# What's in Investors' Information Set?*

**Hagit Levy**†
Baruch College

**Alberto G. Rossi**§
Georgetown University

**Ron Shalev**‡
University of Toronto

**Emanuel Zur**¶
University of Maryland

November 16, 2022

**Abstract**

We exploit a unique dataset to study how the information sets of retail and institutional investors evolve over time. In line with theoretical models of rational inattention, investors' information set is positively related to their sophistication and negatively related to their opportunity cost of time, with more sophisticated investors tracking more stocks and industry sectors. Positive earnings announcements, positive trading news, and large stock price movements are the primary drivers of stock following, with few differences across investor types. Changes in stock following are positively related to future stock returns at horizons shorter than one month but negatively related to stock returns at longer horizons. Consistently, stocks with high levels of stock following tend to be over-valued and have lower expected returns. Finally, stock following tends to exacerbate the effect of news on returns and volatility, and does not result in the quicker incorporation of financial information into prices. Overall, our results suggest that stock following—particularly from non-professional investors—may have destabilizing effects on financial markets.

*We would like to thank Azi Ben-Rafael, Tony Cookson, Zhi Da, Joey Engelberg, David Hirshleifer, Peter Kelly, Yaron Levi, Marina Niessner, Lin Peng, Laura Veldkamp, and seminar participants at Bocconi University for their suggestions on earlier drafts of the paper. All errors are our own.

†Baruch College,

‡University of Toronto

§McDonough School of Business, Georgetown University. Email: agr60@georgetown.edu.

¶University of Maryland

# 1 Introduction

Modern information technology has made financial information overwhelmingly abundant. As a result, rather than being limited by what information they can acquire (asymmetric information), modern investors are limited by their attention: the time and cognitive energy they invest in processing available information (Kahneman, 1973). This makes attention choices central to portfolio choices and, in turn, an important determinant of asset prices.

A number of studies investigated the drivers of investors' attention and the effect of investors' inattention on financial markets. Early research used indirect measures of attention such as stock returns, trading volume, day of the week, and news (e.g., Barber and Odean (2008), Gervais, Kaniel, and Mingelgrin (2001), Hou, Xiong, and Peng (2009), Della Vigna and Pollet (2009), and Yuan (2015)) or aggregate measures of attention such as Google's Search Volume Index (Da, Engelberg, and Gao, 2011). More recent studies use proprietary data to directly observe attention in Bloomberg terminals (Ben-Rephael, Da, and Israelsen, 2017; Liu, Peng, and Tang, Forthcoming), website logins to retirement accounts (Sicherman et al., 2016), and web page searches in brokerage accounts (Gargano and Rossi, 2018). These data sources identify how investors collect specific pieces of information. However, they do not allow the researchers to observe which stocks individuals keep in their information set at a given point in time.

As modeled by Merton (1987), Van Nieuwerburgh and Veldkamp (2009) and Van Nieuwerburgh and Veldkamp (2010), individual and professional investors are unlikely to keep the universe of stocks traded in the market in their information set. They are instead likely to "follow" a limited number of stocks. Understanding how certain stocks are followed while others are not is fundamental to understanding how financial markets incorporate new information into prices. At one extreme, if investors' information set does not evolve over time, news about stocks that no one follows in the market may be ignored and have no impact on stock prices until much later (Huberman and Regev, 2001). In this case, the inclusion of a stock in investors' information set is a very important factor affecting how information is incorporated into asset prices. On the other extreme, if investors' information set evolves instantaneously with the release of new information, the inclusion of a stock in investors' information set should not affect how information is incorporated into asset prices.

Furthermore, the degree to which stock following allows for efficient incorporation of information into asset prices is ultimately an empirical question. If investors correctly incorporate new information, broader stock following may result in greater informational efficiency. If investors over-extrapolate (Barberis et al., 2018) or are characterized by diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018), we should expect a wider following to result in a greater degree of extrapolation and, potentially, a lower degree of informational efficiency.

In this paper, we study the economic effects of stock following using data from *Seeking Alpha*. *Seeking Alpha* is a crowdsourced content service provider for financial markets. The content in *Seeking Alpha* is crowdsourced from a large number of contributors who provide opinion pieces that cover a broad range of securities, asset classes, and investment strategies. Over the years, *Seeking Alpha* has become extremely popular among investors and is now the world's largest investing community, with over 20 million users logging into the website every month. Academic researchers have investigated the value and the effects of opinions in social media using the content created by contributors on the platform (e.g. Chen et al. (2014), Campbell, DeAngelis, and Moon (2019), Kogan, Moskowitz, and Niessner (2021), Dyer and Kim (2021)). Starting in 2011, registered users have been able to create watchlists of securities they want to follow. For the securities they include in their watchlists, subscribers receive emails containing all the content produced by contributors and breaking news, such as earnings releases and merger announcements, and transcripts of conference calls. We were given access to the watchlists data and our study is based on it.

Our data is uniquely suited to study investors' information sets as we observe, at the user level and for over 6 million users, all the initiations and changes (additions and deletions) in the watchlists since the introduction of the service and until March 2019.[1] The data hence allows us to *directly* observe the complete process of creating and maintaining an individual's investment information set, from the time an individual starts to follow a specific security to the time they decide to remove a security from the watchlist. The data also contains the self-designation of the individual subscribers. When

[1] Whereas investors likely collect information on stocks on their watchlist from multiple sources, such as StockTwits and Google, they are not likely to create different watchlists for different platforms they collect information from. For this reason, the watchlists we observe on Seeking Alpha are likely fair representations of investors' information sets. At the same time, we expect to measure the information sets of professional investors with more noise, compared to retail investors, because the former are likely to use other professional tools, like the watchlist features in Bloomberg terminals, to formulate their investment decisions. We thank Azi Ben-Rafael for this comment.

signing up for the *Seeking Alpha* service, users self-select their investor category from a list of over seventy designations. Designations include—among others—part-time investors, full-time investors, students, retirees, executives, hedge fund employees, mutual fund employees, and academics. The users' self-designations allow us to study how individual investors with differing financial expertise form their information sets.

We start by providing novel facts regarding stock following and testing whether investors' information sets relate to their attention capacity constraints. According to the theoretical frameworks developed in Van Nieuwerburgh and Veldkamp (2009) and Van Nieuwerburgh and Veldkamp (2010), investors' information acquisition can result in *generalized learning*, whereby investors learn about multiple assets or multiple sources of risk, or *specialized learning*, where information acquisition focuses on a single asset or a single source of risk. Whether investors opt for the first or second information acquisition strategy depends on their preferences and information capacity constraints. Our results suggest that investors tend to be generalized learners as very few investors focus their attention on individual stocks or individual sources of risk.

We also find that investors who are more sophisticated (full-time investors) or have a lower opportunity cost of time (retirees) form larger watchlists than those who are less sophisticated (occasional investors) or have a greater opportunity cost of time (executives), irrespective of whether we use the individual stocks or individual sources of risk to compute our results.

Investors with greater information capacity make more frequent changes to their watchlists. Full-time investors change 5.5 of the watchlist securities in a year while students and executives change 3.5 and 2.7 of their securities, respectively. Investors increase the number of securities on their watchlists over time. On average, investors add 3.7 securities per year and remove 2.1 securities per year. After one year of activity the average (median) investor has 10 (7) stocks on her watchlist and after 8 years (maximum years in our sample), an average (median) investor has 35 (22) stocks on her watchlist. Across all users and years of activity, the average (median) number of stocks in a watchlist in our sample is 13.5 (9). Gargano and Rossi (2018) document that investors hold an average (median) of 6.5 (4) stocks in their portfolio. Assuming investors are drawn from similarly representative populations, the comparison between the number of securities in our sample watchlist and the number reported in Gargano and Rossi (2018) for actual holdings suggests that the number of stocks in investors' information set is significantly larger (more than twice as many) than the number of stocks they actually own. This comparison also highlights the uniqueness of our watchlists database compared to previous studies that focus on investors' actual holdings.

Barber and Odean (2008) hypothesize that investors purchase only stocks that have caught their attention so that preferences determine choices after attention has determined the choice set. Barber and Odean (2008), however, did not have access to a proxy for investors' information set and were able to test their hypothesis only indirectly. We provide a direct test of the hypothesis in Barber and Odean (2008)—and extend the analysis to the types of news that drive investors' attention—by mapping and ranking the news events that drive investors to include stocks in their watchlist. We do so by merging our stock following data with RavenPack news data, which covers a wide range of firm-related news, such as earnings announcements, analysts' actions, M&A announcements, product and service introductions, dividend news, legal actions, executive turnovers, large stock price movements, and many others.

The majority of news types result in watchlist additions and deletions, with the most significant effect related to large movements in stock prices. Earnings releases, trading (stop or resumption of trading in firm's stock), analysts' actions, credit ratings, M&A's, executive turnovers, introductions of products and services, as well as firm marketing events, all drive investors to start following stocks. News on executive compensation, on the other hand, does not. Given that linear regressions do not provide a natural way to rank the relative importance of various news types, we consider Boosted Regression Trees (BRTs) that do allow us to compute the ranking and relative importance of different attention-grabbing news on stock following.

BRTs' relative influence measures show that watchlist changes capture different variation compared to the instantaneous attention measures that have been proposed so far in the literature. Instantaneous attention measured using Google's Search Volume Index, stock returns, or volatility is mainly driven by stock price movements and trading news. Combined, these two types of news explain more than 90% of the variation in these quantities. In contrast, five news types explain 90% of the variation in bloomberg's AIA, introduced in Ben-Rephael, Da, and Israelsen (2017) as a measure of institutional investors' attention, and 7 news types explain 90% of the variation in watchlists. The news types, however, that drive Bloomberg's AIA attention are different from the ones that explain watchlist changes. The top five news types that drive AIA are earnings releases, guidance, firm events, stock price movements, and analysts' actions. The top five news types that explain watchlist changes are stock price movements, earnings releases, trading, firm events, and analysts' actions. The other two major news events that explain watchlist changes are firm guidance and news on transactions. These results show that the decision to add a security to an individual's information set is conceptually different from the decision to pay instantaneous attention to a given stock.

As further confirmation that watchlist changes behave differently than instantaneous attention, we show that the relative ranking of news events on watchlist changes is virtually unaffected by controlling for all other measures of instantaneous attention—Google trends (Da, Engelberg, and Gao, 2011), Bloomberg's AIA (Ben-Rephael, Da, and Israelsen, 2017), returns, and realized volatility.

Finally, BRTs also allow us to analyze the differential influence of positive and negative news on attention. On this front, we find that positive news increases stock following more than negative news, with positive earnings release news having the most influence on following, consistent with investors believing that stock prices under-react to positive news and stocks are more likely to be undervalued after good news, compared to stocks over-reacting to negative news.[2]

In the second part of the paper, we investigate the asset pricing implications of stocks following. We start by assessing whether stock following predicts future returns. A large body of literature (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011; Cookson, Engelberg, and Mullins, 2021) shows that individual investors tend to focus their trades and attention on stocks that have appreciated in the past, but once they purchase the stocks, they realize negative returns going forward. To test whether individuals start paying attention to stocks while still appreciating (increasing in value), we form investment portfolios based on the changes in following at the stock level. The portfolio of stocks with significant and positive daily changes in following greatly outperforms portfolios of stocks with low changes in following, with rather large economic magnitudes. A $1 investment in the portfolio

[2]An alternative economic channel that could explain these results is "motivated beliefs" (Cassella et al., 2021), whereby stockholders might be more inclined to follow news about the stocks they hold in good periods rather than bad periods. Unfortunately, we cannot distinguish between these two competing explanations because we do not observe the investment portfolios of Seeking Alpha users. We thank Peter Kelly for this comment.

with large positive changes in following in 2011 results in \$2.2 at the beginning of 2019, representing a total return of (2.2-1)/1=120%, compared to a return of (1.6-1)/1=60% over the same period for the low following-changes portfolio. These results persist after controlling for size, value, and momentum using DGTW benchmarks.

To reconcile our findings with the body of research that finds a negative relation between retail investors purchasing specific stocks and their subsequent performance, we relate changes in attention to future returns up to two years. We find a positive and significant relation between changes in stock following and returns up to a few weeks. Subsequently, the relation reverses and becomes negative and significant. Quantitatively, the short-term appreciation associated with positive changes in following is rather large (4% on an annualized basis) and quickly declines to approximately -1% after a few months and remains unchanged after that. These results suggest two complementary takeaways. First, changes in stock following contain important information as they are associated with short-term stock appreciations. Second, combining our results with those uncovered by prior literature suggests that individuals start following stocks while they are still appreciating. Still, they do not add them to their investment portfolios quickly enough to realize profitable trades.

Next, we test whether—in addition to changes in attention—levels of attention relate to stock returns. Stocks with low following are likely to have higher expected returns if stock following and the probability of a stock being over-valued are positively related. The opposite is true if higher stock following is related to stock under-valuation. We partition the sample into stocks with low and high attention levels and find that low-attention stocks deliver higher returns than high-attention stocks. We also find that, whereas changes in attention are correlated with returns in both groups, they have a stronger effect when stocks have a low following.

We are also interested in exploring the cross-sectional implications of attention with respect to stocks movements. In particular, we investigate whether commonality in stock following translates to co-movement of stocks. By showing that pairs of stocks that are frequently included together in the same watchlists co-move above and beyond what their fundamental commonality entails, our study provides direct evidence in support of the model of Peng and Xiong (2006).[3]

In the last part of the paper, we assess whether and how stock following is related to how infor-

[3]We would like to thank Zhi Da for this suggestion.

mation is incorporated into stock prices. We show it in three distinct ways. First, high-attention stocks have a much stronger short-term reversal than low-attention stocks, consistent with higher overreaction to news in high-attention stocks. Second, stock following amplifies the effect of news on stock returns in direct panel regression tests. Third, we focus on one piece of value-relevant information that companies routinely disclose—earnings announcements—and for which we have ex-ante expectations in the form of analysts' earnings forecasts. We find that for both negative and positive earnings surprises, the price reaction on announcement days is stronger for stocks with high attention compared to stocks with low attention. We also find that the gap in returns does not close within a quarter, and, if anything, stocks with high following have a higher post-earnings announcement drift compared to stocks with low following.

The asset pricing implications of attention in a setting where attention is measured directly have been investigated. Ben-Rephael, Da, and Israelsen (2017) provide evidence that spikes in attention mitigate the post-earnings announcement drift among institutional investors. Specifically, they show that stocks exhibiting spikes in attention around earnings announcements adjust more quickly to the post-earnings announcement price and do not exhibit a drift post-earnings. While our analysis is similar to theirs, the insight we provide is different in the following respects: First, the nature of attention we focus on is different. Whereas they measure spikes in attention to a firm's stock compared to normal levels of attention to said stock, our measure captures the degree of stock following associated with each stock. Further, whereas Ben-Rephael, Da, and Israelsen (2017) measure institutional investor attention, we use a measure of individual investors' attention. The different findings between the studies further support the distinction between the different types of attention and their relations to price dynamics.

## 2 Related Literature

The behavioral literature suggests that individuals' cognitive capacities constrain information processing (Kahneman, 1973). The implications of limited attention in the context of investment decisions have been studied mainly theoretically by modeling the decisions of investors who focus on a limited number of stocks and do not keep in their information set the universe of stocks traded on all exchanges— see Merton (1987). In particular, Van Nieuwerburgh and Veldkamp (2009) and Van Nieuwerburgh and Veldkamp (2010) suggest that investors' information acquisition can result in generalized learning, whereby investors learn about multiple assets, or specialized learning, whereby investor focus their information acquisition on a single asset. Whether investors adopt the first or second information acquisition strategy depends on their preferences and information capacity. Gabaix et al. (2006) model the decision to collect and process information on a security as a real option, increasing attention when the value of the option increases and decreasing attention when the value of the option decreases. Another strand of this theoretical literature studies the implications of limited attention. Peng and Xiong (2006) show that strong co-movement in security prices within sectors are consistent with a theoretical framework in which attention-constrained individuals focus on industry and market-wide information rather than firm-specific information. Our work complements this theoretical literature by providing the first empirical evidence on how investors form and modify their information sets.

Whereas the theoretical literature, for the most part, describes attention to securities as a continuous process of information collection and processing, in the spirit of stock following, the empirical literature has, thus far, investigated attention using indirect proxies such as stock returns, trading volume, day of the week, and news (e.g. Barber and Odean (2008), Gervais, Kaniel, and Mingelgrin (2001), Hou, Xiong, and Peng (2009), Della Vigna and Pollet (2009), and Yuan (2015)) or aggregate measures of attention such as Google's Search Volume Index (Da, Engelberg, and Gao, 2011). More recent studies use proprietary data to directly observe attention in Bloomberg terminals (Ben-Rephael, Da, and Israelsen, 2017; Liu, Peng, and Tang, Forthcoming), website logins to brokerage accounts (Sicherman et al., 2016), and web page searches in brokerage accounts (Gargano and Rossi, 2018). In these recent studies, the nature of the data allows only to observe instantaneous attention. Our watchlists data allows us to go beyond the instantaneous attention that has been explored so far in the literature and understand why certain stocks are followed more than others and what are the economic consequences of stock following.

A prominent factor that brings stocks to investors' attention is news. Multiple studies have shown the effect of news on attention, subsequent trading and investment outcomes. Da, Engelberg, and Gao

(2011) offer google search as a measure of individual investor attention and provide evidence that it is affected by news events. Barber and Odean (2008) use brokerage data on individual clients and show that they are net buyers of stocks that are in the news. Gargano and Rossi (2018) analyze data of page searches and the time spent by investors on their broker's website to investigate factors that drive attention. They find that investors tend to spend more time on stocks that are more frequently in the news, and that paying attention rewards investors, with higher investment returns. Sicherman et al. (2016) use data of clients' logins to their brokerage accounts and find that the type of news (good or bad) affects the attention investors pay their portfolio. Ben-Rephael, Da, and Israelsen (2017) analyze data from Bloomberg terminals and find that news coverage also drive institutional investors' attention. Liu, Peng, and Tang (Forthcoming) use Bloomberg and Google Trends data to document that different news drive institutional and individual investors' attention. Their results highlight the importance of considering clientele effects in understanding the effect of news on attention and asset prices. Finally, Madsen and Niessner (2019) provide evidence that even firm advertisements, which hardly provide new information, drive investors' attention. Using a comprehensive news database, as well as data on individuals stock following, we are able to explore which news types affect investors' information set. We are also able to provide evidence on stocks characteristics, other than transient news, that drive investors to follow stocks.

Finally, a number of papers have studied the capital market effects of limited attention. Most relevant to this study are Gargano and Rossi (2018) that provide evidence of a positive relation between attention and performance in attention grabbing stocks, and Ben-Rephael, Da, and Israelsen (2017) that provide evidence that spikes in attention to a stock among institutional investors mitigate the post-earnings announcement drift effect. We contribute to this literature by studying the relation between stock following and price dynamics.

## 3 Data and Summary Statistics

The primary data source of this study is a proprietary database of security watchlists, shared with us by *Seeking Alpha*. *Seeking Alpha* is a crowdsourced internet content service provider for financial markets participants. The informational content of Seeking Alpha is provided by contributors with different levels of expertise, including individual investors, industry experts and finance professionals (financial advisors of financial institutions). In July 2011, Seeking Alpha added a feature to their website, which allowed registered users to create watchlists of securities. For those subscribers who opt to create a watchlist, the service provider sends daily emails with content relevant to the securities in their watchlists. The content includes the contributed pieces that are relevant to the security, breaking news such as earnings announcements, merger announcements, and transcripts of conference calls associated with the firm.

Our data include the personal identification number of every registered user to the service since July 2011, a record of every security ticker added to a registered user's watchlist, and the time and date of each addition to the watchlist. Our data also includes a record of every ticker removed from a registered user's watchlist (i.e., the time and date of the removal) from October 2015 onwards. We identify the initiation of a watchlist as the first date that a registered user adds a security to her watchlist. Typically, users add multiple securities on the initial date and then update their watchlist over time by adding or removing a single security every time.

The data set contains information on 6,068,738 registered users that created one watchlist from July 2011 to December 2018.[4] The data also include about 600,000 registered users that created multiple watchlists over time. To avoid duplicate entries, we remove watchlists of users that created more than one watchlist from our analysis. Of the 6.1 million users, 4,136,697 initiated a watchlist but did not make any changes to it throughout our sample period, and 1,932,041 users added and removed securities over time. Our analyses focus on the latter group of 1,932,041 registered users, which we refer to as active users.

Panel A of Table 1 reports the distribution of users opening accounts over the period July 2011 - December 2018 by the self-designation of watchlist creators. The second year of the service (2012) has the lowest number of watchlist initiations (410,292) and year 2014 has the largest number of watchlist initiations (1,228,074). Panel B shows that annual initiations for active users follow a similar trend. The largest number of watchlist creators in our sample self-designate as occasional investors with 465,000 active watchlists, followed by finance professionals, full time investors and retirees, all with

[4] Our descriptive statistics are compiled from users that registered to the service until the end of December 2018 to allow for classification into the active user category.

more than 100,000 active watchlists.

Table 2 reports statistics on the users' watchlist activity over time. This table, and all following tables, pertain only to the 1.9 million active registered users. Panel A of Table 2 reports the average number of securities in a watchlist as a function of the number of years the watchlist is active. The average (median) number of stocks comprising an initial watchlist is 6 (5). After one year of activity an average (median) investor has 10 (7) stocks on her watchlist and after 8 years (maximum years in our sample), an average (median) investor has 35 (22) stocks in her watchlist. Across all users and years of activity, the average (median) number of stocks in a watchlist in our sample is 13.5 (9). Studies that track individual investors' actual holdings (e.g. Gargano and Rossi (2018)) document that investors hold an average (median) of 6.5 (4) stocks in their portfolio. The comparison between the number of stocks in our sample watchlists and the number of actual portfolio stocks in Gargano and Rossi (2018) suggests that investors follow a significantly larger number of stocks (more than twice as many) than they own. The comparison also highlights the uniqueness of our watchlist database compared to previous studies focusing on investors' actual holdings.

Panel B of Table 2 reports statistics of stock additions to watchlists and Panel C reports statistics on the deletions from watchlists (reported only for watchlists initiated after October 2015). Investors add more securities on average (3.7 securities per year) than they remove from their watchlists (2.1 securities per year).

Our second main source of data comes from the RavenPack news analytics database. RavenPack analyzes unstructured content from thousands of publications to extract company news. The information analyzed by RavenPack includes a wide array of firm related news spanning financial (e.g. earnings release, stock price movements, analysts' actions, and dividends), operational (e.g. CEO turnover, restructuring), and other firm events (e.g. product development and introduction, M&As, litigation, accidents and war crimes). RavenPack's textual analysis provides multiple quantitative measures for the news collected on each firm including relevance, novelty and sentiment measures. Table 4 reports the frequency of news data.[5] Column 1 reports the number of news events per firm per year, columns 2-4 break down the news based on their positivity level (positive, neutral and negative). The most frequent type of news in the database relates to technical analysis averaging 14 news

[5]For details on how we process the RavenPack Data for this study, please refer to Online Appendix A.2.

events per firm per year. The second most frequent relates to insiders buying and selling the stock and averages 7 events per year. Earnings release is next with 5.25 news event per year and analysts actions news occur 3.8 times a year on average. Finally, news on large stock price movements occur on average once a year.

Additional standard data sources used in the analysis include COMPUSTAT for firms financial information, Center for Research in Security Prices (CRSP) data for the estimation of trading volume, number of shares outstanding, prices, and stock returns, and I/B/E/S (IBES) data for analysts' coverage and earnings announcements. Finally, we use data on Web searches from Google and data on institutional investors' attention from Bloomberg.

Details of all the variables used in the empirical estimates are described in Online Appendix A.2.

# 4 Stock Following and Information Capacity Constraints

The first set of tests we provide focuses on how investors' information set relates to their degree of sophistication and their opportunity cost of time. According to the theoretical framework developed in Van Nieuwerburgh and Veldkamp (2009) and Van Nieuwerburgh and Veldkamp (2010), if asset returns are independent, investors' information acquisition can result in generalized learning, whereby investors learn about multiple assets, or specialized learning, whereby investors focus their information acquisition on a single asset. When assets are correlated, specialized learning entails investors learning about one source of risk, such as, for example, tech stocks or only stocks whose returns are highly correlated. Generalized learning entails learning about multiple sources of risk, such as multiple sectors.

Whether investors opt for the first or second information acquisition strategy depends on their preferences and information capacity constraints, where the latter can be thought to depend on investors' sophistication and their opportunity cost of time. Fortunately, in our context, we can use investors' profession to gauge the degree to which they are sophisticated and how much time they can possibly dedicate to researching stocks on Seeking Alpha. For example, full-time investors are likely to be more sophisticated than occasional investors. At the same time, retirees may have a lower opportunity cost of time than executives because they do not have intense full-time jobs.

Because it is difficult to find a perfect mapping between the theoretical framework of Van Nieuwerburgh and Veldkamp (2009) and our *Seeking Alpha* setting, we provide two alternative sets of results. In the first (Table 3), we report results on individual stocks followed by *Seeking Alpha* users, conditional on their occupation. Panel A presents statistics on initial watchlists. Panel B and C present annual additions and deletions, respectively, and Panel D presents statistics on the total number of changes made to watchlists per year (Additions + Deletions). Our results show significant differences in the number of stocks investors include in their watchlist at initiation across different types of investors. Full-time investors, likely the most sophisticated investor group, create the largest watchlists with, on overage, 6.9 stocks. Executives and occasional investors who are, respectively, likely to have a higher opportunity cost of time and who are likely to be less sophisticated than full-time investors, create smaller initial watchlists with 5.9 and 6.1 stocks, respectively. Students, who on average have little to gain from tracking stocks because less wealthy, create the smallest initial watchlists that contain only 5.7 stocks. Our results also show that information capacity constraints play an important role in the maintenance of a watchlist. Investors with greater information capacity make more frequent changes to their watchlists. For example, full-time investors change 5.5 of the watchlist securities in a year while students and executives change 3.5 and 2.7 of their securities, respectively. Overall, these results suggest that investors tend to be generalized learners as virtually no one in our dataset focuses on a single stock.

In the second set of results, we take a broader approach and consider how many sectors are tracked by each user of *Seeking Alpha*, implicitly assuming that different sectors are exposed to potentially different sources of risk. We first assign each stock to one of the 49 Fama-French industry sectors; we then compute how many sectors the users of *Seeking Alpha* track when they first open a watchlist, as well as after 2, 4, and 8 years after opening the watchlist, conditioning on their self-designation. While we report the full set of results in Table Online I, we summarize our key findings in Figure 1 where we plot the cumulative average number of sectors tracked by full-time investors, retirees, occasional investors, students, and executives.

Across all investor groups, *Seeking Alpha* users are generalized learners, in the sense that, from the time they start a watchlist, they track at least three-four sectors of the economy. At the time watchlists are created, more sophisticated investors, such as full-time investors, and investors who have a low opportunity cost of time, such as retirees, are the ones that create the largest watchlists. The ones that create the smallest watchlists are instead the ones who have very little to gain from investing, such as students, and those who have an extremely high opportunity cost of time, such as executives. In the middle, we find occasional investors, who are neither sophisticated nor have a high or low opportunity cost of time. As we expand the time horizon, the total number of sectors tracked by each investor group increases, and the differences across the various groups magnify. After eight years, full-time investors track a total of 14.5 sectors while company executives track as little as 11 sectors.

Overall, both the results based on individual stocks and the ones based on sectors indicate that the majority of investors are generalized learners. Even when we consider the number of sectors tracked by investors' initial watchlists in Panel A of Table Online I we show that, at the 25th percentile, the majority of the groups (except for the undeclared ones) track at least two sectors. In unreported results, we find that the groups with the most specialized learners are executives, occasional investors, and students, with 23%, 19%, and 18% of the group members engaging in specialized learning. In line with Van Nieuwerburgh and Veldkamp (2009) and Van Nieuwerburgh and Veldkamp (2010), we find that, for generalized learners, the number of stocks and sectors followed is closely related to investors' attention capacity constraints.

# 5 Baseline Results on Stock Following and News

Barber and Odean (2008) hypothesize that investors are net buyers of attention-grabbing stocks, which results from investors' difficulty in tracking the thousands of stocks they can potentially buy. The empirical findings in Barber and Odean (2008) provide indirect evidence for this hypothesis and make the authors conclude that preferences determine choices after attention has determined the choice set. In this section, we provide a direct test of the hypothesis in Barber and Odean (2008)—and extend their analysis to the types of news that drive investors' attention—by relating changes in investors' information set to different news types.

We begin by estimating baseline regressions investigating the news-related drivers of individuals' decisions to initiate and/or cease following individual stocks. Our specification reads as follows:

$$ATTENTION_{i,t} = \alpha_i + \alpha_t + \beta_1 NEWS_{i,t} + \epsilon_{i,t}, \tag{1}$$

where the dependent variable, $ATTENTION_{i,t}$ is proxied by eight different attention measures. The four measures based on *Seeking Alpha* data are: $ADDITION-EX$, the number of times a stock $i$ was added to existing watchlists during week $t$; $ADDITION-NEW$, the number of times a stock $i$ was added to newly created watchlists during week $t$; $DELETIONS$, the number of times a stock $i$ was removed from watchlists during week $t$; $NET-ADDITIONS$, the number of times a stock $i$ was added to existing watchlists minus the number of times a stock was removed from watchlists during week $t$. The four additional attention measures are: 1) Google Trend Searches (Da, Engelberg, and Gao, 2011), which are generally thought of as measures of retail-investor attention; 2) Bloomberg's abnormal institutional investor attention (AIA) (Ben-Rephael, Da, and Israelsen, 2017); 3) stock returns; 4) and stock price volatility. To maintain consistency across all attention measures, Google Trends being the constraint, we work at the weekly frequency.

The firm news variable is denoted by "NEWS." It is measured as an indicator variable that takes the value of 1 if the RavenPack database reports a news related event associated with the firm on week $t$. We include the following news events from the database: Trading news (stop and restart of trading in firm stock), stock price news (news about large changes in firms' stock prices), stock restructuring news, additions and removals of a stock from indexes, transactions in firm stocks, legal action involving the firm, firm guidance, mergers and acquisitions, earnings release, announcement of firm event, products and services introduction, credit rating initiation and update, creation of partnerships, insider trading, dividends, executive turnover, executive compensation, labor issues, major shareholder disclosures, technical analysis news, and news about accidents a firm was involved in. We include firm fixed effects to control for time invariant factors that may affect the number of times a stock is added to watchlists in our database, such as firm size. The weekly fixed effects control for time trends affecting all tickers.

Results for the baseline regressions are reported in Table 5. The table contains 8 columns, each representing a different attention measure. We highlight a number of facts. First, except for news about executive compensation, accidents, labor issues and major shareholders disclosure, all news events attract attention to stocks, with results consistent across all attention measures. Also, news events are positively related to both stock additions and deletions, implying that the same news may be interpreted differently by different investors. For some, it is a positive signal to include the stock in the information set. For others, it is a negative signal.

Second, news events have a higher explanatory power for the decision to start or stop collecting and processing information about a stock (watchlist additions and deletions) compared to instantaneous attention, proxied by Google trends. Whereas news events explain 29% of the variation (R-squared) in ADDITION-EX and 21% in DELETIONS, news explains only 10% of the variation in google searches. This first piece of evidence suggests that the type of attention captured by *Seeking Alpha*'s watchlists is different from the attention captured by existing measures such as Google trends.

Third, news events have a much stronger explanatory power for stock additions to existing watchlists compared to stock additions to newly created watchlists: the $R^2$ is 29% in the first case and 17% in the second, suggesting that the marginal decision to add additional stocks to someone's information set is economically distinct from the one of creating a group of initial stocks to follow.

Finally, while Bloomberg's AIA is a dummy variable and hence not directly comparable to Google trends and our watchlist measures of attention, it displays significant differences in terms of the pieces of news that correlate with it. For example, trading and stock price movement news, which explain a large portion of google trends and watchlist changes, are less closely related to Bloomberg's AIA changes compared, for example, to "Guidance News," as evidenced by the coefficient estimates and $t$-statistics. Whereas the coefficient on trading news (guidance) for existing watchlist additions is 0.65 (0.17) with a $t$-statistic of 19.93 (10.95), the coefficient on trading news (guidance) for Bloomberg AIA is 0.09 (0.14) with a $t$-statistic of 15.23 (22.37). These comparisons, however, provide limited insights because they are based on standard panel linear regression estimates.

One major limitation of linear regressions is that, while it is possible to estimate the economic and statistical significance of each source of news in explaining changes in stock following and/or instantaneous attention, there isn't a natural measure of rank ordering of the covariates in the model as well as the relative importance of each explanatory variable. In our context, it is important to understand what is the relative economic significance of the different sources of news in driving investor attention. To accomplish this, we adopt a machine learning tool named Boosted Regression Trees (BRTs).

### 5.1 BRTs and the Relative Importance of Different Types of News

Linear regressions do not provide a natural measure for the relative importance of the various sources of news in explaining investors' attention. We therefore estimate the equivalent of Equation 1 using BRTs (for an introduction to BRTs, see Online Appendix A.1). However, because BRTs do not allow naturally to control for firm and time effects, we estimate our model in three steps. In the first one, we regress $ATTENTION_{i,t}$ on firm and time effects using a standard panel regression model and keep the residuals, which we call "orthogonalized attention." In the second, we do the same for each of the regressors and we keep the residuals, which we call collectively "orthogonalized news." In the final step, we use BRTs to estimate the relation between orthogonalized attention and orthogonalized news. If we were estimating this final step using OLS, we would obtain coefficient estimates that are numerically identical to the ones obtained using one-step panel regression estimates reported in Table 5. Estimating this relation using BRTs instead allows us to compare the relative importance of the various sources of news in explaining investors' attention.

We report the results of this analysis in Figure 2. Subfigure A reports the relative influence measures of news on watchlist changes (NET-ADDITIONS) as the attention variable. Subfigures B, C, D, and E repeat the estimates using Google Trends, Bloomberg's AIA, stock returns, and volatility, respectively. News on large stock price movements is overwhelmingly the single most important driver of individuals' decision to start following a stock by adding it to the watchlist. Large stock price movements are also the most influential factors attracting attention both for Google trends and stock returns. The second most relevant type of news for watchlist changes is earnings releases, which turns out to be the most important driver of Bloomberg's AIA. Recall that Google Trends are usually thought to capture retail investors' attention while Bloomberg's AIA is taken as a measure of institutional investors' attention. The relative influence results suggest that watchlist changes seem to capture some elements of both institutional investor and retail investor attention with greater emphasis on the latter.

Google Trends, stock returns, and volatility have more than 90 percent of the relative influence of news concentrated on 2 news types—large stock price movements and trading news. On the other hand, Bloomberg's AIA is driven by a much wider range of factors with the top five news types, i.e. earnings releases, guidance, firm events, stock price movements, and analysts' actions explaining 90% of the variation in the measure. Watchlist changes, in this respect, are closer to Bloomberg's AIA as they are driven by numerous news types, with the relative influence of large stock price movements only at 30 percent and trading news at 15 percent. Other than large stock price movements and trading news, the release of firms' financial information is a major catalyst for stock following, with earnings releases ranking second with a 20 percent relative influence and firm guidance ranking sixth with an 8 percent relative influence.

Note also that our results on the relative ranking of news events on stock following are virtually unaffected by controlling for other measures used in the literature to capture investors' attention, such as changes in Google trends (Da, Engelberg, and Gao, 2011), Bloomberg's AIA, returns, and realized volatility, as shown in Subfigure (e) of Figure 3. This shows that these types of news have an impact on stock following above and beyond the one they have on instantaneous attention.

A key feature of the *Seeking Alpha* data is the self-designation of investor types, which allows us to test whether investors of varying degrees of sophistication form their information sets on the basis of different types of news. We provide evidence along this dimension in Figure 3. Across all subfigures that categorize investors in 9 different groups, ranging from students and occasional investors to academics and finance professionals, the types of news that drive investors' stock following are—consistently—large stock price movements, trading news, firms events, and earnings releases. We find very little differences in relative influences across different investor categories.

### 5.2 The differential effect of positive and negative news

Given that the vast majority of investors rarely short stocks, individuals are likely to include in their information set securities they consider purchasing (or they already own) as they perceive them to be undervalued. By partitioning news based on whether they convey positive or negative information about the affected firms, we can provide evidence on the extent to which investors believe stock prices under-react to positive news or over-react to negative news.

To this end, we focus on watchlist changes (additions minus deletions from existing watchlists) as the primary measure of stock following, and test whether it is affected differently by positive and negative news. Our specifications control for other measures of attention such as changes in Google trends, Bloomberg's AIA, stock returns, and stock return volatility. Controlling for stock returns is especially important in this setting as many news events impact stock returns, which in turn may affect stock following. We report regression results in Table 6, where we focus on positive, negative and neutral news separately.

While the table is very dense, a few facts arise. First, whereas legal news attract stock following only when they are negative, firms' index related news (addition to and removal from an index), insider trading, product and service announcements, partnerships and dividends news generate more following only when they are positive. Second, news related to mergers and acquisitions, earnings announcements, and technical analysis attract following irrespective of the sentiment of the news, but the effect is stronger when news are positive. Third, news on strong stock price movements on the downside increase stock following more than news about movements on the upside. The same is true for news on credit ratings and guidance. Finally, analysts' actions news exhibit no differential effect depending on whether they are positive or negative, both of which increase stock following.

Overall the linear results suggest that both negative and positive news affect investors' attention and stock following, with some differences related to the type of news. Once again, the linear regression results do not help us understand the relative importance of the different news types, so we turn to BRTs. In subfigures (a) and (b) of Figure 4, we report BRT relative influence measures separately for positive and negative news. Positive earnings release news are much more influential in attracting stock following than positive trading news (45% relative influence vs 20%) while negative stock price movements news are more important than negative earnings release news (35% relative influence vs 22%).

Finally, in subfigure (c) of Figure 4 we estimate a BRT specification that includes both positive and negative news. Positive earnings release news are the most important source of news, followed by positive trading news, negative stock price news and positive stock price news. Aggregating across news types shows positive news are much more influential in affecting changes in investors' watchlists, with a total relative influence of 65%, compared to 35% for negative news, consistent with investors believing stocks under-react to positive news rather than stocks over-reacting to negative ones.

### 5.3 Stocks Fundamental Characteristics and Following

Results presented thus far focused on investors' marginal decision to add or delete stocks to and from their information set. As shown in Table 5, this decision is different from the one of creating an initial watchlist, which is less influenced by transient news: the $R^2$ for existing watchlists (first column) is 29%, almost double the one of new watchlists (second column), which equals instead 17%.

In this subsection, we consider stock-specific, non-news-related drivers of stock following and focus on watchlist initiations. We estimate the baseline specification:

$$Initial\_Attention_{i,t} = \alpha + x'_{i,t}\beta + \epsilon_{i,t}, \tag{2}$$

where $Initial\_Attention_{i,t}$ is the logged number of times a stock is included in a newly created watchlist in a given week divided by the overall lagged following of the stock, and $x_{i,t}$ is a vector of non-news-related stock characteristics. We follow Gargano and Rossi (2018) and include in $x_{i,t}$ measures firm fundamental covariates, such as size, profitability, leverage, age, and R&D spending. We also include stock price information, such as volatility and higher moments of stock returns, as well as trading-related covariates, such as trading volume, turnover and the ratio of a firm's stock price to its book value of equity. Finally, we include analysts' coverage and institutional ownership. Table 7 reports the results.

Firm size and profitability positively relate to attention, whereas firms' R&D spending, age, and leverage do not relate to following. Past returns, volatility, skewness, market-to-book ratio, and stock turnover all increase the likelihood that a stock is included in an initial watchlist, while kurtosis and volume do not relate to attention. Finally, analysts' coverage do not play a role, while large institutional holdings negatively relate to stock following.

Overall the results are consistent with the notion that individuals tend to follow larger, more profitable firms that performed well in the past and that past stock price movements are perceived to reflect larger expected gains and losses, consistent with investors engaging in trend-chasing when adding stocks to their information sets.

# 6 Stock Following and Price Dynamics

Our analysis thus far showed that stock following is different from instantaneous direct attention—as captured by Google trends and Bloomberg's AIA—or instantaneous indirect attention—as measured by stock returns and stock price volatility. We also presented evidence consistent with similar reactions to news by different types of investors. In this section, we analyze the economic consequences of stock following. We first analyze the information content of stock following by constructing investment portfolios on the basis of changes in investors' stock following and show that stock following contains valuable, but short-lived, information that allows for the construction of profitable investment strategies.

We then examine whether stock following relates to how information is incorporated into stock prices. High-attention stocks have a much stronger short-term reversal compared to low-attention stocks, consistent with high-attention stocks overreacting to news more. Second, we show stock following amplifies the effect of news on stock returns. Finally, we zoom into earnings announcement news, possibly one of the most consequential news releases, and show that stocks with high following not only experience larger price changes on earnings release days, but also experience more pronounced drifts thereafter, consistent with stock following having de-stabilizing effects on financial markets.

## 6.1 Stock Following and Stock Returns

As a first validation exercise and to make sure that, indeed, changes in stock following are related to stock price dynamics, we regress the cumulative annualized trading volume (logged) on changes in stock following, controlling for stock fixed effects and time effects. We find that positive watchlist changes are associated with positive and significant increases in trading volumes at horizons that range from two to four weeks, consistent with investors adding stocks to their watchlist in preparation to trade them in the future. The effect is not present at the one-week horizon and dissipates on the fifth week, where it stays positive but becomes statistically insignificant. Negative watchlist changes, on the other hand, are associated with negative but not statistically significant changes in trading volume.[6]

Next, we estimate whether stock following predicts future returns. A large literature, including Barber and Odean (2008), Da, Engelberg, and Gao (2011), and Cookson, Engelberg, and Mullins (2021), shows that individual investors focus their trades and attention on stocks that have appreciated in the past. The same investors realize negative returns after purchasing such stocks. An important consideration is that a significant amount of time may elapse between the time individuals start following a stock and the time they actually purchase it. To test whether individuals start paying attention to stocks while they are still appreciating, we form investment portfolios on the basis of the changes in attention at the stock level.

We use data from July 2011 to March 2019 and compute daily changes in stock following across all users in our dataset.[7] Every day we divide the stocks into three quantiles on the basis of the changes in stock following. In order to make changes in stock following comparable across all stocks, we make sure to divide the changes in stock following by the level of stock following on the previous day. We then compute the next day returns across all stocks in each of the three portfolios. Finally, we cumulate the daily returns of each portfolio from the first date to the last date available. We report the results in Figure 5.

In Subfigure (a), we show that the portfolio of stocks with large, positive daily changes in attention greatly outperforms the portfolio of stocks with low (or negative) changes in attention. The economic magnitudes are large. A $1 investment in the portfolio with high attention changes in 2011 results in $2.2 on March 2019, for a total return of (2.2-1)/1=120%. The portfolio with low attention changes delivers instead a return of (1.6-1)/1=60% over the same period.

A concern with the results reported in Subfigure (a) is that the outperformance may simply be compensation for risk, that is, individuals may be paying attention to high risk stocks (which is consistent with the evidence in Table 7) and the high return may represent compensation for risk. To overcome this limitation, we recompute the results using DGTW adjusted returns, rather than simple

[6]We would like to thank Zhi Da for suggesting this validation exercise.

[7]Note that for the first four years of our Dataset, Seeking Alpha users could not delete stocks from their Watchlist, so we start showing negative changes in following only starting in 2015.

stock returns. The advantage of using DGTW returns rather than simple returns is that we control for three sources of outperformance that have been widely documented in the literature, size, value and momentum. Momentum is likely the most important effect we want to control for, because our results in the previous sections show that an important determinant of stock following rests in past positive price appreciations. The results, reported in Subfigure (b) of Figure 5, show that indeed portfolios of stocks with high positive changes in investor attention deliver positive abnormal cumulative returns, while portfolios of stocks with low attention changes deliver negative abnormal cumulative returns.

A second natural question is whether not only the changes, but also the levels of attention predict future returns. We perform this exercise in the remaining subfigures of Figure 5, where we focus on the stocks with low attention levels in Subfigure (c) and stocks with high attention levels in Subfigure (d). Stocks with little following are likely to have higher expected returns if stock following and the probability of a stock being over-valued are positively related. Consistent with this hypothesis, our findings highlight two facts. First, low-following stocks deliver higher returns than high-following stocks: the black and red lines in Subfigure (c) are much higher than the corresponding black and red lines in Subfigure (d). Second, whereas the changes in attention are important in differentiating between high and low returns in both groups, they seem to have a stronger effect when stocks have low following.

Together, these findings suggest, first, that stocks with high following deliver lower returns compared to stocks with low following. Second, positive changes in stock following predict higher stock returns, at least in the short run. To analyze in detail these seemingly contradictory findings, we estimate the panel regressions:

$$Ret_{i,t:t+k} = \alpha_i + \alpha_t + \beta \; W\_Change_{i,t} + \epsilon_{i,t:t+k} \tag{3}$$

where $Ret_{i,t:t+k}$ is the return realized by stock $i$ from the end of week $t$ to the end of week $t+k$, $\alpha_i$ represent stock fixed effects and $\alpha_t$ represent time effects—which absorb stock-level differences in returns over the period as well as time variations in stock market returns. Finally, $\beta$ measures the relation between stock returns and $W\_Change_{i,t}$, the weekly percentage change in stock following associated with stock $i$ over the course of week $t$. We let $k$ range from 1 to 100, meaning that the future returns are computed over an horizon of almost two years.

Subfigure (a) of Figure 6 reports the beta coefficient estimates and associated 95% confidence intervals for the different regressions. There is a positive and significant relation between stock following and returns up to 6 weeks. Subsequently, returns revert and the relation becomes insignificant from week 9 to week 20. Finally, the relation between attention changes and returns becomes negative and significant from week 21 to week 100.

The results in Subfigure (a) cumulate the returns, but do not annualize them, so the positive returns we uncover in the first weeks may seem economically small. Once we annualize the returns in Subfigure (b), their magnitude becomes more comparable across horizons and our results show that the short-term effects we document are economically sizable. The appreciation associated with positive changes in following is rather large (4% on an annualized basis) and quickly declines to approximately -1% after four months and remains unchanged thereafter.

Overall, the results reported in this section show two complementary findings. First, changes in stock following contain important information in that they are associated with short-term stock appreciations. Second, over longer horizons high following is generally associated with stocks being overvalued and having lower returns.

### 6.2 Commonality in Attention and Stock Returns Co-movement

Peng and Xiong (2006) show that the co-movement in security prices within sectors is consistent with a theoretical framework in which attention-constrained individuals do not focus on stock-specific news but sector-wide news. Previous research has shown empirical support for the model of Peng and Xiong (2006). For example, using large jackpot lotteries as exogenous shocks that attract investors' attention away from the stock market, Huang, Huang, and Lin (2019) show that stock returns co-move more with the market on large jackpot days, consistent with investors being less attentive on jackpot days.

Most of the empirical evidence on the effect of attention/inattention in the literature, however, is *indirect* because information on the stocks tracked by individual investors was not available so far. Our data provides a very natural setting to provide a *direct* test of the Peng and Xiong (2006) model because it allows us to estimate directly the degree to which commonalities in attention patterns across stocks relate to stock return correlations.

We start by focusing on the stocks in the S&P 500 index, computing the correlation in daily returns across all stock pairs, and reporting the cross-sectional variation across all stocks in Figure 7. In Subfigure (a), we compare the correlations of stocks that are in the same industries in blue and different industries in red where industries are categorized using 1-digit SIC codes. Subfigure (b) through (d) repeat the exercise with 2-digit, 3-digit, and 4-digit SIC codes, respectively. We highlight two facts. First, stocks in the same SIC codes are more strongly correlated with each other compared to stocks in different SIC codes. Second, as we use a finer SIC classification in subfigures (b) through (d), the correlation between stocks in the same industry becomes larger and larger.

In the second step, we compute—for each pair of stocks in the S&P 500—the degree to which they are jointly held by investors in their watchlists. We compute this measure by counting the total number of watchlists where a specific pair of stocks are held together (over the full sample) and scaling it by the total number of watchlists that include one of the two stocks in the pair. This scaling guarantees that the commonality in watchlist presence is bounded between zero and one.

In the third step, we estimate cross-sectional regressions relating the degree of commonality in attention across stocks to their correlations in returns. We start by reporting results across all stocks in column (1) of Table 8. The average correlation among the stocks in the sample that have no commonality in attention is 12.3% (the constant), and a unit increase in the commonality of attention is associated with a 22.5% increase in stocks' return correlations.

The results in the first column do not consider the fact that certain industry-wide news are likely to affect similarly stocks in the same industry. In the second and third columns, we repeat the exercise but focus on stocks that belong to the same industries (column 2) or different industries (column 3) using stocks' 1-digit SIC code. We highlight two facts. First, and as expected, stocks that belong to the same industry have a higher correlation compared to those that belong to different industries: the constant is 17.7% in column (2) and 11.6% in column (3). Second, the relation between commonality in attention and returns correlation is much stronger for stocks in the same industry: the coefficient on the commonality in attention is 39.5% in column (2) and 14.8% in column (3).

One may be concerned that using 1-digit SIC codes may be too broad to define companies' membership in a given industry. In columns (4) through (9), we re-estimate our results using 2-digit, 3-digit, and 4-digit industry classifications. In all cases, we find robust results. Even though the baseline for stocks that do not have commonality in attention increases as we consider more and more granular industry specifications, the coefficient on commonality in attention is strongly related to stock returns correlation, in line with the predictions of the model by Peng and Xiong (2006).

# 7 Stock Following and the Incorporation of News into Asset Prices

In the previous section, we explored the time series and cross-sectional implications of stock following. In this section, we continue exploring the real effects of stock following by studying the degree to which stock following affects how new information is incorporated into asset prices.

## 7.1 Stock Following and Short Term Reversal

If investors correctly incorporate new information, greater stock following is likely to improve the speed and precision with which new information is incorporated into asset prices. If investors over-react to news, stock following may result in de-stabilizing effects. Short-term reversal is often thought to be generated by investors' over-reaction to news and price trends (Shiller, Fischer, and Friedman, 1984; Black, 1986; Da, Liu, and Schaumburg, 2014; Summers and Summers, 1989). Seeking Alpha data provides an ideal laboratory to study these effects.

We take the universe of stocks in our data and compute, at the daily frequency, the cumulated returns for each stock over the previous 22 days. We then double-sort stocks into quantiles on the basis of their past returns and level of following, controlling for the firms' market capitalization to avoid confounding stock following for company size. Third, for each attention level, we construct long-short equally-weighted reversal portfolios that go long in the loser stocks and short in the winner stocks. Finally, we cumulate the returns of these long-short portfolios over time. Results are reported in Figure 8.

The reversal of high attention stocks (blue dotted line) is much more pronounced than that of low attention stocks (black solid line), consistent with high attention stocks over-reacting to news more than low attention stocks. The economic magnitudes are large. The long-short high attention portfolio is associated with a 300% cumulated return. The low attention portfolio instead features a cumulated return of only 120% over the same period, almost one third. The short-term reversal strategy on the full set of stocks is instead associated with a cumulative return of a little over 200%.

The results provided in this section are consistent with high attention stocks being associated with over-reaction to news and price trends, suggesting that stock following may exacerbate the impact of news on financial markets and not increase the efficiency with which news are incorporated into asset prices.

## 7.2 Stock Following and Price Reaction to News

In this section, we test the degree to which information capacity constraints affect investors' reaction to news. Intuitively, investors characterized by limited attention are bound to react more to the release of novel information related to the stocks they follow, because 1) they are more likely to observe such information (i.e. they receive emails about it) and 2) they are likely knowledgeable enough to understand the relevance of such news for the value of the company and the associated stock, place trades and consequently affect asset prices. To estimate the degree to which limited attention matters in asset markets, when it comes to the incorporation of news, we estimate panel regressions of the following form at the weekly frequency:

$$\begin{aligned} Outcome_{i,t} = \alpha_i + \alpha_t + &\sum_{l \in \{G,B,N\}} \beta_l \times News_{i,l,t} \times Following_{i,t-1} \\ &+ \sum_{l \in \{G,B,N\}} \gamma_l \times News_{i,l,t} + \delta \times Following_{i,t-1} + \eta \times X_{i,t-1} + u_{i,t} \end{aligned} \tag{4}$$

where $Outcome_{i,t}$—the dependent variable of interest—is, alternatively, stock returns, stock price volatility, trading volume or searches measured via Google trends; $News_{i,l,t}$ is the number of news of type $l$ regarding stock $i$ that occurs in week $t$. News are categorized into three groups: Good (G), Bad (B) and Neutral (N). Finally, $Following_{i,t-1}$ is investors' following of the stock as of week $t-1$, divided by the market capitalization of the stock, and $X_{i,t-1}$ is a vector containing the following control variables: market leverage, book leverage, firm income, Market-to-Book Ratio, asset tangibility, R&D, variation in analysts recommendations, variation in analysts EPS forecasts, the fraction of institutional investors holdings out of total shares outstanding, institutional investors breadth, the Herfindahl index of institutional investors, log price, log number of analysts covering the stock, log market capitalization, past risk-adjusted returns, past skewness and past kurtosis.

Note that the presence of firm-specific and time fixed-effects means that our coefficients are estimated only exploiting within-firm variation in stock following. The coefficients of interest are the $\beta$s, which measure the differential effect of news on stock returns, volatility, and google trends as a function of the stock following on *Seeking Alpha*. The results are reported in Table 9. In columns 1 and 2 we report results for the effect of attention around news events on stock returns. We then repeat the analysis for stock returns volatility (columns 3 and 4), trading volume (columns 5 and 6), and google trend searches (columns 7 and 8). Odd columns do not include control variables while even columns do.

Consistent with the results in Figures 5 and 6, the coefficient on stock following $\delta$ is positive and significant: an increase in stock following over the previous week predicts positively returns on week $t$. Positive news are also related to returns: the $\gamma_G$ coefficient is strongly positive and significant. The coefficients $\beta_G$ on the interaction between lagged following and positive news is significant at the 1 percent level for stock returns, stock returns volatility, trading volume and Google searches. The effect is significantly stronger for all outcome variables—except for Google trend searches—after controlling for a slew of company characteristics. In all cases, an increase in stock following amplifies the effect of news on stock returns, volatility, trading volume and google searches.

Results around negative news are insignificant for returns, trading volume and google searches, but they are positive and significant for volatility. Together with the positive news results, these coefficient estimates suggest two findings. First, investors' stock following increases disagreement around news, irrespective of whether they are positive or negative. Second, investors' stock following is related asymmetrically to stock returns, depending on the sign of the news, in that the effect of positive news on returns is amplified but the effect of negative news is not. These results provide yet another piece of evidence that stock following may have destabilizing effects on stock prices.

### 7.3 Stock Following and Earnings Announcements

The results reported so far focus on all news types, and therefore mix the effect of news of high and low relevance. In addition, while we can observe news events, it is hard to quantify the ex-ante expectation of the majority of investors prior to the news release. For example, knowing that GM laid off 50,000 employees may be interpreted as good news by those expecting a bigger layoff and bad news for those that were expecting zero layoffs. While Ravenpack categorize news as positive, negative or neutral, it is unclear how precise their algorithm is. Furthermore, the algorithm for the classification may have been trained on the very same data we are using to evaluate the effect of news on stock prices.

To circumvent these potential limitations, we focus on one piece of value-relevant information that is routinely disclosed by companies—earnings announcements—and for which we have ex-ante expectations: analysts earnings forecast. The fundamental question we address is whether stock following affects the flow of new information into asset prices. On the one hand, high level of stock following implies better baseline information about the firm and quicker processing of the news, which could increase the impact of news on asset prices and therefore reduce or eliminate the post-earnings announcement drift that has been documented in the literature. On the other hand, higher following could also have the reverse effect of leading to an over-reaction to earnings news and a subsequent drift. A third possibility is that stock following does not have any impact on asset prices around earnings news. This is likely to occur if investors react similarly to stocks they follow and stocks they do not follow.

Ben-Rephael, Da, and Israelsen (2017) find that institutional investors' attention measured as Bloomberg's AIA accelerates incorporation of information into prices and reduces the drift in prices following the announcement, consistent with the first possibility described above. Consistently, Della Vigna and Pollet (2009) find that earnings announcements that occur on Friday, when investor inattention is more likely, have a 15% lower immediate response and a 70% higher delayed response. We follow Della Vigna and Pollet (2009) and Ben-Rephael, Da, and Israelsen (2017) and track how prices react when earnings are released as a function of how the earnings announced relate to analysts' expectations and the stock following of the company. Specifically, we take all earnings announcements and compute Standardized Unexpected Earnings (SUE) by subtracting mean earnings analysts' forecasts from the actual earnings announced and dividing by the standard deviation of the earnings forecasts. We then construct SUE quintiles across earnings announcement dates and companies, where the first quintile contains the earnings announcements with the most negative earnings surprises, and the fifth quintile, earnings announcements with the most positive earnings surprises. We also separate the stocks into two groups based on their size-adjusted stock following, defined as the log of the ratio between stock following and market capitalization, with the least followed stocks in the first group and the most followed stocks in the second group. Finally, we track the cumulative DGTW-adjusted daily returns of the two groups of stocks around earnings announcements with Standardized Unexpected Earnings in the lowest and highest quintile.

The results of this analysis are reported in Figure 9. The results in Subfigure (a) are associated with negative earnings surprises, and the results in Subfigure (b) are associated with positive earnings surprises. The figures highlight a few important facts. First, for both negative and positive earnings surprises, the price reaction on announcement days is stronger for stocks with high attention (red line) compared to stocks with low attention (blue line). Second, the gap in response to the news does not close for 90 days (the full quarter until the next quarter's earnings announcement). In fact, firms with high stock following exhibit more drift than stocks with low stock following, essentially increasing the gap in returns between highly followed stocks and sparsely followed stocks.[8] Finally, the gap in announcement returns between highly followed stocks and sparsely followed ones is larger for positive earnings surprises, while the drift is larger for negative surprises. Taken together, these results suggest that higher stock following is associated with a larger stock price reaction to news. In the case of positive news, this translates to a larger price jump on announcement day and a larger drift. For negative news, probably because of limits to short-selling, higher attention translates into a more pronounced drift over the three months following the announcement, but not differential price jumps on announcement days.

These results, together with the ones in the previous sections, suggest that stock following may have de-stabilizing effects on stock prices rather than increasing the efficiency with which information incorporates into asset prices.

[8] When we follow Della Vigna and Pollet (2009) and compute the ratio between the post-earnings announcement drift and the earnings announcement returns, we indeed find that it is greater for stocks with higher following.

### 7.3.1 Distinguishing Between Professional and Non-Professional Following

The results reported so far do not distinguish between investor types. For example, we would expect more sophisticated investment professionals to arbitrage away mis-pricing in financial markets because they have the investment capital to do so. Professional investors' experience in financial markets is also likely to limit the extent to which they over-extrapolate (Barberis et al., 2018) or are characterized by diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018).

Even though our results in Section 5.1 show that similar news attract the attention of different types of investors, stock following may have a differential effect on stock prices depending on how sophisticated and deep-pocketed the followers are. We test for this in Figure 10 where we group our Seeking Alpha users into two groups - non-professional investors—a category that includes students, retirees, journalists, executives and academics—and professional investors, which includes full time investors and professional investors.[9]

Subfigures (a) and (b) reports the results for non-professional and professional investors for the fifth quintile of earnings news, i.e. the most positive ones. Two facts emerge. First, differential stock following by non-professional investors (Subfigure (a)) is strongly related to stock price dynamics in that the red line (high stock following) experiences a bigger jump on announcement day and a stronger subsequent drift compared to the blue line (low stock following). Also, the confidence intervals show that the stock price dynamics are statistically different from each other, suggesting that stock following by non-professional investors is related to significantly different stock price dynamics.

Second, differential stock following by professional investors (Subfigure (b)) does not seem to have an impact on stock price dynamics post earnings announcements. This result is consistent with professional investors being able to incorporate new stocks in their information set in real time as new information is released and stock following by professional investors having very little impact on how stocks behave around news releases.

In unreported results, we repeat the analysis for the first quantile of earnings news—the most negative ones. In this case, we do not observe any differential effect of stock following by either non-professional or professional investors, probably because short-selling constraints limit the extent to

[9]We exclude from the analysis the undeclared investors.

which investors of either type can impact stock prices through trading.

These results, together with the ones reported above, show that stock following may have de-stabilizing effects on stock prices, particularly when it is associated with non-professional investors and positive news.

## 8 Conclusions

We exploit a unique dataset to study how the information sets of retail and institutional investors evolve over time. In line with theoretical models of rational inattention, investors' information set is related to their sophistication, with more sophisticated investors tracking more stocks and industry sectors. Similar types of news—i.e., positive earnings announcements, positive trading news and large stock price movements—are the primary drivers of stock following across all types of investors. Changes in stock following are positively related to stock returns at horizons shorter than one month, but negatively related to stock returns at longer horizons. Consistently, stocks with high levels of following tend to be over-valued and have lower expected returns. Finally, stock following tends to exacerbate the effect of news on returns and volatility and does not result in the quicker incorporation of financial information into prices. Overall, our results suggest that stock following—particularly from retail investors—may have de-stabilizing effects on financial markets.

## References

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez. 2015. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118:135–67.

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies* 21:785–818.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2018. Extrapolation and bubbles. *Journal of Financial Economics* 129:203–27.

Ben-Rephael, A., Z. Da, and R. D. Israelsen. 2017. It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies* 30:3009–47.

Bianchi, D., M. Buchner, and A. Tamoni. 2020. Bond risk premia with machine learning. Working Paper.

Black, F. 1986. Noise. *The journal of finance* 41:528–43.

Bordalo, P., N. Gennaioli, and A. Shleifer. 2018. Diagnostic expectations and credit cycles. *The Journal of Finance* 73:199–227.

Breiman, L., J. Friedman, C. J. Stone, and R. Olshen. 1984. *Classification and regression trees*. Taylor & Francis.

Campbell, J. L., M. D. DeAngelis, and J. R. J. Moon. 2019. Personal stock holdings and investors' response to stock analysis on social media. *Review of Accounting Studies* 24:731–79.

Cassella, S., B. Golez, H. Gulen, and P. Kelly. 2021. Horizon bias in expectations formation. *Available at SSRN 3759035* .

Chen, H., P. De, Y. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27:1367–403.

Cookson, J., J. Engelberg, and W. Mullins. 2021. Echo chambers. Working Paper.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

Da, Z., Q. Liu, and E. Schaumburg. 2014. A closer look at the short-term return reversal. *Management science* 60:658–74.

Della Vigna, S., and J. M. Pollet. 2009. Investor inattention and friday earnings announcements. *The Journal of Finance* 64:709–49.

Dyer, T., and E. Kim. 2021. Anonymous equity research. *Journal of Accounting Research* 59:575–611.

Frank, M. Z., and V. K. Goyal. 2003. Testing the pecking order theory of capital structure. *Journal of Financial Economics* 67:217–48.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189–232.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg. 2006. Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review* 96:1043–68.

Gargano, A., and A. G. Rossi. 2018. Does it pay to pay attention? *The Review of Financial Studies* 31:4595–649.

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

Gu, G., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies* 33:2223–73.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning*. Springer.

Hou, K., W. Xiong, and L. Peng. 2009. A tale of two anomalies: The implications of investor attention for price and earnings momentum. *Available at SSRN 976394* .

Huang, S., Y. Huang, and T.-C. Lin. 2019. Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics* 132:369–83.

Huberman, G., and T. Regev. 2001. Contagious speculation and a cure for cancer: a nonevent that made stock prices soar. *Journal of Finance* LVL:387–96.

Kahneman, D. 1973. *Attention and effort*, vol. 1063. Citeseer.

Kogan, S., T. J. Moskowitz, and M. Niessner. 2021. Social media and financial news manipulation. Working Paper.

Liu, H., L. Peng, and Y. Tang. Forthcoming. Retail attention, institutional attention. *Journal of Financial and Quantitative Analysis* .

Madsen, J., and M. Niessner. 2019. Is investor attention for sale? the role of advertising in financial markets. *Journal of Accounting Research* 57:763–95.

Merton, C. R. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42:483–510.

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

Shiller, R. J., S. Fischer, and B. M. Friedman. 1984. Stock prices and social dynamics. *Brookings papers on economic activity* 1984:457–510.

Sicherman, N., G. Loewenstein, D. Seppi, and S. Utkus. 2016. Financial attention. *Review of Financial Studies* 29:863–97.

Summers, L. H., and V. P. Summers. 1989. When financial markets work too well: A cautious case for a securities transactions tax. *Journal of financial services research* 3:261–86.

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

———. 2009. Information immobility and the home bias puzzle. *Journal of Finance* 64:1187–215.

Yuan, Y. 2015. Market-wide attention, trading, and stock returns. *Journal of Financial Economics* 116:548–64.

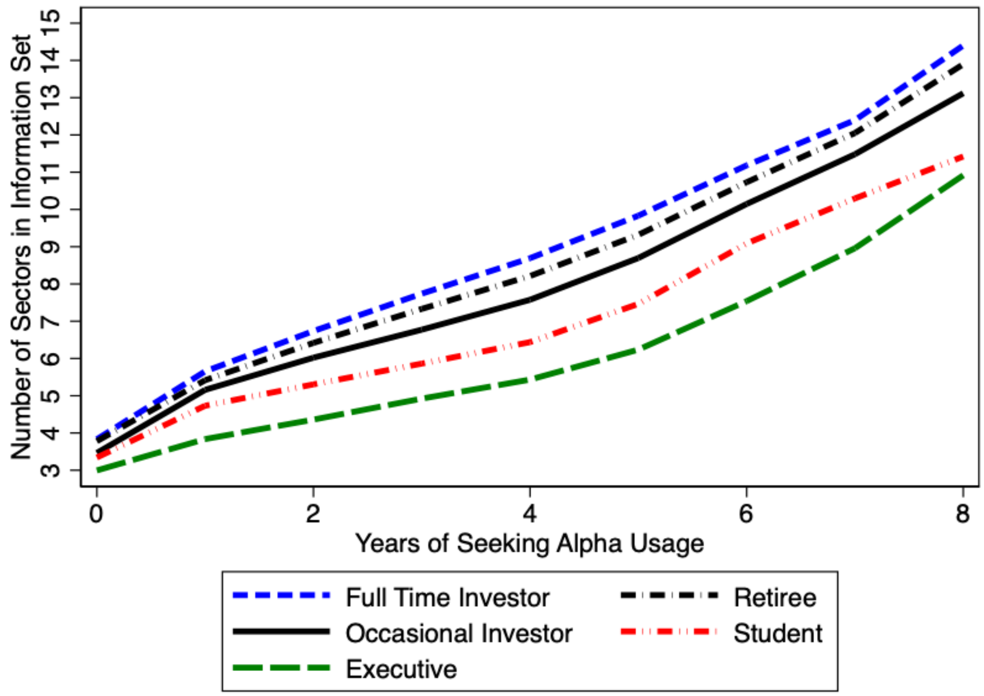
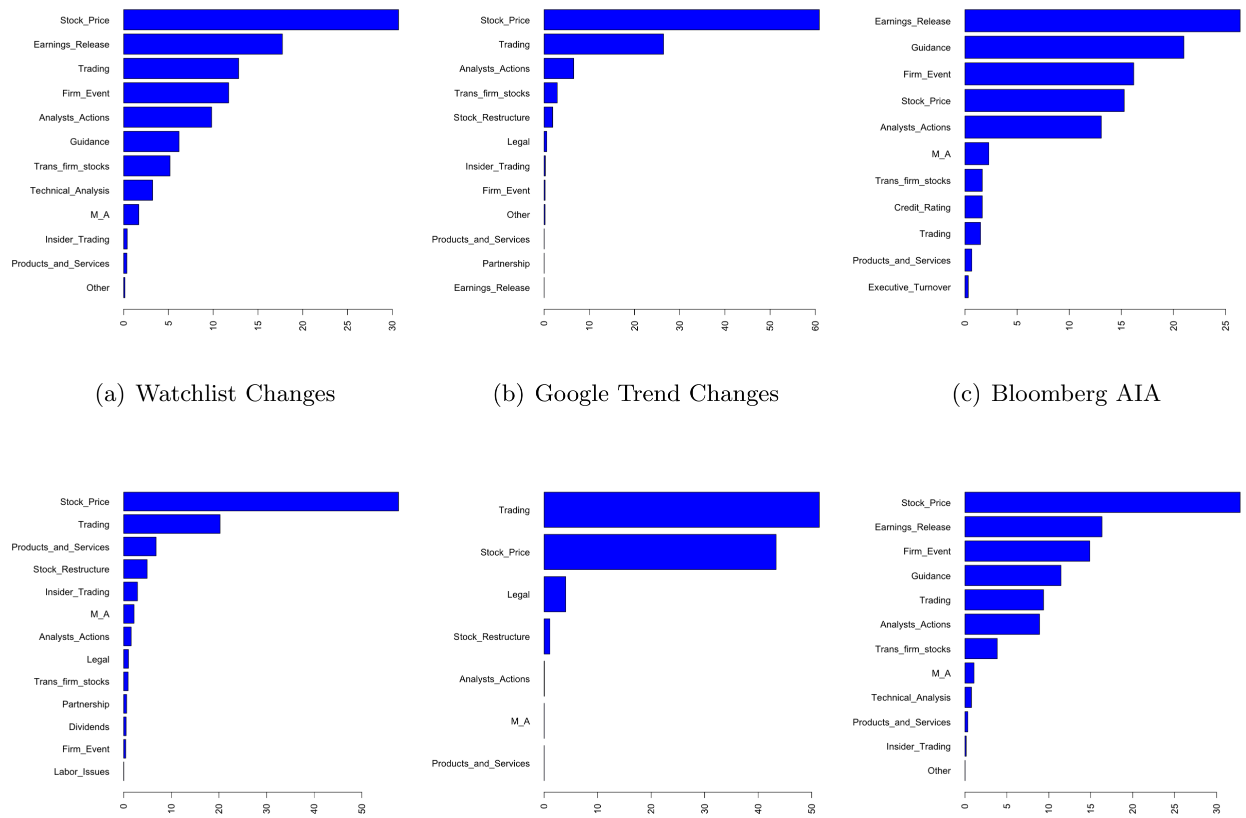


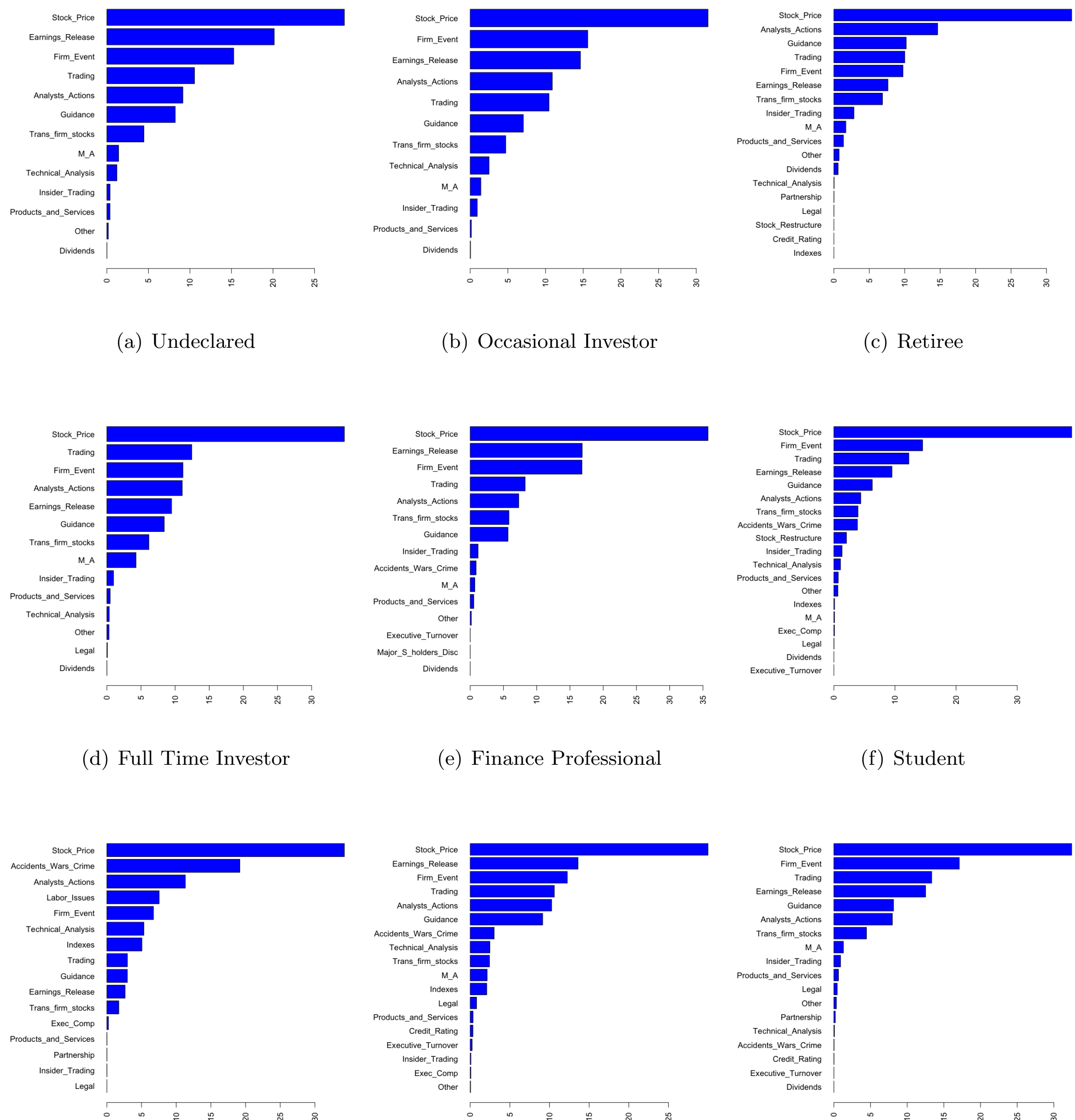

Figure 1: This figure reports the average number of sectors followed by investors on *Seeking Alpha* for five groups of investors: full-time investors, retirees, occasional investors, students, and executives. For each group of investors, we compute the cumulated average number of sectors followed from the initiation of the watchlist all the way to 8 years after starting the watchlist.




(a) Watchlist Changes (b) Google Trend Changes (c) Bloomberg AIA

(d) Returns (e) Realized Variance (f) Watchlist Changes, with Controls

Figure 2: **Relative importance of News in Explaining Stock Following and Attention.** This figure reports the relative influence measures of the different news types in explaining changes in stock following (Subfigure A), instantaneous attention by retail investors measured by Google Trends changes (Subfigure B), instantaneous attention by sophisticated investors measured by Bloomberg's AIA indicator (Subfigure C), stock returns (Subfigure D) and stocks' realized variance (Subfigure E). Finally, in Subfigure F we report the relative influence measures of the different news types in explaining changes in stock following *controlling* for Google Trends, Returns, and Volatility. The results in each panel are computed in three steps. In the first one, we regress each measure of following/attention on firm and time effects using a standard panel regression model and keep the residuals, which we call "orthogonalized attention." In the second, we do the same for each of the regressors, and we keep the residuals, which we call collectively "orthogonalized news." In the final step, we use BRTs to estimate the relation between orthogonalized attention and orthogonalized news.

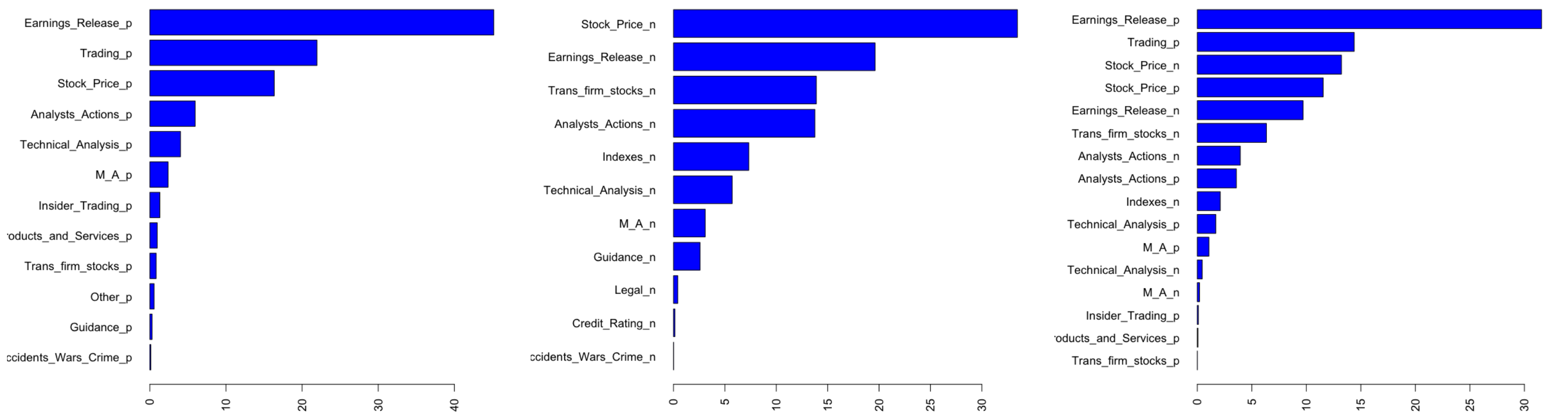



(a) Undeclared (b) Occasional Investor (c) Retiree

(d) Full Time Investor (e) Finance Professional (f) Student

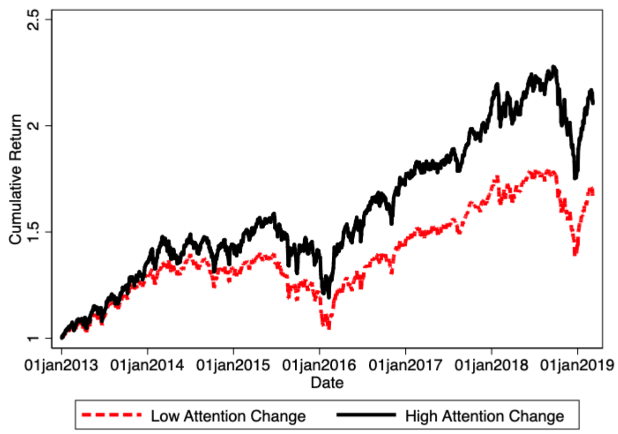
(g) Academic (h) Executive (i) Journalist

Figure 3: **Relative importance of News across Different Investor Groups.**
This figure reports the relative influence measures of the different news types in explaining changes in stock following across nine different investor groups: undeclared (A), occasional investors (B), retirees (C), full-time investors (D), finance professionals (E), students (F), academics (G), business executives (H) and journalists (I). The results in each panel are computed in three steps. In the first one, we regress each measure of following/attention on firm and time effects using a standard panel regression model and keep the residuals, which we call "orthogonalized attention." In the second, we do the same for each of the regressors and we keep the residuals, which we call collectively "orthogonalized news." In the final step, we use BRTs to estimate the relation between orthogonalized attention and orthogonalized news.

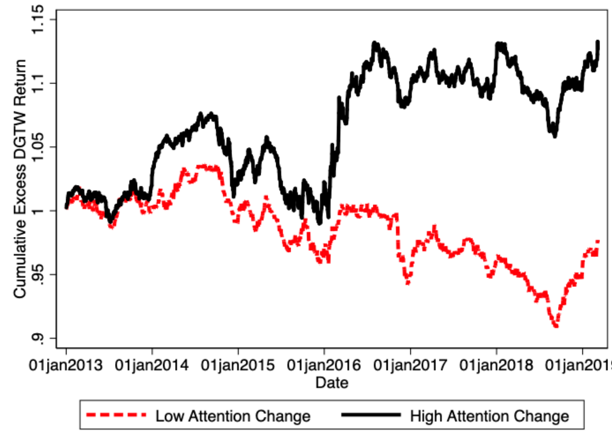



(a) Positive News, with Controls (b) Negative News, with Controls (c) Positive and Negative News, with Controls

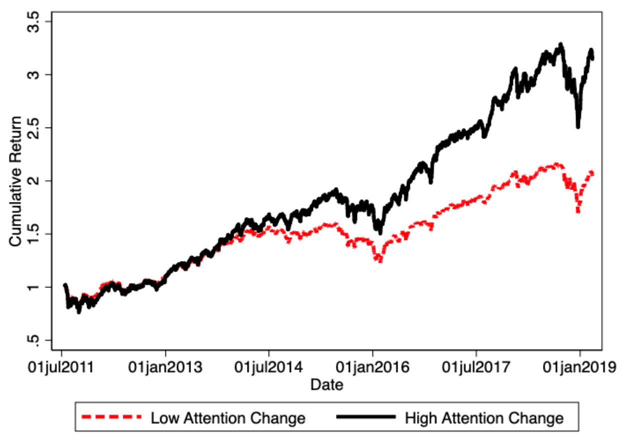
Figure 4: **Relative importance of News for investors' following, controlling for Google Trends, Returns, and Volatility.** This figure reports the relative influence measures of the different news types in explaining changes in stock following controlling for Google Trends, Bloomberg's AIA, Returns, and Volatility. Subfigures A and B focus on positive and negative news, respectively, while Subfigure C divides all news categories into positive and negative. The results in each panel are computed in three steps. In the first one, we regress each measure of following/attention on firm effects, time effects, Google Trends changes, stock returns, and stocks' realized variance, using a standard panel regression model and keep the residuals, which we call "orthogonalized attention." In the second, we do the same for each of the regressors, and we keep the residuals, which we call collectively "orthogonalized news." In the final step, we use BRTs to estimate the relation between orthogonalized attention and orthogonalized news.



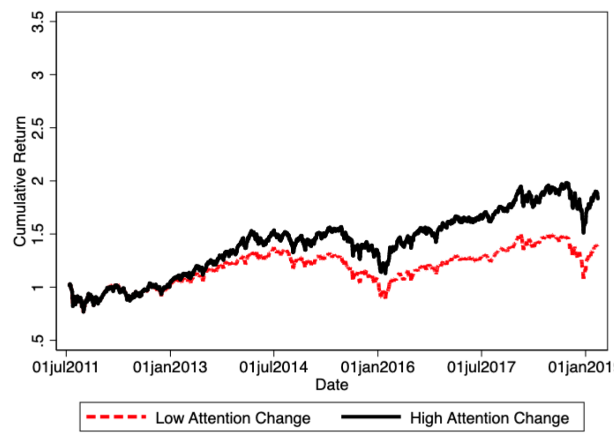

(a) All Stocks. Cumulative Returns




(b) All Stocks. Abnormal DGTW Returns

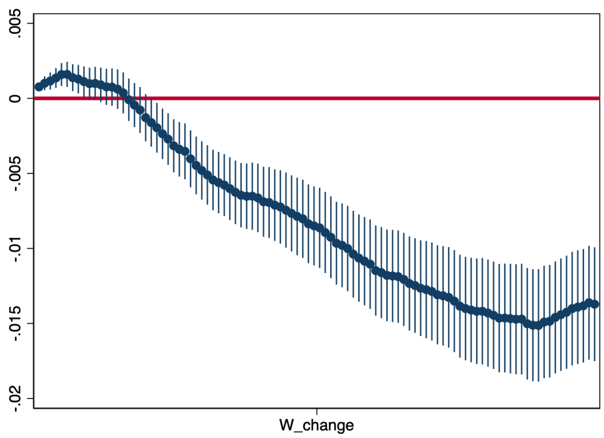



(c) Low Following Stocks. Cumulative Returns

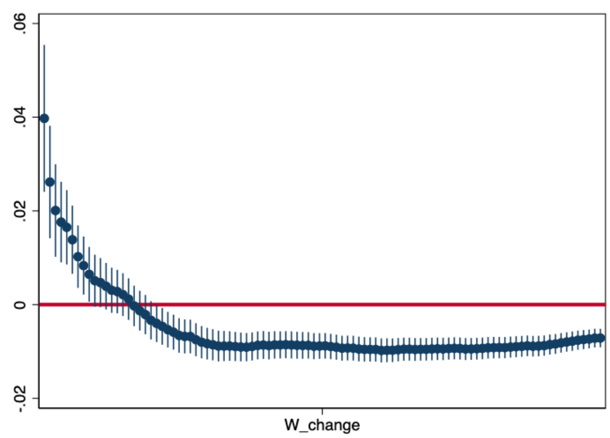



(d) High Following Stocks. Cumulative Returns

Figure 5: **Changes in Stock Following and Stock Returns.**
This figure reports results relating changes in stock following and stock performance. We use data from 2013 to 2019 and compute daily changes in stock following across all users in our dataset. Every day we divide the stocks in three quantiles on the basis of the changes in stock following. In order to make changes in stock following comparable across all stocks, we make sure to divide the changes in stock following by the level of stock following on the previous day. We then compute the next day returns across all stocks in each of the three portfolios. Finally, we cumulate the daily returns of each portfolio from the first date to the last date available. In Subfigure (a) we report the results for stocks with low and high changes in attention. In Subfigure (b) we repeat the exercise but using DGTW-adjusted returns, rather than simple stock returns. Subfigure (c) and (d) repeat the analysis in subfigure (a) but focus on stocks that have low and high overall attention, respectively.




(a) Coefficient Estimates: Cumulated Returns

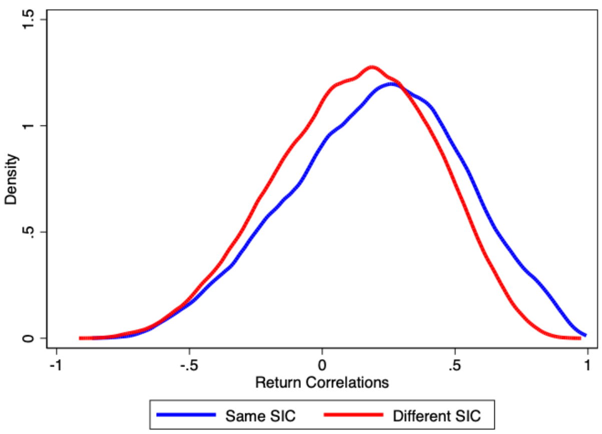



(b) Coefficient Estimates: Cumulated Annualized Returns

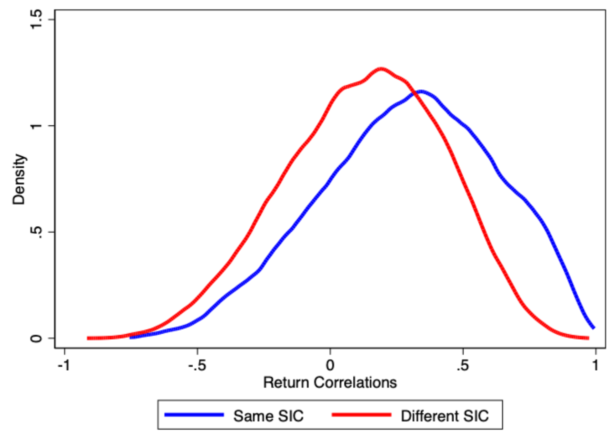
Figure 6: **Changes in Stock Following and Returns at Different Horizons.**
This figure reports the coefficient estimates from the following baseline panel regression:

$$Ret_{i,t:t+k} = \alpha_i + \alpha_t + \beta\ W\_Change_{i,t} + \epsilon_{i,t:t+k}$$

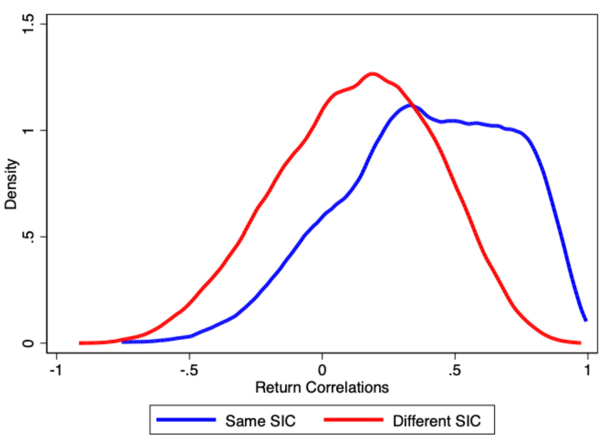
where $Ret_{i,t:t+k}$ is the return realized by stock $i$ from the end of week $t$ to the end of week $t+k$, $\alpha_i$ represent stock fixed effects and $\alpha_t$ represent time effects—which absorb stock-level differences in returns over the period as well as time variations in stock market returns. Finally, $\beta$ measures the relation between stock returns and $W\_Change_{i,t}$, the weekly percentage change in stock following associated with stock $i$ over the course of week $t$. We let $k$ range from 1 to 100, meaning that the future returns are computed over an horizon of almost two years. Subfigure (a) reports the beta coefficient estimates and associated 95% confidence intervals for the different regressions. Subfigure (b) repeats the exercise annualizing the returns at different horizons.



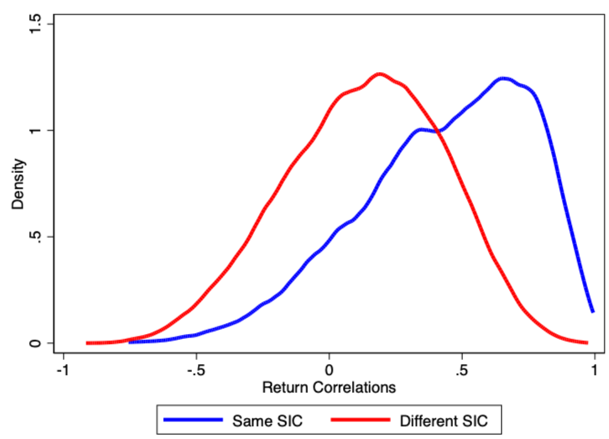

(a) 1-digit SIC codes




(b) 2-digits SIC codes




(c) 3-digits SIC codes




(d) 4-digits SIC codes

Figure 7: **Daily Stock Returns Correlations Within and Across Industries**
This plot reports the cross-sectional distribution of pairwise stock return correlations for companies that are in the same or different industries. We focus on the stocks in the S&P 500 index and compute the correlation in daily returns across all stock pairs. In Subfigure (a), we compare the correlations of stocks that are in the same industries in blue and different industries in red, where industries are categorized using 1-digit SIC codes. Subfigure (b) through (d) repeat the exercise with 2-digit, 3-digit, and 4-digit SIC codes, respectively.

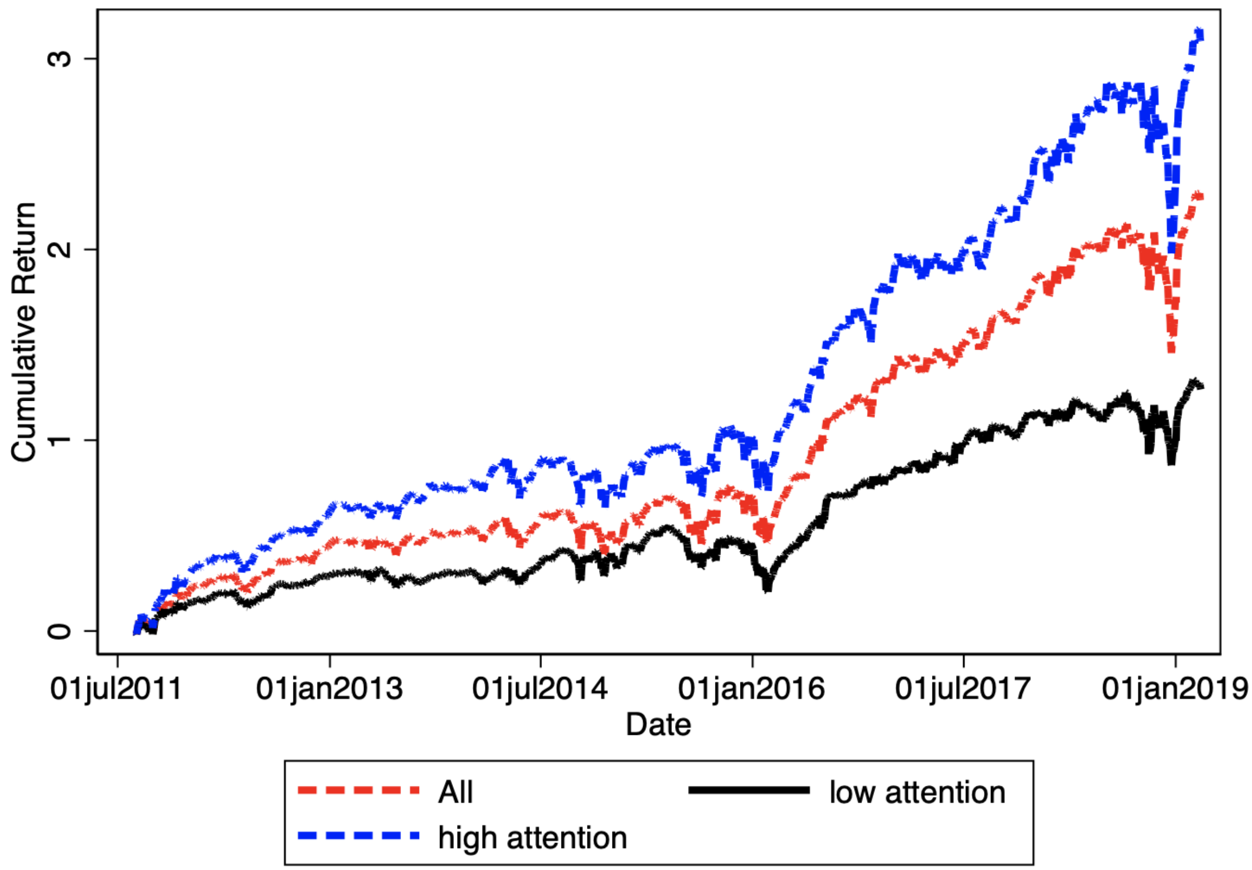
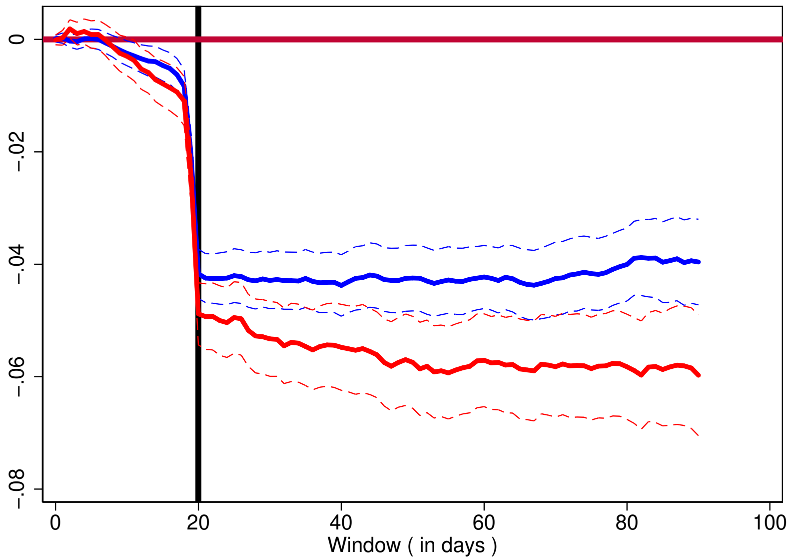
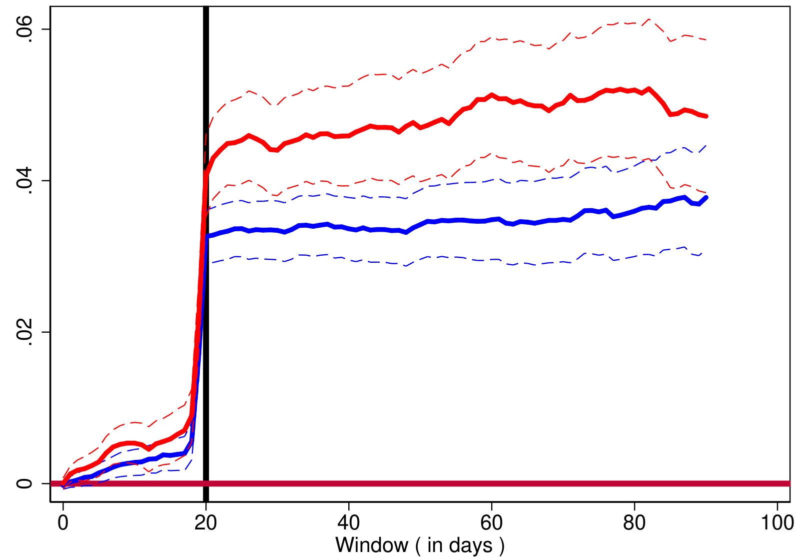



Figure 8: **Short Term Reversal and Stock Following**
This figure shows the performance of short-term reversal, conditional on investor stock following. We take the universe of stocks in our data and compute, at the daily frequency, the cumulated returns for each stock over the previous 22 days. We then double sort stocks into quantiles on the basis of their past returns and level of following, controlling for the firms' market capitalization to avoid confounding stock following for company size. Third, for each attention level, we construct long-short equally-weighted reversal portfolios that go long in the stocks with the lowest returns and short in the stocks with the highest returns. Finally, we cumulate the returns of these long-short portfolios over time.

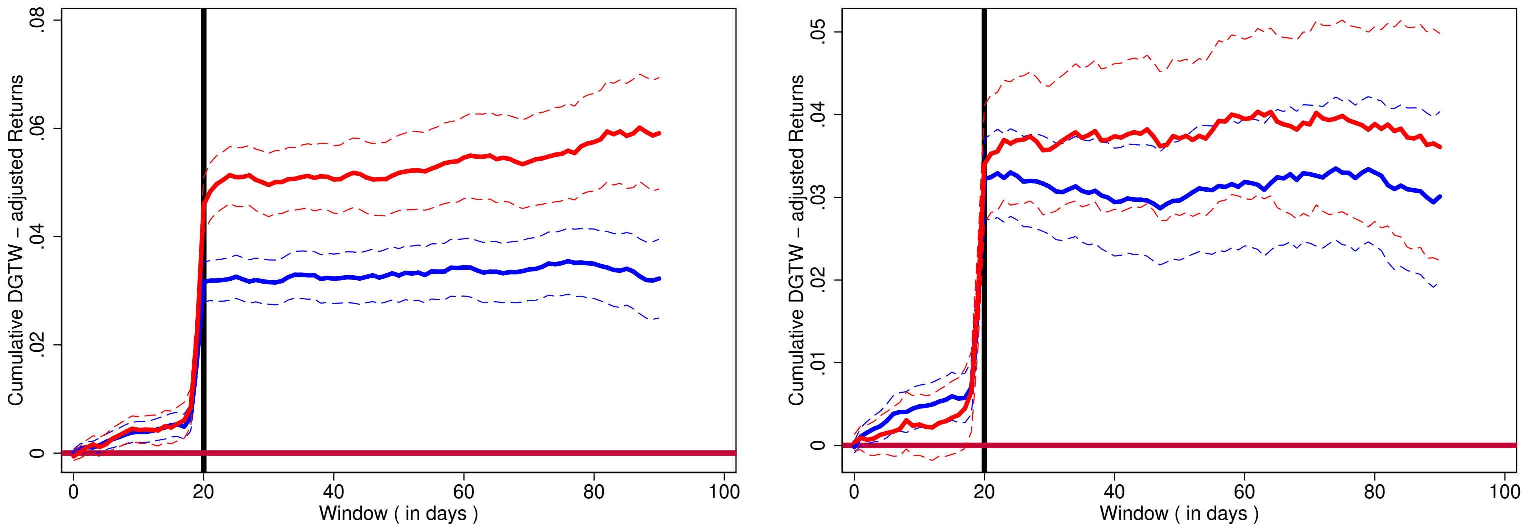



(a) First Quantile of Earnings News




(b) Fifth Quantile of Earnings News

Figure 9: **Stock Following and Earnings Announcement Returns.**
This figure reports results on the cumulated abnormal returns around earnings announcement news. We consider all earnings announcements and compute Standardized Unexpected Earnings (SUE) by subtracting mean earnings analysts' forecasts from the actual earnings announced, and dividing by the standard deviation of the earnings forecasts. We then construct SUE quintiles across earnings announcement dates and companies, where the first quintile contains the earnings announcements with the most negative earnings surprises and the fifth quintile, earnings announcements with the most positive earnings surprises. We also separate the stocks into two groups based of their size-adjusted stock following, defined as the log of the ratio between stock following and market capitalization, with the least followed stocks in the first group and the most followed stocks in the second group. Finally, we track the cumulative DGTW-adjusted daily returns of the two groups of stocks around earnings announcements with Standardized Unexpected Earnings in the lowest and highest quintile. The results in Subfigure (a), are associated with negative earnings surprises and results in Subfigure (b), are associated with positive earnings surprises. The red lines show results for high size-adjusted stock following, whereas blue lines show results for low size-adjusted stock following.




(a) Non-Professional—Fifth Quantile of Earnings News

(b) Professional—Fifth Quantile of Earnings News

Figure 10: **Stock Following by Different Investor Types and Earnings Announcement Returns.** This figure reports results on the cumulated abnormal returns around earnings announcement news. We consider all earnings announcements and compute Standardized Unexpected Earnings (SUE) by subtracting mean earnings analysts' forecasts from the actual earnings announced, and dividing by the standard deviation of the earnings forecasts. We then construct SUE quintiles across earnings announcement dates and companies, where the first quintile contains the earnings announcements with the most negative earnings surprises and the fifth quintile, earnings announcements with the most positive earnings surprises. We also separate the stocks into two groups based of their size-adjusted stock following, defined as the log of the ratio between stock following and market capitalization, with the least followed stocks in the first group and the most followed stocks in the second group. Finally, we track the cumulative DGTW-adjusted daily returns of the two groups of stocks around earnings announcements with Standardized Unexpected Earnings in the lowest and highest quintile. The results in Subfigure (a) compute stock following among non-professional investors—a category that includes students, retirees, journalists, executives, and academics—and positive earnings surprises. The results in Subfigure (b) compute stock following among professional investors—a category that includes full-time investors and professional investors—and positive earnings surprises. The red lines show results for high size-adjusted stock following, whereas blue lines show results for low size-adjusted stock following.

## Table 1. Account types on Seeking Alpha

### Panel A. All Users

| Year | Occasional Investor | Retiree | Full Time Investor | Finance Professional | Student | Academic | Executive | Journalist/ Other | Undeclared | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 2011 | 95,787 | 40,347 | 55,446 | 49,225 | 25,655 | 3,726 | 15,218 | 43,053 | 103,785 | 432,242 |
| 2012 | 105,534 | 42,995 | 37,834 | 32,266 | 20,843 | 3,122 | 15,290 | 7,394 | 145,014 | 410,292 |
| 2013 | 322,003 | 74,486 | 92,575 | 130,618 | 107,334 | 13,582 | 50,299 | 27,458 | 370,690 | 1,189,045 |
| 2014 | 272,400 | 55,457 | 48,060 | 111,404 | 80,370 | 5,534 | 34,902 | 99,449 | 520,498 | 1,228,074 |
| 2015 | 132,178 | 25,521 | 19,216 | 63,373 | 46,378 | 2,304 | 18,045 | 64,727 | 362,675 | 734,417 |
| 2016 | 126,354 | 23,082 | 16,460 | 58,658 | 39,726 | 2,117 | 18,859 | 65,414 | 391,580 | 742,250 |
| 2017 | 124,728 | 21,483 | 16,539 | 48,253 | 34,469 | 1,594 | 14,243 | 54,748 | 406,251 | 722,308 |
| 2018 | 84,792 | 15,402 | 12,599 | 32,671 | 24,926 | 1,053 | 8,645 | 33,662 | 396,360 | 610,110 |
| Total | 1,263,776 | 298,773 | 298,729 | 526,468 | 379,701 | 33,032 | 175,501 | 395,905 | 2,696,853 | 6,068,738 |

### Panel B. Active Users

| Year | Occasional Investor | Retiree | Full Time Investor | Finance Professional | Student | Academic | Executive | Journalist/ Other | Undeclared | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 2011 | 34,184 | 17,531 | 20,615 | 14,232 | 4,963 | 917 | 4,591 | 11,834 | 30,910 | 139,777 |
| 2012 | 49,778 | 22,081 | 19,777 | 15,347 | 7,709 | 1,268 | 6,747 | 3,323 | 57,392 | 183,422 |
| 2013 | 144,354 | 38,906 | 45,662 | 51,106 | 32,332 | 4,752 | 20,321 | 9,695 | 151,602 | 498,730 |
| 2014 | 96,859 | 22,752 | 20,637 | 29,360 | 16,586 | 1,362 | 10,282 | 23,521 | 156,610 | 377,969 |
| 2015 | 45,165 | 9,189 | 7,977 | 16,319 | 9,519 | 482 | 5,028 | 15,020 | 110,220 | 218,919 |
| 2016 | 40,842 | 7,800 | 6,401 | 13,759 | 7,562 | 377 | 4,634 | 14,151 | 114,286 | 209,812 |
| 2017 | 36,087 | 6,712 | 5,740 | 9,108 | 5,137 | 222 | 2,886 | 9,703 | 108,450 | 184,045 |
| 2018 | 17,070 | 3,580 | 3,094 | 3,751 | 2,022 | 84 | 1,067 | 3,971 | 84,728 | 119,367 |
| Total | 464,339 | 128,551 | 129,903 | 152,982 | 85,830 | 9,464 | 55,556 | 91,218 | 814,198 | 1,932,041 |

This table reports in Panel A the number of users in our dataset that created watchlists over the period July 2011—December 2018. Panel B repeats the exercise for active users only, defined as those individuals who change their watchlist at least once after 30 days from initiation.

## Table 2. Number of Tickers in Watchlist

**Panel A. Number of Stocks Followed by Years of Attention**

| | # Obs | Mean | Std. Dev. | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| # At initiation | 1,932,041 | 5.9 | 5.0 | 4 | 5 | 6 |
| 1 year | 1,932,041 | 10.3 | 11.0 | 5 | 7 | 12 |
| 2 years | 1,279,949 | 12.8 | 15.0 | 5 | 8 | 14 |
| 3 years | 920,536 | 14.8 | 17.8 | 6 | 9 | 17 |
| 4 years | 647,341 | 16.9 | 20.71 | 6 | 10 | 19 |
| 5 years | 411,733 | 19.7 | 24.4 | 7 | 12 | 22 |
| 6 years | 213,162 | 23.7 | 28.8 | 8 | 15 | 27 |
| 7 years | 81,158 | 27.9 | 32.8 | 10 | 18 | 32 |
| 8 years | 23,339 | 34.7 | 40.1 | 13 | 22 | 40 |

**Panel B. Number of Stocks Added to Watchlists by Years of Attention**

| | # Obs | Mean | Std. Dev. | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| 1 year | 1,932,041 | 5.1 | 10.0 | 0 | 1 | 5 |
| 2 years | 1,279,949 | 2.9 | 6.8 | 0 | 1 | 3 |
| 3 years | 920,536 | 2.7 | 6.6 | 0 | 1 | 2 |
| 4 years | 647,341 | 2.8 | 7.0 | 0 | 1 | 2 |
| 5 years | 411,733 | 3.3 | 7.9 | 0 | 1 | 2 |
| 6 years | 213,162 | 3.7 | 8.3 | 1 | 1 | 3 |
| 7 years | 81,158 | 3.9 | 8.8 | 1 | 1 | 3 |
| 8 years | 23,339 | 4.4 | 8.7 | 1 | 1 | 4 |

**Panel C. Number of Stocks Removed from Watchlists by Years of Attention**

| | # Obs | Mean | Std. Dev. | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| 1 year | 559,138 | 2.2 | 4.9 | 0 | 0 | 2 |
| 2 years | 229,147 | 1.8 | 4.4 | 0 | 0 | 1 |
| 3 years | 76,411 | 2.1 | 4.9 | 0 | 0 | 2 |
| 4 years | 8,627 | 1.6 | 4.2 | 0 | 0 | 1 |

This table reports summary statistics on the watchlist activity over time of the active users, defined as those individuals who change their watchlist at least once after 30 days from initiating it. Panel A reports the average number of securities in a watchlist as a function of the number of years the watchlist is active. Panel B (Panel C) reports statistics of stock additions (deletions) to (from) watchlists.

## Table 3. Stock Following by Investor Category

| | **Panel A. Initial Stocks** | | | | | **Panel B. Stocks Additions Per Year** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | 25% | Median | 75% | Mean | Std. Dev. | 25% | Median | 75% |
| Occasional Investor | 6.1 | 4.7 | 4 | 5 | 7 | 4.0 | 7.0 | .67 | 1.7 | 4.4 |
| Retiree | 6.5 | 5.5 | 3 | 5 | 8 | 3.9 | 6.6 | .67 | 1.8 | 4.4 |
| Full Time Investor | 6.9 | 6.2 | 4 | 5 | 8 | 4.5 | 7.6 | .8 | 2 | 5 |
| Professional Finance | 6.7 | 5.9 | 5 | 5 | 7 | 3.5 | 6.2 | .5 | 1.3 | 4 |
| Student | 5.7 | 4.2 | 5 | 5 | 6 | 3.0 | 5.4 | .5 | 1 | 3.1 |
| Academic | 5.7 | 4.3 | 4 | 5 | 6 | 2.9 | 5.2 | .4 | 1 | 3 |
| Executive | 5.9 | 4.3 | 5 | 5 | 6 | 2.3 | 4.3 | .5 | 1 | 2.5 |
| Journalist/Other | 6.4 | 5.5 | 5 | 5 | 6 | 3.3 | 6.4 | .5 | 1 | 3 |
| Undeclared | 5.4 | 4.7 | 3 | 5 | 6 | 5.3 | 8.7 | 1 | 2 | 6 |

| | **Panel C. Stock Deletions Per Year** | | | | | **Panel D. Stock Changes Per Year** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | 25% | Median | 75% | Mean | Std. Dev. | 25% | Median | 75% |
| Occasional Investor | 2.3 | 4.8 | 0 | 0 | 2 | 5.0 | 9.3 | .75 | 2 | 5 |
| Retiree | 2.4 | 5.1 | 0 | 0 | 2.3 | 4.8 | 8.9 | .8 | 2 | 5 |
| Full Time Investor | 2.8 | 5.7 | 0 | 0 | 3 | 5.5 | 10.0 | 1 | 2.1 | 5.7 |
| Professional Finance | 1.3 | 3.5 | 0 | 0 | 1 | 4.1 | 7.8 | .5 | 1.5 | 4.4 |
| Student | 1.7 | 3.7 | 0 | 0 | 2 | 3.5 | 6.8 | .5 | 1 | 4 |
| Academic | 1.4 | 3.6 | 0 | 0 | 1 | 3.3 | 6.5 | .5 | 1 | 3.6 |
| Executive | 0.9 | 2.8 | 0 | 0 | .5 | 2.7 | 5.5 | .5 | 1 | 3 |
| Journalist/Other | 1.3 | 3.7 | 0 | 0 | 1 | 4.0 | 8.4 | .5 | 1 | 4 |
| Undeclared | 2.3 | 4.8 | 0 | 0 | 2 | 6.6 | 11.0 | 1 | 2.7 | 7 |

This table reports the number of stocks followed by investors on *Seeking Alpha*. In Panel A, we report the cross-sectional averages, standard deviations, as well as the 25th, 50th, and 75th percentiles of the initial number of stocks in each investor's watchlist, computed for each investor category. Panel B, C, and D report the same quantities for the number of stocks added every year, the number of stocks deleted every year, and the number of stocks changed in the watchlist every year. All quantities are computed using only active users, defined as those individuals who change their watchlist at least once after 30 days from initiating it.

**Table 4. Positive, Negative and Neutral News**

| | All | Positive | Neutral | Negative |
|---|---|---|---|---|
| Trading | 0.518 | 0.487 | 0.000 | 0.031 |
| Stock_Price | 1.094 | 0.621 | 0.000 | 0.473 |
| Stock_Restructure | 0.021 | 0.015 | 0.000 | 0.006 |
| Indexes | 0.034 | 0.033 | 0.000 | 0.001 |
| Trans_firm_stocks | 0.749 | 0.442 | 0.006 | 0.301 |
| Legal | 0.255 | 0.053 | 0.000 | 0.201 |
| Guidance | 1.630 | 0.499 | 0.989 | 0.142 |
| M_A | 0.653 | 0.092 | 0.000 | 0.561 |
| Analysts_Actions | 3.752 | 2.331 | 0.060 | 1.362 |
| Firm_Event | 4.144 | 0.048 | 4.096 | 0.000 |
| Earnings_Release | 5.256 | 2.644 | 1.345 | 1.267 |
| Products_and_Services | 1.326 | 1.260 | 0.004 | 0.062 |
| Credit_Rating | 0.871 | 0.457 | 0.076 | 0.338 |
| Partnership | 0.532 | 0.532 | 0.000 | 0.000 |
| Insider_Trading | 7.172 | 2.791 | 0.000 | 4.381 |
| Dividends | 1.411 | 0.267 | 1.131 | 0.013 |
| Executive_Turnover | 1.981 | 1.642 | 0.002 | 0.336 |
| Exec_Comp | 0.064 | 0.055 | 0.000 | 0.008 |
| Accidents_Wars_Crime | 0.011 | 0.001 | 0.000 | 0.010 |
| Labor_Issues | 0.072 | 0.025 | 0.000 | 0.046 |
| Major_S_holders_Disc | 3.487 | 0.000 | 3.487 | 0.000 |
| Technical_Analysis | 13.829 | 6.301 | 2.995 | 4.532 |
| Other | 8.399 | 1.960 | 4.554 | 1.884 |

This table reports the average number of news weeks per firm every year, categorized by the type of news. In Column 1, we report the total number of news. In columns 2, 3, and 4 we break down news in positive, negative and neutral.

**Table 5. News, Stock Following and Instantaneous Attention**

| | Additions_ex | Additions_new | Deletions | Net_Additions | Google Trend | Bloomberg | Returns | Volatility |
|---|---|---|---|---|---|---|---|---|
| Trading | 0.65*** | 0.26*** | 0.33*** | 0.60*** | 0.57*** | 0.09*** | 2.46*** | 8.14*** |
| | (19.93) | (11.29) | (14.50) | (18.65) | (9.50) | (15.23) | (5.36) | (17.79) |
| Stock_Price | 0.51*** | 0.16*** | 0.19*** | 0.49*** | 0.43*** | 0.16*** | 0.58*** | 4.29*** |
| | (21.83) | (11.34) | (15.34) | (21.79) | (13.34) | (20.17) | (2.77) | (18.85) |
| Stock_Restructure | 0.31*** | 0.15** | 0.26*** | 0.24*** | 0.54*** | 0.04** | -5.67*** | 4.17*** |
| | (5.17) | (2.43) | (3.97) | (4.06) | (3.75) | (2.19) | (-3.70) | (5.33) |
| Indexes | 0.35*** | 0.15** | 0.18*** | 0.34*** | 0.18** | 0.12*** | 1.31** | 0.89*** |
| | (5.46) | (2.52) | (3.23) | (4.84) | (2.26) | (5.36) | (2.31) | (5.29) |
| Trans_firm_stocks | 0.35*** | 0.14*** | 0.13*** | 0.33*** | 0.17*** | 0.08*** | -0.75*** | 1.63*** |
| | (19.61) | (10.43) | (10.35) | (18.87) | (7.07) | (15.98) | (-4.97) | (13.80) |
| Legal | 0.22*** | 0.13*** | 0.23*** | 0.17*** | 0.23*** | 0.04*** | 0.02 | 1.40*** |
| | (8.63) | (5.42) | (10.59) | (6.79) | (4.76) | (5.15) | (0.07) | (4.97) |
| Guidance | 0.17*** | 0.05*** | 0.05*** | 0.16*** | 0.01 | 0.14*** | 0.01 | 1.03*** |
| | (10.95) | (5.36) | (5.61) | (10.86) | (0.61) | (22.37) | (0.05) | (12.11) |
| M_A | 0.21*** | 0.05*** | 0.03*** | 0.21*** | 0.05*** | 0.10*** | 0.74*** | 0.43*** |
| | (14.92) | (4.60) | (3.22) | (14.81) | (3.01) | (17.45) | (5.28) | (4.27) |
| Analysts_Actions | 0.18*** | 0.04*** | 0.05*** | 0.17*** | 0.07*** | 0.09*** | -0.01 | 0.85*** |
| | (23.81) | (6.53) | (11.43) | (22.71) | (7.31) | (24.59) | (-0.22) | (19.10) |
| Firm_Event | 0.14*** | 0.10*** | 0.07*** | 0.14*** | 0.03*** | 0.08*** | 0.02 | 0.96*** |
| | (15.18) | (9.80) | (10.94) | (15.22) | (4.58) | (18.23) | (0.28) | (20.25) |
| Earnings_Release | 0.19*** | 0.08*** | 0.08*** | 0.17*** | 0.01** | 0.08*** | 0.04 | 1.03*** |
| | (20.19) | (8.54) | (11.72) | (19.17) | (2.03) | (20.12) | (0.58) | (17.74) |
| Products_and_Services | 0.09*** | 0.04*** | 0.01* | 0.09** | 0.05*** | 0.05*** | 0.90*** | 0.64*** |
| | (10.10) | (5.34) | (1.82) | (10.15) | (4.09) | (13.00) | (10.21) | (7.56) |
| Credit_Rating | 0.07*** | 0.04*** | 0.05*** | 0.06*** | 0.04*** | 0.08*** | -0.14 | 0.29*** |
| | (7.95) | (5.06) | (6.55) | (6.93) | (3.63) | (15.93) | (-1.46) | (4.92) |
| Partnership | 0.07*** | 0.01 | 0.01 | 0.07*** | 0.06*** | 0.04*** | 0.48*** | 0.27*** |
| | (6.14) | (1.48) | (0.94) | (5.87) | (2.75) | (8.04) | (3.83) | (2.92) |
| Insider_Trading | 0.05*** | 0.03*** | 0.02*** | 0.05*** | 0.05*** | -0.01*** | 0.38*** | 0.22*** |
| | (11.83) | (5.83) | (4.60) | (11.38) | (6.29) | (-6.13) | (10.89) | (8.83) |
| Dividends | 0.06*** | 0.02*** | 0.02*** | 0.06*** | -0.02*** | 0.01** | 0.16*** | -0.19*** |
| | (6.75) | (3.52) | (3.30) | (6.67) | (-4.09) | (2.51) | (3.13) | (-5.26) |
| Executive_Turnover | 0.02*** | 0.02*** | 0.01** | 0.02*** | 0.01* | 0.03*** | -0.09 | 0.27*** |
| | (4.40) | (2.87) | (2.10) | (4.02) | (1.86) | (13.47) | (-1.36) | (5.52) |
| Exec_Comp | 0.03 | 0.00 | 0.01 | 0.03 | 0.05 | 0.05*** | 0.48 | 0.39* |
| | (1.50) | (0.32) | (0.54) | (1.58) | (1.46) | (3.44) | (1.58) | (1.77) |
| Accidents_Wars_Crime | 0.09 | 0.11 | 0.04 | 0.10 | 0.11 | 0.14*** | -0.49 | -0.10 |
| | (0.85) | (1.33) | (0.63) | (0.97) | (1.36) | (3.03) | (-1.22) | (-0.44) |
| Labor_Issues | -0.02 | 0.04* | 0.03 | -0.02 | 0.03 | 0.03** | -0.99*** | 0.03 |
| | (-0.77) | (1.70) | (1.63) | (-0.67) | (1.05) | (2.12) | (-3.33) | (0.18) |
| Major_S_holders_Disc | -0.00 | 0.00 | 0.00 | -0.00 | 0.01** | 0.01** | -0.02 | 0.05* |
| | (-0.48) | (0.24) | (0.10) | (-0.28) | (2.30) | (2.09) | (-0.41) | (1.87) |
| Technical_Analysis | -0.11*** | -0.03** | -0.05*** | -0.10*** | 0.01 | 0.00 | -0.02 | 0.02 |
| | (-8.74) | (-2.30) | (-4.50) | (-8.16) | (0.77) | (0.63) | (-0.26) | (0.35) |
| Other | 0.03*** | 0.01*** | 0.03*** | 0.02*** | 0.04*** | 0.01*** | 0.06 | 0.13*** |
| | (7.82) | (2.93) | (8.11) | (6.06) | (6.03) | (4.14) | (1.65) | (5.43) |
| Constant | -0.07*** | -0.03*** | -0.03*** | -0.06*** | 0.07*** | 0.09*** | -0.00 | 4.45*** |
| | (-16.75) | (-7.77) | (-7.77) | (-15.96) | (12.22) | (52.90) | (-0.08) | (183.62) |
| R-Square | 0.29 | 0.17 | 0.21 | 0.23 | 0.10 | 0.33 | 0.10 | 0.31 |
| Obs | 687,060 | 687,060 | 687,060 | 687,060 | 580,983 | 653,749 | 633,944 | 633,944 |

This table reports results on the relation between news and individuals' decision to pay attention to individual stocks. We estimate the following baseline specification:

$$ATTENTION_{i,t} = \alpha_i + \alpha_t + \beta_1 \cdot NEWS_{i,t} + \epsilon_{i,t}$$

where $ATTENTION_{i,t}$ is proxied by seven different attention measures as described in Section 5. The firm news variable "NEWS" is measured as an indicator variable that takes the value of 1 if the RavenPack database reports a news related event associated with the firm on week $t$. The types of news we use are described in Section 5. The coefficients $\alpha_i$ and $\alpha_t$ denote stock and time fixed effects. Standard errors are double-clustered at the stock and time levels.

**Table 6. Watchlist Changes and Positive, Negative and Neutral News— with Controls**

| | All News | Negative | Positive | Neutral |
|---|---|---|---|---|
| Trading | 0.20*** | -0.27*** | 0.26*** | 0.00 |
| | (8.26) | (-4.87) | (9.65) | (.) |
| Stock_Price | 0.26*** | 0.33*** | 0.23*** | 0.00 |
| | (13.22) | (14.59) | (9.84) | (.) |
| Stock_Restructure | 0.04 | 0.34*** | -0.08 | -0.68** |
| | (0.64) | (3.03) | (-1.13) | (-2.54) |
| Indexes | 0.26*** | 0.03 | 0.26*** | 0.08 |
| | (3.95) | (0.13) | (3.87) | (0.24) |
| Trans_firm_stocks | 0.22*** | 0.36*** | 0.13*** | 0.50* |
| | (14.52) | (13.20) | (7.75) | (1.74) |
| Legal | 0.07*** | 0.11*** | -0.00 | 0.18 |
| | (3.30) | (3.90) | (-0.25) | (0.86) |
| Guidance | 0.14*** | 0.23*** | 0.07*** | 0.22*** |
| | (9.42) | (8.39) | (3.87) | (11.59) |
| M_A | 0.17*** | 0.15*** | 0.43*** | 0.28 |
| | (12.97) | (12.44) | (8.85) | (0.72) |
| Analysts_Actions | 0.11*** | 0.13*** | 0.12*** | -0.03 |
| | (18.79) | (15.84) | (15.72) | (-1.24) |
| Firm_Event | 0.09*** | 0.00 | -0.02 | 0.19*** |
| | (9.17) | (0.00) | (-0.53) | (14.38) |
| Earnings_Release | 0.12*** | 0.16*** | 0.28*** | -0.00 |
| | (12.45) | (6.67) | (16.93) | (-0.19) |
| Products_and_Services | 0.05*** | 0.04 | 0.06*** | 0.23 |
| | (6.30) | (1.35) | (7.20) | (1.64) |
| Credit_Rating | 0.04*** | 0.09*** | 0.02* | 0.04* |
| | (4.99) | (5.68) | (1.94) | (1.84) |
| Partnership | 0.04*** | 0.00 | 0.04*** | 0.00 |
| | (3.44) | (0.00) | (3.88) | (0.00) |
| Insider_Trading | 0.04*** | 0.01** | 0.07*** | 0.00 |
| | (9.17) | (2.10) | (9.22) | (0.00) |
| Dividends | 0.06*** | 0.05 | 0.09*** | 0.08*** |
| | (7.52) | (0.41) | (5.12) | (8.50) |
| Executive_Turnover | 0.00 | 0.02 | 0.01 | -0.16 |
| | (0.67) | (1.24) | (1.33) | (-1.59) |
| Exec_Comp | 0.01 | 0.04 | -0.00 | -0.11*** |
| | (0.50) | (0.76) | (-0.02) | (-6.78) |
| Accidents_Wars_Crime | 0.10 | 0.11 | -0.05 | 0.00 |
| | (1.00) | (1.02) | (-0.76) | (.) |
| Labor_Issues | -0.02 | 0.02 | -0.03 | 0.00 |
| | (-1.28) | (0.77) | (-1.03) | (.) |
| Major_S_holders_Disc | -0.01 | 0.00 | 0.00 | -0.01 |
| | (-0.77) | (0.00) | (0.00) | (-0.72) |
| Technical_Analysis | -0.10*** | -0.02*** | -0.04*** | -0.06*** |
| | (-9.12) | (-2.77) | (-6.41) | (-11.02) |
| Other | 0.01** | -0.01 | 0.05*** | 0.01** |
| | (2.45) | (-0.88) | (6.87) | (2.26) |
| Change in Google Trend | 0.09*** | 0.09*** | 0.09*** | 0.09*** |
| | (12.17) | (11.93) | (12.08) | (12.03) |
| Bloomberg AIA | 0.05*** | 0.11*** | 0.07*** | 0.10*** |
| | (7.91) | (14.46) | (11.12) | (13.94) |
| W_ret | 0.00*** | 0.01*** | 0.00*** | 0.00*** |
| | (5.38) | (5.71) | (3.80) | (4.54) |
| W_real_var | 0.04*** | 0.04*** | 0.04*** | 0.04*** |
| | (17.86) | (18.35) | (18.41) | (18.89) |
| Constant | -0.24*** | -0.25*** | -0.26*** | -0.26*** |
| | (-21.96) | (-22.61) | (-24.41) | (-23.56) |
| R-Square | 0.30 | 0.29 | 0.29 | 0.29 |
| Obs | 563,919 | 563,919 | 563,919 | 563,919 |

This table reports results on the relation between news and investors' decision to pay attention to specific stocks. The specifications are similar to the ones described in the caption to Table 5, but features the following differences. First, we use watchlist changes (addition minus deletions from existing watchlists) as a measure of attention. Second, we partition news on the basis of whether they convey positive, negative, or neutral information about the affected firms. Finally, we control for other measures of attention such as changes in Google Trends, Bloomberg' AIA, stock returns, and stock volatility.

## Table 7. Initial Attention and Stock Characteristics

| | Spec 1 | Spec 2 | Spec 3 | Spec 4 |
|---|---|---|---|---|
| Past Returns | 0.085*** | 0.079*** | 0.089*** | 0.101*** |
| | (4.54) | (4.43) | (5.19) | (5.43) |
| Past Variance | 0.218*** | 0.283*** | 0.150*** | 0.065** |
| | (9.01) | (10.26) | (5.73) | (2.08) |
| Past Skewness | -0.005* | -0.003 | 0.005* | 0.008** |
| | (-1.78) | (-1.01) | (1.70) | (2.52) |
| Past Kurtosis | 0.000 | -0.000 | -0.001 | 0.000 |
| | (0.13) | (-0.77) | (-1.65) | (0.27) |
| Market-to-Book Ratio | | 0.015*** | 0.017*** | 0.017*** |
| | | (4.45) | (4.79) | (4.47) |
| R&D | | -0.623** | -0.472* | -0.323 |
| | | (-2.13) | (-1.69) | (-1.12) |
| Profitability | | 0.664*** | 0.341*** | 0.445*** |
| | | (6.94) | (3.98) | (4.27) |
| Book Leverage | | 0.036* | -0.024 | -0.034 |
| | | (1.67) | (-1.05) | (-1.43) |
| Assets | | | 0.020*** | 0.024*** |
| | | | (4.58) | (4.95) |
| Analyst Coverage | | | -0.002** | -0.001 |
| | | | (-2.39) | (-1.47) |
| Turnover | | | 0.013*** | 0.013*** |
| | | | (15.07) | (12.59) |
| Volume | | | 0.000** | 0.000 |
| | | | (2.04) | (0.79) |
| Age | | | | -0.000 |
| | | | | (-0.56) |
| Frac. Institutional | | | | -0.060*** |
| | | | | (-2.69) |
| Constant | -0.107*** | -0.146*** | -0.359*** | -0.338*** |
| | (-6.28) | (-7.35) | (-12.17) | (-10.77) |

This table reports results of pooled regressions relating individuals' decision to include stocks in their initial watchlist and non-news related stock characteristics. The baseline specification we estimate is:

$$Initial\_Attention_{i,t} = \alpha + x_{i,t}'\beta + \epsilon_{i,t},$$

where $Initial\_Attention_{i,t}$ is the logged number of times a stock is included in a newly created watchlist in a given week divided by the overall lagged following of the stock. The vector of regressors $x_{i,t}$ includes fundamental covariates, such as size, profitability, leverage, age, and R&D spending; stock price information, such as volatility and higher moments of stock returns; and trading related covariates, such as trading volume, turnover and the ratio of a firm's stock price to its book value of equity. Finally, $x_{i,t}$ includes analysts' coverage and institutional ownership. Standard errors are double-clustered at the stock and time levels.

**Table 8. Commonality in Attention and Returns Correlations**

| | All | Same 1-digit SIC | Different 1-digit SIC | Same 2-digit SIC | Different 2-digit SIC | Same 3-digit SIC | Different 3-digit SIC | Same 4-digit SIC | Different 4-digit SIC |
|---|---|---|---|---|---|---|---|---|---|
| Attention | 0.225*** | 0.395*** | 0.148*** | 0.546*** | 0.164*** | 0.465*** | 0.171*** | 0.444*** | 0.187*** |
| | (27.91) | (20.62) | (16.61) | (17.13) | (19.61) | (11.96) | (20.78) | (9.28) | (22.80) |
| Constant | 0.123*** | 0.177*** | 0.116*** | 0.246*** | 0.120*** | 0.343*** | 0.121*** | 0.385*** | 0.122*** |
| | (163.90) | (81.61) | (145.73) | (58.09) | (158.00) | (55.00) | (161.09) | (45.02) | (162.68) |
| R-Square | 0.003 | 0.012 | 0.001 | 0.030 | 0.002 | 0.032 | 0.002 | 0.036 | 0.002 |
| N | 234,740 | 33,784 | 200,956 | 9,466 | 225,274 | 4,326 | 230,414 | 2,286 | 232,454 |

This table reports results relating commonalities in stock following and stock return co-movements. The results are computed as follows. We start by focusing on the stocks in the S&P 500 index and computing the correlation in daily returns across all stock pairs. In the second step, we compute—for each pair of stocks in the S&P 500—the degree to which they are jointly held by investors in their watchlists. We compute this measure by counting the total number of watchlists where a specific pair of stocks are held together (over the full sample) and scaling it by the total number of watchlists that include one of the two stocks in the pair. This scaling guarantees that the commonality in watchlist presence is bounded between zero and one. In the third step, we estimate cross-sectional regressions relating the degree of commonality in attention across stocks to their correlations in returns. The results in column (1) are computed for all stock pairs. In the second and third columns, we repeat the exercise but focus on stocks that belong to the same industries (column 2) or different industries (column 3) using stocks' 1-digit SIC code. In columns (4) through (9), we re-estimate our results using 2-digit, 3-digit, and 4-digit industry classifications.

**Table 9. The Effect of Being in Investors' stock following on Return, Volatility, Volume and Google Attention**

| Dependent var at $t$ | Ret | Ret | Risk | Risk | Volume | Volume | Google | Google |
|---|---|---|---|---|---|---|---|---|
| Following× Pos-News | 0.53*** | 1.31*** | 0.86*** | 1.45*** | 0.02*** | 0.04*** | 0.06*** | 0.05*** |
| | (3.48) | (2.77) | (4.91) | (2.67) | (4.64) | (5.31) | (4.38) | (3.65) |
| Following× Neg-News | -0.13* | 0.34 | 0.16** | 0.55*** | 0.00 | 0.00 | 0.01* | 0.01 |
| | (-1.84) | (1.65) | (2.10) | (2.65) | (1.12) | (0.44) | (1.77) | (0.80) |
| Following× Neut-News | -0.05 | -0.19* | -0.11* | -0.19* | 0.00 | -0.00 | -0.01 | -0.01 |
| | (-1.32) | (-1.80) | (-1.96) | (-1.82) | (0.54) | (-0.21) | (-1.32) | (-1.44) |
| Positive-News | 0.19*** | 0.12*** | 0.28*** | 0.22*** | 0.03*** | 0.02*** | 0.02*** | 0.01*** |
| | (14.45) | (5.40) | (18.45) | (9.39) | (11.82) | (10.52) | (10.27) | (10.31) |
| Negative-News | -0.17*** | -0.19*** | 0.27*** | 0.21*** | 0.03*** | 0.02*** | 0.02*** | 0.01*** |
| | (-12.61) | (-13.99) | (24.28) | (17.15) | (15.98) | (15.05) | (13.38) | (11.54) |
| Neutral-News | -0.03*** | -0.03*** | 0.10*** | 0.12*** | 0.01*** | 0.00*** | 0.00*** | 0.00** |
| | (-3.27) | (-3.25) | (14.29) | (15.63) | (5.15) | (3.45) | (4.30) | (2.55) |
| Following | 0.37*** | 1.01*** | 0.66*** | 0.22 | 0.02 | 0.01 | 0.00 | 0.02 |
| | (3.65) | (3.09) | (6.12) | (1.03) | (1.53) | (0.40) | (0.12) | (0.33) |
| Constant | 0.09*** | 3.68 | 4.50*** | 10.52*** | 14.16*** | 6.04*** | 0.09*** | -3.48*** |
| | (5.55) | (1.52) | (221.98) | (5.62) | (3656.99) | (11.13) | (40.34) | (-4.52) |
| Other Controls | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ |
| R-Square | 0.11 | 0.17 | 0.31 | 0.37 | 0.86 | 0.92 | 0.11 | 0.12 |
| Obs | 624,297 | 303,766 | 624,297 | 303,766 | 656,899 | 303,766 | 555,704 | 276,721 |

This table reports results on the real effects of stock following. We estimate panel regressions of the following form at the weekly frequency:

$$Outcome_{i,t} = \alpha_i + \alpha_t + \sum_{l\in\{G,B,N\}} \beta_l \times News_{i,l,t} \times Following_{i,t-1}$$
$$+ \sum_{l\in\{G,B,N\}} \gamma_l \times News_{i,l,t} + \delta \times Following_{i,t-1} + \eta \times X_{i,t-1} + u_{i,t}$$

where $Outcome_{i,t}$—the dependent variable of interest—is, alternatively, stock returns, stock price risk (volatility), trading volume or stock searches measured via Google trends; $News_{i,l,t}$ is the number of news of type $l$ regarding stock $i$ that occurs in week $t$. News are categorized into three groups: Good (G), Bad (B) and Neutral (N). Finally, $Following_{i,t-1}$ is investors' following of the stock as of week $t-1$, divided by the market capitalization of the stock, and $X_{i,t-1}$ is a vector containing the following control variables: market leverage, book leverage, firm income, Market-to-Book Ratio, asset tangibility, R&D, variation in analysts recommendations, variation in analysts EPS forecasts, the fraction of institutional investors holdings out of total shares outstanding, institutional investors breadth, the Herfindahl index of institutional investors, log price, log number of analysts covering the stock, log market capitalization, past risk-adjusted returns, past skewness and past kurtosis. In columns 1 and 2 we report results for the effect of stock following around news events on stock returns. We repeat the analysis for stock returns volatility in columns 3 and 4, trading volume in columns 5 and 6, and google trend searches in columns 7 and 8. Even columns add the stock-specific controls mentioned above. Standard errors are double-clustered at the stock and time levels.

# Online Appendix

(Not for publication)

## Online Appendix A.1 A Primer on BRTs

We work with BRTs for a number of reasons. First, BRTs have exhibited strong predictive performance in various fields. For example, BRTs routinely place at the very top in many Kaggle machine learning competitions (see Machine Learning Challenge Results for examples of competition results). In financial settings, extensive horse races show tree-based methods perform as well as neural networks and outperform other linear and non-linear methods (Gu, Kelly, and Xiu, 2020; Bianchi, Buchner, and Tamoni, 2020). Second, BRTs can handle large, high-dimensional datasets because they perform both variable selection and shrinkage in an automated fashion. They are also robust to outliers and can handle missing values. Third, although most machine learning methods, including neural networks, focus only on predictive performance and are criticized as "black boxes," one advantage of BRTs is their good interpretability inherited from regression trees (Hastie, Tibshirani, and Friedman, 2009).[10] For example, we can estimate which covariates matter, among the many available, using *relative-influence measures* and obtain non-parametric estimates of the relation between funds' expected returns and each of their characteristics using *partial dependence plots*. We present below a more formal treatment of BRTs. Section Online Appendix A.1.1 describes regression trees, Section Online Appendix A.1.2 describes boosting.[11]

### Online Appendix A.1.1 Regression Trees

Suppose we have $P$ potential predictor ("state") variables and a single dependent variable over $T$ observations, i.e., $(x_t, y_{t+1})$ for $t = 1, 2, ..., T$, with $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$. Fitting a regression tree requires deciding (i) which predictor variables to use to split the sample space and (ii) which split points to use. The regression trees we use employ recursive binary partitions, so the fit of a regression

[10] See Gu, Kelly, and Xiu (2020) for some explainable machine learning techniques.

[11] Our description draws on Friedman (2001), who provides a more in-depth coverage of the approach.

tree can be written as an additive model:

$$f(x) = \sum_{j=1}^{J} c_j I\{x \in S_j\},$$

where $S_j,\ j = 1, ..., J$ are the regions we split the space spanned by the predictor variables into, $I\{\}$ is an indicator variable, and $c_j$ is the constant used to model the dependent variable in each region. If the $L^2$ norm criterion function is adopted, the optimal constant is $\hat{c}_j = mean(y_{t+1}|x_t \in S_j)$.

The globally optimal splitting point is difficult to determine, particularly in cases where the number of state variables is large. Hence, we use a sequential greedy algorithm. Using the full set of data, the algorithm considers a splitting variable $p$ and a split point $s$ so as to construct half-planes,

$$S_1(p, s) = \{X|X_p \leq s\} \qquad and \qquad S_2(p, s) = \{X|X_p > s\},$$

that minimize the sum of squared residuals:

$$\min_{p,s} \left[ \min_{c_1} \sum_{x_t \in S_1(p,s)} (y_{t+1} - c_1)^2 + \min_{c_2} \sum_{x_t \in S_2(p,s)} (y_{t+1} - c_2)^2 \right]. \tag{5}$$

For a given choice of $p$ and $s$, the fitted values, $\hat{c}_1$ and $\hat{c}_2$, are

$$\begin{aligned} \widehat{c}_1 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_1(p, s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_1(p, s)\}, \\ \widehat{c}_2 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_2(p, s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_2(p, s)\}. \end{aligned} \tag{6}$$

The best splitting pair $(p, s)$ in the first iteration can be determined by searching through each of the predictor variables, $p = 1, .., P$. Given the best partition from the first step, the data is then partitioned into two additional states and the splitting process is repeated for each of the subsequent partitions. Predictor variables that are never used to split the sample space do not influence the fit of the model, so the choice of splitting variable effectively performs variable selection.

Regression trees are generally employed in high-dimensional datasets where the relation between predictor and predicted variables is potentially non-linear. This feature is important in our context, because which variables may be more or less relevant ex-ante is unclear. Furthermore, it is difficult to know in our context whether there is a linear relation between predictor and predicted variables. On the other hand, the approach is sequential, and successive splits are performed on fewer and fewer observations, increasing the risk of fitting idiosyncratic data patterns. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

### Online Appendix A.1.2 Boosting

Boosting is based on the idea that combining a series of simple prediction models can lead to more accurate forecasts than those available from any individual model. Boosting algorithms iteratively re-weight data used in the initial fit by adding new trees in a way that increases the weight on observations modeled poorly by the existing collection of trees. From above, recall that a regression tree can be written as:

$$\mathcal{T}\left(x; \{S_j, c_j\}_{j=1}^{J}\right) = \sum_{j=1}^{J} c_j I\{x \in S_j\}. \tag{7}$$

A boosted regression tree is simply the sum of regression trees:

$$f_B(x) = \sum_{b=1}^{B} \mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^{J}\right),$$

where $\mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^{J}\right)$ is the regression tree used in the $b$-th boosting iteration and $B$ is the number of boosting iterations. Given the model fitted up to the $(b-1)-th$ boosting iteration, $f_{b-1}(x)$, the subsequent boosting iteration seeks to find parameters $\{S_{j,b}, c_{j,b}\}_{j=1}^{J}$ for the next tree to solve a problem of the form

$$\{\hat{S}_{j,b}, \hat{c}_{j,b}\}_{j=1}^{J} = \min_{\{S_{j,b}, c_{j,b}\}_{j=1}^{J}} \sum_{t=0}^{T-1} \left[y_{t+1} - \left(f_{b-1}(x_t) + \mathcal{T}_b\left(x_t; \{S_{j,b}, c_{j,b}\}_{j=1}^{J}\right)\right)\right]^2.$$

For a given set of state definitions ("splits"), $S_{j,b}$, $j = 1, .., J$, the optimal constants, $c_{j,b}$, in each state are derived iteratively from the solution to the problem

$$\begin{aligned} \hat{c}_{j,b} &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} \left[ y_{t+1} - (f_{b-1}(x_t) + c_{j,b}) \right]^2 \\ &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} \left[ e_{t+1,b-1} - c_{j,b} \right]^2, \end{aligned} \tag{8}$$

where $e_{t+1,b-1} = y_{t+1} - f_{b-1}(x_t)$ is the empirical error after $b-1$ boosting iterations. The solution to this problem is the regression tree that most reduces the average of the squared residuals $\sum_{t=1}^{T} e_{t+1,b-1}^2$, and $\hat{c}_{j,b}$ is the mean of the residuals in the $j$th state.

Forecasts are simple to generate from this approach. The boosted regression tree is first estimated using data from $t = 1, ..., t^*$. Then, the forecast of $y_{t^*+1}$ is based on the model estimates and the value of the predictor variable at time $t^*$, $x_{t^*}$. Boosting makes it more attractive to employ small trees (characterized by few terminal nodes) at each boosting iteration, reducing the risk that the regression trees will overfit. Moreover, by summing over a sequence of trees, boosting performs a type of model averaging that increases the stability and accuracy of the forecasts.

### Online Appendix A.1.3 Relative Influence Measures and Partial Dependence Plots

One criticism of machine learning algorithms is that they are "Black Boxes" that do not provide a lot of intuition to the researcher and the reader. This criticism is hardly applicable to BRTs that instead feature very useful and intuitive visualization tools.

#### Online Appendix A.1.3.1 Relative Influence Measures

The first measure commonly used is generally referred to as "relative influence" measures. Consider the reduction in the empirical error every time one of the covariates $x_{l,\cdot}$, is used to split the tree. Summing the reductions in empirical errors (or improvements in fit) across the nodes in the tree gives a measure of the variable's influence (Breiman et al., 1984):

$$I_l(\mathcal{T}) = \sum_{j=2}^{J} \Delta e(j)^2 I(x(j) = l),$$

where $\Delta e(j)^2 = T^{-1}\sum_{t=1}^{T}\left(e_t(j-1)^2 - e_t(j)^2\right)$ is the reduction in the squared empirical error at the $j^{th}$ node and $x(j)$ is the regressor chosen at this node, so $I(x(j) = l)$ equals 1 if regressor $l$ is chosen, and 0 otherwise. The sum is computed across all observations, $t = 1, ..., T$, and over the $J-1$ internal nodes of the tree.

The rationale for this measure is that at each node, one of the regressors gets selected to partition the sample space into two sub-states. The particular regressor at node $j$ achieves the greatest reduction in the empirical risk of the model fitted up to node $j-1$. The importance of each regressor, $x_{l,\cdot}$, is the sum of the reductions in the empirical errors computed over all internal nodes for which it was chosen as the splitting variable. If a regressor never gets chosen to conduct the splits, its influence is zero. Conversely, the more frequently a regressor is used for splitting and the bigger its effect on reducing the model's empirical risk, the larger its influence.

This measure of influence can be generalized by averaging over the number of boosting iterations, $B$, which generally provides a more reliable measure of influence:

$$\bar{I}_l = \frac{1}{B}\sum_{b=1}^{B} I_l(\mathcal{T}_b).$$

This is best interpreted as a measure of relative influence that can be compared across regressors. We therefore report the following measure of relative influence, $\overline{RI}_l$, which sums to 1:

$$\overline{RI}_l = \bar{I}_l / \sum_{l=1}^{L} \bar{I}_l.$$

## Online Appendix A.2 Variables Definition

### Attention Variables

- *ADDITION-EX*: The number of times stock $i$ was added to existing watchlists during week $t$. Data Source: Seeking Alpha.
- *ADDITION-NEW*: The number of times stock $i$ was added to newly created watchlists during week $t$. Data Source: Seeking Alpha.
- *DELETIONS*: The number of times stock $i$ was removed from watchlists during week $t$. Data Source: Seeking Alpha.
- *NET-ADDITIONS*: The number of times stock $i$ was added to existing watchlists minus the number of times it was removed from watchlists during week $t$. Data Source: Seeking Alpha.
- *Following*: The number of Seeking Alpha users following a stock in week $t$ divided by the market capitalization of the stock.
- *Google Trends*: Weekly Search Volume Index (Google SVI) based on stock ticker scaled by its time-series average following Da, Engelberg, and Gao (2011).
- *Bloomberg's AIA*: Weekly abnormal institutional investor attention (AIA) on Bloomberg (Ben-Rephael, Da, and Israelsen, 2017).
- *Stock Returns*: The returns realized by a company at the weekly or daily frequency, in line with the frequency at which the results are computed.
- *Stock Price Volatility*: The sum of daily squared returns, computed at the weekly or daily frequency, in line with the frequency at which the results are computed.

### News

We use the Dow Jones Raven Pack Edition, which comprises "relevant information from Dow Jones Newswires, regional editions of the Wall Street Journal, Barron's, and MarketWatch."

Every news in Raven Pack contains a score between 0-100 that indicates how strongly related the entity is to the underlying news story, with higher values indicating greater relevance. For any news story that mentions an entity, RavenPack provides a relevance score. A score of 0 means the entity was passively mentioned, while a score of 100 means the entity was prominent in the news story. Values above 75 are considered significantly relevant. A value of 100 indicates that the entity identified plays a key role in the news story and is considered highly relevant. We include in the analysis only news with a score of 100.

Every news is also associated with an "event novelty score" (ENS), which represents how "new" or novel a news story is within a 24-hour time window across all news stories in a particular package (Dow Jones, Web or PR Editions). We only keep the cases where ENS=100, that is, only the first time a story is run.

Finally, every piece of news is associated with an ESS variable—news sentiment—defined as "A granular score between 0 and 100 that represents the news sentiment for a given entity by measuring various proxies sampled from the news. The score is determined by systematically matching stories typically categorized by financial experts as having a short-term positive or negative financial or economic impact. The strength of the score is derived from a collection of surveys where financial experts rated entity-specific events as conveying positive or negative sentiment and to what degree. Their ratings are encapsulated in an algorithm that generates a score ranging from 0-100 where 50 indicates neutral sentiment, values above 50 indicate positive sentiment, and values below 50 show negative sentiment."

We use this raw data information to construct the following variables:

- *NEWS type*: Indicator variable that takes the value of one if RavenPack database reports a news related event associated with the firm on week $t$, zero otherwise. News type is defined by the Category variable in RavenPack and is aggregated into 23 distinct types:

  1. **Technical Analysis** contains: *technical-view-bullish; technical-view-bearish; technical-view; technical-view-oversold; relative-strength-index-overbought; relative-strength-index-oversold; technical-price-level-resistance-bearish.*

  2. **Trading** contains: *stop/start-trading-in-firm-stocks.*

3. **Insider Trading** contains: *insider-sell; insider-buy.*
4. **Major Shareholder Disclosure** contains: *major-shareholders-disclosure.*
5. **Firm Event** contains: *conference-call; board-meeting.*
6. **Earnings Release** contains: *earnings; earnings-positive; earnings-negative; earnings-up; earnings-down; earnings-above-expectations; earnings-below-expectations; earnings-meet-expectations; earnings-per-share-positive; earnings-per-share-negative; earnings-per-share; earnings-per-share-above-expectations; earnings-per-share-up; earnings-per-share-down; earnings-per-share-below-expectations; earnings-per-share-meet-expectations; pretax-earnings-up; pretax-earnings-down; pretax-earnings-positive; pretax-earnings-negative; operating-earnings-positive; operating-earnings; operating-earnings-negative; operating-earnings-up; operating-earnings-down; ebitda-positive; ebitda-up; ebitda; ebitda-negative; ebitda-down; ebit-positive; ebit-up; ebit-down; ebita-positive; ebita-up; ebita; revenues; revenue-up; revenue-down; revenue-above-expectations; revenue-below-expectations; revenue-meet-expectations; revenue-volume; revenue-volume-up; revenue-volume-down; same-store-sales-up; same-store-sales-down; same-store-sales; same-store-sales-below-expectations; same-store-sales-meet-expectations; same-store-sales-above-expectations.*
7. **Guidance** contains: *earnings-guidance; earnings-guidance-meet-expectations; earnings-guidance-up; earnings-guidance-down; earnings-guidance-below-expectations; earnings-guidance-above-expectations; earnings-guidance-suspended; earnings-per-share-guidance; ebitda-guidance; ebitda-guidance-up; ebitda-guidance-down; revenue-guidance; revenue-guidance-up; revenue-guidance-down; revenue-guidance-meet-expectations; revenue-guidance-above-expectations; revenue-guidance-below-expectations; same-store-sales-guidance-up; same-store-sales-guidance; same-store-sales-guidance-down; dividend-guidance; dividend-guidance-up; dividend-guidance-down.*
8. **Analysts Actions** contains: *earnings-estimate; earnings-estimate-downgrade; earnings-estimate-upgrade; operating-earnings-guidance; operating-earnings-guidance-up; operating-earnings-guidance-down; evenue-estimate; evenue-estimate-downgrade; analyst-ratings-change-negative; analyst-ratings-change-positive; analyst-ratings-change-neutral; analyst-ratings-*

*set-positive; analyst-ratings-set-neutral; analyst-ratings-set-negative; analyst-ratings-history-positive; analyst-ratings-history-negative; analyst-ratings-history-neutral; price-target-upgrade; price-target-downgrade; price-target-set.*

9. **Executive Turnover** contains: *executive-appointment; executive-resignation; executive-firing; executive-death; executive-health; executive-scandal.*

10. **Executive Compensation** contains: *executive-salary; executive-salary-increase; executive-salary-cut; executive-shares-options; executive-compensation.*

11. **Labor Issues** contains: *layoffs; hirings; strike; strike-ended; union-pact; union-pact-rejected; workforce-salary-increase; workforce-salary-decrease.*

12. **Transaction Firm Stocks** contains: *ownership-increase-held; ownership-decrease-held; public-offering; buybacks; fundraising; investment-recipient; private-placement; ipo; rights-issue; spin-off; ipo-pricing; ipo-price-decrease; ipo-price-increase; stake-acquiree.*

13. **Stock Restructure** contains: *reverse-stock-splits; stock-splits*

14. **Products and Services** contains: *business-contract; business-contract-terminated; product-release; product-delayed; regulatory-product-approval-granted; regulatory-product-approval-denied; regulatory-product-approval-conditional; supply-increase; supply-decrease; supply-unchanged; government-contract; market-entry; product-price-raise; product-price-cut; product-recall; demand-increase; demand-decrease; demand-unchanged; market-share-loss; market-share-gain; market-share; product-discontinued; demand-guidance-increase; demand-guidance-decrease.*

15. **M&A** contains: *business-combination; acquisition-acquirer; acquisition-interest-acquirer; acquisition-rumor-acquiree; acquisition-rumor-acquirer; unit-acquisition-acquirer; unit-acquisition-interest-acquirer; unit-acquisition-rumor-acquirer; stake-acquirer; merger; merger-rumor.*

16. **Dividends** contains: *dividend; dividend-up; dividend-down; dividend-suspended; dividend-meet-expectations; dividend-above-expectations; dividend-below-expectations.*

17. **Stock Price** contains: *stock-gain; stock-loss.*

18. **Credit Rating** contains: *credit-rating-affirmation; credit-rating-set; credit-rating-downgrade;*

*credit-rating-upgrade; credit-rating-unchanged; credit-rating-provisional-rating; credit-rating-confirmation; credit-rating-action; credit-rating-publish; credit-rating-corrected; credit-rating-no-rating; credit-rating-outlook-stable; credit-rating-outlook-negative; credit-rating-outlook-positive; credit-rating-outlook-revision; credit-rating-outlook-developing; credit-rating-watch-negative; credit-rating-watch-positive; credit-rating-watch-removed; credit-rating-watch-unchanged; credit-rating-watch-developing; credit-rating-watch; credit-rating-revision-enhancement.*

19. **Partnership** contains: *partnership; joint-venture.*

20. **Legal** contains: *legal-issues-defendant; legal-issues-dismissed-defendant; settlement; sanctions-target; sanctions-lifted-target; legal-verdict-disfavored; legal-verdict-favored; patent-infringement-defendant; fraud-defendant; fraud; antitrust-suit-defendant; discrimination-defendant; antitrust-settlement; sanctions-guidance-target; antitrust-investigation; corruption; corruption-defendant; defamation-defendant; tax-evasion; embezzlement-defendant; regulatory-investigation; regulatory-investigation-completed-sanction; exchange-noncompliance.*

21. **Indexes** contains: index-listing; index-delisting; index-delisting-issuer; index-rebalance.

22. **Accident, Wars, Crime** contains: *power-outage; facility-accident; force-majeure; force-majeure-lifted; aircraft-accident; mine-accident; spill; public-transport-accident; pipeline-accident; factory-accident; automobile-accident; cyber-attacks; cyber-attacks-threat; weapons-testing; explosion; transportation-disruption; embargo-lifted-target; embargo-issuer; embargo-target; embargo-lifted-issuer; bombing-attack-target; bombing-threat-target; violent-attack-target; violent-attack-threat-target; shooting; protest-ended-protester; protest-protester; protest-protestee; evacuation-authority.*

23. **Other** contains the remaining unclassified events.

- *Positive-News*: Indicator variable that takes the value of one if the average news sentiment in week t is positive (above 50).

- *Negative-News*: Indicator variable that takes the value of one if the average news sentiment in week t is negative (below 50).

- *Neutral-News*: Indicator variable that takes the value of one if the average news sentiment in week t is neutral (equal to 50).

### Stock Characteristics

- *Past Returns*: Realized excess return of stock $i$ in the previous year.

- *Past Variance*: Realized variance of stock $i$ in the past year - computed using daily squared returns. We follow Amaya et al. (2015) and use the following expression where $r_{j,i}$ is return of stock $i$ on day $j$ and $N$ is the number of trading days:

$$RVar_i = \sum_{j=1}^{N} r_{j,i}^2$$

- *Past Skewness*: Realized skewness of stock $i$ in the past year - computed using daily returns. We follow Amaya et al. (2015) and use the following expression where $r_{j,i}$ is return of stock $i$ on day $j$ and $N$ is the number of trading days:

$$RSkew_i = \frac{\sqrt{N}\sum_{j=1}^{N} r_{j,i}^3}{RVar_i^{3/2}}$$

- *Past Kurtosis*: Realized kurtosis of stock $i$ in the past year - computed using daily returns. We follow Amaya et al. (2015) and use the following expression where $r_{j,i}$ is return of stock $i$ on day $j$ and $N$ is the number of trading days:

$$RKurt_i = \frac{N\sum_{j=1}^{N} r_{j,i}^4}{RVar_i^2}$$

- *Market-to-Book Ratio*: The ratio of market value of assets and book value of assets. Market value of assets is computed as the sum of the closing stock price (Compustat item: PRCC) multiplied by the common shares (Compustat item: CSHPR), debt in current liability (Compustat item: DLC), long term debt (Compustat item: DLTT), and preferred stocks (Compustat item: PSTK), minus deferred taxes and investment tax credit (Compustat itme: TXDITC).

- *R&D*: The ratio of research and development expenses (Compustat item: XRD) and sales (Compustat item: SALE). Following Frank and Goyal (2003), we replace R&D with 0 if the entry related to research and development expense is missing.
- *Profitability*: The ratio between operating income before depreciation (Compustat item: OIBDP) and total assets (Compustat item: AT).
- *Book Leverage*: Computed as the sum of long term debt and debt in current liabilities, divided by the market value of assets.
- *Assets*: Log total assets.
- *Analyst Coverage*: The (log) number of analysts covering the stock.
- *Turnover*: The Stock's volume divided by the shares outstanding.
- *Volume*: The stock's volume.
- *Age*: The (log) number of years a company has been in the CRSP data set.
- *Frac. Institutional*: The fraction of the shares outstanding owned by institutional investors.

## Table Online I. Industry Sector Following by Investor Category

| | Panel A. Sectors Followed when Starting the Watchlist | | | | | Panel B. Sectors Followed After 1 Year of Starting the Watchlist | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | 25% | Median | 75% | Mean | Std. Dev. | 25% | Median | 75% |
| Occasional Investor | 3.5 | 2.4 | 2 | 3 | 4 | 5.2 | 3.9 | 3 | 4 | 6 |
| Retiree | 3.8 | 2.8 | 2 | 3 | 5 | 5.4 | 4.0 | 3 | 5 | 7 |
| Full Time Investor | 3.8 | 3.1 | 2 | 3 | 5 | 5.7 | 4.3 | 3 | 5 | 7 |
| Professional Finance | 3.6 | 2.9 | 2 | 3 | 4 | 5.0 | 4.0 | 3 | 4 | 6 |
| Student | 3.3 | 2.3 | 2 | 3 | 4 | 4.7 | 3.3 | 3 | 4 | 6 |
| Academic | 3.4 | 2.4 | 2 | 3 | 4 | 4.7 | 3.3 | 3 | 4 | 6 |
| Executive | 3.0 | 2.2 | 2 | 3 | 4 | 3.8 | 3.0 | 2 | 3 | 5 |
| Journalist/Other | 3.3 | 2.8 | 2 | 3 | 4 | 4.6 | 3.9 | 2 | 4 | 5 |
| Undeclared | 3.2 | 2.5 | 1 | 3 | 4 | 5.5 | 4.5 | 3 | 4 | 7 |

| | Panel C. Sectors Followed After 4 Years of Starting the Watchlist | | | | | Panel D. Sectors Followed After 8 Years of Starting the Watchlist | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | 25% | Median | 75% | Mean | Std. Dev. | 25% | Median | 75% |
| Occasional Investor | 7.6 | 5.6 | 4 | 6 | 10 | 13.1 | 7.3 | 8 | 12 | 17 |
| Retiree | 8.2 | 5.7 | 4 | 7 | 11 | 13.9 | 7.3 | 8 | 13 | 18 |
| Full Time Investor | 8.7 | 6.3 | 4 | 7 | 11 | 14.4 | 8.0 | 8 | 13 | 19 |
| Professional Finance | 7.7 | 6.1 | 4 | 6 | 10 | 14.0 | 8.6 | 7 | 12 | 19 |
| Student | 6.4 | 4.8 | 4 | 5 | 8 | 11.4 | 7.1 | 6 | 10 | 14 |
| Academic | 6.8 | 5.1 | 4 | 5 | 8 | 13.3 | 8.2 | 7 | 12 | 17 |
| Executive | 5.4 | 4.4 | 3 | 4 | 6 | 10.9 | 6.8 | 6 | 9 | 15 |
| Journalist/Other | 7.2 | 5.8 | 3 | 5 | 9 | 13.5 | 8.1 | 7 | 12 | 18 |
| Undeclared | 8.0 | 6.5 | 4 | 6 | 11 | 12.9 | 8.1 | 7 | 11 | 18 |

This table reports the number of sectors followed by investors on *Seeking Alpha*. In Panel A, we report the cross-sectional averages, standard deviations, as well as the 25th, 50th, and 75th percentiles of the initial number of sectors (out of the 49 Fama-French sectors) in each investor's watchlist, computed for each investor category. Panel B, C, and D report the same quantities computed after 1, 4, and 8 years of opening a watchlist. All quantities are computed using only active users, defined as those individuals who change their watchlist at least once after 30 days from initiating it.