Market Efficiency in the Age of Machine Learning

Citation for published version:
Barbopoulos, LG, Dai, R, Putniš, T & Saunders, A 2021 'Market Efficiency in the Age of Machine Learning' NYU Stern School of Business Research Paper Series, pp. 1-56.

Market Efficiency in the Age of Machine Learning

Leonidas G. Barbopoulos[*] Rui Dai[†] Tālis J. Putniņš[‡] Anthony Saunders[§]

This version
April 7, 2021

[*]University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: leonidas.barbopoulos@ed.ac.uk. Tel: +44 131 650 8344. Corresponding author.

[†]Wharton Research Data Services (WRDS), The Wharton School, University of Pennsylvania, 3819 Chestnut Street Philadelphia PA, 19104. Email: rdai@wharton.upenn.edu.

[‡]University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia and University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: Talis.Putnins@uts.edu.au.

[§]Department of Finance, Kaufman Management Center, Stern School of Business, New York University, New York, N.Y. 10012. Email: asaunder@stern.nyu.edu.

Market Efficiency in the Age of Machine Learning

Abstract

As machines replace humans in financial markets, how is informational efficiency impacted? We shed light on this issue by exploiting unique data that allow us to identify when machines access company information (8-K filings) versus when humans access the same information. We find that increased access by machines, particularly from cloud computing services, significantly improves informational efficiency and reduces the price drift following information events. We address identification through instrumental variables, exogenous power and cloud outages, and a quasi-natural experiment. We show that machines are better able to handle linguistically complex filings and are less susceptible to bias from negative sentiment, whereas humans are better at combining incremental information.

Keywords: Market efficiency; Information acquisition; Artificial intelligence; Algorithmic trading.

JEL Classification Codes: G10; G12; G14.

"Thirty years ago the best fund manager was the one with the best intuition... Now those who take a "scientific approach", using machines, data and AI, can have an edge."

David Siegel, co-chairman of Two Sigma (*The Economist, October 5th, 2019*)

1. Introduction

As machines replace humans in financial markets, how is informational efficiency impacted? One possibility is that human biases are reduced, information is processed faster and in larger volumes, leading to improved price discovery and higher informational efficiency. Yet another possibility is that as machines replace humans, some of the "soft" information that humans (but not machines) are able to interpret is lost and consequently prices become less informative. While humans make a range of errors in interpreting information, machines are also not infallible, and they too can make different errors. Which of these opposing effects dominates is an empirical question that has broad implications for the functioning of financial markets and ultimately the efficient allocation of resources and risk.

We analyze the impact of machines on informational efficiency using a unique dataset in which we can distinguish between when a machine reads information released by a company versus when a human accesses the same information. We take the event-level Form 8-K viewership data from the SEC's EDGAR server. The Form 8-K filings are the means through which companies notify the market of important unscheduled corporate events including changes to a material agreement, financial information, acquisitions, substantial impairments, and any other events deemed important to shareholders. What distinguishes our paper from other studies of these company filings is that we separate each viewership of an 8-K filing into machine viewers and human viewers based on the downloading behavior captured by the server logs.[1] We further partition viewership at the level of the viewer's computing facility using IP addresses. No prior study to the best of our knowledge has examined material information acquisition at this granular level.

[1] Following the previous literature (Drake et al., 2015; Ryans, 2017), we identify information acquisition by a machine when the acquiring entity consistently downloads a large volume of 8-K filings beyond human comprehension within a short period of time.

First we show that machine viewership of 8-K filings has experienced significant growth in recent years.[2] Total machine viewership increases from 0.57 million views in 2008 to 8.78 million views in 2016, while total human viewership remains stable with an average of 0.3 million views per year. The viewership from cloud computing services, such as Amazon Web Services, was only 1% of the machine viewership before 2008, and the same figure increased to 14% in 2008, to 35% in 2012, and to 62% in 2016. This rapid growth coincides with the increase in popularity of both quantitative investment strategies, whose trading volume grew from 20% to 36% of institutional volume between 2014 and 2019, and cloud computing services, which offer many unique benefits to machine-based investment strategies, such as anonymity, stability, and minimum on-site maintenance.[3]

To understand the differences between how machines and humans access information, we compare their determinants. As expected, both human and machine viewers pay more attention to timely and complex 8-K. However, human viewers tend to focus on large and value firms, suggesting that human viewers pay more attention to well-known and mature firms. On the other hand, machine viewers have no preference for the type of firms, which is consistent with machines having less capacity constraints compared to humans that are forced to ration their scarce cognitive ability and attention. Human viewers also focus more on firms with fewer institutional shareholders that put considerable effort into monitoring firms and enhancing their disclosure transparency. The number of analysts following a firm does not have a significant impact on machine versus human viewership, suggesting that analyst coverage does not necessarily affect investors' attention and demand for information. Interestingly, human viewers have a tendency for increased viewership of 8-K filings that have negative sentiment content, but machines are indifferent to the sentiment of the 8-K.

A noticeable difference between machines and humans that helps explain our findings is the type of information that they choose to focus on. A useful feature of the 8-K filings is that they arrange events into topical categories. We find that while human viewers pay significantly more attention to earnings information (item 2.02), machine viewers pay significantly more attention to

[2] An article in *The Economist* on December 21, 2020, states that "For most of the past decade more trades have been done at high frequency by complex algorithms than by humans."

[3] Due to technological advances and market structure developments in recent decades, computer-based trading including algorithmic and high frequency trading account for the majority of trading in current financial markets (e.g., 75% of trading volume in the US in 2009 (Hendershott et al., 2011)).

corporate events not listed by the SEC but deemed material non-public information by the firm (item 8.01). These results are consistent with the notion that the greater computational capacity of machine-based strategies allow them to digest a broader set of unanticipated but important events that human viewers may neglect due to limited attention or limited cognitive capacity.

To quantify the impact of machines on informational efficiency, we measure the magnitude of absolute price drift following the 8-K publication date over the windows $(2, 10)$, $(2, 20)$, and $(2, 40)$ (relative to the 8-K publication date). Relatively more human viewership during the window $t \in \{0, 1\}$ is significantly associated with increased price drift following the 8-K publication date across various windows, consistent with an inefficient assimilation of the information. On the contrary, relatively more machine viewership, particularly from cloud machine services, is associated with more efficient incorporation of the information and no significant price drift following the 8-K publication date. This evidence is robust in all three event windows, various fixed effects, and remain qualitatively similar with alternative ways to define drifts. The results are also robust to an alternative measure based on a variance decomposition that separates noise and information. In terms of economic significance, a one-standard-deviation increase in the human viewership (holding fixed the level of machine viewership) yields about a 13.23% aggregate increase in the price drift in the month following the 8-K publication.

We further partition machine viewership (using IP addresses) into six categories: cloud computing services, financial institutions, data & media publishers, auditing & law firms, internet services providers, and a diverse group of other entities not related to the investment industry. This allows us to separate machine viewership that is unrelated machine-based investment strategies. We find that machine viewership from cloud computing services (most likely to be associated with machine-based investment strategies) has the strongest impact on stock price informational efficiency, being negatively related to post 8-K price drift. A one standard deviation increase in machine viewership from cloud services leads to a 10.49% aggregate reduction in price-drift over the month after 8-K filing date. Surprisingly, the machine viewership from financial institutions is not significantly related to post-event price drift, suggesting that the more sophisticated machine-driven investment strategies tend to be deployed via cloud computing servers and not via IP addresses known to belong to financial firms.

To examine whether the relations between machine viewership and informational efficiency reflect causal effects of machines, we use three identification strategies: instrumental variables, exogenous cloud and power outages, and a quasi-natural experiment. Our first approach is a two-stage least-squares (2SLS) estimation in which human (and machine) viewership is instrumented by market sentiment and macroeconomic news, which disproportionately affect human ability to assimilate information. Our second approach exploits exogenous disruptions to machines from cloud service and major power outage events. Our third approach uses the firm's inclusion in the S&P500 index as a quasi-natural experiment. Inclusion events lead to a sudden surge in viewership primarily from human analysts and fund managers who, unlike machines, have limited capacity to follow stocks outside of major indices (Farboodi et al., 2020). The results from all three of these identification strategies support the notion that machines cause an improvement in informational efficiency.

What is it that machines do better than humans to result in improved informational efficiency? We test several potential mechanisms. First, we investigate whether machines are better than humans at comprehending information when it is presented in a complex manner. Consistent with our conjecture, we find that machine viewership lessens the post-event price drift among the 8-Ks that are linguistically more complex from the perspective of a human. Second, machines may be less susceptible to human biases. Previous literature, such as Tetlock (2007), documents that media pessimism leads to inefficient price reactions as human traders show behaviors expected of noise traders. We find support for the notion that machines improve informational efficiency by being less susceptible to such biases — i.e., machine views have greater market efficiency improvements among the 8-Ks containing more negative sentiment.

The third mechanism that we examine is how well machines perform relative to humans in combining incremental signals in an 8-K filing with existing signals outside a given 8-K. Dugast and Foucault (2018) show that there is a trade-off between fast and automated decision-making versus slower but deeper analysis, which is potentially where humans might have an edge over machines. Our tests use item 2.02 of the 8-Ks, which usually reveals more precise information regarding the same earnings events that have been preliminarily disclosed in a press release. Our analysis indicates that the human, not machine, viewership of item 2.02 is associated with less price drift, consistent with our conjecture that humans have an advantage in deeper analysis that involves combining multiple sequential signals that have a degree of overlap.

Lastly, we examine whether the positive effects of cloud computing machines are due to better information processing (machines making informed decisions) or faster information processing, or both. We find that machine viewership is significantly positively related to measures of informed trading following the 8-K publication date, whereas for humans we get the opposite result. Our analysis also suggests that the cloud-machine viewership of 8-Ks is significantly associated with increased algorithmic trading activity around the information release, based on odd lot and trade size proxies. Therefore, the evidence is consistent with machines having an advantage in both the accuracy of information processing and speed.

Our paper makes two important contributions to the information disclosure literature. Prior research on information acquisition via SEC filings considers human heterogeneity and how it affects the informational efficiency of prices (Drake et al., 2015; Dyer, 2019). Another strand of literature proposes various machine learning models to show that computationally-intensive quantitative strategies can generate superior performance compared to traditional asset pricing models (Gu et al., 2020). They do not, however, investigate how information-based quantitative investment strategies impact the price discovery process. We fill this void by examining the impact of machine viewership of information on the price discovery process and find that quantitative investment strategies, especially those implemented through cloud computing services, help improve market efficiency.

Our paper also contributes to the literature on price informativeness by examining the specific mechanism of information acquisition by different types of investors in a casual setup. Prior studies suggest that sophisticated investors improve stock price efficiency (Akbas et al., 2015; Kokkonen and Suominen, 2015; Cao et al., 2018). Recent literature focuses on investors' capability of handling information and studies their impact on market efficiency. Chen et al. (2020b), for example, document that aggressive trades by hedge funds, based on acquired information from the SEC's EDGAR server, mitigates the impairment of market efficiency caused by analyst coverage reductions. Begenau et al. (2018) argue that certain investors can process larger data more effectively and therefore help price assets of firms with big data more accurately. Our paper expands this line of research by showing that information acquisition by machine-based investment strategies impacts post-announcement price drift differently from information acquisition by humans.

2. Theoretical Framework

Several branches of the literature help understand how machines could impact informational efficiency and the channels through which those effects occur. The related theory includes rational expectations models of endogenous information acquisition at one end, through to recent models of competing algorithmic investors and traders at the other. Also relevant are studies that analyze the relative efficacy of machine learning methods in assimilating information about stocks.

2.1. How do machines impact informational efficiency?

A key insight from rational expectations models is that when information acquisition and processing is costly, stock prices only partially reflect the available information and hence investor efforts in becoming informed are compensated (Grossman and Stiglitz, 1980; Kyle, 1985, 1989; Admati and Pfleiderer, 1988).[4] A direct consequence is that a reduction in the cost of information acquisition and processing will tend to increase the amount of information acquisition and improve stock price informativeness. Advances in computing and automation reduce the marginal costs of information acquisition and processing. Therefore, the first effect we expect as machines become more prevalent in investment settings is more information acquisition, likely accompanied by better informational efficiency of prices.

Recent studies of the role of machines and machine learning in financial markets argue that the cost of gathering and analyzing information has declined as machines replace humans. For example, machines have greater economies of scale in gathering and processing information (Chen et al., 2020a). Increased computational power and the use of machine learning algorithms can allow investors to outperform traditional empirical asset pricing models (Gu et al., 2020), make use of

[4] Sims (2003, 2006) further argues that investors' limited information-processing capacity discourages the information assimilation process into stock prices.

highly unstructured textual data (Bybee et al., 2020), and front-run slower traders (Van Kervel and Menkveld, 2019).

A wide range of behavioural and information processing effects are also relevant in understanding the effects of replacing humans with machines in financial markets. Machine-based strategies that make investment decisions following rules extrapolated from data may be less susceptible to factors that bias human decision-making including emotions, heuristics, and impulsiveness. They are also likely to be faster than humans, especially for demanding and complex tasks (Sims, 2003, 2006; Bradshaw et al., 2020).[5] Studies of heterogeneity among humans show that investors have different information-processing models and, in turn, respond to identical public signals differently (Xiong, 2013). Humans also trade for non-investment rationales.[6] Presumably due to these reasons, quantitative and machine-driven investors have grown rapidly and in 2019 account for approximately 36% of institutional stock trading volume and manage around 35% of US equity.[7]

In sum, given the potential for machines and machine learning methods to reduce information acquisition and processing costs (Chen et al., 2020a) and produce faster, less biased trading decisions based on firm-specific information (Sims, 2003, 2006; Peng, 2005; Veldkamp, 2006), our *main hypothesis* is: *Machine viewership of company information (SEC's Form 8-K) increases informational efficiency (relative to human viewership) and decreases the price drift following the 8-K publication.*

2.2. Shortcomings of machines and potential negative effects

Machines, like humans, are not infallible. Practitioners and academics have expressed concerns that computationally intensive black-box algorithms may end up identifying spurious correlations or data biases, and under-performing traditional methods in the long-run.[8] Critics have also argued that machine learning methods produce strategies that appear profitable, but load on difficult-to-arbitrage

[5]Bradshaw et al. (2020) argue that soft news is relatively more costly to process than hard news and markets are likely to react more strongly to analysts' processing of soft news than of hard news as hard news is more likely to have already been factored into prices.

[6]For example, Barber and Odean (2002) argue that overconfident traders can cause markets to underreact to the informational content of rational traders. Another strand of studies has identified numerous behavioral rationales for over-trading, including entertainment (Dorn and Sengmueller, 2009), sensation seeking (Barber and Odean, 2008; Grinblatt and Keloharju, 2009), gambling (Kumar, 2009; Cookson, 2018), and learning by doing (Linnainmaa, 2011).

[7]Anonymous. "March of the machines. The stock market is now run by computers" *The Economist*, Oct 5th 2019.

[8]For example, see Mark Hulbert (2020, Jan 5). Use AI for Picking Stocks? Not So Fast. *Wall Street Journal.* Retrieved from http://online.wsj.com.

stocks and therefore often fail when trading costs and other frictions are taken into consideration (Avramov et al., 2020). Moreover, machines are limited in their capacity to process soft information, compared to humans (Bradshaw et al., 2020).

Dugast and Foucault (2018) show that there is a trade-off between using information to make quick and automated decisions versus making slower and more deliberated decisions. The quick and automated decisions (reflecting machine-driven investment) have the advantage of acting on information faster and thereby reaping the profits of acting ahead of slower investors, but at the cost of sometimes making substantial errors that are more severe than those of slower and more deliberated decision-making. The slower and more deliberated decision-making of humans can benefit from the use of soft information and from combining low and high precision signals received at different points in time. It is therefore possible that while machines may speed up price discovery, they also introduce new types of errors into the price discovery process.

Another potential downside, which has been explored in the algorithmic trading literature, such as Weller (2017), is that automated models may have an adverse effect on stock price informativeness, despite their importance for transmitting available information into prices. For example, Baldauf and Mollner (2020) find that faster speeds enable high-frequency traders (HFT) to be more successful at order anticipation, which can have a negative effect on information production due to informed traders having less time to trade before HFTs react. Lee and Watts (2020) find that when the SEC's "Tick Size Pilot" increased the tick size and reduced algorithmic trading, pre-announcement stock returns are better able to predict the news of the upcoming earnings release, consistent with improved price discovery.[9]

Automated investment models that rely on fast processing of big data may not improve the price efficiency for stocks without large quantities of machine-readable data. For example, Begenau et al. (2018) investigate whether investors more efficiently price firms with big data. They find that big data disproportionately benefits large firms, allowing investors to produce more accurate forecasts and reduce uncertainty, which ultimately reduces the firm's cost of capital. In a similar spirit,

[9] On the contrary, Chordia and Miao (2020) use comprehensive intraday data and find that low-latency trading improves the long-term informational efficiency of stock prices.

Farboodi et al. (2020) find that the informativeness of stock prices, measured using the method proposed by Bai et al. (2016), varies significantly across groups of stocks, with large and growth stock prices increasingly reflecting information about future earnings. Therefore, there are potential negative effects from the tendency for machines to replace humans making it even more important to empirically test our main hypothesis.

2.3. Mechanisms and channels

Given that there are potentially both positive and negative effects of machines on informational efficiency, an important question is under what conditions the positive effects are likely to dominate? We begin with the issue of machine-readability. Hwang and Kim (2017) document that issuing financial disclosure documents with low readability causes firms to trade at significant discounts relative to their fundamentals. They argue that firms with difficult-to-read documents could evoke feelings of uncertainty and distrust among investors. In contrast, automated strategies lack such subjective components in their decision making processes. Cao et al. (2020) find that increased machine downloading activity motivates firms to prepare filings according to machines' readership and processing capability. This would further differentiate the impact of readability on the information assimilation speed of automated models from that of human readers. Therefore, we expect the positive effect of automated strategies on stock price efficiency to be pronounced when machines gather and analyze the information in 8-Ks that are difficult for humans to read.

Another strand of the literature suggests that investor sentiment affects stock prices. Tetlock (2007) uses the daily content from a popular *Wall Street Journal* column to proxy for investor sentiment and finds that high media pessimism predicts short-run downward pressure in stock prices. García (2013) uses the Loughran and Mcdonald (2011) dictionaries to measure the sentiment of *New York Times* articles, and finds that sentiment predicts stock returns during recessions. While the above studies show that investor sentiment impacts stock prices, it remains an open question whether machines are less susceptible to being influenced by negative sentiment. We hypothesize that the positive impact of machines on stock price efficiency is stronger when the information disclosure contains negative sentiment or takes place on negative sentiment days, which is when human decisions are most likely to be biased by sentiment.

Lastly, recent studies have theoretically investigated whether lower information processing cost and increased information transparency lead to more precise signals. Dugast and Foucault (2018) suggest that lower costs of gathering corporate disclosures (by machines) can cause a decline in the equilibrium demand for more precise signals about fundamentals, such as those that require more effort and deeper analysis. Humans may have an advantage over machines in slower but deeper analysis that combines sequential pieces of information. Hence, we expect that machines will not improve stock price efficiency or could even harm it when it comes to the sequential arrival of information that lends itself to a deeper analysis by human experts.

3. Data and Summary Statistics

We first describe our data sources and the variables used in our analysis. We then provide an overview of the cloud computing services that have revolutionised modern financial systems. Lastly, we provide descriptive statistics of the key variables.

3.1. Data sources and identifying machines vs humans

Our sample is made up of all firms in the CRSP and Compustat universe during a 14 year period from January 2003 to December 2016. We collect data on each viewing (referred to as a visit) of each 8-K filing (a notification to investors of an unscheduled material or extraordinary event that is important to shareholders) from the SEC's EDGAR server, giving us nearly 4 billion visits.[10] Each observation contains a partially anonymized IP address, time stamp, HTTP status codes (e.g., 200 for successful delivery), and crawler flag.[11] We obtain 8-K filings from WRDS SEC Analytics Suite (approximately 1.2 million 8-K filings). We restrict our sample to the filings that match firms in the CRSP and Compustat universe through the CIK-GVKEY link provided by WRDS SEC-Suite.

[10] To be precise, we have 3,966,935,088 visits to 1,203,881 8-K filings issued between January 2003 and December 2016. We remove the visits that fail to retrieve documents from EDGAR server (HTTP code$\neq$200), only visit index page (index dummy= 1) that come from search engines, and other similar web crawlers (crawler= 1).

[11] The raw files also include filing-specific information: Central Index Key (CIK) used in SEC's EDGAR server to identify filers, accession numbers that uniquely match a specific SEC filing, and files visited (e.g., exhibits or index file).

To protect the privacy and intellectual property of the filing viewers, the SEC redacts the last octet in their IP addresses. Similar to Chen et al. (2020a), we match IP addresses to organizations using a dataset from MaxMind by the first three octets in the IP address. In our sample, the proportion of the IP addresses that share the same first three IP octets, but come from different organizations, is small (3.46%). When multiple organizations are assigned to a redacted IP address, we use the organizations associated with most IP addresses within the anonymized octet for the IP address.

Next, we categorize the organizations of 8-K viewers into seven groups using data from Thomson Reuters Global Ownership and Capital IQ. The former data set contains approximately 1.6 million (financial) institutional investors and investment vehicles, such as mutual funds and hedge funds. The latter contains approximately 14.2 million public and private firms, as well as their subsidiaries. We label institutional investors and funds as "Institutional Investors" based on industry type from both databases, while we group the remaining organizations into "Auditing & Law Firms", "Clouding Services", "Media & Data Vendors", "Educations & Regulators", "Internet Services Providers" and "Others".[12]

We use two approaches to separate human viewership from machine viewership. We first follow Drake et al. (2015) to define visits from the IP addresses with any downloads per minute greater than 5, or IP addresses with more than 1,000 downloads per day as machine visits. Then, we follow Dechow et al. (2015) to count the visits from IP addresses with any downloads per minute greater than 25, or with the number of CIK's downloaded per minute being greater than 3, or with more than 500 downloads during the day as machine visits.

For each 8-K filing in our sample, we obtain the report date (Conformed Period of Report), the

[12]To facilitate the matching between the MaxMind's organization and entity names and the other two data sets, we use a fuzzy-name matching algorithm developed by WRDS, which involves human validation through Amazon's Mechanical Turk. To separate Internet Service Provider and Cloud Computing from other technology firms (in order to identify unmatched entitles), we first rely on Google to obtain the URL and short text results in the top 10 hits. We only search the organizations that have visited the EDGAR server for 8-K at least 100 times in our sample period. For the ones we cannot identify, we rely on Wikipedia and organizations' website that is obtained from the first 10 Google search hits. Our manual search allows us to reduce the visits from IP addresses from entities categorized as "Others", including entities from IP addresses that remain unidentified after manual searching, as well as the industries outside the remaining six categories. These exclusions account to be less than 3.86% of total 8-K visits in our final sample. The Appendix Table B reports top 10 Organizations of "Cloud Computing", "Institutional Investors", "Internet Provider Services", and "Others".

filing date, the SEC release date, the item type, and the text of the 8-K filing, including the text from any attached exhibits (Item 9.01). Almost all 8-K filings must be reported within *four* business days of the event date, with a very small number of exceptions (Ben-Rephael et al., 2020). To focus on the 8-K filings that carry new information, we limit the sample to filings that contain events that are released within four business days from the event date.[13]

We also use a number of other data sources. We obtain stock price data from CRSP and fundamental data from Compustat. We use I/B/E/S to for analyst coverage, and 13F data from Refinitive for ownership. We obtain the S&P500 inclusion dates from CRSP, quarterly earnings announcement dates from Compustat, could service from Gunawi et al. (2016) and power outage event data from Mukherjee et al. (2018).[14] To calculate the directional trading and algorithmic trading measures, we use NYSE Trade and Quote (TAQ) and the SEC's Market Information Data and Analytics System (MIDAS) available through WRDS. All variables are defined in the Appendix Table A.

3.2. Construction of efficiency measures

The absolute cumulative abnormal return (ACAR) of Ball and Brown (1968) and Fama et al. (1969) is a standard measure of the incorporation of information into prices. This measure constructs post-event price drift net of a predicted returns from a factor model, for example,

$$CAR_{i,t}^{0,T} = \sum_{t=0}^{T} \left(r_{i,t} - \alpha_i - \sum_{k=1}^{k} \beta_{i,k} f_{m,t} \right) = \sum_{t=0}^{T} \varepsilon_{i,t} \tag{1}$$

where $r_{i,t}$ is the raw return of stock i on date t and α_i and $\beta_{i,k}$ are estimated from a regression of k mimicking portfolios. The cumulative abnormal return (CAR) cumulates the abnormal return from announcement date $t = 0$ to T, and the $ACAR$ is the absolute value of this number. The $ACAR(0, T)$ measures the announcement information that enters prices through the announcement day to the day T. To measure the drift components of the information integrated into price, we proceed along

[13] On August 23, 2004, the SEC redefined qualifying events that trigger 8-K filers and accelerated the 8-K publication date. For more information about this rule see the SEC Financial Reporting Releases Nos. 33-8400 and 34-49424. To ensure data consistency, we match the items defined in the previous version of 8-K filings to the ones in the post-2004 version over our sample period.

[14] Gunawi et al. (2016) collects the cloud service outage from 2009 to 2015. We follow their method to obtain complementary cloud service outage events from 2003 through to 2016. Besides, we drop the outage events cloud service related to gaming (e.g., Play Station), business storage (e.g., Salesforce), and social media (e.g., Facebook). We also manually collect the major power outage events in the second half of 2016 from Google to complement Mukherjee et al. (2018), which provides the outage data from January 2000 to July 2016.

the lines of Meulbroek (1992) and take the difference between total post-event cumulative abnormal return and those over a short window post to 8-K announcement. Specifically, we use the absolute difference between post-announcement price variation and variation over a short window right after the announcement:

$$DRIFT(2,T) = \left| CAR_{i,t}^{0,T} - CAR_{i,t}^{0,1} \right| \tag{2}$$

with $T > 1$ to capture post announcement drift.

For robustness, as an alternative measure of inefficiency in how information is incorporated into prices, we use the noise return variance from Brogaard et al. (2021), which is based on a variance decomposition model. Furthermore, we also use alternative machine viewership measures, i.e. the fraction of machine viewership. All results remain qualitatively similar.

3.3. Summary statistics

Figure (1) plots annual number of visits in our sample by machine and human viewers, as well as the total number of visits. Figure (1) depicts the exponential growth of machine viewership of 8-K since 2007, whereas human viewership of 8-K has maintained its relatively low level over time.

Cloud computing services, such as Amazon Web Services, enable largely anonymous, on-demand network access to computing services over the Internet. These services include data backup, disaster recovery, machine learning, and big data analytics. Cloud computing services can protect users' identities, avoid internet service disruption, and access servers from the desired geographic locations. Therefore, cloud computing services are appealing to various financial institutions that otherwise need to maintain an in-house computing center. The technology hurdle to deploying investment strategies through cloud computing and the cost associated with its features mentioned above establish barriers to entry for less sophisticated investors to benefit from cloud computing. The trends in viewership of 8-K filings by cloud computing services are consistent with the overall trends of cloud computing development in recent years. As illustrated in Figure (2), cloud computing viewership represented only 1% of the machine viewership before 2008, 14% in 2008, 35% in 2012, and 62% in 2016. Therefore, cloud computing is an important means of information acquisition and processing, which is expected to directly affect market efficiency.

Table (1) provides the number of observations, mean, standard deviation, median, and the 25th and 75th percentiles of our key variables. To focus on the immediate viewership right after the 8-K announcement date, we restrict our sample to the visits on the SEC's 8-K filing day and the day following ($t \in \{0, 1\}$). This leaves us with 551,054 8-K filings with all main control variables available. Among other notable statistics reported in Panel A, machines view more 8-K filings (log viewership of 3.34 and 3.29, based on Drake et al. (2015) and Dechow et al. (2015) methods, respectively), on average, compared to humans (log viewership of 1.05 and 1.34, based on Drake et al. (2015) and Dechow et al. (2015) methods, respectively). The median and standard deviation of both groups indicate that machines are far more active in viewing 8-Ks than humans. Among other statistics reported in Panel A, the average firm size is about \$6.5m while the average book-to-market ratio is 0.75. The average institutional holding is 61%, with 4.1 analysts covering the firm, on average. Panel B reports viewership broken down into six item types that appear in 8-K filings. The most frequently viewed is item 2.02, which is largely related to quarterly and annual earnings announcements. The mean (median) 2-day viewership of a filing with item 2.02 is 69.91 (41) views.

Panel C reports viewership broken down by the viewer type using seven categories defined in Section (*3.1*). Viewership from "Cloud Computing" and "Internet Provider" are considerably greater than those from the "Rest" of organization types. The proportion of machine viewership from Cloud Computing (Others) is 98.33% (46.02%), the highest (lowest) among all seven categories. The total viewership per organization is heavily skewed. For example, the average number of 8-K views per Cloud Computing organization is 13,780, while the median is only 10 views with a standard deviation of 204,000. This indicates the 8-K viewership is highly concentrated – there is a small number of highly active viewers. Figure (2) plots the number of views by organization.

4. Machine and Human Viewership

We begin by investigating the factors that affect the variation in machine and human viewership of firm i's 8-K on days $t \in \{0, 1\}$, relative to the 8-K publication date. We then analyze the impact of machine and human viewership on informational efficiency. Next, we address endogeneity concerns using a quasi-natural experiment, exogenous shocks, and instrumental variables. Finally, we examine the channels driving the results.

4.1. Determinants of machine or human viewership activity of 8-K

To understand the drivers of $Machine$ and $Human$ viewership we estimate the following regression:

$$V_{i,t}^{M,H} = \alpha_0 + \sum_{j=1}^{k} \varphi_j \Gamma_{i,j,t} + \tilde{f} + \tilde{\tau} + \varepsilon_{i,t} \quad (3)$$

where $V_{i,t}^{M,H}$ is our measures of the number of 8-K views by machines or humans or both (total visits) on days $t \in \{0, 1\}$. The control variables, $\Gamma_{i,j,t}$, include several firm characteristics, market conditions, and 8-K characteristics (indexed by j), at time t, such as the negative sentiment ($FinNeg$) content in the 8-K, the Fog readability (FOG), the word count of text in the 8-K ($WordCount$), the days-to-release that captures the number of days between the event date and the 8-K filing date ($DaysRelease$), the number of items in the 8-K ($\#Item$), the firm $'sbook - to - marketratio$(BM), $thefirm$'s size ($SIZE$), firm i's institutional ownership ($InstOwn$), and the number of analysts following the firm i ($Analysts$). All variables are defined in the Appendix Table A. $\tilde{f}$ and $\tilde{\tau}$ are firm and year fixed effects, respectively.

Table (2) reports the results of these regressions. Column (1) shows that 8-K filings with considerable negative sentiment content tend to receive more total visits, as reflected by the positive, though marginally significant, $FinNeg$ coefficient. This tendency is strongest for humans (Columns 3 and 5) and is insignificant for machines (Columns 2 and 4). These results suggest that negative sentiment content tends to attract more attention from human viewers, possibly due to emotional effects for human investors. Column (1) also shows that larger 8-Ks are more likely to have a significantly larger total number of visits, as proxied by the word count in the 8-K ($WordCount$) and the number of items included in the 8-K ($\#Item$). Larger 8-Ks are likely to reveal more material information. As a result, both machines and humans more actively view larger 8-Ks (Columns 2-5). Column (1) also shows that timely 8-Ks are likely to receive more attention, as the number of days between the event date and the 8-K publication date ($DayRelease$) is negatively related to total viewership.

Table (2) also shows that larger firms ($SIZE$) have greater human viewership (Columns 3 and

5) but not significantly greater machine viewership (Columns 2 and 4). This result suggests that human readers tend to download the filings of larger and perhaps more established corporations with longer histories than those of smaller firms, consistent with Cao et al. (2020). Columns (3 and 5) also reveal that value firms (BM) have greater human viewership but not significantly greater machine viewership (Columns 2 and 4). Similarly, humans pay more attention to the 8-Ks of firms with low institutional ownership ($InstOwn$), which are likely to lack information transparency, yet machines show no preference over those firms. These findings are consistent with prior studies highlighting the implications of two commonly understood differences between humans and machines, which are due to their capacity and rationality (Abis, 2020).

We partition 8-K filings into topical categories based on the items contained within the 8-K and examine how human and machine viewership varies across the topical categories.[15] We regress the measures of human or machine viewership on a set of dummy variables for the different 8-K item types and control for the same factors as in Table (2).[16]

Table (3) reports the results of those regressions. They show that there is significant heterogeneity in the viewing activity of the different items in the 8-K by machines and humans. For example, column (1) shows that machines do not pay significant attention to item 2.02 that mostly refers to non-GAAP earning disclosures, as reflected in the insignificant item 2.02 coefficient. In contrast, humans pay significant attention to item 2.02 (column 6). Humans may have a particular preference for events that convey certain and well-known types of information, but machine viewers do not have such preference. Columns (2) and (7) show that machines, but not humans, pay significantly more attention to item 8.01 (i.e., Other Events). These results are consistent with the notion that machines rely on their ample computational capacity to identify material information that is not standardized in a typical, anticipated 8-K. Human viewers, in contrast, are more likely to target specific information disclosure events rather unspecified ones.

Humans tend to pay less attention to item 5.02, relative to other items in 8-K filings (Column 9).

[15] We focus on the five most common categories. As suggested by Panel B of Table 1, the rest of the categories altogether are around 13.15% of 8-K categories in our sample.

[16] While the main results are based on the measures of machine and human viewership proposed by Ryans (Ryans, 2017), in an unreported table we repeat the tests with the DRT measures (Drake et al., 2015) and find similar results.

Therefore, human readers appear to be less attentive to unanticipated events related to the departure or election of officers or directors. In contrast, machines do not significantly under allocate attention to item 5.02. Lastly, the total viewership by both machines and humans is higher for voluntary disclosure items 7.01 (i.e., Regulation FD Disclosure) and mandatory disclosure item 1.01 (i.e., Entry into a Material Definitive Agreement) as reported in columns (3, 5, 8, and 10).

In summary, our findings are consistent with the notion that humans and machines scrutinize specific 8-K items differently. Our results suggest that automated models are more attentive to new information disclosed in non-standardized 8-K items.

4.2. Viewership and market efficiency

In this section we investigate the effect of machine and human viewership activity of 8-K on the underlying firm's stock price efficiency. We hypothesize that machines will contribute to a faster and more efficient reaction of stock prices to new information. Therefore, we expect that the price drift following 8-K releases will be larger when there is a greater prevalence of humans relative to machines reading the filing. Among machines, we expect that the investors using cloud computing services are likely to be among the most sophisticated and therefore will have the largest positive impact on informational efficiency.

We estimate the following regression:

$$DRIFT(2,T)_{i,t} = \alpha_0 + \sum_{j=1}^{2} \beta_j V_{i,t}^{M,H} + \sum_{j=1}^{k} \varphi_j \Gamma_{i,j,t} + \tilde{f} + \tilde{\tau} + \varepsilon_{i,t} \tag{4}$$

where $DRIFT(2,T)_{i,t}$ is a measure of *inefficiency* defined as the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $V_{i,t}^{M,H}$ is our measures of the 8-K viewership by machines or humans. The control variables, $\Gamma_{i,j,t}$, include several firm and 8-K filing characteristics, such as firm i's book-to-market ratio, size, return-on-assets, leverage, standard deviation of monthly return over the year prior to the 8-K, institutional ownership concentration, and analyst coverage. All variables are defined in the Appendix Table A. $\tilde{f}$ and $\tilde{\tau}$ are firm and year fixed effects, respectively.

Table (4) reports the results. Columns (1-3) provide the first evidence that the price drift following the 8-K publication date is mainly associated with human viewership of the 8-K, rather than by aggregated machine viewership. In particular, column (1) shows that the *aggregate* machine viewership (including all categories of machines) of 8-K has no significant effect on the price drift following the 8-K publication date, as shown by the insignificant $Machine$ coefficient. In contrast, column (2) shows that 8-K viewership by humans has a positive and highly significant effect on the price drift. Column (3) simultaneously accounts for viewership by machines and humans and confirms the findings reported in columns (1-2).

In columns (4-6) of Table (4) we extend our analysis by increasing the $DRIFT(2,T)$ window from $DRIFT(2,10)$ to $DRIFT(2,20)$, and in column (7-9) to $DRIFT(2,40)$, with the same control variables. We find consistent and even stronger evidence regarding the impact of viewership by machines and humans on the price drift.[17] We also examine an alternative measure of inefficiency in Section (*4.4*) and find similar results. The effects of human viewership on the price drift are also economically significant. For example, a one standard deviation increase in human viewership is associated with a 13.23% aggregate increase in the price drift over the post-publication month.[18] Therefore, the well-established inefficiencies from overreaction and underreaction to information in financial markets is driven by humans and not machines.

The regressions control for several firm characteristics. Across all columns (1-9) we find that firm size and ROA are negatively related to the price drift post to 8-K filing date suggesting that larger and more profitable firms have more efficient prices. Consistent with Ben-Rephael et al. (2020), columns (1-9) show a negative and highly significant relation between the institutional ownership and the price drift post to 8-K filing date, with the coefficient of $InstOwn$ ranging between -0.012 and -0.022 (significant at the 1% confidence level).[19]

[17]We also repeat our analysis with an alternative machine viewership, which is defined as the fraction of machine or cloud machine viewership to total visits. The results, reported in the Appendix Table C are qualitatively similar. We also repeat all tests with the machine and human viewership measured by the DRT (Drake et al., 2015) method and the results are qualitatively similar.

[18]The standard deviation of raw human visits in the first two days is 51.96, and its log is $ln(1 + 51.96) = 3.97$. Given average number of post-event price drift, $DRIFT(2,20)$ is 0.09, one standard deviation increase in human visits leads a 13.23% (0.003\$ $\times$ \$3.97/0.09) increase in post-event price drift.

[19]In their analysis of abnormal institutional attention and price discovery during the filing period, Ben-Rephael et al. (2020) find that price discovery prior to the filing period is 9.7% higher when institutional investors are paying attention, which ultimately results in a reduction of 9.7% in the subsequent price discovery during the filing period.

4.3. Machine viewership, hosting facilities, and price drift

Machines are far from a monolithic group. In fact, a considerable proportion of machine-based viewership is not directly associated with sophisticated investors, nor with investment strategies (such as viewership from university, regulators, and auditing firms). Furthermore, as we discussed above, we expect that the investors using cloud computing are likely to be among the most sophisticated and therefore will have the largest positive impact on informational efficiency. We test these differences between types of machines by using the different categories that we identified on the basis of the machine's IP address: (a) cloud computing services, (b) financial institutions, (c) data & media publishers, and (d) a combined type (others) for those entities not associated with the professional investment practice, including auditing & law firms, internet services providers, and a diverse group for the entities. We regress the inefficiency measure, $DRIFT(2,T)_{i,t}$, on measures of viewership by each of the machine types. We include the same control variables as before: BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$.

Table (5) reports the results from this analysis. The results (columns (1) and (5)) show that machine viewership from cloud computing services does indeed have a significant positive association with informational efficiency (negative association with drift). In terms of economic significance, a one standard deviation increase in the machine-based viewership of 8-K from cloud computing services is associated with a 10.49% aggregate decrease in the price drift over the post-filing month.[20] Extending the price drift window up to 40 trading days following the 8-K publication date, the machine viewership of 8-K from cloud computing services is no longer statistically significant although the point estimate remains the same. These results are consistent with more variability and less statistical power in the longer horizon measure, or that investors using cloud computing services tend to exploit their informational advantage over a short-term horizon.

The estimates in columns (2), (6), and (10) of Table (5) show that machine viewership from financial institutions other than those using cloud computing facilities ($InstMachine$) and machine

[20]The unreported standard deviation of raw machine visits from clouding services in the first two days is 110.98, and its log is $ln(1 + 110.98) = 4.72$. Given average number of post-event price drift, $DRIFT(2,20)$, is 0.09, one standard deviation increase in machine visits from cloud services leads a 10.49% ($= 0.002 \times 4.72/0.09$) decrease in post-event price drift.

viewership from data & media publishers ($DataMachine$) does not have a significant effect on the price drift. Viewership from other machine types ($OtherMachine$) is positively associated with inefficiency as measured by price drift.[21]

4.4. Evidence from alternative measures

To test the robustness of our main results, we re-estimate the regressions from the previous subsections but using a different measure of information (in)efficiency – the $Noise$ measure developed by Brogaard et al. (2021). This measure is the standard deviation of estimated pricing errors obtained from a variance decomposition model that separates information and noise. It captures both underreaction and overreaction to information as two forms of inefficiency that contribute to "noise" in prices. In a series of validation tests, Brogaard et al. (2021) show that the $Noise$ measure captures informational inefficiency in prices.

Similar to the previous subsections, we regress $Noise$ on measures of viewership by each of the machine types and include the same control variables as before. The results support the conclusions made based on the drift measure.[22] Specifically, human viewership has a significant positive association with noise in prices, while cloud computing machine viewership has a significant negative association with noise. These results therefore support the notion that access to information by machines operating from cloud computing services tends to improve the informational efficiency of prices.

5. Addressing Endogeneity

While the evidence so far is consistent with our main hypothesis, to more formally examine the causality between human/machine access to information and informational efficiency, we exploit three identification strategies: (a) instrumental variables, (b) exogenous shocks, and (c) a quasi-natural experiment.

[21] Once again, in an unreported table in which we repeat our tests with the machine-based viewership defined by DRT (Drake et al., 2015), the results are qualitatively similar.

[22] The results are reported in the Appendix Table D.

5.1. Instrumental variables approach

We first address potential endogeneity using a two-stage least-squares (2SLS) framework in which machine or human viewership is instrumented by various market sentiment proxies and macroeconomic news. Our tests are based on the notion that investor sentiment (measured using the index of Baker and Wurgler (2006)), macroeconomic news announcements (Peng and Xiong, 2006; Kacperczyk et al., 2016), changes in the VIX during the window $(-20,-1)$ relative to 8-K publication date, and the CRSP value-weighted market return over the same window $(-20,-1)$ ($CRET(-20,-1)$) are likely to disproportionately affect humans compared to machines. For example, humans are more likely to be affected by sentiment and more likely to be constrained in their ability to process relevant information during periods of intense information arrival. While these instrumental variables are likely to affect market efficiency, their effects are through the ability for market participants (machines and humans) to access and assimilate relevant company-specific information. Therefore, our first-stage regression is:

$$\begin{aligned} V_{i,t}^{M,H} = \alpha_0 \quad &+ \quad \beta_1 InvSent + \beta_2 MacroNews + \beta_3 VIX(-20,-1) + \beta_4 CRET(-20,-1) \\ &+ \quad \sum_{j=1}^{k} \varphi_j \Gamma_{i,j,t} + \tilde{f} + \tilde{\tau} + \varepsilon_{i,t} \end{aligned} \tag{5}$$

where $V_{i,t}^{M,H}$ is machine or human viewership of firm i's 8-K on days $t \in \{0,1\}$, measured by the Ryans (2017) method. We estimate the first-stage equation for machines or humans separately. $InvSent$, $MacroNews$, $VIX(-20,-1)$, and $CRET(-20,-1)$ are our instruments. $\Gamma_{i,j,t}$ is a vector of controls j including BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$ and $Analysts$. $\tilde{f}$ and $\tilde{\tau}$ are firm and year fixed effects.

Table (6) reports the results. Columns (1) and (2) show that our instruments significantly affect human, but not on machine, viewership of 8-K filings as hypothesized. Human viewership is negatively related to investor sentiment, suggesting that humans are less likely to access information related to unscheduled events when investor sentiment is high. Human viewership is also negatively related to the release of macroeconomic news at the time of the 8-K publication date. This result is consistent with humans having attention and cognitive constraints that make them less attentive to firm-specific information when there is substantial market-wide information, e.g., (Peng and

Xiong, 2006; Kacperczyk et al., 2016). Lastly, human viewership is also negatively impacted by the $VIX(-20,-1)$ and the market return $CRET(-20,-1)$, consistent with humans having a tendency to read more 8-Ks in rising markets and during lower volatility. The Stock-Yogo test of the null hypothesis: $\hat{\beta}_3 = \hat{\beta}_4 = \hat{\beta}_5 = \hat{\beta}_6 = 0$ is rejected as F-statistic is in excess of 10 for human viewership, but as expected, we cannot reject the same null hypothesis for machine viewership. Overall, the first-stage regressions confirm that the four instruments disproportionately affect human 8-K viewership.

The second-stage regressions are:

$$DRIFT(2,T)_{i,t} = \alpha_0 + \sum_{j=1}^{2} \beta_j \widehat{V}_{i,t}^{M,H} + \sum_{j=1}^{k} \varphi_j \Gamma_{i,j,t} + \tilde{f} + \tilde{\tau} + \varepsilon_{i,t} \tag{6}$$

where $DRIFT(2,T)_{i,t}$ is our baseline measure of *inefficiency*, defined earlier in Equation (2), $\widehat{V}_{i,t}^{M,H}$ are the fitted values of machine and human viewership, $\Gamma_{i,j,t}$ is the same vector of control variables, and $\tilde{f}$ and $\tilde{\tau}$ are firm and year fixed effects.

The second-stage results are also in Table (**6**), columns (3-8). They are consistent with our baseline results discussed earlier. Namely, human viewership is significantly positively related to price drift whereas machine viewership, pooling across all the machine types, is not detrimental to market efficiency. Hence the instrumental variables models reinforce the estimates reported in Table (**4**) and discussed in Section (*4.2*).

5.2. Evidence from exogenous disruptions

Our second approach is based on exogenous cloud service and major power outages that disproportionately disrupt machines. We use major power outage events in the US as the disruption to machine-based viewership. To focus on the events that are likely to affect many investors using automated models, we consider a day as having an outage (an *Outage Day*) if it hits one of major cloud service providers, such as Amazon Web Services, or impacted at least 500,000 customers and lasted for at least 10 hours. Such outages, by interrupting the connectivity of computing facilities running automated investment models, are expected to reduce the ability for machines to improve price discovery.

We regress $DRIFT(2,T)_{i,t}$ on machine and human viewership and a dummy variable that is one on *outage* days and zero otherwise. We include interactions between cloud computing, machine and human viewership and the *outage* days dummy, and control for our standard list of stock characteristics: BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$.

Table (7) reports the results. As expected, we find that the coefficients of the interaction between $CouldMachine$ or $Machine$ viewership and *Cloud Outage Day*, $\hat{\beta}_2$ and $\hat{\beta}_4$, are positive and significant in columns (1) and (2). These estimates suggest that during unexpected cloud outage days the viewership through some clouds is less effective in reducing the price drift post to 8-K publication day, and hence there is a corresponding deterioration in market efficiency. The relationship remains consistent up to 20 trading days after the 8-K publication day (columns 4 and 5), though it becomes statistically weaker. In contrast, we find no significant effects for the interaction term between human viewership and major cloud outages (insignificant $\hat{\beta}_6$ coefficients in columns (3) and (6)).

Similarly, the coefficient of the interaction between $Machine$ viewership and *Power Outage Day* ($\hat{\beta}_4$) is also positive and significant in column (8), though the coefficient of interaction between cloud viewership and power outage is statistically insignificant. Besides, the significant relationship is only present up to 10 trading days after the 8-K publication day. Lastly, no significant effects are found for the interaction term between human viewership and major power outages (insignificant $\hat{\beta}_6$ coefficients in columns (9) and (12)). These results suggest that the power outages are more likely to interrupt the functionality of computing facilities that are not as geographically diversified or not as technologically advanced as major cloud service providers.

5.3. S&P500 inclusion as a quasi-natural experiment

Our third approach uses a firm i's inclusion in the S&P500 index as a quasi-natural experiment (Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Bennett et al., 2020). A stock is added to the S&P500 index only when another stock is excluded, mainly due to major corporate actions such as mergers, bankruptcy, and spin-offs. The sudden surge in institutional investors' attention for the newly added stock creates an exogenous shock in viewership, disproportionately from human analysts and fund managers due to their limited capacity to follow stocks outside of their mandate

(Farboodi et al., 2020).[23] Human attention is a scarce cognitive resource (Barber and Odean, 2008; Kahneman, 1973) and influences the market's reaction to corporate announcements (e.g., Hirshleifer et al. (2009); Kempf et al. (2017)). The existing literature also shows that a stock's inclusion in the S&P500 index attracts more sophisticated and skilled investors (Chen et al., 2004). This increase in sophistication and skill is another reason why inclusion in the S&P500 index is expected to improve the impact of humans on informational efficiency.

In contrast, machines are less likely to be constrained in their capacity to process information and therefore, as we showed earlier, show less of a tendency to favor particular types of stocks in their information acquisition activities. We therefore expect inclusion in the S&P500 index to have little or no effect on how machines impact informational efficiency.

We use difference-in-differences regressions to examine the joint effects of 8-K viewership by humans or machines and the firm's inclusion in the S&P500 index on the price drift following 8-K filings. The dependent variable is the $DRIFT(2,T)_{i,t}$ measure. As the key independent variables, we have interactions of machine or human viewership of firm i's 8-K with a dummy variable for S&P500 index inclusion ($InIndex_t$). This allows us to also include the $InIndex_t$ variable by itself to capture any index effects that are not associated specifically with human or machine viewership of 8-Ks. We also include the human and machine viewership variables by themselves to capture their effects that are not related to index inclusion, firm and year fixed effects, and a range of control variables. We also limit the sample in these tests to stocks that are included in the S&P500 index at some stage during our sample period to reduce the effect of other firm characteristics associated with the price drift.

Table (8) reports the results of the difference-in-difference tests. The primary coefficients of interest are for the interaction term between human viewership and index inclusion, $V^H_{i,t} \times InIndex_t$. These coefficients are negative in all specifications (columns 2, 4 and 6) and statistically significant at the 5% or 10% levels. Consistent with our hypothesis, the results suggest that index inclusion improves the impact of humans viewership on informational efficiency. That is, all else equal, hu-

[23] Begenau et al. (2018) examine whether big data disproportionately benefits large firms by splitting firms by S&P500 index membership. Similarly, Farboodi et al. (2020) examine whether firms in the S&P500 index have different stock price informativeness.

man viewership in S&P500 stocks is more likely to improve efficiency and reduce drift than human viewership in non-S&P500 stocks.

In contrast, the coefficient of $V_{i,t}^{M} \times InIndex_t$, is insignificant, suggesting that S&P500 inclusion does not significantly affect the impact of machines on price drift. This result is also consistent with our hypothesis that, unlike human viewership, automated models are less resource-limited and are not concentrated in index stocks.

Overall, the results from the instrumental variables tests, the exogenous outages, and the quasi-experimental tests of index inclusion events support a causal interpretation of our baseline results about how humans and machines impact price drift following 8-K filings. In an unreported table, we repeat our tests with the machine and human viewership measured by DRT (Drake et al., 2015), and find qualitatively similar results.

6. The Mechanisms

In this section, we explore a few key mechanisms through which machines and humans impact informational efficiency. First, if automated models are built on predefined rules or extrapolated by machine learning, they should be capable of handling complicated textual records better than humans. Second, machines should also be less susceptible to emotions and therefore should be better able to handle information in the presence of negative sentiment content that is known to elicit bias in human decisions. Third, we explore how well machines and humans process and combine incremental signals for the same events, drawing on the theoretical predictions of Dugast and Foucault (2018) about the trade-off between fast and noisy versus slow and accurate information processing. Finally we examine whether the positive effects of cloud-based machines is associated with more accurate processing of information in 8-Ks, faster processing of the information, or both, by examining measures of informed trading and algorithmic trading.

6.1. Readability

We first test how well humans and machines impound hard-to-read information by examining linguistically complex 8-Ks. We use two proxies of *readability* to test the effectiveness in processing and assimilating such information: the *Gunning FOG*, and the *Flesch-Kincaid*. We label them as *DifficultToRead*. We regress $DRIFT(2,T)_{i,t}$ on machine and human viewership activities of 8-K filings, the two measures of 8-K *DifficultToRead*, and importantly interactions between machine and human viewership and *DifficultToRead*. We control for BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$.

The results are reported in Table (**9**) using *Gunning FOG* in columns (1-4) and *Flesch-Kincaid* in columns (5-8). Consistent with our conjecture, we find that machine 8-K viewership (in columns 1 and 5) lessens the price drift following 8-K filings by a larger amount for harder-to-read 8-Ks, which is information that will be more challenging for human readers to interpret. Another possible interpretation of this finding is consistent with recent evidence suggesting that the increasing machine downloading activity motivates firms to prepare filings according to machines' readership, processing capacity and capability (Cao et al., 2020). In contrast, the interactions between human viewership of 8-Ks and *DifficultToRead* measures are not statistically different from zero (columns (3-4) and (7-8)).

6.2. Negative sentiment

We examine whether machine viewers, in particular those associated with automated investment decisions, are better able to handle negative sentiment content in 8-Ks. We regress $DRIFT(2,T)_{i,t}$ on (a) machine viewership from cloud computing services, (b) machine viewership from financial institutions, and (c) negative sentiment in a 8-K defined by Loughran and Mcdonald (2011), as well as interactions between cloud machine and institutional investor machine with negative sentiment. We include a wide range of controls such as BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$.

Table (**10**) reports the results. The results support the notion that machines are better at handling negative sentiment without bias and therefore have a larger positive effect on informational efficiency (larger negative effect on drift) when information is more likely to induce bias in humans due to negative sentiment. Specifically, columns (1), (3), and (5) show that machine viewership from cloud computing services reduces the price drift more among the 8-K with considerable negative sentiment content. The interaction term coefficients are negative and significant up to two months following the 8-K publication date. Along similar lines, results in columns (2), (4), and (6) show that machine viewership from financial institutions also reduces the price drift following 8-Ks with higher negative sentiment content, even though the machine viewership from institutional investors does not affect price drift overall. When we replace machine viewership from cloud computing services with aggregated machine viewership, the interaction variables are no longer significant, suggesting machine viewership from cloud computers and financial institutions play an important role in correcting the emotional biases of humans.

6.3. Incremental information and earnings announcement drift

Guided by the theoretical predictions of Dugast and Foucault (2018), we examine whether machines or humans are better at combining sequential signals to produce more precise information. The theory shows there is a trade-off between fast and noisy responses to early signals and slower but more accurate responses that potentially combine more signals. We expect that the incremental signals are less likely to affect the information processing preciseness of automated models, which are heavily relying on the information provided in an 8-K filing, than human viewers, including experienced analysts who focus on a small number of firms. To examine this difference, we use item 2.02 of the 8-Ks, which largely reveals the same earnings information that has already been disclosed in the press release in an abbreviated format. We regress $DRIFT(2,T)_{i,t}$ on machine and human viewership activities, a dummy variable that is one for 8-K filings that include $item2.02$, and zero otherwise, as well as interactions between machine and human viewership and the $item2.02$ dummy. We include standard controls: BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$. All variables are defined in the Appendix Table A.

The results reported in Table (11) indicate no evidence that incremental information in $item2.02$

affects the impact of machine viewership on price drift. In contrast, the interaction term between human viewership and $item2.02$ has a negative and significant effect on the price drift up to 10 or 20 days following the 8-K publication date. Therefore, based on the theoretical prediction of Dugast and Foucault (2018), our analysis shows that automated models are limited in their ability to use incremental signals to produce more precise information, whereas humans are better able to use incremental signals regarding earning announcements more efficiently.

6.4. Viewership and informed trading

Is the positive effect of cloud computing machines on informational efficiency the result of machines being better at interpreting the information (more informed), faster at interpreting the information, or both? To shed some light on this issue, we first test the relation between human and machine viewership and measures of informed trading (this subsection) and then the relation between viewership type and algorithm trading (next subsection).

The microstructure-based Probability of Informed Trading (PIN) is widely used as a measure of the time-varying information asymmetry in financial markets. It captures the information advantage held by better informed investors (Bharath et al., 2009) based on the properties of the order flow. We use the PIN to test how the level of informed trading varies depending on the levels of human and machine viewership of 8-Ks. We regress $PIN(0,T)$, which is the average daily Probability of Informed Trading of firm i over the window $(0,T)$ on machine and human viewership, controlling for BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$.

Table (**12**) presents our findings from this analysis. Machine viewership on days $t \in \{0,1\}$ is positively related to $PIN(0,1)$ and $PIN(0,5)$ (columns 1 and 5). In contrast, we find a significantly negative relation between human viewership and $PIN(0,1)$ and $PIN(0,5)$ (columns 2 and 6). We further explore the results by machine category. We find that machine viewership from cloud computing services is predominant positive driver of the position relation with $PIN(0,1)$ and $PIN(0,5)$ (columns 3 and 7).

In summary, the results indicate that the machine (human) viewership is significantly positively (negatively) related to the probability of informed trading. These results support the earlier evidence that cloud computing machines in particular contribute to price discovery and improve informational efficiency, while humans tend to be less well informed traders.

6.5. Viewership and algorithmic trading

Automated investment decisions implemented through algorithmic trading are likely to be faster in responding to new information than manual trading decisions and in particular manual handling of order execution. This speed advantage of machines relative to humans is possibly one of the contributors to the positive effect of cloud computing machines on informational efficiency. To examine this notion, we test the relation between algorithmic trading, proxied by (a) the Odd Lot Ratio, and (b) the Trade Size, and contemporary machine and human 8-K viewership.[24] Algorithmic trading usually splits large trades into smaller order to spread out trade over time, blend in with other order flow, and avoid a large price change. For this reason, a smaller trade size and larger proportion of odd lots is often considered to reflect algorithmic trading. An automated model that improves market efficiency would involve both textual data analysis and algorithmic trading.

We regress either the Odd Lot Ratio or the Trade Size (as dependent variables in separate estimations) on machine and human viewership. We control for BM, $SIZE$, ROA, LEV, STD_{RET}, $InstOwn$, and $Analysts$. Table (**13**) presents our findings from this analysis. There is interesting heterogeneity in the impact of human and machine viewership of 8-Ks on days $t \in \{0, 1\}$ on the distribution of $OddLotRatiot(0, 1)$. As expected we find that human 8-K viewership is negatively related to the amount of algorithmic trading proxied by the $OddLotRatio(0, 1)$ (column (2)). In contrast, we find that the machine 8-K viewership from cloud computing services is positively related to the $OddLotRatio(0, 1)$ as indicated in column (3). Hence machine viewership from cloud computing services leads to more algorithmic trading.

We also examine how human and machine viewership of 8-Ks affects $TradeSize(0, 1)$ in columns (5-8). We find that human 8-K viewers are positively related to the trade size (column 6). In contrast,

[24] As algorithmic trading based on newly arrived 8-K filings is likely to involve fast trading as investors race to trade on the basis of the information, we expect to see predominantly informed market orders. These should be reflected in the Odd Lot Ratio and the Trade Size suggested by Weller (2017), but not necessarily in other proxies of algorithmic trading that are more targeted at measuring algorithmic market making, such as the Order-to-Trade Ratio or the Cancel-to-Trade Ratio.

machine 8-K viewership from cloud computing services is negatively related to the trade size (column 7). Therefore, cloud machine (human) viewership is significantly and positively (negatively) related to algorithmic trading around the publication date of the 8-K.

7. Conclusion

Advances in computing and machine learning are rapidly changing how investors acquire and use firm-specific information. We find that machine viewership of company 8-K filings has increased exponentially during recent years and now represents a significant fraction of total visits of 8-Ks in the SEC's EDGAR server.

Using novel data, we show that these changes in how information is accessed and used affect how the information is impounded into stock prices. Information acquired by machines from cloud computing servers significantly improves informational efficiency and decreases price drift following information releases. In contrast, humans accessing the same information do not benefit informational efficiency as much and can even harm efficiency. We overcome identification problems by examining the causal effect of machine and human viewership on the price drift using instrumental variables, exogenous cloud service and major power outages, and a quasi-natural experiment.

We find that machine viewership improves informational efficiency the most when the information contained in the 8-K filing is more linguistically complex and therefore harder for a human to read. Machine viewers from cloud computing services and also financial institution are less susceptible to bias from negative sentiment than humans and therefore significantly improve efficiency in how markets digest news that has a negative content or tone. We also find that sequential/incremental signals may be more difficult for automated models to combine than they are for a human and this is one aspect of information processing where humans may currently still have an advantage. Our analysis further shows that the machine (human) viewership is significantly and positively (negatively) related to the daily probability of informed trading. Finally, we find that information viewership by cloud computing machines is significantly associated with the level of algorithm trading activity.

Overall, our findings uncover the important role of machine-based quantitative investment or trading strategies in assimilating and incorporating information in the underlying firm's stock prices. Our findings contribute to the ongoing debate on the role of machines in financial markets and their impact on informational efficiency.

References

Abis, S., 2020. Man vs . machine : Quantitative and discretionary equity management. Columbia University Working Paper pp. 1–79.

Admati, A. R., Pfleiderer, P., 1988. A Theory of Intraday Patterns: Volume and Price Variability. The Review of Financial Studies 1, 3–40.

Akbas, F., Armstrong, W. J., Sorescu, S., Subrahmanyam, A., 2015. Smart money, dumb money, and capital market anomalies. Journal of Financial Economics 118, 355 – 382.

Avramov, D., Cheng, S., Metzker, L., 2020. Machine learning versus economic restrictions: Evidence from stock return predictability. Working Paper pp. 1–72.

Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative? Journal of Financial Economics 122, 625–654.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680.

Baldauf, M., Mollner, J., 2020. High-frequency trading and market performance. The Journal of Finance 75, 1495–1526.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of Accounting Research 6, 159–178.

Barber, B., Odean, T., 2002. Online investors: Do the slow die first? Review of Financial Studies 15, 455–488.

Barber, B. M., Odean, T., 2008. 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.

Begenau, J., Farboodi, M., Veldkamp, L., 2018. Big data in finance and the growth of large firms. Journal of Monetary Economics 97, 71–87.

Ben-Rephael, A., Da, Z., Easton, P., Israelsen, R. D., 2020. Who pays attention to sec form 8-k? Working Paper pp. 1–51.

Bennett, B., Stulz, R. M., Wang, Z., 2020. Does joining the s&p 500 index hurt firms? Working paper series, Ohio State University, Charles A. Dice Center for Research in Financial Economics.

Bharath, S. T., Pasquariello, P., Wu, G., 2009. Does asymmetric information drive capital structure decisions? The Review of Financial Studies 22, 3211–3243.

Bradshaw, M. T., Lock, B., Wang, X., Zhou, D., 2020. Soft information in the financial press and analyst revisionsanalyst revisions. The Accounting Review .

Brogaard, J., Nguyen, T. H., Putniņš, T. J., Wu, E., 2021. What moves stock prices? the role of news, noise, and information. Review of Financial Studies Forthcoming.

Bybee, L., Kelly, B. T., Manela, A., Xiu, D., 2020. The structure of economic news. Tech. rep., National Bureau of Economic Research.

Cao, C., Liang, B., Lo, A. W., Petrasek, L., 2018. Hedge fund holdings and stock market efficiency. The Review of Asset Pricing Studies 8, 77–116.

Cao, S., Jiang, W., Yang, B., Zhang, A. L., 2020. How to talk when a machine is listening: Corporate disclosure in the age of ai. Available at SSRN 3683802 .

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

Chen, H., Noronha, G., Singal, V., 2004. The price response to s&p 500 index additions and deletions: Evidence of asymmetry and a new explanation. The Journal of Finance 59, 1901–1930.

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

Chordia, T., Miao, B., 2020. Market efficiency in real time: Evidence from low latency activity around earnings announcements. Journal of Accounting and Economics 70, 101335.

Cookson, J. A., 2018. When saving is gambling. Journal of Financial Economics 129, 24–45.

Dechow, P. M., Lawrence, A., Ryans, J. P., 2015. SEC Comment Letters and Insider Sales. The Accounting Review 91, 401–439.

Dorn, D., Sengmueller, P., 2009. Trading as entertainment? Management Science 55, 591–603.

Drake, M. S., Roulstone, D. T., Thornock, J. R., 2015. The determinants and consequences of information acquisition via edgar. Contemporary Accounting Research 32, 1128–1161.

Dugast, J., Foucault, T., 2018. Data abundance and asset price informativeness. Journal of Financial Economics 130, 367 – 391.

Dyer, T., 2019. Does public information acquisition level the playing field or widen the gap? an analysis of local and non-local investors. Working Paper pp. 1–63.

Fama, E. F., Fisher, L., Jensen, M. C., Roll, R., 1969. The adjustment of stock prices to new information. International Economic Review 10, 1–21.

Farboodi, M., Matray, A., Veldkamp, L., Venkateswaran, V., 2020. Where Has All the Data Gone? NBER Working Papers 26927, National Bureau of Economic Research, Inc.

García, D., 2013. Sentiment during recessions. The Journal of Finance 68, 1267–1300.

Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. Journal of Finance 64, 549–578.

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33, 2223–2273.

Gunawi, H. S., Hao, M., Suminto, R. O., Laksono, A., Satria, A. D., Adityatama, J., Eliazar, K. J., 2016. Why does the cloud stop computing? lessons from hundreds of service outages. In: Proceedings of the Seventh ACM Symposium on Cloud Computing, pp. 1–16.

Hendershott, T., Jones, C. M., Menkveld, A. J., 2011. Does algorithmic trading improve liquidity? The Journal of Finance 66, 1–33.

Hirshleifer, D., Lim, S. S., Teoh, S. H., 2009. Driven to distraction: Extraneous events and underreaction to earnings news. The Journal of Finance 64, 2289–2325.

Hwang, B.-H., Kim, H. H., 2017. It pays to write well. Journal of Financial Economics 124, 373–394.

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2016. A rational theory of mutual funds' attention allocation. Econometrica 84, 571–626.

Kahneman, D., 1973. Attention and Effort. Prentice-Hall series in experimental psychology, Prentice-Hall.

Kempf, E., Manconi, A., Spalt, O., 2017. Distracted Shareholders and Corporate Actions. The Review of Financial Studies 30, 1660–1695.

Kokkonen, J., Suominen, M., 2015. Hedge funds and stock market efficiency. Management Science 61, 2890–2904.

Kumar, A., 2009. Who gambles in the stock market? The Journal of Finance 64, 1889–1933.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A. S., 1989. Informed speculation with imperfect competition. The Review of Economic Studies 56, 317–355.

Lee, C., Watts, E. M., 2020. Tick size tolls: Can a trading slowdown improve price discovery? The Accounting Review Forthcoming.

Linnainmaa, J. T., 2011. Why do (some) households trade so much? Review of Financial Studies 24, 1630–1666.

Loughran, T., Mcdonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance 66, 35–65.

Meulbroek, L. K., 1992. An empirical analysis of illegal insider trading. The Journal of Finance 47, 1661–1699.

Mukherjee, S., Nateghi, R., Hastak, M., 2018. Data on major power outage events in the continental u.s. Data in Brief 19, 2079 – 2083.

Peng, L., 2005. Learning with information capacity constraints. The Journal of Financial and Quantitative Analysis 40, 307–329.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563–602.

Ryans, J. P., 2017. Using the edgar log file data set. Working Paper pp. 1–36.

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

Sims, C. A., 2003. Implications of rational inattention. Journal of Monetary Economics 50, 665 – 690.

Sims, C. A., 2006. Rational inattention: Beyond the linear-quadratic case. The American Economic Review 96, 158–163.

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance 62, 1139–1168.

Van Kervel, V., Menkveld, A. J., 2019. High-frequency trading around large institutional orders. The Journal of Finance 74, 1091–1137.

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

Weller, B. M., 2017. Does Algorithmic Trading Reduce Information Acquisition? The Review of Financial Studies 31, 2184–2226.

Wurgler, J., Zhuravskaya, E., 2002. Does arbitrage flatten demand curves for stocks? The Journal of Business 75, 583–608.

Xiong, W., 2013. Bubbles, Crises, and Heterogeneous Beliefs, Cambridge University Press, p. 663–713.

Figure 1: **8-K viewership activity by machines and humans**

This figure presents total, machine, and human viewership activities of firm i's 8-K files on days $t \in \{0, 1\}$ (relative to the 8-K publication date). Total visits is the natural logarithm of one plus the number of views by both machines and humans. Machine visits and Human visits denote machine and human viewership activities of 8-K filings of firm i on days $t \in \{0, 1\}$, measured by methods proposed by Ryans (2017).

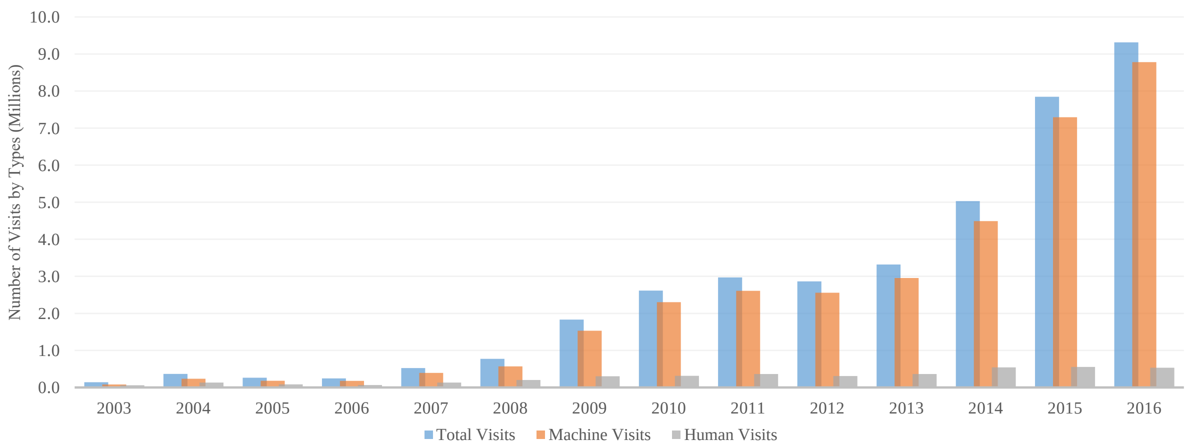

Figure 2: **8-K viewership activity by organization.**

This figure presents machine viewership activity of firm i's 8-K filings by Cloud Computing, Institutional Investor, and Internet Provider on days $t \in \{0, 1\}$ (relative to the 8-K publication date), measured by methods proposed by Ryans (2017).

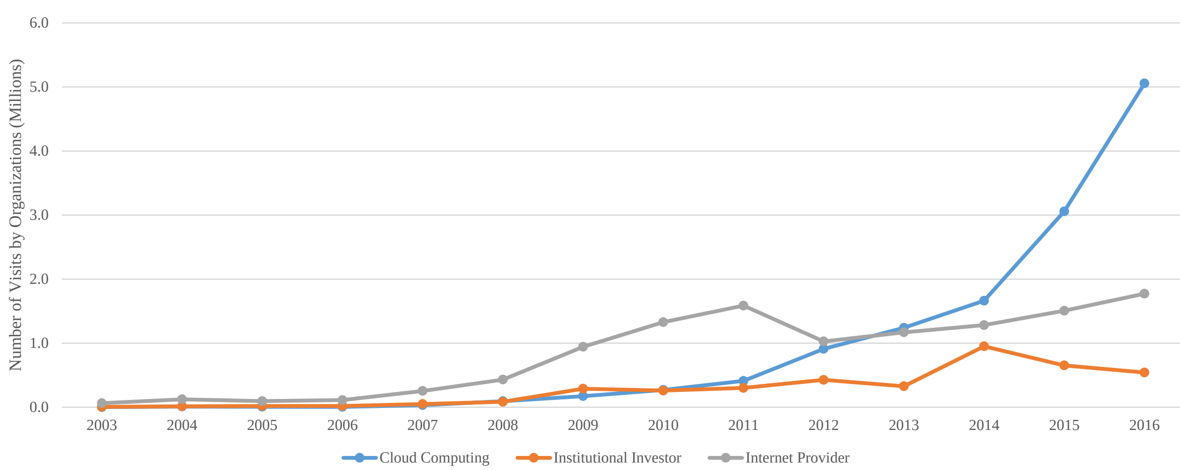

Table **1**: **Summary Statistics**

This table reports summary statistics of the main variables used in our analysis. $Machine$ and $Human$ (Panel A) denote our main variables of machine and human viewership activities of firm i's 8-K filings on days $t \in \{0, 1\}$, relative to 8-K publication date, which are measured based on methods proposed by Ryans (2017). Alternatively, our machine and human viewership activities of 8-K filings are measured by methods proposed by Drake et al. (2015), which are denoted by $Machine_{DRT}$ and $Human_{DRT}$, respectively. Our main (control) variables, which are winsorized at the 1st and 99th percentiles, include: BM refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA is the return on assets; LEV is the total debt over total assets; STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ is the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ is the number of analyst covering the firm in the quarter prior to its 8-K filing disclosure. $DRIFT(2, 10)$, $DRIFT(2, 20)$, and $DRIFT(2, 40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the 8-K announcement, as described in Equation (2). $Item\#.\#$ (Panel B) refers to a dummy variable equal to one if an 8-K filing containing a 8-K item#.#, zero otherwise. Accordingly, the $Item1.01$ refers to *Entry into a Material Definitive Agreement*, the $Item2.02$ refers to *Results of Operations and Financial Condition*, the $Item5.02$ refers to *Departure of Directors or Certain Officers Election of Directors Appointment*, the $Item7.01$ refers to *Regulation FD Disclosure*, and the $Item8.01$ refers to *Other Events*. Lastly, visiting organizations (Panel C) refer to the number of 8-K visits by organizational types on days $t \in \{0, 1\}$ relative to 8-K publication date. All variables are defined in the Appendix Table A.

	N	Mean	Std Dev	25th Pctl	Median	75th Pctl
			Panel A: Summary Statistics			
$Machine$	551,136	3.29	1.42	2.08	3.61	4.41
$Machine_{DRT}$	551,136	3.34	1.41	2.08	3.66	4.45
$Human$	551,136	1.34	1.07	0.69	1.39	2.08
$Human_{DRT}$	551,136	1.05	0.96	0.00	1.10	1.61
BM	551,136	0.75	0.99	0.29	0.55	0.89
$SIZE$	551,136	6.44	2.05	4.99	6.40	7.80
ROA	551,136	-0.03	0.20	-0.02	0.02	0.06
LEV	551,136	0.23	0.22	0.04	0.19	0.37
STD_{RET}	551,136	0.12	0.10	0.07	0.10	0.15
$InstOwn$	551,136	0.61	0.32	0.34	0.66	0.87
$Analysts$	551,136	1.63	0.98	0.69	1.79	2.40
$DRIFT(2, 10)$	551,054	0.06	0.09	0.02	0.04	0.07
$DRIFT(2, 20)$	551,054	0.09	0.10	0.02	0.05	0.11
$DRIFT(2, 40)$	551,054	0.13	0.15	0.04	0.08	0.17
			Panel B: 8-K Items			
	#Filings			*#Visits per Filing*		
Item 1.01	62,396	77.58	143.79	9	40	101
Item 2.02	193,330	69.91	138.98	9	41	90
Item 5.02	70,920	74.17	121.78	12	47	97
Item 7.01	117,747	78.84	184.95	9	45	102
Item 8.01	144,925	71.28	159.56	8	39	91
Other Items	89,218	82.72	147.72	15	52	108
			Panel C: Visiting Organizations(1000)			
	#Org.			*#Visits per Organization (Org.)*		
Auditing & Law Firms	130	3.49	19.49	0.03	0.45	2.29
Data Cloud Services	938	13.78	204.00	0.00	0.01	0.07
Data Vendor & Media	381	4.92	39.22	0.00	0.01	0.08
Education and Regulator	3,172	0.10	1.50	0.00	0.00	0.01
Institutional Investor	992	3.98	34.14	0.01	0.04	0.32
Internet Service Provider	1,153	10.14	77.68	0.01	0.03	0.48
Others	18,930	0.07	0.79	0.00	0.00	0.02

Table **2**: **Determinants of machines and human viewership of 8-Ks**

This table reports OLS regression estimates from our analysis on the determinants of 8-K viewership activity on days $t \in \{0, 1\}$, relative to 8-K publication date. $TotalVisits$ is the natural logarithm of one plus the number of 8-K views by both machines and humans. $Machine$ and $Human$ denote our main variables of machine and human viewership activities of firm i's 8-K filings on days $t \in \{0, 1\}$, relative to 8-K publication date, which are measured based on methods proposed by Ryans (2017). Alternatively, our machine and human viewership activities of 8-K filings are measured by methods proposed by Drake et al. (2015), which are denoted by $Machine_{DRT}$ and $Human_{DRT}$, respectively. $FinNeg$ is the proportion of negative words defined by Loughran and Mcdonald (2011) in an 8-K filing, FOG is the Gunning fog index, a readability measure, is computed based on the words used in an 8-K filing, $WordCount$ is the number of words used in an 8-K filing, *DayRelease* is the number of days between 8-K event date to the 8-K publication date, *#Item* is the number of topics included in an 8-K filing, BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, *InstOwn* is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and *Analysts* is the number of analyst covering a firm in the quarter prior to an 8-K filing, which are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. All $t-$statistics reported in parentheses are computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$TotalVisits$	$Machine$	$Human$	$Machine_{DRT}$	$Human_{DRT}$
	(1)	(2)	(3)	(4)	(5)
$FinNeg$	0.984*	-0.028	4.756***	0.169	5.500***
	(1.76)	(-0.05)	(9.02)	(0.30)	(10.11)
FOG	0.000	0.000	-0.002	0.000	-0.001
	(0.31)	(0.49)	(-1.21)	(0.41)	(-1.07)
$WordCount$	0.050***	0.051**	0.060***	0.051***	0.047***
	(3.21)	(2.99)	(5.85)	(3.18)	(5.06)
DayRelease	-0.021***	-0.014***	-0.055***	-0.017***	-0.042***
	(-7.09)	(-5.00)	(-7.87)	(-5.72)	(-8.06)
#Item	0.098***	0.081***	0.150***	0.092***	0.100***
	(11.58)	(8.41)	(16.56)	(9.72)	(12.59)
BM	0.009**	0.001	0.037***	0.003	0.037***
	(2.28)	(0.22)	(5.99)	(0.86)	(6.52)
$SIZE$	0.012	0.007	0.033***	0.008	0.032***
	(1.37)	(0.89)	(3.08)	(0.93)	(3.21)
InstOwn	0.004	-0.007	-0.148***	-0.001	-0.167***
	(0.09)	(-0.17)	(-3.12)	(-0.02)	(-3.61)
Analysts	0.000	-0.002	0.009	-0.002	0.009
	(0.04)	(-0.19)	(0.72)	(-0.16)	(0.82)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Nobs	556,283	556,283	556,283	556,283	556,283
Adj. R^2	0.845	0.348	0.404	0.862	0.872

Table **3**: **Machine and human viewership of different 8-K items**

This table reports OLS regression estimates from our analysis on the determinants of viewership activities of 8-K filings on days $t \in \{0, 1\}$ (relative to the 8-K publication date) by $Machine$ and $Human$. $Machine$ and $Human$ denote machine and human viewership activities of firm i's 8-K filings on days $t \in \{0, 1\}$ (relative to the 8-K publication date), measured by methods proposed by Ryans (2017). $Item\#.\#$ refers to a dummy variable equal to one if an 8-K filing containing a 8-K item#.#, zero otherwise. Accordingly, the $Item1.01$ refers to *Entry into a Material Definitive Agreement*, the $Item2.02$ refers to *Results of Operations and Financial Condition*, the $Item5.02$ refers to *Departure of Directors or Certain Officers Election of Directors Appointment*, the $Item7.01$ refers to *Regulation FD Disclosure*, and the $Item8.01$ refers to *Other Events*. Control variables include: $FinNeg$ that refers to the proportion of negative words defined by Loughran and Mcdonald (2011) in an 8-K filing; FOG that refers to the Gunning fog index, a readability measure, is computed based on the words used in an 8-K filing; $WordCount$ that refers to the number of words used in an 8-K filing; $DayRelease$ that refers to the number of days between 8-K event date to the 8-K publication date; BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$Machine$					$Human$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Item2.02$	-0.006					0.233***				
	(-0.30)					(6.82)				
$Item8.01$		0.082***					0.016			
		(3.06)					(1.69)			
$Item7.01$			0.102***					0.137***		
			(8.79)					(8.84)		
$Item5.02$				-0.030					-0.205***	
				(-1.37)					(-7.62)	
$Item1.01$					0.114***					0.175***
					(5.65)					(5.33)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	533,160	533,160	533,160	533,160	533,160	533,160	533,160	533,160	533,160	533,160
Adj. R^2	0.863	0.863	0.864	0.863	0.864	0.393	0.383	0.385	0.387	0.386

Table 4: **Impact of machine and human viewership on price drift**

This table reports OLS regression estimates from our analysis of the impact of $Machine$ and $Human$ viewership activities of firm i's 8-K filings on days $t \in \{0,1\}$ (relative to the 8-K publication date) on $DRIFT(2,10)$, $DRIFT(2,20)$, and $DRIFT(2,40)$. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $Machine$ and $Human$ denote machine and human viewership activities, measured by methods proposed by Ryans (2017). BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, ROA is the return on assets, LEV is the total debt over total assets, STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing, $InstOwn$ is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and $Analysts$ is the number of analyst covering a firm in the quarter prior to an 8-K filing, are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. *Nobs* refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$			$DRIFT(2,20)$			$DRIFT(2,40)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Machine$	0.003		0.001	0.004		0.002	0.006		0.003
	(0.93)		(0.22)	(1.00)		(0.45)	(1.21)		(0.61)
$Human$		0.003***	0.003***		0.004***	0.003***		0.005***	0.004***
		(4.12)	(6.20)		(3.86)	(7.00)		(4.40)	(6.22)
BM	0.004***	0.004***	0.004***	0.005***	0.005***	0.005***	0.007***	0.007***	0.007***
	(6.40)	(6.37)	(6.35)	(5.76)	(5.66)	(5.71)	(5.51)	(5.41)	(5.44)
$SIZE$	-0.010***	-0.010***	-0.010***	-0.012***	-0.013***	-0.013***	-0.019***	-0.019***	-0.019***
	(-9.43)	(-9.63)	(-9.55)	(-13.22)	(-13.62)	(-13.46)	(-13.22)	(-13.59)	(-13.49)
ROA	-0.030***	-0.029***	-0.029***	-0.041***	-0.040***	-0.040***	-0.061***	-0.061***	-0.061***
	(-7.02)	(-7.09)	(-7.05)	(-9.04)	(-9.20)	(-9.11)	(-9.56)	(-9.63)	(-9.59)
LEV	0.020***	0.019***	0.019***	0.030***	0.028***	0.029***	0.047***	0.045***	0.045***
	(4.01)	(3.88)	(3.94)	(5.63)	(5.54)	(5.58)	(5.59)	(5.49)	(5.55)
STD_{RET}	0.027***	0.026**	0.027**	0.047***	0.046***	0.046***	0.078***	0.077***	0.077***
	(3.02)	(2.94)	(2.95)	(4.11)	(4.03)	(4.06)	(4.99)	(4.90)	(4.93)
$InstOwn$	-0.013***	-0.012***	-0.012***	-0.018***	-0.017***	-0.017***	-0.022***	-0.022***	-0.022***
	(-4.13)	(-4.03)	(-4.07)	(-4.98)	(-4.85)	(-4.93)	(-3.49)	(-3.39)	(-3.45)
$Analysts$	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.12)	(0.11)	(0.11)	(-0.58)	(-0.61)	(-0.60)	(-1.22)	(-1.26)	(-1.25)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	550,928	550,928	550,928	550,928	550,928	550,928	550,928	550,928	550,928
Adj. R^2	0.174	0.174	0.174	0.237	0.237	0.237	0.257	0.258	0.258

Table 5: Machine types and price drift

This table reports OLS regression estimates from our analysis of the impact of $CloudMachine$, $InstMachine$, $DataMachine$, $OtherMachine$ viewership activities of firm i's 8-K filings on days $t \in \{0, 1\}$ (relative to the 8-K publication date) on $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $CloudMachine$ is the machine viewership by cloud computing services; $InstMachine$ is the machine viewership by financial institutions; $DataMachine$ is the machine viewership by data and media publishers; finally $OtherMachine$ is the machine viewership by the remaining categories. Control variables across all models include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. *Nobs* refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$				$DRIFT(2,20)$				$DRIFT(2,40)$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$CloudMachine$	-0.002***				-0.002**				-0.002			
	(-3.68)				(-2.27)				(-1.42)			
$InstMachine$		0.002				0.002				0.003		
		(1.24)				(1.23)				(1.26)		
$DataMachine$			-0.001				-0.001				-0.002	
			(-0.39)				(-0.46)				(-0.50)	
$OtherMachine$				0.002**				0.003**				0.003**
				(2.68)				(2.59)				(2.82)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848
Adj. R^2	0.180	0.180	0.180	0.180	0.235	0.235	0.235	0.235	0.258	0.258	0.258	0.258

Table 6: **Instrumental variables tests**

The table presents estimates from our two-stage least squares analysis using four market sentiment metrics as instrumental variables. $Machine$ and $Human$ denote machine and human viewership activities of 8-K filings on days $t \in \{0, 1\}$. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). Our instruments are: $InvSent$, $MacroNews$, $VIX(-20,-1)$, and $CRET(-20,-1)$. BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, ROA is the return on assets, LEV is the total debt over total assets, STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing, $InstOwn$ is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and $Analysts$ is the number of analyst covering a firm in the quarter prior to an 8-K filing, are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Weak ID $F-$stat is for the Stock-Yogo test and $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	First Stage		Second Stage					
	$Machine$	$Human$	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,40)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Machine$			0.113		0.152		0.202	
			(1.71)		(1.75)		(1.60)	
$Human$				0.096**		0.124**		0.163**
				(2.31)		(2.27)		(2.64)
$InvSent$	-0.260	-0.225*						
	(-1.09)	(-1.90)						
$MacroNews$	-0.020	-0.060***						
	(-1.39)	(-5.18)						
$VIX(-20,-1)$	0.007	-0.103*						
	(0.12)	(-1.72)						
$CRET(-20,-1)$	-0.573	-0.720***						
	(-1.24)	(-2.65)						
BM	-0.001	0.045***	0.004***	-0.001	0.005***	-0.001	0.007***	-0.000
	(-0.15)	(6.86)	(5.23)	(-0.35)	(5.14)	(-0.25)	(5.06)	(-0.10)
$SIZE$	0.004	0.054***	-0.010***	-0.015***	-0.013***	-0.019***	-0.020***	-0.028***
	(0.49)	(4.55)	(-9.15)	(-5.98)	(-9.09)	(-5.86)	(-10.24)	(-7.47)
ROA	0.009	-0.138***	-0.030***	-0.016**	-0.042***	-0.023***	-0.063***	-0.039***
	(0.47)	(-3.25)	(-7.25)	(-2.93)	(-8.77)	(-3.07)	(-9.43)	(-3.97)
LEV	0.009	0.311***	0.018***	-0.011	0.028***	-0.009	0.045***	-0.004
	(0.33)	(6.16)	(4.16)	(-0.95)	(5.17)	(-0.63)	(5.28)	(-0.26)
STD_{RET}	0.007	0.286***	0.030***	0.004	0.046***	0.012	0.077***	0.032
	(0.22)	(5.28)	(4.03)	(0.26)	(4.04)	(0.60)	(4.90)	(1.25)
$InstOwn$	0.007	-0.118***	-0.012**	-0.000	-0.018**	-0.002	-0.022**	-0.002
	(0.18)	(-2.67)	(-2.72)	(-0.06)	(-2.80)	(-0.30)	(-2.24)	(-0.22)
$Analysts$	0.001	0.005	0.000	-0.000	-0.000	-0.001	-0.001	-0.002
	(0.17)	(0.51)	(0.34)	(-0.03)	(-0.16)	(-0.71)	(-0.56)	(-1.15)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	550,411	550,411	550,411	550,411	550,411	550,411	550,411	550,411
Weak ID F-stat	1.30	51.83						

Table 7: **Evidence from cloud and power outages**

The table presents estimates from our analysis of the impact of disruption proxies on the price drift post to 8-K publication date as a test to our information-assimilation channel. *Cloud Outage Day* refers to the day characterised by a reported service disruption to one or more major cloud service providers; *Power Outage Day* refers to the day characterised by the loss of the electrical power network supply to an end user. We label them as *Outage day*. $DRIFT(2,10)$ and $DRIFT(2,20)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $CloudMachine$, $Machine$ and $Human$ denote cloud machine, machine and human viewership activities of 8-K filings on days $t \in \{0,1\}$. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Cloud Outage Day						*Power Outage Day*					
	$NDRIFT(2,10)$			$NDRIFT(2,20)$			$NDRIFT(2,10)$			$NDRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$CloudMachine$	-0.002***			-0.002**			-0.002***			-0.002**		
	(-4.11)			(-2.40)			(-3.70)			(-2.26)		
$CloudMachine\times$	0.004***			0.004**			0.000			-0.001		
$OutageDay$	(3.63)			(2.50)			(0.32)			(-0.40)		
$Machine$		0.003			0.005			0.003			0.005	
		(0.95)			(1.06)			(0.97)			(1.09)	
$Machine\times$		0.002***			0.002**			0.002**			0.002	
$OutageDay$		(3.20)			(2.22)			(2.35)			(1.12)	
$Human$			0.003***			0.004***			0.003***			0.004***
			(4.35)			(4.12)			(4.84)			(4.42)
$Human\times$			0.001			0.001			0.004			0.003
$OutageDay$			(1.21)			(0.70)			(1.48)			(1.06)
$OutageDay$	-0.008**	-0.006*	0.000	-0.010*	-0.009	-0.001	0.001	-0.006*	-0.003*	0.003	-0.005	-0.002
	(-2.83)	(-1.86)	(0.19)	(-2.02)	(-1.70)	(-0.20)	(0.34)	(-2.01)	(-1.93)	(0.41)	(-1.04)	(-0.88)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739
Adj. R^2	0.206	0.206	0.207	0.235	0.235	0.236	0.205	0.206	0.207	0.235	0.235	0.236

Table 8: **The effects of S&P500 inclusion on machines and humans**

This table reports regression estimates from difference-in-difference models that examine the joint effects of 8-K viewership and S&P500 inclusion on post-event price drift. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $Machine$ and $Human$ denote machine and human viewership activities of firm i's 8-K filings on days $t \in \{0,1\}$. *InIndex* is a binary variable that is assigned the value of one if a stock is added into the S&P500 index, and zero otherwise. BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, ROA is the return on assets, LEV is the total debt over total assets, STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing, *InstOwn* is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and *Analysts* is the number of analyst covering a firm in the quarter prior to an 8-K filing, are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,40)$	
	(1)	(2)	(3)	(4)	(5)	(6)
$Machine$	0.002		0.004		0.005	
	(1.27)		(1.50)		(1.04)	
$Machine \times InIndex$	-0.001		-0.001		-0.001	
	(-1.56)		(-1.33)		(-0.35)	
$Human$		0.003**		0.004**		0.005*
		(2.70)		(2.17)		(2.14)
$Human \times InIndex$		-0.002**		-0.002**		-0.003*
		(-2.35)		(-2.24)		(-1.99)
$InIndex$	0.003	0.004*	0.008*	0.007**	0.010	0.013*
	(1.43)	(1.87)	(2.10)	(2.43)	(1.34)	(2.11)
BM	0.002*	0.002	0.003	0.003	0.010**	0.010**
	(1.78)	(1.71)	(1.31)	(1.27)	(2.94)	(2.98)
$SIZE$	-0.004**	-0.004**	-0.006***	-0.007***	-0.005	-0.005
	(-2.58)	(-2.78)	(-3.28)	(-3.30)	(-1.18)	(-1.30)
ROA	-0.033***	-0.032***	-0.037***	-0.036***	-0.053***	-0.052***
	(-4.03)	(-4.10)	(-4.27)	(-4.24)	(-5.24)	(-5.15)
LEV	0.004	0.004	0.009	0.009	0.012	0.011
	(0.77)	(0.73)	(1.38)	(1.34)	(1.44)	(1.43)
STD_{RET}	0.047**	0.045**	0.065**	0.062**	0.108	0.105
	(2.27)	(2.19)	(2.32)	(2.22)	(1.75)	(1.69)
InstOwn	-0.012**	-0.012**	-0.017***	-0.017***	-0.015*	-0.014
	(-2.76)	(-2.67)	(-3.93)	(-3.85)	(-1.83)	(-1.72)
Analysts	0.000	0.000	0.000	0.000	-0.006	-0.006
	(0.22)	(0.21)	(0.03)	(0.05)	(-1.73)	(-1.74)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
$Nobs$	35,020	35,020	35,020	35,020	35,020	35,020
Adj. R^2	0.157	0.158	0.168	0.168	0.192	0.193

Table 9: **Readability and price drift**

The table presents estimates from our analysis of the impact of readability ($DifficultToRead$) proxies on the price drift post to 8-K filings date as a test to our information-assimilation channel. *Gunning FOG* and *Flesch-Kincaid* are readability measures computed based on the textual contents of a 8-K to proximate how hard the filing is to read. $DRIFT(2,10)$ and $DRIFT(2,20)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $Machine$ and $Human$ denote machine and human viewership activities of 8-K filings on days $t \in \{0,1\}$. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and *Analysts* that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Gunning FOG				*Flesch-Kincaid*			
	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,10)$		$DRIFT(2,20)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Machine \times$	-0.006**		-0.003		-0.006**		-0.004	
$DifficultToRead$	(-2.48)		(-0.75)		(-2.52)		(-0.96)	
$Machine$	0.004		0.005		0.004		0.005	
	(1.24)		(1.10)		(1.20)		(1.12)	
$Human \times$		-0.003		-0.003		-0.004		-0.004
$DifficultToRead$		(-1.37)		(-1.29)		(-1.47)		(-1.51)
$Human$		0.004***		0.005***		0.004***		0.005***
		(3.95)		(3.79)		(4.05)		(3.90)
$DifficultToRead$	0.009	-0.001	0.006	0.004	0.010*	-0.001	0.007	0.004
	(1.76)	(-0.19)	(0.69)	(0.84)	(1.91)	(-0.32)	(0.85)	(0.89)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848
Adj. R^2	0.206	0.207	0.235	0.236	0.205	0.206	0.237	0.237

Table **10**: **Negative sentiment and price drift**

The table presents estimates from our analysis of the impact of negative sentiment on the price drift post to 8-K filing date as a test to our information-assimilation channel. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $CloudMachine$ is the machine viewership by cloud computing services. $InstMachine$ is the machine viewership by financial institutions. $FinNeg$ is the proportion of negative words defined by Loughran and Mcdonald (2011) in an 8-K filing. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,40)$	
	(1)	(2)	(3)	(4)	(5)	(6)
$CloudMachine\times$	-0.037**		-0.056**		-0.080*	
$FinNeg$	(-2.66)		(-2.67)		(-1.98)	
$CloudMachine$	-0.002***		-0.002*		-0.002	
	(-3.27)		(-1.95)		(-1.18)	
$InstMachine\times$		-0.062**		-0.093*		-0.136*
$FinNeg$		(-2.19)		(-2.13)		(-1.99)
$InstMachine$		0.002		0.003		0.004
		(1.43)		(1.38)		(1.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,828	517,828	517,828	517,828	517,828	517,828
Adj. R^2	0.206	0.206	0.235	0.235	0.258	0.258

Table **11**: **Incremental earnings information and price drift**

The table presents estimates from our analysis of the impact of negative sentiment on the price drift post to 8-K filing date as a test to our information-assimilation channel. $DRIFT(2,10)$ and $DRIFT(2,20)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $Machine$, $Human$, and $CloudMachine$ denote machine and human viewership activities of 8-K filings on days $t \in \{0,1\}$. $Item2.02$ is a dummy variable assigned the value of one if a 8-K contain $Item2.02$, zero otherwise. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$			$DRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$Machine$	0.003			0.005		
	(0.98)			(1.07)		
$Machine\times$	0.000			0.000		
$Item2.02$	(0.37)			(0.57)		
$Human$		0.004***			0.005***	
		(4.73)			(4.61)	
$Human\times$		-0.001**			-0.001**	
$Item2.02$		(-2.44)			(-2.42)	
$CloudMachine$			-0.002***			-0.002**
			(-3.67)			(-2.39)
$CloudMachine\times$			-0.000			-0.000
$Item2.02$			(-0.48)			(-0.04)
$Item2.02$	-0.001	-0.000	-0.000	-0.005*	-0.004**	-0.004**
	(-0.68)	(-0.41)	(-0.35)	(-1.82)	(-2.56)	(-2.22)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,743	517,743	517,743	517,743	517,743	517,743
Adj. R^2	0.206	0.207	0.206	0.236	0.236	0.235

Table **12**: **Viewership and informed trading**

The table presents estimates from our analysis of the impact of machine (and sub-categories including *cloud machine* and *financial institution machine*) and human viewership activities on the average daily Probability of Informed Trading on days $(0,1)$ and $(0,5)$, relative to the 8-K publication date, as a validation test. Daily Average $PIN(0,1)$ and Daily Average $PIN(0,5)$ is the average of daily probability of information-based trading (PIN) over windows $(0,1)$ and $(0,5)$, relative to the 8-K publication date. $Machine$ and $Human$ denotes machine and human viewership activities of 8-K filings on days $t \in \{0,1\}$. $CloudMachine$ is the machine viewership by cloud computing services, and $InstMachine$ is the machine viewership by financial institutions. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; *InstOwn* that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and *Analysts* that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. *Nobs* refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Daily Average $PIN(0,1)$				Daily Average $PIN(0,5)$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Machine$	0.002**				0.002***			
	(2.61)				(3.06)			
$Human$		-0.002***				-0.002***		
		(-5.86)				(-4.59)		
$CloudMachine$			0.001*				0.002**	
			(1.88)				(2.40)	
$InstMachine$				-0.000				0.000
				(-0.02)				(0.11)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	500,817	500,817	471,747	471,747	500,817	500,817	471,747	471,747
Adj. R^2	0.591	0.591	0.597	0.597	0.674	0.674	0.679	0.679

Table **13**: **Viewership and algorithmic trading**

The table presents estimates from our analysis of the impact of machine (and sub-categories including *cloud machine* and *financial institution machine*) and human viewership activities on algorithmic trading (including the $OddLotRatio(0,1)$ and the $TradeSize(0,1)$) as a validation test. $OddLotRatio(0,1)$ is the fraction of volume associated with abnormally small trades (less than 100 shares) over the day 0 and 1 relative to 8-K publication date. $TradeSize(0,1)$ is the number of shares traded divided by the number of trades over the day 0 and 1 relative to 8-K publication date. $Machine$ and $Human$ denote machine and human viewership activities of 8-K filings on days $t \in \{0,1\}$. $CloudMachine$ is the machine viewership by cloud computing services, and $InstMachine$ is the machine viewership by financial institutions. Control variables across all columns include: BM that refers to the ratio of book value over the market value of common equity in the year end prior to an 8-K filing; $SIZE$ that refers to the natural logarithm of market value of common equity in the year end prior to an 8-K filing; ROA that refers to the return on assets; LEV that refers to the total debt over total assets; STD_{RET} that refers to the standard deviation of monthly return over the year prior to an 8-K filing; $InstOwn$ that refers to the institutional ownership percentage calculated in the quarter prior to an 8-K filing; and $Analysts$ that refers to the number of analyst covering a firm in the quarter prior to an 8-K filing. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$OddLotRatio(0,1)$				$TradeSize(0,1)$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Machine$	0.000				0.000			
	(0.19)				(0.45)			
$Human$		-0.009**				0.005***		
		(-2.93)				(7.47)		
$CloudMachine$			0.005**				-0.002**	
			(3.92)				(-2.78)	
$InstMachine$				-0.002				0.002
				(-0.57)				(1.67)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	165,821	165,821	165,821	165,821	165,821	165,821	165,821	165,821
Adj. R^2	0.101	0.101	0.101	0.101	0.534	0.534	0.534	0.534

Appendix Table A: Variables, definitions, and sources

Variable	Definition	Source
	Panel A: *Viewership Variables*	
$Machine$	The natural logarithm of non-human visits to an 8-K filing in day 0 and 1, defined by Ryans (2017).	SEC Log File
$Human$	The natural logarithm of human visits to an 8-K filing in day 0 and 1, defined by Ryans (2017).	SEC Log File
$TotalVisit$	The natural logarithm of both machine and human visits to an 8-K filing in day 0 and 1.	SEC Log File
$Machine_{DRT}$	The natural logarithm of non-human visits to an 8-K filing in day 0 and 1, defined by Drake et al. (2015).	SEC Log File
$Human_{DRT}$	The natural logarithm of human visits to an 8-K filing in day 0 and 1, defined by Drake et al. (2015).	SEC Log File
$CloudMachine$	The natural logarithm of non-human visits (defined by Ryans (2017)) to an 8-K filing in day 0 and 1 from cloud computing services.	SEC Log File, MaxMind, Thomson Reuters Ownership, & CIQ
$Machine/Total$	The ratio of machine viewership to total visits of 8K filings on days $t \in \{0,1\}$ relative to 8K publication date.	Our estimates
$CloudMachine/Total$	The ratio of cloud machine viewership to total visits of 8K filings on days $t \in \{0,1\}$ relative to 8K publication date.	Our estimates
$InstMachine$	The natural logarithm of non-human visits (defined by Ryans (2017)) to an 8-K filing in day 0 and 1 from institutional investors.	SEC Log File, MaxMind, Thomson Reuters Ownership, & CIQ
$DataMachine$	The natural logarithm of non-human visits (defined by Ryans (2017)) to an 8-K filing in day 0 and 1 from data and media vendor.	SEC Log File, MaxMind, Thomson Reuters Ownership, & CIQ
$OtherMachine$	The natural logarithm of non-human visits (defined by Ryans (2017)) to an 8-K filing in day 0 and 1 from the IP we cannot identify entity type by Capital IQ, Thomson data, and Google.	SEC Log File, MaxMind, Thomson Reuters Ownership, & CIQ
	Panel B: *Dependent Variables / Research outcome variables*	
$DRIFT(2,T)$	The absolute difference between post-announcement price variation CAR(0, T) and short variation right after the announcement CAR(0,2), where cumulative abnormal returns (CAR) are from Fama French 3 Factor + Momentum model.	CRSP & Compustat
$PIN(0,T)$	Average of daily probability of information-based trading (PIN) over window (0, T). The PIN is calculated daily over 5-minute interval	TAQ
$Noise$	The share of stock return variance that is attributable to noise.	CRSP & Compustat
$OddLotRatio(0,1)$	The fraction of volume associated with abnormally small trades (less than 100 shares) over the day 0 and 1 relative to 8-K publication date.	SEC MIDAS
$TradeSize(0,1)$	The number of shares traded divided by the number of trades over the day 0 and 1 relative to 8-K publication date.	SEC MIDAS

Appendix Table A: Variables, definitions, and source (Continued)

Variable	Definition	Source
	Panel C: *8-K variables*	
$\#Item$	Number of topics included in an 8-K filing	WRDS SEC Suite
$WordCount$	Number of words used in an 8-K filing	WRDS SEC Suite
$FinNeg$	The proportion of negative words defined by Loughran and Mcdonald (2011) in an 8-K filing.	WRDS SEC Suite
$DayRelease$	The number of days between 8-K event date to publication date	WRDS SEC Suite
$Item\#.\#$	A dummy variable equal to one if an 8-K filing containing a 8-K item#.#, zero otherwise.	WRDS SEC Suite
	Panel D: *Control Variables*	
LEV	Total debt over total assets	Compustat
BM	The ratio of book value over the market value of common equity in the year end prior to an 8-K filing	CRSP & Compustat
$SIZE$	The natural logarithm of market value of common equity in the year end prior to an 8-K filing	CRSP
$InstOwn$	The institutional ownership percentage calculated in the quarter prior to an 8-K filing	Factset 13F
$Analysts$	The number of analyst covering a firm in the quarter prior to an 8-K filing	I/B/E/S
STD_{RET}	Standard Deviation of Monthly Return over the year prior to an 8-K filing	CRSP
$InIndex$	A dummy variable equal to one if a stock has been added into S&P500 index, zero otherwise.	CRSP
	Panel E: *Instrumental Variables*	
$InvSent$	Investor sentiment index based on first principal component of FIVE (standardized) sentiment proxies where each of the proxies has first been orthogonalized with respect to a set of six macroeconomic indicators	Jeffrey Wurgler's website
$MacroNews$	A dummy variable equal to one if a major macroeconomic indicator update is announced the 8-K publication data, zero otherwise.	Bloomberg
$VIX(-20,-1)$	Return based on VIX level over window (-20, -1) relative to 8-K publication day	OptionMetrics
$CRET(-20,-1)$	CRSP value weighted return over window (-20, -1) relative to 8-K publication day	CRSP
	Panel F: *Disruption and Readability Measures*	
Cloud Outage Day	A day characterised by a reported service disruption to one or more major cloud service providers.	Gunawi et al. (2016)
Power Outage Day	A day characterised by loss of the electrical power network supply to an end user.	Mukherjee et al. (2018)
Gunning FOG	The Gunning fog index, a readability measure, is computed based on the words used in an 8-K filing.	WRDS SEC Suite
Flesch-Kincaid	Flesch–Kincaid grade level designed to indicate how difficult a passage in English is to understand.	WRDS SEC Suite

Appendix Table B: Top 8-K Viewers by Organisation Types

Organization Type	Organization Name	# Visits (Millions)
Cloud Computing	AMAZON.COM	5.635
Cloud Computing	1&1 INTERNET AG	1.836
Cloud Computing	VICTORY NETWORKS	1.476
Cloud Computing	SOFTLAYER TECHNOLOGIES	1.098
Cloud Computing	FUSIONSTORM	0.482
Cloud Computing	CORESITE	0.320
Cloud Computing	RACKSPACE LTD.	0.300
Cloud Computing	DIGITAL OCEAN	0.274
Cloud Computing	NET ACCESS CORPORATION	0.203
Cloud Computing	SUNGARD AVAILABILITY NETWORK SOLUTIONS	0.181
Institutional Investor	BARCLAYS CAPITAL	0.656
Institutional Investor	TWO SIGMA INVESTMENTS, LLC	0.496
Institutional Investor	THE GOLDMAN SACHS GROUP	0.403
Institutional Investor	HUTCHIN HILL CAPITAL	0.285
Institutional Investor	SCHONFELD TOOLS, LLC.	0.236
Institutional Investor	HARTFORD LIFE INSURANCE COMPANY	0.230
Institutional Investor	HUTCHIN HILL CAPITAL, LP	0.222
Institutional Investor	D. E. SHAW & CO.	0.180
Institutional Investor	WILLIAM O'NEIL & COMPANY	0.164
Institutional Investor	KNIGHT CAPITAL GROUP	0.098
Internet Provider	CENTURYLINK	1.765
Internet Provider	COMCAST	1.042
Internet Provider	SPECTRUM	0.922
Internet Provider	VERIZON	0.584
Internet Provider	OPTIMUM ONLINE	0.523
Internet Provider	AT&T SERVICES	0.518
Internet Provider	OXFORD NETWORKS	0.411
Internet Provider	ROGERS CABLE	0.357
Internet Provider	ILIAD-ENTREPRISES	0.331
Internet Provider	RESILANS AB	0.330
Rest/Other	DOW JONES & COMPANY	0.588
Rest/Other	MARKIT ON DEMAND	0.366
Rest/Other	BLOOMBERG, LP	0.258
Rest/Other	MORRIS, NICHOLS, ARSHT AND TUNNEL	0.221
Rest/Other	THOMSON REUTERS U.S. LLC	0.147
Rest/Other	MCGRAW-HILL	0.127
Rest/Other	CHENGXI MIDDLE SCHOOL	0.061
Rest/Other	GOOGLEBOT*	0.060
Rest/Other	REGUS GROUP SERVICES LTD	0.053
Rest/Other	GODADDY.COM, LLC	0.049

* Googlebot is the generic name for Google's web crawler.

Appendix Table C: Alternative Measure of Machine Viewership

This table reports OLS regression estimates from our analysis of the alternative measures of machine and cloud machine viewership activities of firm i's 8-K filings on days $t \in \{0,1\}$ (relative to the 8-K publication date) on $DRIFT(2,10)$, $DRIFT(2,20)$, and $DRIFT(2,40)$. $DRIFT(2,10)$, $DRIFT(2,20)$ and $DRIFT(2,40)$ are measured by the absolute difference between post-announcement price variation and variation over a short window right after the announcement, as described in Equation (2). $Machine/Total$ and $CloudMachine/Total$ denote viewership fraction from machine and cloud machine, and $TotalVisit$ is the logarithm of total viewership. BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, ROA is the return on assets, LEV is the total debt over total assets, STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing, $InstOwn$ is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and $Analysts$ is the number of analyst covering a firm in the quarter prior to an 8-K filing, are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. $Nobs$ refers to the number of observations. Adj.R^2 is the adjusted R^2 value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,40)$	
	(1)	(2)	(3)	(4)	(5)	(6)
$Machine/Total$	-0.010**		-0.011		-0.015	
	(-2.18)		(-1.61)		(-1.63)	
$CloudMachine/Total$		-0.017**		-0.020*		-0.023
		(-2.36)		(-1.98)		(-1.50)
$TotalVisit$	0.003	0.003	0.005	0.006	0.008	0.008*
	(1.14)	(1.45)	(1.23)	(1.51)	(1.52)	(1.91)
BM	0.003***	0.003***	0.005***	0.005***	0.007***	0.007***
	(6.64)	(6.68)	(5.58)	(5.54)	(5.63)	(5.67)
$SIZE$	-0.008***	-0.008***	-0.012***	-0.012***	-0.018***	-0.018***
	(-12.93)	(-13.32)	(-13.58)	(-13.82)	(-13.41)	(-13.57)
ROA	-0.026***	-0.026***	-0.039***	-0.040***	-0.061***	-0.062***
	(-7.91)	(-8.01)	(-9.24)	(-9.36)	(-9.68)	(-9.74)
LEV	0.015***	0.015***	0.027***	0.027***	0.045***	0.045***
	(4.19)	(4.15)	(5.53)	(5.51)	(5.51)	(5.49)
STD_{RET}	0.031***	0.031***	0.045***	0.046***	0.079***	0.079***
	(4.66)	(4.70)	(3.83)	(3.85)	(4.62)	(4.66)
$InstOwn$	-0.010***	-0.009***	-0.016***	-0.015***	-0.019**	-0.019**
	(-3.45)	(-3.44)	(-4.12)	(-4.12)	(-2.66)	(-2.65)
$Analysts$	-0.000	-0.000	-0.001	-0.001	-0.003**	-0.003**
	(-0.35)	(-0.33)	(-1.19)	(-1.17)	(-2.21)	(-2.19)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,848	517,848	517,848	517,848	517,848	517,848
Adj. R^2	0.218	0.219	0.237	0.238	0.260	0.260

Appendix Table D: Impact of viewership on price variance due to noise trading

This table reports OLS regression estimates from our analysis of the impact of $Machine$, $Human$ and $CloudMachine$ viewership activities on $Noise$. $Noise$ (the dependent variable) is the share of stock return variance that is attributable to noise, which is developed by Brogaard et al. (2021). $Machine$, $Human$ and $CloudMachine$ denote machine, human, and cloud machine viewership activities, respectively, of firm i's 8-K filings on days $t \in \{0, 1\}$. BM is the ratio of book value over the market value of common equity in the year end prior to an 8-K filing, $SIZE$ is the natural logarithm of market value of common equity in the year end prior to an 8-K filing, ROA is the return on assets, LEV is the total debt over total assets, STD_{RET} is the standard deviation of monthly return over the year prior to an 8-K filing, *InstOwn* is the institutional ownership percentage calculated in the quarter prior to an 8-K filing, and *Analysts* is the number of analyst covering a firm in the quarter prior to an 8-K filing, are control variables across all columns. All variables are winsorized at the 1st and 99th percentiles. All variables are also defined in the Appendix Table A. *Firm FE* and *Year FE* denote firm and year fixed effects. *Nobs* refers to the number of observations. Adj.R^2 is the adjusted R-squared value. $t-$statistics are reported in parentheses and computed based on adjusted standard errors clustered at the firm-year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	$Noise$		
	(1)	(2)	(3)
$Machine$	0.116		
	(0.94)		
$Human$		0.101***	
		(4.38)	
$CloudMachine$			-0.081**
			(-2.38)
BM	0.127***	0.123***	0.127***
	(8.08)	(8.05)	(8.12)
$SIZE$	-0.243***	-0.247***	-0.242***
	(-8.62)	(-8.80)	(-8.39)
ROA	-0.642***	-0.628***	-0.640***
	(-8.53)	(-8.64)	(-8.44)
LEV	0.576***	0.547***	0.575***
	(4.90)	(4.76)	(4.85)
STD_{RET}	0.771***	0.746**	0.776***
	(3.05)	(2.96)	(3.05)
$InstOwn$	-0.196**	-0.186**	-0.197**
	(-2.52)	(-2.39)	(-2.50)
$Analysts$	-0.000	-0.002	0.001
	(-0.01)	(-0.07)	(0.04)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Nobs	455,607	455,607	455,607
Adj. R^2	0.486	0.487	0.485