Institutional Investor Attention

Alan Kwan, Yukun Liu, and Ben Matthies*

March 22, 2022

Abstract

We study institutional investor attention using their daily internet news reading. We measure fund-level investor attention to both aggregate and firm-specific information, and relate it to portfolio allocation decisions. During economic downturns, institutional investors shift their attention away from firm-specific news towards aggregate news and the cross-sectional dispersion of their attention to firm-specific news increases. Investor attention is significantly and positively related to portfolio holdings, and this relationship is stronger for more sophisticated funds and those which have limited attention capacity. Fund attention to news exhibits clear habitats with the majority of variation in attention driven by fund and firm effects. Lastly, attention by sophisticated investors predicts future stock returns.

JEL Classification: G10, G20, G40

Keywords: Investor Attention, Portfolio Allocation

*Alan Kwan is with the University of Hong Kong, apkwan@hku.hk, Yukun Liu is with the University of Rochester, yliu229@simon.rochester.edu and Ben Matthies is with the University of Notre Dame, bmatthie@nd.edu. We thank Nicholas Barberis, Zhi Da, Kent Daniel, Joseph Engelberg, Paul Gao, Stefano Giglio, Ben Golez, Ron Kaniel, Andrew Karolyi, Tim Loughran, Stefano Pegoraro, Paul Schultz, John Shim, Denis Sosyura, and Chen Wang for helpful comments.

1 Introduction

The struggles of the canonical portfolio allocation model to explain key empirical regularities in portfolio allocation decisions, such as under diversification (e.g., Friend and Blume (1975); Goetzmann and Kumar (2008)), home bias (e.g., French and Poterba (1991); Coval and Moskowitz (1999)), and style investing (e.g., Brown and Goetzmann (1997); Chan et al. (2002)), have sparked a large theoretical literature. A potential mechanism proposed by the literature to address these puzzles is limited attention (e.g., Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2009, 2010); Kacperczyk et al. (2016); Mackowiak et al. (2021)). In these models, investors need to decide how to allocate their scarce attention to acquire investment relevant information. A major barrier in testing these theories is that typically investor attention to different information is unobservable. In this paper, we lever a new dataset to directly observe fund-level attention to information of both aggregate and firm-specific news, allowing us to test existing theoretical predictions that link investor attention, portfolio allocation, and business cycles.

We conduct our study using a proprietary dataset on internet activity from a data analytics company. The data analytics company maintains a large publisher partnership network and aggregates visitor activities on the websites of these publishers. In particular, major news organizations comprise a substantial fraction of these publishers. We identify visits from each institutional investor and therefore observe its real-time reading of news articles across all partnering publishers. We gather detailed information on the content of each article and identify whether the article is about aggregate or firm-specific news, the latter of which we match to the financial information about the company. Our final database consists of the institutional investor, time, content and source of the article, and the stock information if the article is about a specific company. For conciseness, we refer to these institutional investors as "funds" and the companies the funds read about (or hold) as "firms". We join this database to investor holdings data from Factset and to CRSP and COMPUSTAT to obtain a final sample. Our sample contains 167,153,700[1] fund-firm-quarter observations comprised of 3,953 distinct funds that hold assets,[2] 3,859 distinct firms from the last quarter of 2017 to the second quarter of 2021.

We design a set of tests guided by existing theories of limited attention (e.g., Van Nieuwerburgh and Veldkamp (2009, 2010); Mackowiak et al. (2021)). We focus on four components of investor attention: time-series properties of investor attention (e.g., the relationship with

[1] There are 8,987,609 observations with non-zero reading activity and 8,620,396 fund-firm-quarter observations with non-zero holdings.

[2] We use FactSet fund classifications to separate funds into four broad groups: Investment Advisers (e.g., mutual funds); Hedge Funds; Broker-Dealers (e.g., Banks and Brokerage firms); and Other.

macroeconomic conditions and the business cycle), the relationship between investor attention and portfolios, the determinants of investor attention, and the predictive power of investor attention for future firm returns.

We first study the time-series properties of funds' attention to different news. It has been shown that how funds provide value changes over the business cycle (e.g., Glode (2011); Kacperczyk et al. (2014)). On the theory side, Kacperczyk et al. (2016) link this time-series pattern of funds' value creation to investors' limited attention over the business cycle. A key prediction of this theory is that investors will allocate more attention to aggregate news during economic downturns and more attention towards firm-specific information during booms. We construct the following fund-level measures of investor attention: aggregate conditions, COVID, firm-specific news, and the heterogeneity in investor attention across funds. We examine how these measures evolve over time and through the COVID crisis. The pattern we find is consistent with the predictions of these models. At the onset of the COVID crisis, attention to aggregate signals increases and attention to firm-specific news falls sharply. The results hold true both in raw reading counts and controlling for production of news. We also find that the cross-sectional dispersion of investor attention to firm-specific news increases during the crisis. While not a direct prediction of Kacperczyk et al. (2016), this finding mirrors the prediction of their model that the dispersion of portfolio holdings should increase during downturns.

Next, we turn to the relationship between investor attention and portfolio allocations. When attention capacity is limited, fund managers need to decide how to allocate their attention efficiently across different assets (e.g., Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2009); Kacperczyk et al. (2014); Kacperczyk et al. (2016)). A common theoretical prediction of the literature is a positive association between investor attention and portfolio holdings. In these models, investors optimally choose their attention and portfolio allocations, subject to the constraint of limited attention capacity. In equilibrium, investors would prefer to hold assets they are more informed about on average.

We study the empirical relationship between funds' portfolio holdings and their attention to firm-specific information. First, we compare their news reading about firms they hold compared with those they do not. We find that funds allocate a substantial fraction of their attention to stocks they hold: the ratio of reading about held firms to total reading about all firms is 38.6 percent for mutual funds and 13.6 percent for hedge funds compared to the respective average number of firms held of 191 and 103. In a given quarter on average, a mutual fund (hedge fund) reads about 21.6 (16.7) percent of the firms in its portfolio and 2.8 (3.1) percent of the tradeable firms not currently held.

Furthermore, investor attention and holding are positively related on the intensive margin. We measure the attention of a fund to a company as the fraction of its reading of the firm in a given quarter compared to the fund's total reading activity about all firms in the quarter. We construct an analogous measure of the fraction of the dollar holdings of a firm compared to the total holdings of the fund and run a regression of attention on fund holdings. The relationship between a fund's attention and its portfolio holding is positive and significant at the 1-percent level under a variety of specifications. That is, attention scales with the portfolio holdings. These results hold even when we include firm×time fixed effects, which remove effects from production of news. We use fund characteristics to construct proxies for fund attention capacity and sophistication and document a stronger relationship between attention and holding for more funds with less attention capacity and for more sophisticated funds. These results are consistent with the predictions of the rational inattention theory developed in Van Nieuwerburgh and Veldkamp (2010).

Next, we test whether investor attention predicts future changes in fund holdings. We run a specification based on the sample of held firms and another one based on non-held firms. For the sample of held firms, we find that changes in fund attention predict a decline in holdings with significance at the 1 percent level. In the baseline specification without fixed effects, the coefficient estimate on changes in attention of -0.412 indicates that 1 percentage point increase in the fraction of a fund's reading about a firm corresponds to a 0.4 percent decrease in the fund's holding of the firm in the next quarter. In the non-held specification, we document a positive relationship between attention and future holdings, which is significant at the 1 percent level under most sets of fixed effects. The coefficient estimate on changes in attention in the baseline no fixed effect specification of 0.173 indicates that a 1 percentage point increase in the fraction of reading by a fund about a firm corresponds with a 0.17 percent increase in future holdings.

In the next section of our paper, we study the determinants of investor attention. We begin by documenting the role of firm characteristics by regressing investor attention on lagged firm variables, including beta, size, book-to-market, and past returns. We find that attention is higher for large firms, value firms, firms with high beta, low return-to-asset, and high absolute value of returns. During the COVID period, funds decrease their attention for high beta companies. That is, in downturns, while investors increase attention about information of the aggregate economy, they do not do so by learning more about high beta firms. On the other hand, they increase attention for companies with high absolute value of returns, suggesting that the attention-grabbing effect of return volatility becomes stronger during downturns (e.g., Barber and Odean (2008)).

Furthermore, we decompose the variance of fund attention into the following components: fixed fund characteristics, fixed firm characteristics, common variation in investor attention over time, and a residual component. We find that among the sample of held firms, fund fixed effects explain 11.7 percent of the variation in investor attention, firm fixed effects explain 33.7 percent of the variation and time fixed effects explain 0.0 percent. We include specifications with pairs of fixed effects from each combination of $Fund$, $Firm$, and $Time$. $Fund \times Time$ fixed effects explain 29.7 percent of the variation in investor attention which is higher than the fund fixed effect R-squared of 11.7 percent, indicating the importance of time-series variation of investor attention within fund. $Firm \times Time$ fixed effects explain 47 percent of the variation in investor attention, an increase over the firm fixed effect R-squared of 33.7 percent. $Fund \times Firm$ fixed effects explain 55.1 percent of the variation which indicates that the majority of investor attention variation is explained by fund-firm pairs.

We find a similar pattern of results in fund attention about not held firms. Overall, these results suggest that funds have attention habitats: attention is concentrated on a small set of firms and exhibits little variation over time. This result echoes patterns of under-diversification in fund portfolios studied by Van Nieuwerburgh and Veldkamp (2010). This pattern may arise when funds hold an informational edge in a particular set of firms, such as in the home-bias literature (French and Poterba (1991); Coval and Moskowitz (1999); Van Nieuwerburgh and Veldkamp (2009)) or due to positive feedback effects between attention and holdings, leading to specialization in both holdings and information production.

Lastly, we test whether investor attention by funds convey valuable information about future firm returns. A key assumption for the theories of limited attention is that funds are able to acquire value-relevant news about companies and trade on it. We show that investor attention when funds increase portfolio holdings of a company positively predict its stock returns, while attention when funds decrease portfolio holdings negatively predict its stock returns. Furthermore, we find that the results are more significant for attention by hedge funds, which are commonly believed to be the more sophisticated institutional investors, instead of mutual funds. The findings are consistent with the view that investor attention to news produces value-relevant information, which is reflected in subsequent stock returns.

Our paper relates to the literature that studies limited attention and constrained information processing capacity. Sims (2003) theoretical studies implications of rational inattention in macroeconomics. Moscarini (2004), Mackowiak and Wiederholt (2009) and Maćkowiak and Wiederholt (2015) study the business cycle implications of rational inattention. limited attention is proposed to explain phenomenons and puzzles in finance, such as and under diversification (e.g., Van Nieuwerburgh and Veldkamp (2010); Kacperczyk et al. (2016)), and style investing (e.g., Barberis and Shleifer (2003); Peng and Xiong (2006)). The relation of limited attention and portfolio allocation is a focus in the literature (e.g., Van Nieuwerburgh and Veldkamp (2009); Van Nieuwerburgh and Veldkamp (2010); Abel et al. (2013); Kacperczyk et al. (2016)).

On the empirical side, various proxies for investor attention, such as Google searches and the news reading activities on Bloomberg, are proposed and shown to be related to stock trading activities. Da et al. (2011) find that investor attention measured by Google searches is associated with higher stock returns. Ben-Rephael et al. (2017) measure aggregate institutional attention using news reading activities on Bloomberg terminals. Several recent studies examine institutional investors' readings of companies' filings by unmasking IPs accessing EDGAR. Chen et al. (2020) show that mutual fund managers acquire Form 4 filings to follow trades by company insiders. Dyer (2021) finds that institutional investors are more likely to read filings by companies that are in the same geographic area. Crane et al. (2022) show that hedge funds also read companies' filings and that these hedge funds tend to outperform hedge funds that do not read companies' filings. Barber and Odean (2008), Kempf et al. (2017), and Hirshleifer and Sheng (2021) study how attention can be attracted or distracted by different attention-grabbing events.

In this paper, we lever a novel dataset that identifies individual institutional investor's online readings of news from major publishers. We study the relations of investor attention to information about both the aggregate and firm-specific news, the content and sources of the news, and the portfolio allocations jointly. The results shed light on the existing theories that link investor attention, portfolio allocation, and business cycles.

More broadly, this paper relates to the literature that examines patterns of portfolio allocation decisions. It has been shown that investors' portfolio allocation decisions significantly deviate from predictions from canonical portfolio allocation models such as Markowitz (1952) and Sharpe (1964). Among others, investors' portfolio allocation decisions exhibit under diversification (e.g., Friend and Blume (1975); Goetzmann and Kumar (2008)), home bias (e.g., French and Poterba (1991); Coval and Moskowitz (1999)), style investing (e.g., Brown and Goetzmann (1997); Chan et al. (2002)), non-participation (e.g., Mankiw and Zeldes (1991); Vissing-Jørgensen (2002)), and tradings induced by attention-grabbing events (e.g., Barber and Odean (2008); Hirshleifer and Sheng (2021)). Giglio et al. (2021) document facts about investors' portfolio holdings and their beliefs.

This paper also relates to the literature on news media and stock market. Tetlock (2007) shows that the tone of news predict stock returns. Fang and Peress (2009) find stocks with no media coverage tend to have higher returns. Peress (2014) identifies causal impact of media on stock market using newspaper strikes. Manela and Moreira (2017) extract information about the VIX from newspaper text and show that it captures investors' concern about disaster risks. Fisher et al. (2021) capture investor attention about various macroeconomics variables using newspaper coverage. Liu and Matthies (2021) use newspaper texts to capture investor concerns about long-run risks and show that it is a priced factor in the cross-section of returns.

Our paper is organized as follows. Section 2 discusses the data construction and and presents summary statistics. Section 3 examines the relationship between investor attention and the macroeconomy. Section 4 examines the relationship between attention and fund portfolios and trading. Section 5 studies the determinants of investor attention to firm-specific information. Section 6 relates investor attention with asset returns. Section 7 concludes.

2 Data

2.1 Financial News

Our measures of investor attention are based on news reading activity by institutional investors. We partner with a data analytics company from the marketing technology space, the "Data Partner". The Data Partner maintains a large network of partnerships with online publishers, focused primarily (but not exclusively) on business content and news. As part of the partnership, participating publishers contribute to the Data Partner's pooled dataset via a technology mechanism, which shares information about web content consumption, including the external IP address of the network originating the HTTP request and the URL of content accessed. Overall, the platform aggregates around 1 billion content consumption events per day. From this large dataset, the Data Partner performs two steps: (1) it associates visitors with companies, when possible, and (2) quantifies the "topics" of the content visitors read about. From these two steps, the Data Partner produces an indicator that aims to quantify the business topics that companies are reading about. These analytics are primarily sold to companies to facilitate sales and marketing – by identifying companies with heightened research interest in a specific business topic, one in principle may be able to narrow down likelier customers for a specific product or service. The Data Partner does not sell these analytics products to financial institutions for the purpose of financial trading. Participating publishers receive some of the Data Partner analytics in return.

We rely on the Data Partner's efforts to map visitors to publisher websites to the firms the visitors work for. For each visitor to a Data Partner website, the Data Partner creates a profile through the use of first and third-party cookies. This enables the publisher, and in turn, the Data Partner, to observe when a visitor returns to a website. Over time, the Data Partner infers the association between the profile and a firm through a wide ensemble of industry-accepted methods. For example, user profiles are associated with a company when visitors use a work email to log into a member's website. Another example is through IP addresses. That is, if a profile consistently logs onto a publisher website from a work-associated IP address, this gives a strong indication that the profile belongs to a particular company. The Data Partner also receives data from third-party sources who also perform identity resolution of visitors. Through its proprietary processes, the Data Partner assembles these various sources of data and determines whether a reliable association between a profile and a company can be inferred, and when it can be, what that association is. Crucially, once a visitor has been associated with a Data Partner reliably, the visitor is associated with that Data Partner even though the visitor may traverse different IP addresses. This allows the data to remain effective even when a visitor is working remotely.

2.2 Aggregation and Link with Holdings

We construct metrics on how much an investor reads about a particular company on a given day. To start, we obtain "event-level" data from the Data Partner from the period of October 2017 to July 2021. Event-level data describes an instance of an article being consumed, including the timestamp, company associated with the visitor, and the URL being read. The Data Partner does not sell this event-level data commercially and was made specially available for academic research. The data are scrubbed of personally identifiable information and were accessed remotely.

From the event dataset, we focus on over a dozen financial publishers. The names of the publishers cannot be disclosed but are among some of the largest financial publishers in the world. The publishers we observe total over 100 million content interactions per day. Although not exclusively, these platforms are primarily English-language.

From the event level dataset, we merge with news data from Ravenpack in order to identify the financial news topic, subject, and sentiment of articles in our dataset. Historically, the Data Partner has primarily tracked content in terms of broad business topics. While some companies are themselves topics, due to the design of its algorithm, it has not been able to comprehensively record companies referenced in articles. Moreover, its topic taxonomy consists of broad business topics, and does not consist of certain topics that might be useful to financial economists such as corporate events (for example, capital raising, mergers, or payout announcements) and textual sentiment. Therefore, we obtain these features from Ravenpack, a leading data analytics provider widely used in academic studies and in practice by high-frequency trading firms, hedge funds, banks, and asset managers.

We access data from Ravenpack 1.0, which includes Ravenpack's most detailed offering inclusive of major financial publishers it licenses content from as well as open-access content across blogs, social media posts, news sites, and regulatory filings. We describe our merge process in the Appendix. Joining our data to Ravenpack allows us to obtain the stock tickers associated with each article. For conciseness, we refer to financial institutions as "funds" and the assets they read about (or hold) as "firms". The final dataset consists of the count of articles read by a fund on a specific date, broken down by stock tickers and topics defined by Ravenpack.

We join this reading data to investor holdings data from Factset, which records a snapshot of the portfolio holdings of each fund at the end of each quarter. Finally, we join this data to CRSP and COMPUSTAT to obtain a quarterly data set at the Fund-Firm-Quarter level with fund holdings of each firm, fund reading about each firm, fund characteristics such as fund returns and fund size, and lagged firm characteristics such as prior quarter returns, book-to-market, market capitalization, trailing CAPM beta, and leverage. The holdings data reflects fund holdings at the end of the quarter. Readings data is based on the total of reading throughout the quarter. CRSP and COMPUSTAT are based on data publicly available at the end of the previous quarter.

2.3 Creating Metrics of Investor Attention

We design aggregated and fund-level measures of investor attention. We describe the construction of key measures below which we utilize in later tests.

2.3.1 Aggregated Attention Measures

We construct a set of attention measures aggregated across all funds for our time-series tests.

Macroeconomic News First, we construct measures of attention to macroeconomic conditions: the United States stock market; and COVID. We construct a measure of attention to the market conditions, $read_{it}^{m}$, as the total reading of fund i in quarter t of articles about the following topics: "NASDAQ 100 Index", "NASDAQ Composite Index", "NASDAQ 100 Index", "New York Stock Exchange", "Nasdaq Stock Exchange", "CBOE Stock Exchange", "New York Stock Exchange", "NYSE Arca", "NYSE American", and "S&P 500". This set comprises all available aggregate U.S. financial market topics. We aggregate $read_{it}^{m}$ across all funds as

$$read_{t}^{m} = \frac{1}{N_t} \sum_{i} read_{it}^{m} \tag{1}$$

where N_t is the number of funds in quarter t. We construct separate versions of $read_t^m$ using reading by mutual funds and reading by hedge funds respectively.

We construct measures of investor attention about news regarding COVID following a similar procedure. We define $read_{it}^{cdc}$ as the total reading of fund i of articles about the "U.S. Center for Disease Control and Prevention". We define $read_t^{cdc} = \frac{1}{N_t}\sum_i read_{it}^{cdc}$.

We construct normalized versions of each macroeconomic attention measure by dividing the reading per fund by the total number of articles written in the same quarter about each topic:

$$read_t^{m,norm} = \frac{read_t^m}{\sum articles_t^m} \tag{2}$$

where $\sum articles_t^m$ is the total number of articles written about the United States stock market in quarter t. Normalized versions of the COVID attention measures, $read_t^{cdc,norm}$, are constructed similarly.

Firm-specific News Next, we measure the fraction of fund reading about firm-specific news relative to total reading. We obtain the total reading of each fund about all articles (firm-specific signals, aggregate signals, and other) and define:

$$fs_t = \frac{\sum_i \sum_j read_{ijt}}{\sum_i read_{it}^{all}} \tag{3}$$

where $\sum_i \sum_j read_{ijt}$ is the total reading of all funds about all firms during quarter t and $\sum_i read_{it}^{all}$ is the total reading of all news (firm-specific, aggregate, leisure) by all funds during quarter t.

We construct a measure of firm-specific reading scaled by news production as:

$$fs_t^{norm} = \frac{\frac{1}{N}\sum_i \sum_j read_{ijt}}{\sum_j articles_{jt}} \tag{4}$$

where $\frac{1}{N}\sum_i \sum_j read_{ijt}$ is the average number of firm-specific articles read by each fund in quarter t and $\sum_j articles_{jt}$ is the total number of firm-specific articles written in quarter t.

Attention Dispersion We construct a measure of average fund attention dispersion in the same spirit of the portfolio weight deviations measure from Kacperczyk et al. (2016). For fund types, $f \in \{Mutual\ Fund, Hedge\ Fund\}$, we construct the "aggregate attention portfolio" as the sum across all funds of reading about each firm in each quarter. The weight of firm j in quarter t in the aggregate attention portfolio of funds of type f is:

$$w_{jt}^{f,*} = \frac{\sum_i read_{ijt}}{\sum_j \sum_i read_{ijt}} \tag{5}$$

where $w_{jt}^{f,*}$ is the weight of firm j at time t based on the total reading of all funds of type f about firm j in quarter t divided by the total reading of all funds about all firms in quarter t. The aggregate attention portfolio captures the total reading of all funds for each firm. We calculate the reading weight of fund i for firm j in time t as:

$$w_{jt}^{i} = \frac{read_{ijt}}{\sum_j read_{ijt}} \tag{6}$$

where w_{jt}^{i} is the weight fund i places on firm j in time t in terms of reading about firm j divided by the total reading of fund i about all firms. Finally, for each fund i, we calculate the sum of squared deviations in the reading weight of the fund from the aggregate attention portfolio weight:

$$w_{it}^{dev} = \sum_j \left(w_{jt}^{f,*} - w_{jt}^{i} \right)^2 \tag{7}$$

Conceptually, this measure captures deviations in fund-level attention patterns from the aggregate attention portfolio.

2.3.2 Fund-level Measures

Attention and Holdings For the investor attention of firm-specific news, we construct our main measure of investor attention of a fund about a firm, ia_{ijt}, which is defined as:

$$ia_{ijt} = \frac{read_{ijt}}{\sum_j read_{ijt}} \tag{8}$$

where $read_{ijt}$ is the total number of articles fund i reads about firm j throughout quarter t, and $\sum_j read_{ijt}$ is the total reading of fund i about all firms j throughout quarter t. ia_{ijt} is the fraction of reading of fund i about firm j in quarter t compared to the fund's total reading activity about all firms in the quarter. Similarly, we construct our fund holding measure, h_{ijt}, as the following:

$$h_{ijt} = \frac{shares_{ijt} \times p_{jt}}{\sum_j shares_{ijt} \times p_{jt}} \tag{9}$$

where $shares_{ijt}$ is the number of shares fund i holds of firm j at the end of quarter t, p_{jt} is the share price of firm j at the end of quarter t, and $\sum_j shares_{ijt} \times p_{jt}$ is the total dollar value of fund i's portfolio at the end of quarter t. h_{ijt} is the fraction that firm j comprises of fund i's portfolio at the end of quarter t.

Attention on the Extensive Margin We compare the fraction of reading about held stocks compared to total reading. For each fund in each quarter, we calculate $rh_{it} = \sum_j read_{ijt} \times 1_{h_{ijt}>0}$ where rh_{it} is the total number of articles read by fund i throughout quarter t about firms held in its portfolio as of the end of quarter t, $read_{ijt}$ is the reading by fund i about firm j throughout quarter t, h_{ijt} is the holdings of fund i of firm j at the end of quarter t, and $1_{h_{ijt}>0}$ is a dummy variable that is 1 if the fund holds the firm in its portfolio at the end of the quarter and 0 otherwise. We similarly define $rnh_{it} = \sum_j read_{ijt} \times 1_{h_{ijt}=0}$ as the total number of articles read by fund i throughout the quarter about firms not held in fund i's portfolio at the end of quarter t. We calculate the fraction of reading about held firms compared to all financial reading and then compute the average across all fund-quarter observations within each fund type, $f \in \{Mutual\ Fund, Hedge\ Fund, Bank/Broker\ Other\}$, as:

$$read\ ratio^f = \frac{1}{N_{ft}} \sum_i \frac{rh_{it}}{rh_{it} + rnh_{it}} \times 1_{fund\ type_i = f} \tag{10}$$

where $1_{fund\ type_i=f}$ is a binary variable equal to 1 if the fund type of fund i is f and 0 otherwise, N_{ft} is the total number of funds of type f in quarter t, and $\frac{rh_{it}}{rh_{it}+rnh_{it}}$ is the fraction of reading about held firms compared to total financial reading for fund i in quarter t.

Moreover, we calculate the fraction of held firms that a fund reads about, rph_{it}, and the fraction of not held firms that a fund reads about, $rpnh_{it}$. rph_{it} is calculated for each fund in each quarter as:

$$rph_{it} = \frac{\sum_j 1_{read_{ijt}>0} \times 1_{h_{ijt}>0}}{\sum_j 1_{h_{ijt}>0}} \tag{11}$$

where $1_{read_{ijt}>0}$ is a dummy variable equal to 1 if fund i reads about firm j during quarter t and 0 otherwise. $rpnh_{it}$ is calculated for each fund in each quarter as

$$rpnh_{it} = \frac{\sum_j 1_{read_{ijt}>0} \times 1_{h_{ijt}=0}}{\sum_j 1_{h_{ijt}=0}} \tag{12}$$

where $1_{h_{ijt}=0}$ equals 1 if fund i does not hold firm j in quarter t and is 0 otherwise.

2.4 Summary Statistics

We present summary statistics in Table 1. Our quarterly panel dataset of fund-firm-quarter reading and holding contains 167,153,700 observations comprised of 3,953 distinct funds, 3,859 distinct firms, and 14 quarters from the fourth quarter of 2017 to the first quarter of 2021. There are 8,987,609 observations with non-zero reading activity and 8,620,396 fund-firm-quarter observations with non-zero holdings. There are 2,139,002 observations with both non-zero holdings and non-zero readings. For the held companies, the funds read 24.8 percent of them while for the non-held companies, they only read 4.3 percent of the companies. The sample mean for ia is 0.173 percent and the sample standard deviation is 1.54 percent for the hold companies. For the non-held companies, the sample mean for ia is 0.031 percent and the sample standard deviation is 0.67 percent.

We use FactSet fund classifications to divide funds into four broad groups: Investment Advisers (such as mutual funds), Hedge Funds, Broker-Dealers (such as Banks and Brokerage firms), and Other.[3] We tabulate statistics on reading activity and holdings for each fund type in Table 2. The Investment Advisers in our sample have the following summary statistics: an average (median) number of firms held each quarter of 191 (86); a mean (median) portfolio dollar value of 7.6 billion (289 million) dollars; an average (median) number of financial articles read each quarter of 9,078 (430); and an average (median) number of articles read per firm each quarter of 17 (7).

The Hedge Funds in our sample have the following summary statistics: an average (median) number of firms held each quarter of 103 (18); a mean (median) portfolio dollar value of 4.8 billion (287 million) dollars; an average (median) number of financial articles read each quarter of 23,447 (551); and an average (median) number of articles read per firm each quarter of 23 (8).

The Broker-Dealers in our sample have the following statistics: an average (median) number of firms held each quarter of 530 (360); a mean (median) portfolio dollar value of 14.8 billion (838 million) dollars; an average (median) number of financial articles read each quarter of 66,604 (1,690); and an average (median) number of articles read per firm each quarter of 55 (12).

3 Time-Series Properties of Investor Attention

We examine the time-series properties of investor attention, with a particular focus on funds' attention to different news during economic downturns. Empirically, it has been

[3]Funds that are not available in FactSet are grouped into the "Unmatched" category.

shown that how funds provide value changes over the business cycle (e.g., Glode (2011); Kacperczyk et al. (2014)). On the theory side, Kacperczyk et al. (2016) link this time-series pattern of funds' value creation to investor attention over the business cycle. Their model predicts that investors will allocate more attention to aggregate versus firm-specific news during economic downturns and more attention towards firm-specific information during booms. During downturns, the volatility of aggregate shocks and the price of risk are higher, which increases the return to information about aggregate news.

In this section, we study how investor attention changes over the business cycle. Our sample features an important recent economic downturn – the COVID crisis. We refer to the peak of the COVID crisis as the first two quarters of 2020. We construct fund-level measures of investor attention about the aggregate versus firm-specific information[4] and study how these measures evolved over time.

Investor Attention Measures at the Time-Series

First, we document how aggregate fund attention to macroeconomic and firm-specific news evolves across our entire sample. Figure 1 presents a time-series of the total reading, total firm-specific reading, and firm-specific reading percentage across all funds each quarter from the last quarter of 2017 to the first quarter of 2021. Total reading includes all articles read by funds (financial, aggregate, and leisure). The firm-specific reading includes all articles about firms read by funds. Firm-specific reading percentage is the average firm-specific reading divided by total reading across all funds. The peak of the COVID crisis, defined as the first two quarters of 2020, is highlighted in blue. Total reading ranges from 100 million to 300 million each quarter, with a peak in 2018. Total reading remains similar during the peak of COVID compared to the periods before and after. The firm-specific reading count and the firm-specific reading percentage fall during the COVID crisis. In summary, funds' overall news reading remains the same during the economic downturn, but their attention shifts from firm-specific information to aggregate information.

Next, we focus on aggregate attention patterns by fund type (mutual fund and hedge fund) during the period surrounding the COVID crisis. Figure 2 presents the time-series of attention to macroeconomic news throughout the COVID period from 2019Q3 to 2020Q4. Panel A presents measures based on mutual fund reading and Panel B presents measures based on hedge fund reading. Each panel is divided into two parts: "COVID" which measures attention to articles about the CDC; "Aggregate Market" which measures attention to articles about the aggregate United States stock market. The measure construction is outlined in

[4]We describe the construction of our aggregated attention measures in Section 2.3.1.

Section 2.3.1 above. The blue line (left axis) shows the average number of articles read about a topic per fund each quarter. The read line plots the average reading divided by the total number of articles written about the topic in the same quarter (right axis). The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel.

The average reading of articles about the CDC, $read_t^{cdc}$, increased from 48 (24) to 465 (263) for mutual funds (hedge funds) at the onset of the crisis. The average reading of articles about the aggregate stock market, $read_t^m$, increased from 345 (145) to 671 (299) for mutual funds (hedge funds) at the onset of the crisis. News normalized reading for COVID, $read_t^{cdc,norm}$, and the market, $read_t^{m,norm}$, increase for both mutual funds and hedge funds from 2019Q4 to 2020Q1. The number of articles written about both topics increased sharply during the same period, fund reading increased by an even larger margin.

We present measures of fund attention to firm-specific news in Figure 3. Mutual fund measures are presented in the top panel and hedge fund measures are presented in the panel below. The blue line (left axis) shows financial reading, fs_t, from Section 2.3.1, the sum of all financial reading (reading about firms) by all funds in the quarter divided by the sum of fund reading about all topics (firm-specific, macroeconomic, and leisure). The red line (right axis) shows normalized financial reading, fs_t^{norm}, the average financial reading per fund divided by the total number of articles written about firms in the same quarter. The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel.

The average fraction of reading about firm-specific news, fs_t, falls from 72% (63%) to 40% (37%) for mutual funds (hedge funds) at the onset of the crisis. The normalized financial reading, fs_t^{norm}, falls from 8.5% (16.5%) to 6.9% (14.1%) for mutual funds (hedge funds). fs_t remains relatively flat from Q3 to Q4 of 2019 for both mutual funds and hedge funds and declines sharply in Q1 of 2020. The normalized reading measures, fs_t^{norm}, fall for both types of funds from Q3 to Q4 of 2019 reflecting an increase in the number of firm-specific articles written in the fourth quarter. While the production of firm-specific news increased in Q4 of 2020, fund consumption of this news remained the same and only declined with the onset of the COVID crisis.

Next, we study the average cross-sectional dispersion in fund attention to firm-specific news. The theory developed by Kacperczyk et al. (2016) does not predict an increase in the dispersion of fund attention during downturns. In their model, the increased volatility of aggregate shocks reduces the weight a Bayesian investor places on their prior beliefs and increases the weight placed on investor-specific signals which generates an increase in the dispersion of portfolio holdings (but not attention) during downturns. Though not an explicit prediction of the model, we test how cross-sectional dispersion in attention evolves during the COVID crisis. Figure 4 presents the quarterly time-series of the dispersion in fund attention throughout the COVID period from 2019Q3 to 2020Q4. The dispersion of fund attention is the average of w_{it}^{dev} across all funds in the quarter where $w_{it}^{dev} = \sum_j \left(w_{jt}^{*} - w_{jt}^{i}\right)^2$.[5] w_{jt}^{*} is the "attention market weight", the sum of reading by all funds about firm j in quarter t divided by the sum of reading by all funds about all firms in the quarter. w_{jt}^{i} is fund i's reading of firm j in quarter t divided by the total reading of fund i about all firms in quarter t. The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel. The cross-sectional dispersion in attention to firm-specific news increases from 0.092 (0.102) to 0.111 (0.122) for mutual funds (hedge funds) from Q4 of 2019 to Q1 of 2020 and then declines to 0.080 (0.076) by Q4 of 2020. Fund attention exhibits a similar pattern of dispersion during the crisis as seen in portfolio holdings.

Finally, we present measures of attention using more granular measures at the daily level. We examine the daily reading activity of each fund in the year around the start of the COVID crisis from October 2019 to October 2020. We construct $read_t^m$ and $read_t^{cdc}$ at the daily level. Figure 5 in the Appendix presents the time-series for each of these variables from October 2019 to October 2020: the average reading per fund about the CDC, labeled "COVID" and the average reading per fund about aggregate financial markets, labeled "Macroeconomy". The Figure presents the trailing seven-day averages to smooth reading patterns on weekends versus weekdays. We highlight in blue the period between February 11, 2020 (when the World Health Organization announced the name of COVID-19) to March 15, 2020 (the start of the shutdown in the United States). With these granular measures, we see a similar increase in attention about aggregate signals and the macroeconomy.

Overall, we observe that, consistent with theory, investors pay much more attention to aggregate signals during downturns and firm-specific attention falls sharply. Moreover, we observe an increase in the heterogeneity of fund attention to different firm-specific information.

4 Investor Attention and Portfolio Allocation

In this section, we first examine the relationship between funds' investor attention and portfolio allocations. In classic models of portfolio allocation (e.g., Markowitz (1952); Sharpe (1964)), investors have unlimited attention capacity and learn all publicly available information, which can lead to any arbitrary relationship between attention and portfolio holdings. However, when there are constraints in attention capacity, fund managers must decide how to allocate the attention efficiently across different assets (e.g., Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2009); Kacperczyk et al. (2014); Kacperczyk et al. (2016)).

[5] Measure construction is discussed in Section 2.3.1.

A common theoretical prediction of the literature is a positive association between attention and portfolio holding. In these models, investors optimally choose to allocate their attention and assets, subject to the limited attention capacity. In equilibrium, investors on average would prefer to hold assets they are more informed about. Moreover, the returns would be higher for assets an investor expects to hold a large position in, which can generate positive feedback effects between attention and holdings. This latter prediction we test in Section 6.

The literature largely relies on the observable features of assets and portfolio holdings to infer the attention allocations of individual investors. As Van Nieuwerburgh and Veldkamp (2010) note, while investor attention to various informational events are unobservable, their model is able to make predictions about the relationship between observable features of assets and portfolio holdings, which they test. We observe investor attention to various informational events, making it possible for us to directly test the relationship between investor attention and portfolio holding. Therefore, we study the relationship between investor attention and portfolio holdings on both the extensive and intensive margin. On the extensive margin, we study the funds' attention for stocks they currently hold against those they do not hold. On the intensive margin, we examine the relationship between attention and the stock holding percentage among the stocks the funds' currently hold.

4.1 Investor Attention and Portfolio Holdings

Extensive Margin We first compare how investors allocate their limited attention between stocks currently in their portfolio and those they do not hold. We report the *read ratio* for each fund type in the first row of Table 3.[6] The fraction of reading about held firms for Mutual Funds is 38.6 percent. Hedge Fund attention is less focused on existing holdings with a *read ratio* of 13.6 percent. We report the average number of firms held across all fund-quarter observations by fund type in the second row.

We average rph_{it} and $rpnh_{it}$ by fund type and report the values in the third and fourth rows of Table 3, respectively. Mutual fund reading per held firm is higher than hedge fund reading per held firm by 21.6 percent and 16.7 percent, respectively. Hedge fund reading about not held firms is relatively higher than mutual fund reading at 3.1 percent and 2.8 percent, respectively.

Overall, we find that investor attention is highly concentrated on firms in a fund's existing portfolio. When accounting for the number of firms in a fund's portfolio and comparing reading per firm between held firms and non-held firms, the difference becomes even sharper. Reading about held firms is stronger for mutual funds compared to hedge funds – hedge funds

[6] Section 2.3.2 outlines the construction of the measures used in this section.

allocate relatively more attention to stocks they do not currently hold.

Intensive Margin Next, we study the relationship between attention and the relative share of a firm in a fund's portfolio. To study the intensive margin, we restrict our sample to the set of held stocks, $h_{ijt} > 0$, resulting in approximately 8.3 million fund-firm-quarter observations. We run a set of regressions of fund reading on holdings:

$$ia_{ijt} = \alpha + \beta h_{ijt} + \mu + \epsilon_{ijt} \tag{13}$$

where ia_{ijt} and h_{ijt} are the investor attention measure and holding measure defined in Section 2.3.2, and μ denotes different sets of fixed effects.

Columns 1 through 5 of Table 4 present the results from the Equation 13 specification implemented with no fixed effects, time, firm, firm and time, and firm×time fixed effects, respectively. Standard errors are clustered at the firm level across all specifications. We document a strong positive association between ia and fund holdings. The coefficient estimate, β, on h_{ijt} is positive and significant at the 1 percent level in all specifications. Based on the coefficient estimate, β, of 0.067 in the no fixed effect specification in Column 1, a 10 percentage point increase in the fraction of a firm in a fund's portfolio corresponds to a 0.67 percentage point increase in the share of reading about the firm by the fund. The magnitude of the coefficient estimate is similar when time fixed effects are included. The β estimate in the firm fixed effect specification in Column 3 is 0.015, suggesting that a 10 percent increase in the fraction of a firm in a fund's portfolio corresponds to a 0.15 percent increase in the share of reading about the firm. The coefficient estimates are also approximately 0.015 when firm+time and firm×time fixed effects are included, respectively. When firm×time fixed effects are included, effects coming from news production are excluded.

Our results are consistent with the predictions of the theoretical model in Van Nieuwerburgh and Veldkamp (2010). Fund attention is concentrated in the smaller set of firms currently held by the fund. Within the set of held firms, attention is positively associated with the magnitude of exposure to the firm measured by the relative fraction a firm comprises of a fund's overall portfolio.

Cross-Section We study how the relationship between attention and portfolio holdings varies across funds based on fund attention capacity and fund financial sophistication. We measure fund attention capacity based on fund size, the level of fund reading, fund reading per dollar of holdings, and reading concentration. Our measures of fund sophistication are based on the reading breadth. For each measure, we sort fund-firm-quarter observations into two groups: high and low. Table 12 in the Appendix summarizes the different measures and outlines their construction.

For fund size, we calculate the market value of each fund i's portfolio in each quarter t, $mv_{it} = \sum_j shares_{ijt} \times p_{jt}$. We define the sorting variable, $high\ mv_{ijt}$, to be 1 if a fund-firm-quarter observation is above the median fund size in a given quarter and 0 otherwise. For fund reading, we sort observations based on the total reading of each fund i in each quarter t, $\sum_j read_{ijt}$ where j indexes firms. We construct the associated sorting variable, $high\ read_{ijt}$, to be 1 for fund-firm-quarter observations above the median total fund reading in a given quarter and 0 otherwise. We follow a similar procedure to sort firms based on the total reading of the fund divided by the dollar value of fund holdings: $rpd_{it} = \frac{\sum_j read_{ijt}}{\sum_j shares_{ijt} \times p_{jt}}$. We define $high\ rpd_{ijt}$ to be 1 for fund-firm-quarter observations above the median rpd_{it} in the quarter and 0 otherwise. To construct the information capacity measure, we sort observations based on reading concentration: $conc_{it} = \sum_j ia_{ijt}^2$ where ia_{ijt} is the fraction of reading of firm j by fund i in quarter t as defined above in Equation 8. We define $high\ conc_{ijt}$ to be 1 for fund-firm-quarter observations above the median $conc_{it}$ in the quarter and 0 otherwise.

Our measure of fund sophistication sorts observations based on the number of unique articles read divided by the total reading: $read\ breadth_{ijt} = \frac{unique_{ijt}}{read_{ijt}}$, where $read_{ijt}$ is the number of articles read by fund i about firm j during quarter t and $unique_{ijt}$ is the number of unique articles read by fund i about firm j during quarter t. Intuitively, *read breadth* measures the breadth of sources the investors pay attention to and acquire information. We define $high\ breadth_{ijt}$ to be 1 for fund-firm-quarter observations above the median $read\ breadth_{ijt}$ in the quarter and 0 otherwise.

Using our sorting measures, we run the following specification:

$$ia_{ijt} = \alpha + \beta_1 h_{ijt} + \beta_2 high_{ij,t-1} + \beta_3 h_{ijt} \times high_{ij,t-1} + \epsilon_{ijt} \tag{14}$$

where ia_{ijt} is the fraction of reading by fund i about firm j during quarter t, h_{ijt} is the fraction of holdings of firm j in fund i's portfolio at the end of quarter t, and $high_{ij,t-1}$ is one of the sorting variables described above. The sorting variables are constructed using time $t-1$ information from the previous quarter. Each column of Table 5 presents the results from the regression in Equation 14 implemented for a different sorting variable. We cluster standard errors by fund and include firm×time fixed effects to remove effects from production of news.

The estimates for the coefficient on ia_{ijt} remain positive and significant at the 1 percent level in all specifications. We focus our discussion on the interaction term coefficient, β_3. Column 1 presents the results based on the fund size sorting variable, *high mv*. The coefficient on the interaction term is -0.005 and significant at the 1 percent level. The attention-holdings relationship is attenuated for larger funds. Column 2 presents the results based on the total reading sorting variable, *high read*. The coefficient on the interaction term is -0.025 and is significant at the 1 percent level. The interaction term in the reading per dollar sort specification shown in Column 3 is also negative and significant at the 1 percent level. Column 4 presents the results based on the reading concentration sorting variable, *high conc*. The interaction term is 0.057 and is significant at the 1 percent level. The relationship between investor attention and fund holdings is weaker for larger funds, funds with higher total reading, higher reading per dollar invested, and funds with a lower concentration of reading. To the extent that large size, higher reading (in level and per dollar), and less reading concentration proxy for fund attention capacity, these results indicate a weaker relationship between attention and holdings for less attention constrained funds. These results are in line with the theoretical predictions because funds that are more attention constrained would allocate optimally choose to allocate their scarce attention to the most value-relevant news, leading to a pronounced relationship between attention and portfolio holding (e.g., Van Nieuwerburgh and Veldkamp (2010); Kacperczyk et al. (2016)).

Our measure of fund sophistication is *highbreadth*, the unique articles read divided by the reading of the fund about the firm. Columns 5 presents the results of the fund sophistication specification. The coefficients on the interaction terms is 0.018 which is significant at the 1 percent level. To the extent that high breadth of reading about a firm proxy for fund sophistication, the relationship between attention and fund holdings is stronger for more sophisticated funds. These results are consistent with the view that sophisticated funds are better in identifying value-relevant news, leading to a pronounced relationship between attention and portfolio holding.

4.2 Predicting Changes in Fund Holdings

In this subsection, we continue to explore the relationship between attention and holding by studying whether investor attention can predict future changes in fund holdings. We construct the change in investor attention for fund i about firm j from quarter $t-1$ to t as:

$$\Delta ia_{ijt} = ia_{ijt} - ia_{ij,t-1} \tag{15}$$

where ia_{ijt} is the fraction of reading about firm j by fund i over quarter t as defined in Equation 8. Similarly, we define the change in holdings for fund i about firm j from quarter t to $t+1$ as:

$$\Delta hs_{ij,t+1} = log\left(1 + hs_{ij,t+1}\right) - log\left(1 + hs_{ijt}\right) \tag{16}$$

where hs_{ijt} is the number of shares of firm j held by fund i at the end of quarter t.[7]

Held Stocks We first restrict our sample to firms held at time t. We regress the change in fund holdings from the end of quarter t to $t+1$ on the change in attention from quarter $t-1$ to t:

$$\Delta hs_{ij,t+1} = \alpha + \beta \Delta ia_{ijt} + \mu + \epsilon_{ij,t} \tag{17}$$

where Δhs and Δia are changes in holdings and changes in investor attention defined above, and μ denotes different sets of fixed effects.

Columns 1 to 5 in Panel A of Table 6 present the results of specifications with no fixed effects, time, firm, firm and time, and firm×time fixed effects, respectively. The coefficient on Δia is negative and significant at the 1 percent level in all specifications. In the baseline specification without fixed effects, the coefficient estimate for β of -0.412 indicates that 1 percentage point increase in the fraction of reading about firm j corresponds to a 0.4 percent decrease in holdings next quarter. The magnitude of the coefficient estimate, β, is largest in the firm×time fixed effect specification in Column 5.

Increases in attention by funds about held stocks predicts future selling. We focus on changes in shares held in order to avoid confounding stock return performance and investment decisions. The relationship between attention and future trading is not clear ex ante: attention to existing holdings may also signal interest in increasing holdings.

Not Held Stocks Next, we study the relationship between investor attention and trading for stocks not held in a fund's portfolio. We regress changes in holdings on changes in attention following the specification in Equation 17, restricted to the sample of non-held stocks. This sample includes fund-firm-quarter observations with zero holding and zero reading resulting in 136.5 million observations. Columns 1 through 5 in Panel B of Table 6 present the results with the same set of fixed effects used in Panel A: no fixed effects, time, firm, firm and time, and firm×time fixed effects, respectively. The coefficient on Δia_{ijt} is positive in all specifications and significant at the 1 percent level in every specification, except the firm×time regression where it is significant at the 10 percent level. The coefficient estimate in the baseline specification with no fixed effect of 0.173 indicates that a 1 percentage point increase in the fraction of reading by fund i about firm j from quarter $t-1$ to t

[7] We obtain similar results using change in market value of the position: $\Delta h_{ijt} = h_{ijt} - h_{ij,t-1}$ where h_{ijt} is the market value of fund i's holding in firm j at the end of quarter t. We use change in shares held in our main specification to avoid confounding change in holdings with stock returns. We study stock return predictability in the next section.

corresponds with a 0.17 percent increase in holdings of firm j by fund i from the end of quarter t to $t+1$.

Our results are consistent with existing empirical findings in the attention literature, which find a positive association between buying activity and measures of attention such as high trading volume, extreme returns, or news articles (Barber and Odean (2008)). We note that one limitation of the data is that we do not observe short selling. While many of the funds we study are long-only (by mandate or by requirement), the lack of short-selling data may affect our measures of hedge fund trading activity.

Buy Sell Indicators As an additional test, we run similar predictability regressions using buy and sell indicators rather than changes in holdings. We define $sell_{ij,t+1}$ as 1 $\Delta hs_{ij,t+1} < 0$ and 0 otherwise. We define $buy_{ij,t+1}$ as 1 if $\Delta hs_{ij,t+1} > 0$ and 0 otherwise. We study the sample of held stocks consisting of 7.9 million fund-firm-quarter observations and run the following regression of sell indicator on change in investor attention:

$$sell_{ij,t+1} = \alpha + \beta \Delta ia_{ijt} + \epsilon_{ij,t+1} \tag{18}$$

where Δia_{ijt} is the change in investor attention defined in Equation 15. Columns 1 through 5 in Panel A of Table 7 present the results of this specification with no fixed effects, time, firm, firm plus time, and firm$\times$time fixed effects, respectively. The coefficient on Δia_{ijt} is positive and significant at the 1 percent level in all specifications. In the baseline specification with no fixed effects, the coefficient estimate of 0.740 indicates that a 1 percentage point increase in the fraction of fund i reading about firm j corresponds to a 0.74 percentage point increase in the likelihood of fund i buying shares of firm j in the next quarter.

We then study the sample of non-held stocks and run the following specification:

$$buy_{ij,t+1} = \alpha + \beta \Delta ia_{ijt} + \epsilon_{ij,t+1} \tag{19}$$

Columns 1 through 5 of Panel B of Table 7 present the results of this specification with the same sets of fixed effects as in Panel A. The coefficient estimate on Δia_{ijt} is positive in all specifications and significant at the 1 percent level in every specification, except the firm$\times$time fixed effect regression. The coefficient of 0.017 in the baseline specification indicates that a 1 percentage point increase in the fraction of fund reading about a firm corresponds with a 0.017 percentage point increase in the likelihood of the fund buying shares of the firm in the next quarter.

5 Determinants of Investor Attention

In this section, we examine the potential determinants of fund attention, including factors beyond fund holdings. We study the relationship between investor attention and firm characteristics, and whether the relationship changes during economic downturns. Finally, we conduct a variance decomposition of investor attention.

5.1 Firm Characteristics

We examine the relationship between investor attention and firm characteristics. We obtain financial data from COMPUSTAT from firm quarterly reports and pricing data from CRSP. We construct firm variables based on investor attention prior to the end of quarter $t-1$. We regress ia_{ijt} on these lagged firm variables using the sample of held firms:

$$\begin{aligned} ia_{ijt} = \alpha + \delta_1 \beta^{capm}_{j,t-1} + \delta_2 size_{j,t-1} + \delta_3 bm_{j,t-1} \\ + \delta_4 abs\left(r_{j,t-1}\right) + \delta_5 roa_{j,t-1} + \mu + \epsilon_{ijt} \end{aligned} \tag{20}$$

where ia_{ijt} is the fraction of reading of fund i about firm j in quarter t as defined in Equation 8, $\beta^{capm}_{j,t-1}$ is lagged CAPM beta calculated using the trailing 252 daily returns ending on the last day of quarter $t-1$, $size_{j,t-1}$ is lagged log firm market capitalization, $bm_{j,t-1}$ is lagged firm book-to-market, $abs\left(r_{j,t-1}\right)$ is the quarterly returns of firm j in quarters $t-1$, $roa_{j,t-1}$ is the lagged return of assets, and μ denotes different sets of fixed effects.

Columns 1 through 6 of Table 8 present the results from the regression specification in Equation 20 with no fixed effects, time, firm, fund, firm and time, and fund×time fixed effects, respectively. Standard errors are clustered by fund in each specification. Across all specifications, the coefficient on CAPM beta is positive and significant at the 1 percent level. The coefficient estimate of 0.04 in the baseline specification indicates that a one standard deviation increase in beta (0.420 in the sample) is associated with a 0.017 percentage point increase in the fraction of reading about a firm, which is about a 10 percent increase of the sample mean (0.173 percent) for held companies. In other words, funds allocate more attention to higher beta stocks, which is consistent with the prediction of Kacperczyk et al. (2016) as high beta stocks provide information on both the firm itself and the aggregate economy.

The coefficient on size is positive in the specifications without fixed effects but is negative when firm fixed effect is included. The coefficient estimate of 0.142 in the baseline specification indicates that a one standard deviation increase in firm size (2.116 in the sample) is associated with a 0.300 percentage point increase in the share of reading about a firm. The coefficient on book-to-market is positive and significant at the 1 percent level across all specifications indicating that funds tilt attention towards value firms. Based on the magnitude of the coefficient of 0.065 in the baseline specification, a one standard deviation increase in book-to-market (0.735 in the sample) is associated with a 0.048 percentage point increase in the share of reading about a firm. The coefficients on absolute value of lagged returns are positive in the no fixed effect, time, fund, and fund plus time fixed effect specifications. The coefficient become negative in the firm plus time fixed effect specification for $abs\left(r_{j,t-1}\right)$. These results indicate that high $abs\left(r_{j,t-1}\right)$ companies are attention-grabbing, which is assumed in several studies (e.g., Barber and Odean (2008); Kempf et al. (2017)). In the baseline specification, the coefficient on $abs\left(r_{j,t-1}\right)$ is 0.120 so a one standard deviation increase in the absolute value of prior quarter returns (0.289) is associated with a 0.035 percentage point increase in the fraction of reading about a firm. Finally, the coefficient on return on assets is negative in every specification except with firm plus time fixed effect. Based on the coefficient in the baseline specification of -0.139, a one standard deviation increase in return on assets (0.220) is associated with a 0.031 percentage point decrease in the share of reading about a firm – funds allocate more attention to firms with lower roa or firms that are underperforming.

Our results indicate a strong relationship between firm characteristics and investor attention: attention is higher for large firms, value firms, firms with high beta, low return-on-assets, and high absolute value of past returns.

5.2 Firm Characteristics and Crisis

We characterize how the relationship between investor attention and firm characteristic changes during the COVID crisis. We define the dummy variable, $covid_t$, as 1 during the two quarters of 2020 and 0 otherwise. Following a similar specification to Equation 20 in subsection 5.1, we regress ia_{ijt} on firm characteristics interacted with $covid_t$:

$$\begin{aligned} ia_{ijt} = \alpha + \delta_1 \beta^{capm}_{j,t-1} + \delta_2 size_{j,t-1} + \delta_3 bm_{j,t-1} \\ + \delta_4 abs\left(r_{j,t-1}\right) + \delta_5 roa_{j,t-1} + \eta_1 covid_t \times \beta^{capm}_{j,t-1} \\ + \eta_2 covid_t \times size_{j,t-1} + \eta_3 covid_t \times bm_{j,t-1} \\ + \eta_4 covid_t \times abs\left(r_{j,t-1}\right) + \eta_5 covid_t \times roa_{j,t-1} \\ + \gamma covid_t + \mu + \epsilon_{ijt} \end{aligned} \tag{21}$$

We restrict the sample to held firms resulting in 7.9 million observations.

Columns 1 and 2 of Table 9 report the regression results with time, and time plus firm fixed effects respectively. All standard errors are clustered by fund. We focus our discussion on the coefficients, η, of the interaction terms. The cross-term on CAPM beta or η_1, is negative and significant at the 1 percent level in the two specifications. That is, funds decrease their attention for firms with high beta during economic downturns. Kacperczyk et al. (2016) theoretical show that investors would shift their attention from acquiring firm-specific information to aggregate information during downturns. To obtain aggregate information, investors can either directly acquire aggregate news or acquire information about companies with high beta, which contain information about the aggregate economy. We find that, in downturns, investors shift attention toward aggregate information, but they do not pay more attention to companies with high beta. The coefficient estimates on the cross-term of $abs\left(r_{j,t-1}\right)$ with *covid* are positive which shows that the attention grabbing effect of $abs\left(r_{j,t-1}\right)$ is stronger during downturns. The coefficient estimate on the cross-term of *roa* is significantly negative, suggesting that funds increase attention to companies that are underperforming. The cross-terms on firm size, η_2, and on bm, η_3, are positive and significant at the 1 percent level in all specifications, suggesting that funds pay more attention to larger companies and value companies during economic downturns.

5.3 Variance Decomposition of Investor Attention

In this section, we explore patterns in attention by decomposing the variance of fund attention into several components: fixed fund characteristics, fixed firm characteristics, common variation in investor attention over time, and a residual component. We run the following specifications:

$$ia_{ijt} = \phi_i + \epsilon_{1,ijt} \tag{22}$$

$$ia_{ijt} = \mu_j + \epsilon_{2,ijt} \tag{23}$$

$$ia_{ijt} = \rho_t + \epsilon_{3,ijt} \tag{24}$$

Our panel contains three dimensions: fund, firm, and quarter. We include interaction terms across each dimension and explore the joint effect of firm cross time, fund cross time, and fund cross firm fixed effects in the following specifications:

$$ia_{ijt} = \mu_j \times \phi_i + \mu_j + \phi_i + \epsilon_{4,ijt} \tag{25}$$

$$ia_{ijt} = \mu_j \times \rho_t + \mu_j + \rho_t + \epsilon_{5,ijt} \tag{26}$$

$$ia_{ijt} = \phi_i \times \rho_t + \phi_i + \rho_t + \epsilon_{6,ijt} \tag{27}$$

Table 10 presents the results of these specifications. The table is divided into three parts: the first two columns contain results based on the sub-sample of 9.0 million observations with non-zero reading; Columns 3 and 4 present the results based on the sub-sample of 2.1 million observations with non-zero reading and non-zero holdings; the last two columns contain the results for the sub-sample of 6.8 million observations with non-zero holdings. Within each section of the table, the first column and second columns measure investor attention using ia_{ijt} and the log of total reading, $log\,(read_{ijt})$, respectively. Each row presents the adjusted R-squared from the regression specifications above based on the fixed effects indicated on the left.

The first column documents the results for the non-zero reading sub-sample using ia_{ijt}. The adjusted R-squared from the fund fixed effects specification in Equation 23 is 11.7 percent. Firm fixed effects shown in the second row explain 33.7 percent of the variation in ia_{ijt}. The adjusted R-squared with time fixed effects from Equation 24 is 0.0 percent. Firm and fund fixed effects explain a large portion of the variation in investor attention, while time fixed effects explain almost none of the variation in reading.

The fourth, fifth, and sixth rows present the adjusted R-squareds from the interacted fixed effects specifications from Equations 25, 26, and 27. $Fund \times Time$ fixed effects explain 29.7 percent of the variation in investor attention which is higher than the fund fixed effect R-squared of 11.7 percent, indicating the importance of time-series variation in ia_{ijt} within fund. $Firm \times Time$ fixed effects explain 47 percent of the variation in ia_{ijt}, an increase over the firm fixed effect R-squared of 33.7 percent: time-series variation in investor attention within firm can explain a portion of the variation that firm fixed effects alone cannot. $Fund \times Firm$ fixed effects explain 55.1 percent of the variation in ia_{ijt} which indicates that the majority of investor attention variation is explained by fund-firm pairs.

Column 2 presents results measuring investor attention using log of reading, $log\,(read_{ijt})$, instead of ia_{ijt}. Fund fixed effects and firm fixed effects explain 13.5 percent and 11.8 percent of the variation in $log\,(read_{ijt})$, respectively. Similar to the ia_{ijt} result, time fixed effects explain 0.2 percent of the variation in $log\,(read_{ijt})$. Rows 4 through 6 show the results from the interacted fixed effect specifications in Equations 25, 26, and 27. $Fund \times Time$ and

$Firm \times Time$ fixed effects have R-squareds of 15.5 percent and 14.8 percent respectively. Both are only slightly higher than the Fund and Firm fixed effects specifications. $Firm \times Fund$ fixed effects explain 72.1 percent of the variation in $log\left(read_{ijt}\right)$. Fund-firm pairs are an even more important factor in log reading.

Interestingly, the patterns in R-squared documented in the all reading sample for both ia_{ijt} and $log\left(read_{ijt}\right)$ are similar in the held stocks sample and the not held stocks sample. Time fixed effects explain little of the variation in investor attention. Fund and Firm fixed effects play a much larger role and $Fund \times Firm$ fixed effects explain the majority of the overall variation in both ia and $log\left(read\right)$. Overall, these results indicate that funds have attention habitats: funds allocate attention to the same set of firms with limited time-series variation in attention. This result echoes patterns of under-diversification in fund portfolios documented empirically and studied in Van Nieuwerburgh and Veldkamp (2010). This pattern may arise when funds believe they have an informational edge in a small set of firms such as in the home-bias literature (Van Nieuwerburgh and Veldkamp (2009)) or when attention and fund holdings exhibit a positive feedback effect with information making investment more attractive which in turn increases the return to information production.

6 Investor Attention and Stock Returns

In this section, we test whether attention by funds conveys valuable information about future firm returns. One of the key assumption in the theoretical literature of limited attention is that funds may allocate attention to acquire value-relevant information about companies. Because funds may gather both favorable and unfavorable information, we thus test the return predictability for attention when funds increase their portfolio holdings for a company versus when they decrease their portfolio holdings.

We construct separate measures based on the investor attention of investment advisers and of hedge funds, respectively.[8] For each type of investor, we construct two measures: the reading of funds which have recently increased holdings in the stock, ial^f_{jt} (investor attention long) where j indexes firms, t indexes time and f indexes fund type $f \in \{Investment\ Adviser, Hedge\ Fund\}$; and the reading of funds which have recently decreased holdings in the stock, ias^f_{jt} (investor attention sell). Both measures are simple extensions of the baseline ia_{it} measure from Equation 8. Specifically, we construct ial^f_{jt} as:

$$ial^f_{jt} = \frac{1}{N^f_{j,q-1}} \sum_i ia_{ijt} \times 1_{buy_{ij,q-1}} \times 1_{fund\ type_i = f} \tag{28}$$

[8]We use the groupings derived from the Factset fund classification described in subsection 2.4: Investment Advisers; Hedge Funds; Banks-Brokers; and Other.

where ia_{ijt} is the fraction of reading of fund i about firm j in month t compared to the fund's total reading activity about all firms in the month (the monthly version of the baseline measure Equation 8), $1_{fund\ type_i=f}$ is a dummy variable equal to 1 if fund i is type $f \in \{Investment\ Adviser, Hedge\ Fund\}$ and 0 otherwise, $1_{buy_{ij,q-1}}$ is a dummy variable equal to 1 if fund i increased the number of shares held of firm j from quarter $q-2$ to $q-1$ and 0 otherwise, and $N^f_{j,q-1}$ is the number of funds that increased their position in firm j in the prior quarter. We index monthly reading by t and quarterly holdings by q since we construct our reading measure at the monthly frequency while holdings are reported at the quarterly frequency. ial^f_{jt} is the average reading activity, ia, of funds which increased their position in firm j in the previous quarter. We construct ias^f_{jt} as:

$$ias^f_{jt} = \frac{1}{N^f_{j,q-1}} \sum_i ia_{ijt} \times 1_{sell_{ij,q-1}} \times 1_{fund\ type_i=f} \tag{29}$$

where ia_{ijt} and $1_{fund\ type_i=f}$ are as defined above and $1_{sell_{ij,q-1}}$ is a dummy variable equal to 1 if fund i decreased the number of shares held of firm j from quarter $q-2$ to $q-1$ and 0 otherwise, and $N^f_{j,q-1}$ is the number of funds that decreased their position in firm j in the prior quarter. ias^f_{jt} is the average ia of funds which decreased their position in firm j in the previous quarter.

We implement the Fama-MacBeth procedure to test the predictive power of these measures on future stock returns. In our first specification, we test ial^f_{jt} and ias^f_{jt} constructed using mutual fund reading activity. In our second specification, we test these measures constructed using hedge fund reading activity. In both specifications, we include controls for lagged firm market capitalization ($size_{j,t-1}$), lagged firm book-to-market ($bm_{j,t-1}$), lagged firm CAPM beta ($\beta^{capm}_{j,t-1}$), and lagged firm idiosyncratic volatility ($iv^{capm}_{j,t-1}$). Lagged variables are based on firm financial and pricing data available at the end of quarter $t-1$. In each quarter t we run the cross-sectional regression:

$$r_{j,t+1} = \alpha_t + \delta_{1,t} ial^f_{jt} + \delta_{2,t} ias^f_{jt} + \delta_{3,t} size_{j,t-1} + \delta_{4,t} bm_{j,t-1} + \delta_{5,t} \beta^{capm}_{j,t-1} + \delta_{6,t} iv^{capm}_{j,t-1} + \epsilon_{j,t+1} \tag{30}$$

where $r_{j,t+1}$ is the monthly return of stock j and the independent variables are as described above.

Table 11 presents the coefficient estimates with Fama-MacBeth standard errors in parentheses below. The first column presents results based on measures constructed using mutual fund data ($ial^{mf}_{jt}, ias^{mf}_{jt}$). The coefficient estimates ial is 0.147 with a t-statistic of 1.523 and the coefficient estimate on ias -0.119 with a t-statistic of 1.068. The sign of the coefficients indicates that attention by mutual funds which have recently increased (decreased) holdings in a firm is associated with higher (lower) future returns but neither coefficients are significant at the 10 percent level. The second column presents results based on hedge fund data ($ial_{jt}^{hf}, ias_{jt}^{hf}$). The coefficient estimate for ial is 0.144 and significant at the 5 percent level. The coefficient estimate for ias is -0.112 and is significant at the 10 percent level. Overall, hedge fund activities are more value relevant than mutual fund activities.

7 Conclusion

We study investor attention to various informational events using proprietary data on the daily internet news reading of over 3,000 financial firms from October 2017 to June 2021. We measure fund-level investor attention of firm-specific news, aggregate conditions and the macroeconomy, and other topics, and link this data with portfolio holdings.

We document several regularities in investor attention. In the time-series, macroeconomic conditions drive trends in attention with fund's focusing on firm-specific information during normal times and aggregate signals during crisis periods. Moreover, the dispersion in attention to different firm-specific information across funds increases. In the cross-section, funds predominantly pay attention to assets they hold. The relationship between attention and portfolio holdings is stronger for funds with limited attention capacity and more sophisticated funds. Investor attention exhibits clear habitats with the majority of variation driven by fund and firm effects. We also find that investor attention predicts both future fund trading and future stock returns.

References

Abel AB, Eberly JC, Panageas S. 2013. Optimal inattention to the stock market with information costs and transactions costs. *Econometrica* **81**: 1455–1481.

Barber BM, Odean T. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* **21**: 785–818.

Barberis N, Shleifer A. 2003. Style investing. *Journal of Financial Economics* **68**: 161–199.

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

Brown SJ, Goetzmann WN. 1997. Mutual fund styles. *Journal of Financial Economics* **43**: 373–399.

Chan LK, Chen HL, Lakonishok J. 2002. On mutual fund investment styles. *The Review of Financial Studies* **15**: 1407–1437.

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

Coval JD, Moskowitz TJ. 1999. Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance* **54**: 2045–2073.

Crane A, Crotty K, Umar T. 2022. Do hedge funds profit from public information. *Management Science* .

Da Z, Engelberg J, Gao P. 2011. In search of attention. *Journal of Finance* **66**: 1461–1499.

Dyer TA. 2021. The demand for public information by local and nonlocal investors: Evidence from investor-level data. *Journal of Accounting and Economics* **72**: 101417.

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

Fisher AJ, Martineau C, Sheng J. 2021. Macroeconomic attention and announcement risk premia. *Journal of Finance* .

French KR, Poterba JM. 1991. Investor diversification and international equity markets. *The American Economic Review* **81**: 222.

Friend I, Blume ME. 1975. The demand for risky assets. *The American Economic Review* **65**: 900–922.

Giglio S, Maggiori M, Stroebel J, Utkus S. 2021. Five facts about beliefs and portfolios. *American Economic Review* .

Glode V. 2011. Why mutual funds "underperform". *Journal of Financial Economics* **99**: 546–559.

Goetzmann WN, Kumar A. 2008. Equity portfolio diversification. *Review of Finance* **12**: 433–463.

Hirshleifer D, Sheng J. 2021. Macro news and micro news: complements or substitutes? *Journal of Financial Economics* .

Kacperczyk M, Nieuwerburgh SV, Veldkamp L. 2014. Time-varying fund manager skill. *The Journal of Finance* **69**: 1455–1484.

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

Kempf E, Manconi A, Spalt O. 2017. Distracted shareholders and corporate actions. *The Review of Financial Studies* **30**: 1660–1695.

Liu Y, Matthies B. 2021. Long run risk: Is it there? *Journal of Finance* .

Mackowiak B, Matejka F, Wiederholt M. 2021. Rational inattention: A review .

Mackowiak B, Wiederholt M. 2009. Optimal sticky prices under rational inattention. *American Economic Review* **99**: 769–803.

Maćkowiak B, Wiederholt M. 2015. Business cycle dynamics under rational inattention. *The Review of Economic Studies* **82**: 1502–1532.

Manela A, Moreira A. 2017. News implied volatility and disaster concerns. *Journal of Financial Economics* **123**: 137–162.

Mankiw NG, Zeldes SP. 1991. The consumption of stockholders and nonstockholders. *Journal of Financial Economics* **29**: 97–112.

Markowitz H. 1952. Portfolio selection. *The Journal of Finance* .

Moscarini G. 2004. Limited information capacity as a source of inertia. *Journal of Economic Dynamics and control* **28**: 2003–2035.

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

Peress J. 2014. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *the Journal of Finance* **69**: 2007–2043.

Sharpe W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *the Journal of Finance* .

Sims CA. 2003. Implications of rational inattention. *Journal of Monetary Economics* **50**: 665–690.

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

Van Nieuwerburgh S, Veldkamp L. 2009. Information immobility and the home bias puzzle. *Journal of Finance* **64**: 1187–1215.

Van Nieuwerburgh S, Veldkamp L. 2010. Information acquisition and under-diversification. *Review of Economic Studies* **77**: 779–805.

Vissing-Jørgensen A. 2002. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* **110**: 825–853.

8 Tables & Graphs

Table 1: Summary Statistics

Investor - Firm - Quarter	
Observations	167,153,700
Non-zero reading (obs)	8,987,609
Non-zero holdings (obs)	8,620,396
Non-zero Holdings	
Non-zero reading (obs)	2,139,002
Non-zero reading (%)	24.8
$Average\,(ia_{ijt})$ (%)	0.173
$\sigma\,(ia_{ijt})$ (%)	1.54
Zero holdings	
Non-zero reading (obs)	6,848,607
Non-zero reading (%)	4.3
$Average\,(ia_{ijt})$ (%)	0.031
$\sigma\,(ia_{ijt})$ (%)	0.67

Table 1 reports summary statistics of the reading and holding panel data.

Table 2: Fund Type - Characteristics

Panel A: Holdings	Firms Held		MV (MM)	
	Mean	Med	Mean	Med
Investment Adviser	191	86	7,605	289
Hedge Fund	103	18	4,861	287
Broker-Dealer	530	360	14,800	838
Other	257	100	9,020	590
Unmatched	121	61	4,457	86

Panel B: Reading	Articles Read		$\frac{Articles}{Firm}$	
	Mean	Med	Mean	Med
Investment Adviser	9,078	430	17	7
Hedge Fund	23,447	551	23	8
Broker-Dealer	66,604	1,690	55	12
Other	88,617	3,085	72	16
Unmatched	4,606	220	12	6

Table 2 reports summary statistics of reading activity and holdings for all funds that hold assets in our sample split by fund type: Investment Adviser; Hedge Fund; Broker-Dealer; Other; and Unmatched. Panel A presents statistics for the number of firms held in a fund's portfolio and the dollar fund size in millions. Panel B presents the mean and median of quarterly total financial articles read and quarterly total financial articles read per firm.

Table 3: Investor Attention: Held versus not held

	Mutual Fund	**Hedge Fund**	**Bank/Broker**	**Other**
read ratio	0.386	0.136	0.581	0.418
firms held	191	103	530	257
rph	0.216	0.167	0.409	0.429
rpnh	0.028	0.031	0.093	0.120

Table 3 reports statistics describing reading about firms held versus not held in a fund's portfolio. *read ratio* is the average across all funds of the held to total reading percentage: the total number of articles read in the quarter about firms in a fund's portfolio divided by the total number of financial articles read in the quarter by the fund. *firms held* is the average number of firms held in a fund's portfolio. *rph* is the average across all funds of the average number of articles read about firms in the fund's portfolio each quarter. *rpnh* is the average across all funds of the average number of articles read about firms not held by a fund.

Table 4: Contemporaneous Fund Holdings and Reading, Held stocks

	ia_{ijt}				
	(1)	(2)	(3)	(4)	(5)
h_{ijt}	0.067***	0.067***	0.015***	0.015***	0.016***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Constant	0.001***	0.001***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed Effects	N	$Time$	$Firm$	$Firm + Time$	$Firm \times Time$
R-Squared	0.011	0.011	0.322	0.322	0.480
Obs	8,302,763	8,302,763	8,302,763	8,302,763	8,302,715

Table 4 reports the contemporaneous investor attention on holdings results from the specification:

$$ia_{ijt} = \alpha + \beta h_{ijt} + \mu + \epsilon_{ijt}$$

Where ia_{ijt} is the fraction of reading of fund i about firm j in quarter t compared to the fund's total reading activity about all firms in the quarter from Equation 8 and h_{ijt} is the fraction that firm j comprises of fund i's portfolio at the end of quarter t from Equation 9. Each column presents the results of specifications using different sets of fixed effects: $\boldsymbol{N}one$, $Time$, $Firm$, $Firm + Time$, $Firm \times Time$. We report coefficients, standard errors clustered by fund in parentheses, the adjusted R-squareds, and the number of observations. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 5: Reading on Holding: Cross-sectional

	ia_{ijt}				
	(1)	(2)	(3)	(4)	(5)
h_{ijt}	0.017***	0.021***	0.035***	0.009***	0.011***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
$high\ mv_{ijt}$	0.000				
	(0.000)				
$h_{ijt} \times high\ mv_{ijt}$	-0.005***				
	(0.002)				
$high\ read_{ijt}$		0.000			
		(0.000)			
$h_{ijt} \times high\ read_{ijt}$		-0.025***			
		(0.002)			
$high\ rpd_{ijt}$			0.000**		
			(0.000)		
$h_{ijt} \times high\ rpd_{ijt}$			-0.029***		
			(0.002)		
$high\ conc_{ijt}$				0.000***	
				(0.000)	
$h_{ijt} \times high\ conc_{ijt}$				0.057***	
				(0.005)	
$high\ breadth_{ijt}$					0.000***
					(0.000)
$h_{ijt} \times high\ breadth_{ijt}$					0.018***
					(0.002)
Fixed Effects	$Firm \times Time$	$Firm \times Time$	$Firm \times Time$	$Firm \times Time$	$Firm \times Time$
R-squared	0.472	0.473	0.473	0.473	0.473
N	7,965,306	7,965,306	7,965,306	7,965,306	7,965,306

Table 5 presents results from contemporaneous investor attention on holdings regression with fund characteristic dummy variables:

$$ia_{ijt} = \alpha + \beta_1 h_{ijt} + \beta_2 high_{ij,t-1} + \beta_3 h_{ijt} \times high_{ij,t-1} + \epsilon_{ijt}$$

Where ia_{ijt} is the fraction of reading of fund i about firm j in quarter t compared to the fund's total reading activity about all firms in the quarter from Equation 8 and h_{ijt} is the fraction that firm j comprises of fund i's portfolio at the end of quarter t from Equation 9, and $high_{ij,t-1}$ is a dummy variable which sorts funds based on attention capacity or sophistication. Each column presents the results using a different dummy variable (the different sorting variables are described in Table 12). We report coefficients, standard errors clustered by fund in parentheses, the adjusted R-squareds, and the number of observations. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 6: Changes in Holdings and Reading

	Panel A: Held				
	(1)	(2)	(3)	(4)	(5)
Δia_{ijt}	-0.412***	-0.397***	-0.679***	-0.736***	-1.253***
	(0.074)	(0.073)	(0.073)	(0.072)	(0.069)
Constant	-0.983***	-0.983***	-0.983***	-0.983***	-0.983***
	(0.031)	(0.031)	(0.034)	(0.034)	(0.034)
Fixed Effects	N	$Time$	$Firm$	$Firm + Time$	$Firm \times Time$
R-squared	0.000	0.003	0.035	0.038	0.094
N	7,891,684	7,891,684	7,891,684	7,891,684	7,891,654
	Panel B: Not Held				
	(1)	(2)	(3)	(4)	(5)
Δia_{ijt}	0.173***	0.164***	0.088***	0.078***	0.035*
	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)
Constant	0.057***	0.057***	0.057***	0.057***	0.057***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effects	N	$Time$	$Firm$	$Firm + Time$	$Firm \times Time$
R-squared	0.000	0.000	0.005	0.005	0.006
N	136,463,519	136,463,519	136,463,519	136,463,519	136,463,519

Table 6 reports the results from the change in holdings on change in attention regression:

$$\Delta hs_{ij,t+1} = \alpha + \beta \Delta ia_{ijt} + \mu + \epsilon_{ij,t}$$

Where $\Delta hs_{ij,t+1} = log\left(1 + hs_{ij,t+1}\right) - log\left(1 + hs_{ijt}\right)$ is the growth in shares owned (the change in log number of shares of firm j held by fund i from the end of quarter t to $t+1$), $\Delta ia_{ijt} = ia_{ijt} - ia_{ij,t-1}$ is the change in investor attention (the change in the fraction of reading about firm j by fund i from quarter $t-1$ to t), and μ denotes different sets of fixed effects. Panel A documents results from the sample restricted to firms held in a fund's portfolio at time t. Panel B documents results from the sample of firms not held in a fund's portfolio at time t. Each column presents the results of specifications using different sets of fixed effects: $\boldsymbol{N}one$, $Time$, $Firm$, $Firm + Time$, $Firm \times Time$.

We report coefficients, standard errors clustered by fund in parentheses, the adjusted R-squareds, and the number of observations. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 7: Buy Sell Indicators

Panel A: Held, $sell_{ij,t+1}$					
	(1)	(2)	(3)	(4)	(5)
Δia_{ijt}	0.740***	0.743***	0.742***	0.750***	0.815***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Constant	0.493***	0.493***	0.493***	0.493***	0.493***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Fixed Effects	N	$Time$	$Firm$	$Firm + Time$	$Firm \times Time$
R-squared	0.000	0.001	0.007	0.008	0.024
N	7,891,684	7,891,684	7,891,684	7,891,684	7,891,654

Panel B: Not Held, $buy_{ij,t+1}$					
	(1)	(2)	(3)	(4)	(5)
Δia_{ijt}	0.017***	0.016***	0.008***	0.007***	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.007***	0.007***	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed Effects	N	$Time$	$Firm$	$Firm + Time$	$Firm \times Time$
R-squared	0.000	0.000	0.006	0.006	0.007
N	136,463,519	136,463,519	136,463,519	136,463,519	136,463,519

Table 7 reports the results from the *buy* and *sell* indicators on change in investor attention regression:

$$sell_{ij,t+1} = \alpha + \beta \Delta ia_{ijt} + \mu + \epsilon_{ij,t}$$

Where $sell_{ij,t+1}$ is a dummy variable equal to 1 if fund i increased its position in firm j from quarter t to $t+1$ and zero otherwise, $\Delta ia_{ijt} = ia_{ijt} - ia_{ij,t-1}$ is the change in investor attention (the change in the fraction of reading about firm j by fund i from quarter $t-1$ to t), and μ denotes different sets of fixed effects. Panel A documents results from the sample restricted to firms held in a fund's portfolio at time t. Panel B documents results from the sample of firms not held in a fund's portfolio at time t. The regression specification in Panel B uses the dummy variable $buy_{ij,t+1}$ which indicates future buying. Each column presents the results of specifications using different sets of fixed effects: $\boldsymbol{N}one$, $Time$, $Firm$, $Firm + Time$, $Firm \times Time$. We report coefficients, standard errors clustered by fund in parentheses, the adjusted R-squareds, and the number of observations. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 8: Reading on Stock Characteristics

	ia_{ijt} (in percentage)					
	(1)	(2)	(3)	(4)	(5)	(6)
beta	0.040***	0.042***	0.011***	0.039***	0.021***	0.041***
	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
size	0.142***	0.142***	-0.012***	0.141***	-0.022***	0.143***
	(0.003)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
bm	0.065***	0.065***	0.015***	0.064***	0.010***	0.065***
	(0.003)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
$abs\,(r_{t-1})$	0.120***	0.142***	0.007***	0.111***	-0.010***	0.134***
	(0.004)	(0.004)	(0.001)	(0.003)	(0.001)	(0.004)
roa	-0.139***	-0.137***	-0.007**	-0.118***	0.001	-0.116***
	(0.006)	(0.006)	(0.003)	(0.006)	(0.003)	(0.006)
Constant	-3.155***	-3.182***	0.442***	-3.152***	0.668***	-3.183***
	(0.070)	(0.071)	(0.028)	(0.077)	(0.030)	(0.078)
Fixed Effects	N	$Time$	$Firm$	$Fund$	$Firm + Time$	$Fund + Time$
R-squared	0.029	0.029	0.337	0.035	0.337	0.035
N	7,879,319	7,879,319	7,879,319	7,879,288	7,879,319	7,879,288

Table 8 presents the results of the investor attention on firm characteristics regression:

$$
\begin{aligned}
ia_{ijt} = \alpha + \delta_1 \beta^{capm}_{j,t-1} + \delta_2 size_{j,t-1} + \delta_3 bm_{j,t-1} + \\
\delta_4 abs\,(r_{j,t-1}) + \delta_5 roa_{j,t-1} + \mu + \epsilon_{ijt}
\end{aligned}
$$

where ia_{ijt} is the fraction of reading of fund i about firm j in quarter t as defined in Equation 8, $\beta^{capm}_{j,t-1}$ is lagged CAPM beta calculated using the trailing 252 daily returns ending on the last day of quarter $t-1$, $size_{j,t-1}$ is lagged firm market capitalization, $bm_{j,t-1}$ is lagged firm book-to-market, $abs\,(r_{j,t-1})$ is the absolute value of quarterly returns of firm j in quarters $t-1$, $roa_{j,t-1}$ is the lagged return of assets, and μ denotes different sets of fixed effects. Each column presents the results of specifications using different sets of fixed effects: $\boldsymbol{N}one$, $Time$, $Firm$, $Firm+Time$, $Fund+Time$. We report coefficients (multiplied by 100 for readability), standard errors clustered by fund in parentheses, the adjusted R-squareds, and the number of observations. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 9: Reading on Stock Characteristics During the Crisis

	ia_{ijt} (in percentage)	
	(1)	(2)
β^{capm}	0.059***	0.032***
	(0.004)	(0.002)
$size$	0.138***	-0.026***
	(0.003)	(0.001)
bm	0.063***	0.009***
	(0.003)	(0.001)
$abs\,(r_{t-1})$	0.155***	-0.025***
	(0.004)	(0.002)
roa	-0.106***	0.014***
	(0.006)	(0.003)
$covid \times \beta^{capm}$	-0.132***	-0.077***
	(0.005)	(0.004)
$covid \times size$	0.035***	0.028***
	(0.002)	(0.001)
$covid \times bm$	0.060***	0.022***
	(0.003)	(0.002)
$covid \times abs\,(r_{t-1})$	0.091***	0.014***
	(0.006)	(0.003)
$covid \times roa$	-0.242***	-0.229***
	(0.009)	(0.010)
Time Fixed Effects	X	X
Firm Fixed Effects		X
R-squared	0.030	0.338
N	7,879,211	7,879,211

Table 9 presents the results of the investor attention on firm characteristics with $covid_t$ dummy regression:

$$\begin{aligned} ia_{ijt} = \quad & \alpha + \delta_1 \beta^{capm}_{j,t-1} + \delta_2 size_{j,t-1} + \delta_3 bm_{j,t-1} \\ & +\delta_4 abs\,(r_{j,t-1}) + \delta_5 roa_{j,t-1} + \eta_1 covid_t \times \beta^{capm}_{j,t-1} \\ & +\eta_2 covid_t \times size_{j,t-1} + \eta_3 covid_t \times bm_{j,t-1} \\ & +\eta_4 covid_t \times abs\,(r_{j,t-1}) + \eta_5 covid_t \times roa_{j,t-1} \\ & +\gamma covid_t + \mu + \epsilon_{ijt} \end{aligned}$$

where ia_{ijt} is the fraction of reading of fund i about firm j in quarter t as defined in Equation 8, $covid_t$ is a dummy variable equal to 1 during 2020Q1 and 2020Q2 and 0 otherwise., $\beta^{capm}_{j,t-1}$ is lagged CAPM beta calculated using the trailing 252 daily returns ending on the last day of quarter $t-1$, $size_{j,t-1}$ is lagged firm market capitalization, $bm_{j,t-1}$ is lagged firm book-to-market, $abs\,(r_{j,t-1})$ is the absolute value of the last quarterly return, $roa_{j,t-1}$ is the lagged return of assets, the interaction terms include all terms between $covid_t$ and the firm variables, and μ denotes different sets of fixed effects. Coefficients are multiplied by 100 for readability. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 10: Fixed Effects

	All (N=8,987,612)		Held (N=2,139,002)		Not held (N=6,848,610)	
	ia_{ijt}	$log\,(read)$	ia_{ijt}	$log\,(read)$	ia_{ijt}	$log\,(read)$
Fund	0.117	0.135	0.143	0.163	0.117	0.133
Firm	0.337	0.118	0.361	0.134	0.332	0.120
Time	0.000	0.002	0.001	0.002	0.001	0.002
$Fund \times Time$	0.297	0.155	0.282	0.188	0.306	0.157
$Firm \times Time$	0.470	0.148	0.518	0.172	0.457	0.154
$Firm \times Fund$	0.551	0.721	0.566	0.768	0.568	0.697

Table 10 presents the results of fixed effect regression specifications from Equations 22, 23, 24, 27, 26, and 25. The table is divided into three parts: the first two columns contain results based on the sub-sample of 9.0 million observations with non-zero reading; Columns 3 and 4 present the results based on the sub-sample of 2.1 million observations with non-zero reading and non-zero holdings; the last two columns contain the results for the sub-sample of 6.8 million observations with non-zero holdings. Within each section of the table, the first column and second columns measure investor attention using ia_{ijt} and the log of total reading, $log\,(read_{ijt})$, respectively. Each row presents the adjusted R-squared from the regression specifications above based on the fixed effects indicated on the left: $Fund$, $Firm$, $Time$, and each pairwise combination.

Table 11: Return Predictability

	Mutual Fund	Hedge Fund
ial_{jt}	0.147	0.144**
	(1.523)	(1.980)
ias_{jt}	-0.119	-0.112*
	(-1.068)	(-1.779)
$size$	0.000	0.000
	(-0.094)	(0.171)
bm	-0.009*	-0.012***
	(-1.906)	(-2.315)
β^{capm}	0.012	0.013
	(1.444)	(1.362)
iv^{capm}	0.269	0.374
	(1.099)	(1.290)
Firms	2,545	2,476
Months	39	39

Table 11 presents the stock return predictability results based on aggregated attention variables: ial^f_{jt} and ias^f_{jt} where j indexes firms, t indexes time, and f indexes fund type. ial^f_{jt} is constructed as:

$$ial^f_{jt} = \frac{1}{N^f_{j,q-1}} \sum_i ia_{ijt} \times 1_{buy_{ij,q-1}} \times 1_{fund\ type_i = f}$$

where ia_{ijt} is the fraction of reading of fund i about firm j in month t compared to the fund's total reading activity about all firms in the month (the monthly version of the baseline measure Equation 8), $1_{fundtype_i=f}$ is a dummy variable equal to 1 if fund i is type $f \in \{Investment\ Adviser, Hedge\ Fund\}$ and 0 otherwise, $1_{buy_{ij,q-1}}$ is a dummy variable equal to 1 if fund i increased the number of shares held of firm j from quarter $q-2$ to $q-1$ and 0 otherwise, and $N^f_{j,q-1}$ is the number of funds that increased their position in firm j in the prior quarter. We index monthly reading by t and quarterly holdings by q since we construct our reading measure at the monthly frequency while holdings are reported at the quarterly frequency. ial^f_{jt} is the average investor attention, ia, of funds which increased their position in firm j in the previous quarter. ias is constructed analogously. We implement the Fama-MacBeth approach to estimate coefficients and standard errors corrected for cross-sectional correlation. In each quarter t we run the cross-sectional regression:

$$r^m_{j,t+1} = \alpha_t + \delta_{1,t} ial^f_{jt} + \delta_{2,t} ias^f_{jt} + \delta_{3,t} size_{j,t-1} + \delta_{4,t} bm_{j,t-1} + \delta_{5,t} \beta^{capm}_{j,t-1} + \delta_{6,t} iv^{capm}_{j,t-1} + \epsilon_{j,t+1}$$

where $r_{j,t+1}$ is the monthly return of stock j, ial and ias are as described above, and $size$, bm, β^{capm}, and iv^{capm} are control variables for market capitalization, book-to-market, CAPM beta, and CAPM idiosyncratic volatility respectively. For each variable k, we compute coefficient estimates for δ_k as $\delta_k = \frac{1}{T} \sum \delta_{k,t}$ the average of the coefficient estimates across each quarter. Fama-MacBeth standard errors are: $se^{FM}_k = \frac{sd(\delta_{k,t})}{\sqrt{T}}$. Table 11 presents the coefficient estimates with Fama-MacBeth standard errors in parentheses below. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Figure 1: Total Reading

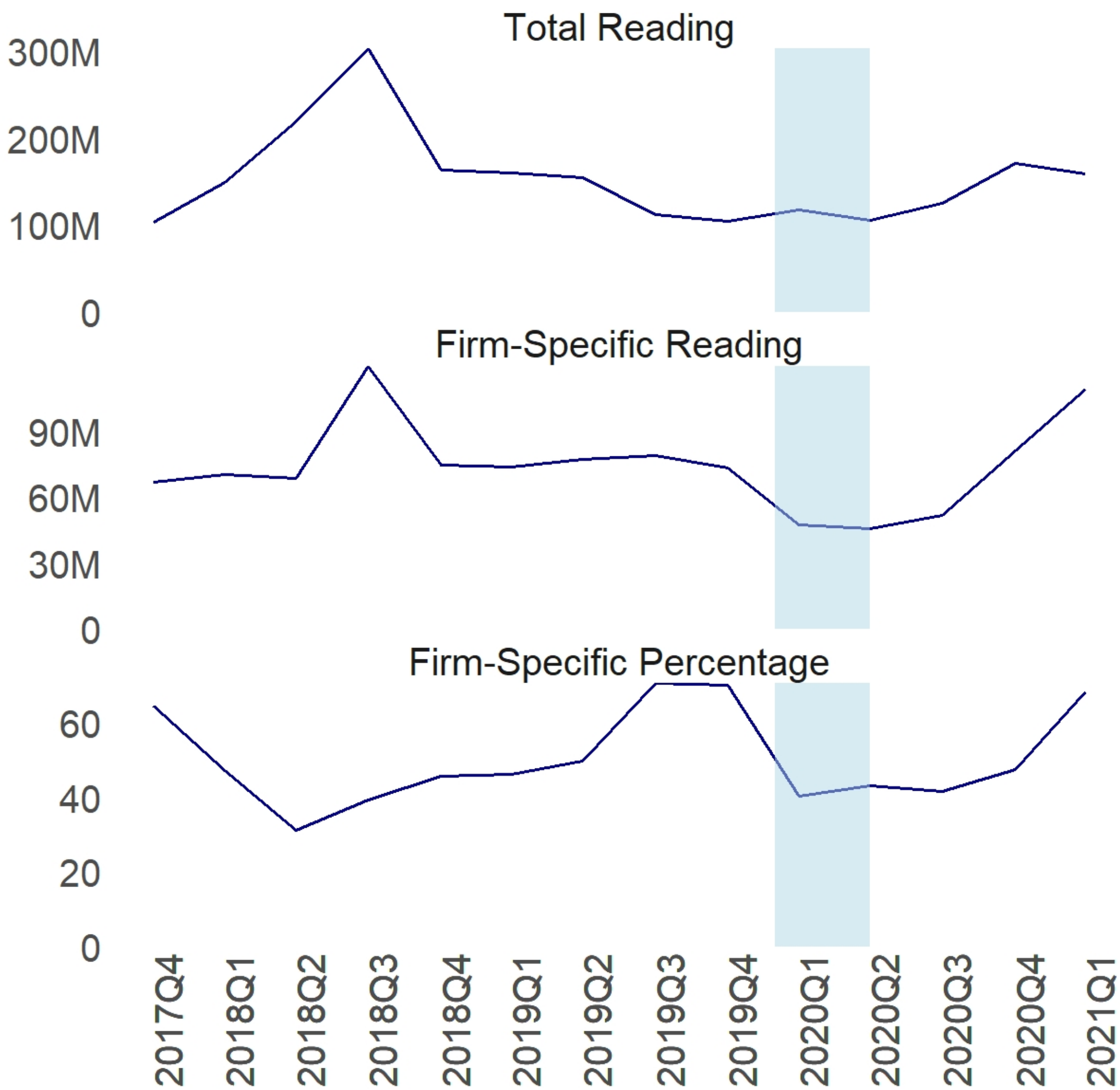
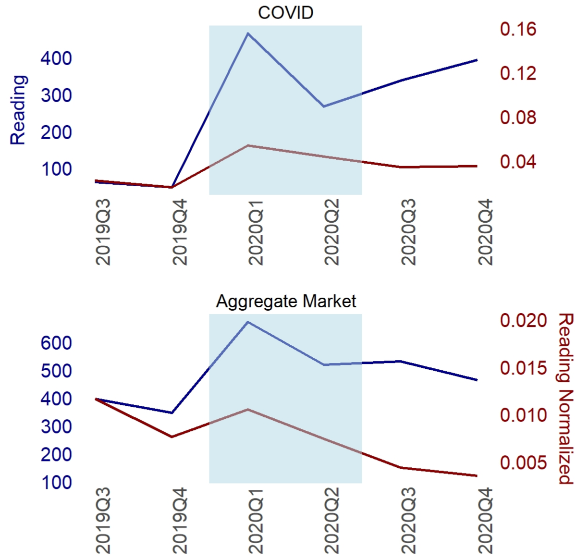


Figure 1 presents the quarterly time-series from 2017Q4 to 2021Q1 of three variables: Total Reading; Firm-Specific Reading; and Firm-Specific Reading Percentage. Total reading is the sum of reading done by all funds each quarter of all topics (financial, aggregate and macro, and other). Firm-Specific Reading is the sum of reading about firms done by all funds each quarter. Firm-Specific Reading Percentage is Firm-Specific Reading divided by Total Reading. The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue.

Figure 2: Reading Activity: Aggregate Signals

(a) Mutual Fund



(b) Hedge Fund

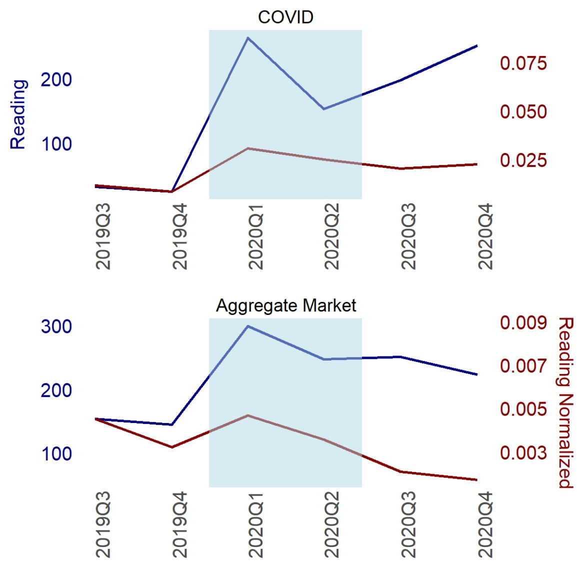


Figure 2 presents the quarterly time-series of fund attention to macroeconomic news throughout the COVID period from 2019Q3 to 2020Q4. Panel A presents measures based on Mutual Fund reading and Panel B presents measures based on Hedge Fund reading. Each panel is divided into two parts: COVID which measures attention to articles about the CDC; Aggregate Market which measures attention to articles about the aggregate United States stock market. The blue line (left axis) shows the average number of articles read about a topic per fund each quarter. The read line plots the average reading divided by the total number of articles written about the topic in the same quarter (right axis). The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel.

Figure 3: Reading Activity: Firm Specific

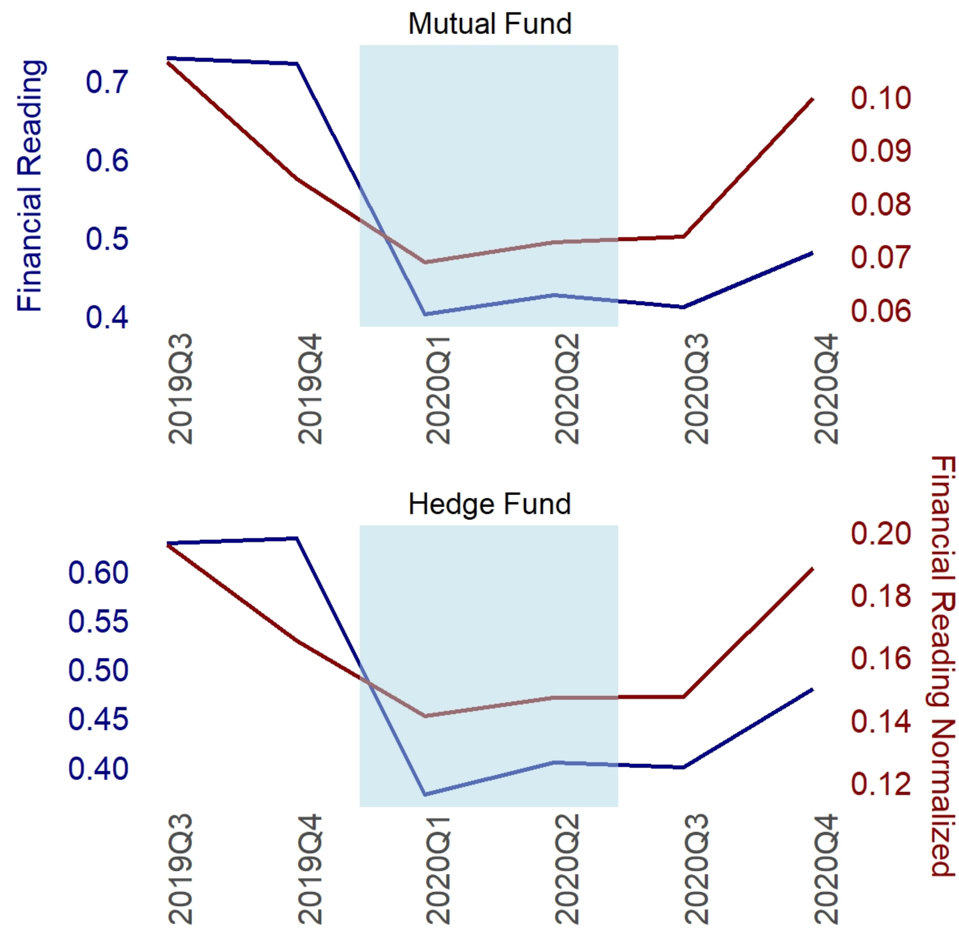


Figure 3 presents the quarterly time-series of fund attention to firm-specific news throughout the COVID period from 2019Q3 to 2020Q4. Mutual fund measures are presented on the top panel and hedge fund measures are presented below. Financial Reading, the blue line (left axis), is the sum of all financial reading (reading about firms) by all funds in the quarter divided by the sum of fund reading about all topics (firm-specific, macroeconomic, and leisure). Financial Reading Normalized, the red line (right axis), is the average financial reading per fund divided by the total number of articles written about firms in the same quarter. The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel.

Figure 4: Reading Activity: Dispersion

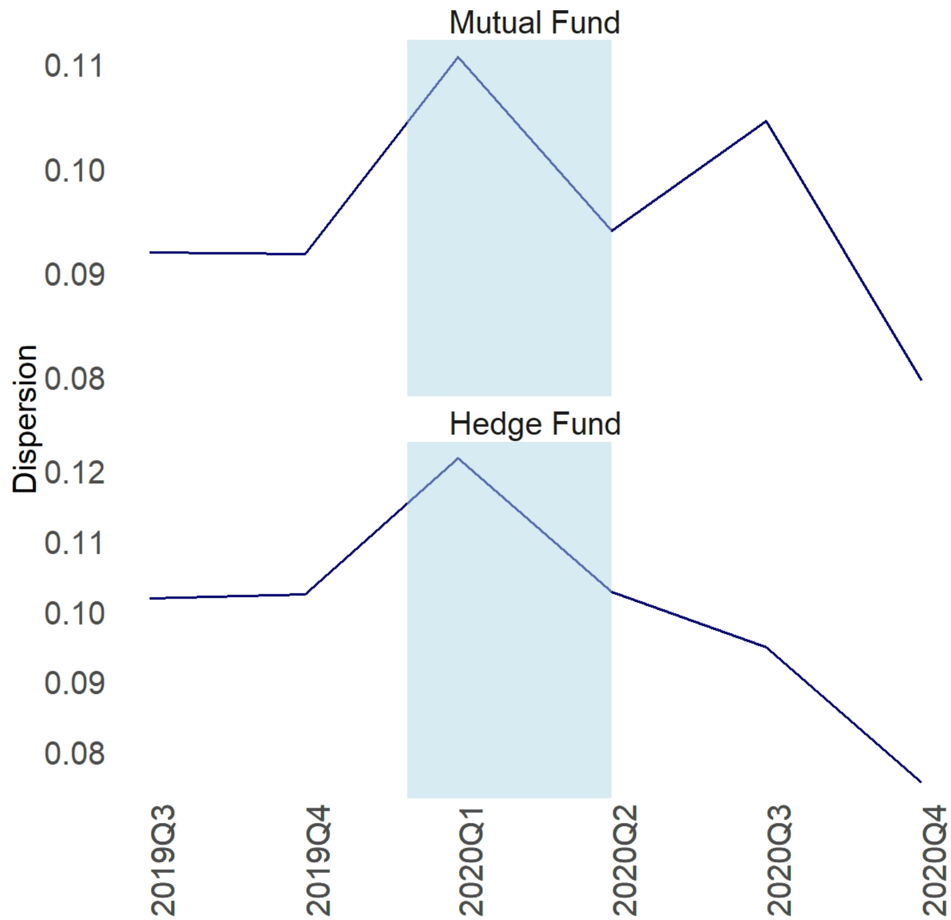


Figure 4 presents the quarterly time-series of the dispersion in fund attention throughout the COVID period from 2019Q3 to 2020Q4. The dispersion of fund attention is the average of w_{it}^{dev} across all funds in the quarter where $w_{it}^{dev} = \sum_j \left(w_{jt}^{*} - w_{jt}^{i}\right)^2$. w_{jt}^{*} is the "attention market weight", the sum of reading by all funds about firm j in quarter t divided by the sum of reading by all funds about all firms in the quarter. w_{jt}^{i} is fund i's reading of firm j in quarter t divided by the total reading of fund i about all firms in quarter t. The start of the COVID crisis (2020Q1 and 2020Q2) is highlighted in blue in each panel.

Appendix

Ravenpack

In this section, we describe the merge between Ravenpack and our URL-level dataset.This merge proceeds in three steps. First, we build a database of URLs and headlines. Our dataset is at the URL level while Ravenpack is at the headline level. Therefore, to merge with Ravenpack, we require an intermediary dataset as Ravenpack does not contain the original URL of an article. Second, we have to merge the URL-headline reading event database with Ravenpack through headline. by exactly matching on headline and day the article was published.Third, we perform data cleaning steps to ensure that the Ravenpack story that we link an article-readership event to is the most appropriate. While our goal is to ensure accuracy and minimize potential for systematic bias in our merge procedure, some of our design choices are informed by computational scale as we must merge several datasets of billions of rows.

Building a Headline-URL database: First, to develop a headline-URL dataset, obtain two sources of data: Global Database of Events, Language, and Tone (GDELT) database and Tiingo. Tiingo is a financial analytics data provider that caters to financial institutions. Institutional clients range from large pension and hedge funds to independent registered investment advisers (RIAs). One of Tiingo's provides is a news feed which records both headlines and URLs for articles across a wide range of financial news sites.

The GDELT Project is an open-source project supported by Google Jigsaw and monitors the world's broadcast, print and web news in over 100 languages. By their own description, their dataset *"identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day."* They collect millions of news articles on a daily basis and also record the URL and title of every article. We also collect various headline-URL datasets made available on Kaggle, a platform where scholars and companies often post datasets for participants to practice machine learning techniques against.We combine these three datasets to form an amalgamated date-URL-headline dataset. If an article appears in two datasets, we use the the headline from Tiingo, then GDELT, and then Kaggle.

We Ignore Frontpage Articles

We focus on non-frontpage articles. It is difficult to know what exact article that is present on a frontpage at any given point in time, given that front pages change often. Moreover, given that investors do not specifically choose to read an article on a frontpage (but rather to check the website itself) it is more dififcult to interpret reading about a firm on the frontpage of a website as the investor intended to pay attention to, or acquire information about, the specific stock. It may be sheer coincidence that the investor happens to read about the firm on the frontpage at that particular time.

Joining to Ravenpack: After joining against our URL-level database, we perform a match to Ravenpack. We proceed in two steps. First, we perform an exact date-headline match between Ravenpack and the master date-URL-headline dataset. We are able to match over half of all non-frontpage reading events via exact match. However, a considerable fraction were not exact matched and require us to perform a fuzzy match between the headline in our headline-URL dataset and Ravenpack.

There are a number of reasons fuzzy matching of headlines may be necessary. First, Ravenpack may record the headline in an article slightly differently. For example, consider the headline "Breaking News: Stocks Slated to End the Quarter on a Historic Run-Up". In one dataset, the term "Breaking News" might be omitted as "Stocks Slated to End the Quarter on a Historic Run-Up". second reason fuzzy matching may be necessary is that headlines change during the day. For example, if the headline is that "Breaking News: Stocks Slated to End the Quarter on a Historic Run-Up", this headline can change to "Stocks Slated to End the Quarter on a Historic Run-Up", or then finally later "Stocks End the Quarter on a Historic Run-Up". Hence two different datasets may parse a given text similarly, but headlines are somewhat mutable.

For all remaining URLs, we perform a fuzzy match between Ravenpack and the amalgamated dataset using 4-gram matching. We choose 4-gram matching because of the availability of comptuationally efficient algorithms to compute this. Give nthat that we must merge tens of millions of headlines in our amalgamated datasets with over 400 million articles scraped in Ravenpack, other approaches are not feasible. We retain all articles above 66%, which means at least 2/3rds of all possible 4-grams match. We perform extensive spot-checking and the results suggest that 66% 4-gram similarity is a reasonable indication the two articles have the same subject.

De-duplication

At this step, for each unique URL, we have all potential Ravenpack stories which could be potential matches for this URL. Even in the case of a headline that is exactly matched, sometimes an we may have two matches from Ravenpack. The first reason is that Ravenpack may record two entries for the same story with the same headline. The second is that an article may be reprinted across different websites. For example, articles from the Associated Press are often re-printed across many different websites. One of our publishers is not actually directly licensed by Ravenpack, but actually re-prints its content through partner publishers with a minor delay.

Therefore, for every of our 11 billion events, we find what we consider to be the best match article in consideration of when the article was read. In principle, we consider the article closest to the event that comes before the event. We consider the Ravenpack article with the closest timestamp to the event, conditional on the Ravenpack article coming before the event. [9]

Finally, we noticed that a number of URLs are not articles but rather search for a specific stock on a financial news site. To the extent it is a quote lookup, we retrieve the ticker embedded in the URL and re-enter into our dataset.

Final assessment

In the end, we match around 85% of reading events of non-frontpage articles. The missing articles are a combination of the inability to find a headline in our master URL-headline database, as well as a corresponding article from Ravenpack. Upon visual inspection of some of the unlinked articles, a substantial fraction related to Covid, political news such as the election, or other non-value relevant events. Therefore, we believe the effective match rate to be much higher.

Return Predictability Timing

Figure 6 shows an example of the timing of the measure construction and stock returns. Fund holdings are reported at the end of each quarter: $Holdings_{q-1}$ is the fund holdings at the end of the fourth quarter of 2019 from December 31, 2019, $Holdings_q$ is the fund holdings from the end of the first quarter of 2020 on March 31, 2020. In the example, we index the three months within each quarter by $1, 2, 3$: $Reading_{q-1,1}$ is the fund reading during the month of October 2019, $Reading_{q-1,2}$ is the reading during the month of November 2019, $Reading_{q-1,3}$ is the reading during the month of December 2019, and $Reading_{q,1}$, $Reading_{q,2}$, and $Reading_{q,3}$ are the reading from January, February, and March of 2020 respectively. Monthly stock returns (i.e. $r_{q-1,1}$, $r_{q-1,2}$) are presented in the same way in the figure. The buy indicator, $1_{buy_{ij,q-1}}$from Equation 28 is 1 if fund i increased holdings in firm j from $q-2$ to $q-1$ (December 31, 2019 to March 31, 2020 in the figure) and 0 otherwise. The sell indicator from Equation 29 is constructed analogously. The reading associated with this indicator is from $reading_{q-1,3}$, $reading_{q,1}$, and $reading_{q,2}$ (months December 2019, January 2020, and February 2020 from the figure) and are used to predict returns $r_{q,1}$, $r_{q,2}$, and $r_{q,3}$ respectively (January 2020, February 2020, and March 2020 monthly returns).

[9] In a small fraction of cases (less than 1%) we could also increase our observation count by including a match to an article disseminated after the reading event. This suggests that we did not match to the original article but a re-print, or that Ravenpack captured the article with a delay.

Tables and Graphs

Table 12: Sorting Characteristics

Attention Capacity	
Fund Size (*high mv*)	Sort based on fund size: $mv_{it} = \sum_j shares_{ijt} \times p_{jt}$ where $shares_{ijt}$ is the number of shares of firm j held by fund i at the end of quarter t, and p_{jt} is the share price of firm j at the end of quarter t. $highmv_{ijt}$ is 1 if the fund-firm-quarter observation is above the median mv_{it} in the quarter and 0 otherwise.
Fund Reading (*high read*)	Sort based on fund reading: $fund\ read_{it} = \sum_j read_{ijt}$ where $read_{ijt}$ is the reading of fund i about firm j throughout quarter t. $high\ read_{ijt}$ is 1 if the fund-firm-quarter observation is above the median $fund\ read_{it}$ in the quarter and 0 otherwise.
Fund Reading per Dollar (*high rpd*)	Sort based on reading per dollar: $rpd_{it} = \frac{\sum_j read_{ijt}}{\sum_j shares_{ijt} \times p_{jt}}$ where $read_{ijt}$ is the reading of fund i about firm j throughout quarter t, $shares_{ijt}$ is the number of shares of firm j held by fund i at the end of quarter t, and p_{jt} is the share price of firm j at the end of quarter t. $high\ rpd_{ijt}$ is 1 if the fund-firm-quarter observation is above the median rpd_{it} in the quarter and 0 otherwise.
Reading Concentration (*high conc*)	Sort based on reading concentration: $conc_{it} = \sum_j ia_{ijt}^2$ where $ia_{ijt} = \frac{read_{ijt}}{\sum_j read_{ijt}}$ is the fraction of reading of fund i about firm j in quarter t defined in Equation 8. $high\ conc_{ijt}$ is 1 if the fund-firm-quarter observation is above the median $conc_{it}$ in the quarter and 0 otherwise.
Sophistication	
Reading Breadth (*high breadth*)	Sort based on reading intensity: $read\ breadth_{ijt} = \frac{unique_{ijt}}{read_{ijt}}$ where $read_{ijt}$ is the reading of fund i about firm j throughout quarter t and $unique_{ijt}$ is the number of unique articles read by fund i about firm j during quarter t. $high\ breadth_{ijt}$ is 1 if the fund-firm-quarter observation is above the median $read\ breadth_{ijt}$ in the quarter and 0 otherwise.

Table 12 summarizes the different measures of fund information capacity and sophistication and outlines the construction of each variable.

Figure 5: Daily Reading during COVID

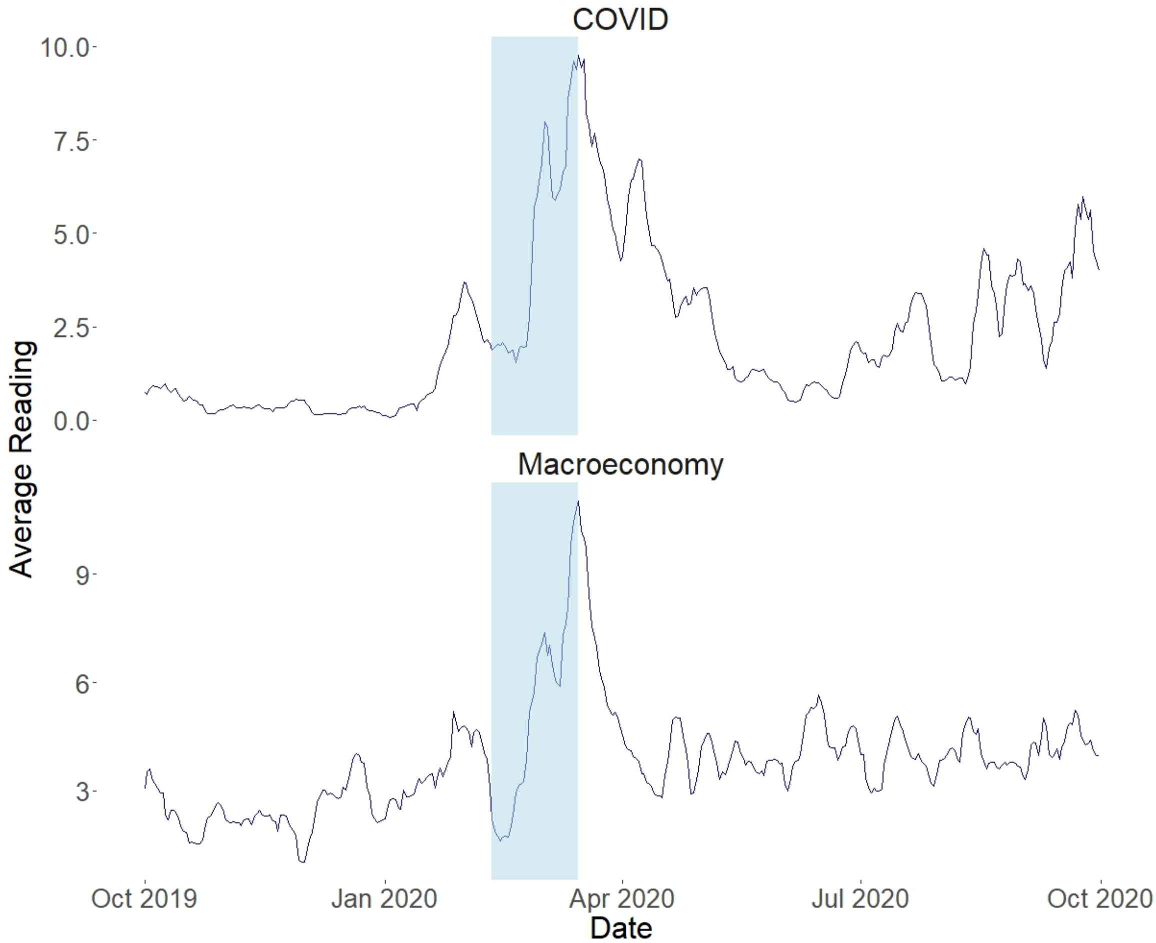


Figure 5 presents a daily time-series October 2019 to October 2020 of fund reading about COVID and the aggregate financial markets. Each variable is calculated as the total reading each day divided by the number of funds. The Figure presents the smoothed trailing seven day averages.

Figure 6: Return Predictability Timing

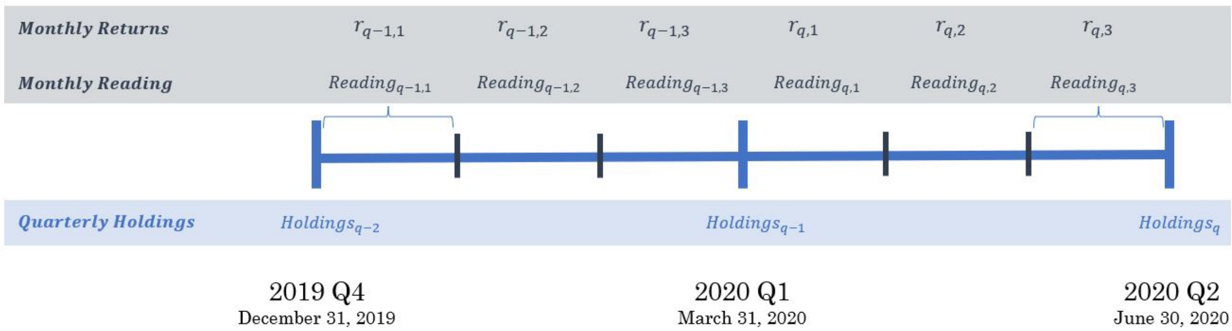


Figure 6 shows and example of the timing of the measure construction and stock returns. Fund holdings are reported at the end of each quarter: $Holdings_{q-1}$ is the fund holdings at the end of the fourth quarter of 2019 from December 31, 2019, $Holdings_q$ is the fund holdings from the end of the first quarter of 2020 on March 31, 2020. In the example, we index the three months within each quarter by $1, 2, 3$: $Reading_{q-1,1}$ is the fund reading during the month of October 2019, $Reading_{q-1,2}$ is the reading during the month of November 2019, $Reading_{q-1,3}$ is the reading during the month of December 2019, and $Reading_{q,1}$, $Reading_{q,2}$, and $Reading_{q,3}$ are the reading from January, February, and March of 2020 respectively. Monthly stock returns (i.e. $r_{q-1,1}$, $r_{q-1,2}$) are presented in the same way in the figure. The buy indicator, $1_{buy_{ij,q-1}}$from Equation 28 is 1 if fund i increased holdings in firm j from $q-2$ to $q-1$ (December 31, 2019 to March 31, 2020 in the figure) and 0 otherwise. The sell indicator from Equation 29 is constructed analogously. The reading associated with this indicator is from $reading_{q-1,3}$, $reading_{q,1}$, and $reading_{q,2}$ (months December 2019, January 2020, and February 2020 from the figure) and are used to predict returns $r_{q,1}$, $r_{q,2}$, and $r_{q,3}$ respectively (January 2020, February 2020, and March 2020 monthly returns).