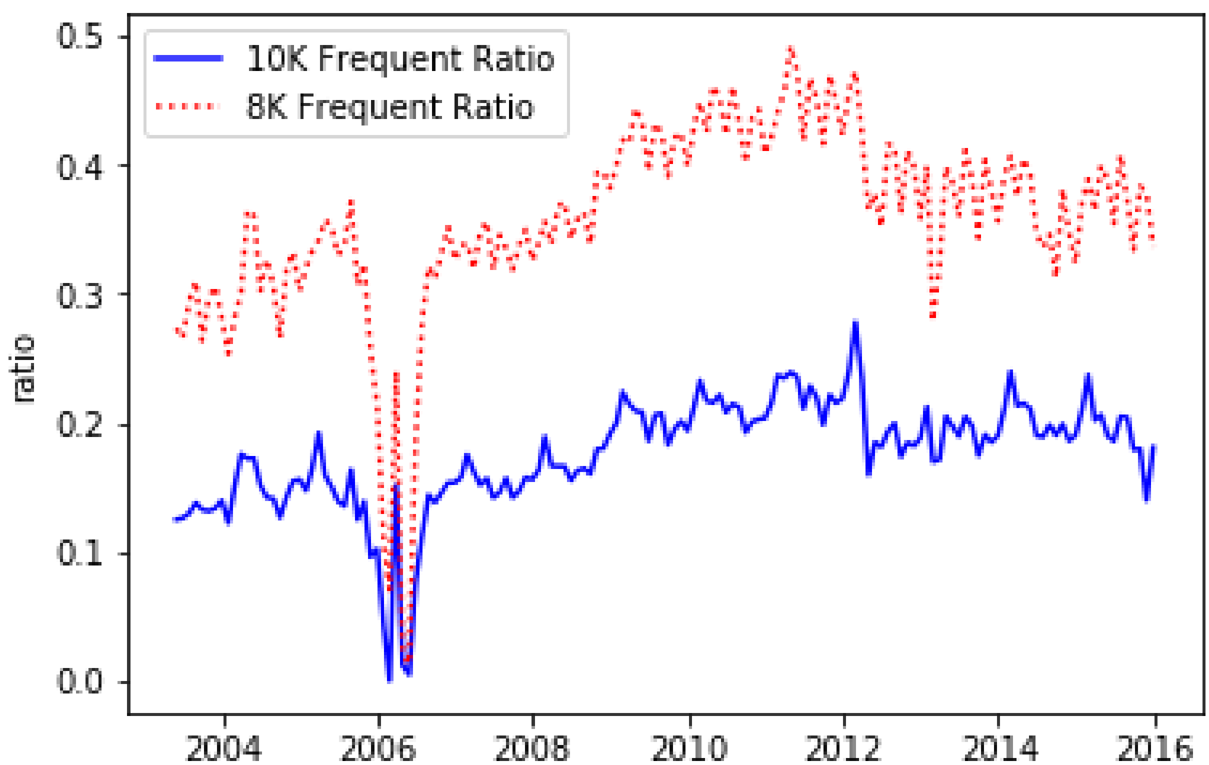


Table A1
Summary statistics

The table shows the summary statistics of main variables at the firm-month level. $views_{10K}$ is the number of 10-K filing downloads. $views_{10Q}$ is the number of 10-Q filing downloads. $views_{8K}$ is the number of 8-K filing downloads. Asset Growth is the annual percentage change in total assets. log(BM) is the natural logarithm of book-to-market ratio. log(ME) is the natural logarithm of market capitalization. Operating Profit is the revenue minus cost of goods sold, SG&A expenses, and interest expense, divided by lagged common shareholders' equity. Abnormal Trading Volume is the difference between trading volume and previous 12-month average trading volume, scaled by the standard deviation of previous 12-month trading volume. SUE is the unexpected quarterly earnings (adjusted by median forecast earnings) divided by fiscal-quarter-end market capitalization. Earning Drift is the sum of daily returns in three days around earnings announcement. Media Coverage is the total number of news in covered by Ravenpack. file 10K/10Q/8K is the number of 10-K/10-Q/8-K filings in the month.

Variable	Obs	Mean	Std. Dev.	Min	Max	P1	P25	P50	P75	P99
$views_{10K}$	502662	133.98	1032.961	0	370231	0	17	44	110	1478
$views_{10Q}$	502662	79.868	2547.154	0	1053239	0	13	32	75	587
$views_{8K}$	502662	75.786	336.343	0	133132	0	11	31	80	655
Asset Growth	502662	.103	.348	-.679	3.197	-.471	-.038	.047	.154	1.748
log(BM)	502662	.642	.622	-1.611	7.644	-.385	.29	.518	.829	3.055
log(ME)	502662	12.979	2.092	5.535	18.626	8.603	11.439	12.908	14.404	17.85
Operating Profit	502662	.694	1.182	-6.469	9.753	-3.027	.285	.537	.925	6.16
Abnormal Trading Volume	502662	.185	1.612	-2.826	19.255	-1.926	-.76	-.22	.647	6.916
SUE	502662	-.006	.16	-6.275	1.528	-.358	-.003	0	.003	.286
Earning Drift	502662	.002	.088	-.464	.524	-.24	-.04	.001	.042	.255
Media Coverage	397780	8.306	9.512	0	407	0	2	6	11	43
file 10K	502662	.089	.319	0	1	0	0	0	0	1
file 8K	502662	1.008	1.147	0	26	0	0	1	2	5
file 10Q	502662	.253	.478	0	1	0	0	0	0	1

Table A2
Univariate Sort on 8-K Crawling Activity
The table shows the univariate sort results for 8-K crawling activity, separated by institutional and retail crawling. To control for firm sizes, I first sort stocks by size into five groups using NYSE breakpoints. Conditional on each NYSE size-group, I then sort stocks by institutional (retail) 8-K crawling into quintiles and form equal-weighted portfolios. Institutions are identified by the name of organizations that own the IP block.

Panel A: Long/Short Institution 8-K Crawling Activity

level	alpha	mktrf	smb	hml	umd	rmw	cma
Low	0.40***	0.848***	0.608***	-0.079	-0.002	-0.393***	-0.134
	(3.19)	(23.4)	(10.19)	(-1.34)	(-0.07)	(-4.92)	(-1.37)
2	0.29***	0.874***	0.643***	0.018	-0.112***	-0.232***	-0.091
	(2.74)	(28.22)	(12.6)	(0.35)	(-4.51)	(-3.39)	(-1.09)
3	0.20**	0.948***	0.671***	-0.008	-0.132***	-0.263***	-0.037
	(2.48)	(39.64)	(17.02)	(-0.21)	(-6.92)	(-4.99)	(-0.57)
4	0.13	0.959***	0.75***	0.037	-0.158***	-0.273***	-0.027
	(1.39)	(35.23)	(16.71)	(0.84)	(-7.29)	(-4.53)	(-0.37)
High	-0.10	1.097***	0.808***	0.158**	-0.348***	-0.118	-0.048
	(-0.74)	(28.0)	(12.51)	(2.44)	(-11.1)	(-1.32)	(-0.45)
H-L	-0.48***	0.246***	0.198**	0.247***	-0.344***	0.272**	0.087
	(-2.67)	(4.69)	(2.28)	(2.83)	(-8.19)	(2.27)	(0.6)

Panel B: Long/Short Retail 8-K Crawling Activity

level	alpha	mktrf	smb	hml	umd	rmw	cma
Low	0.19	0.845***	0.694***	-0.224***	-0.082***	-0.337***	-0.161*
	(1.55)	(23.74)	(11.82)	(-3.87)	(-2.88)	(-4.29)	(-1.69)
2	0.34***	0.888***	0.632***	-0.053	-0.116***	-0.246***	-0.07
	(3.45)	(30.98)	(13.39)	(-1.13)	(-5.07)	(-3.89)	(-0.91)
3	0.18*	0.95***	0.724***	0.019	-0.14***	-0.17***	-0.047
	(1.91)	(33.84)	(15.62)	(0.42)	(-6.27)	(-2.75)	(-0.62)
4	0.15*	0.979***	0.734***	0.091**	-0.171***	-0.211***	0.058
	(1.68)	(36.42)	(16.53)	(2.08)	(-7.98)	(-3.56)	(0.81)
High	-0.02	1.073***	0.805***	0.148**	-0.339***	-0.261***	-0.097
	(-0.14)	(28.02)	(12.74)	(2.38)	(-11.11)	(-3.09)	(-0.95)
H-L	-0.20	0.226***	0.108	0.374***	-0.258***	0.071	0.066
	(-1.25)	(4.74)	(1.37)	(4.86)	(-6.8)	(0.68)	(0.52)

t statistics in parentheses