The Research Behavior of Individual Investors*

Toomas Laarits
NYU Stern School of Business

Jeffrey Wurgler
NYU Stern School of Business and NBER

February 13, 2025

Abstract

Browser data from a representative sample of individual investors provides a detailed account of their search for information, including how much time they spend on stock research, which stocks they research, which categories of information they seek, and when they gather information relative to events and trades. Our best estimate is that the median individual investor spends a total of seven minutes of research per trade on stock research on the traded ticker, primarily just before the trade; the mean investor spends 29 minutes. Individual investors have little specific interest in traditional risk statistics and spend the most time reviewing price charts, followed by analyst opinions. Research interest is correlated with stock size, and salient news and earnings announcements draw more attention. Across individual investors, research styles vary, and investors that focus on short-term information are more likely to trade more speculative stocks.

* We thank Valerie Baldinger, Myeongrok Doh, and Cody Wang for outstanding research assistance and Terry Odean, Raghu Rau, Paul Tetlock, Noah Stoffman (discussant) and participants at NYU Stern, the NBER Conference on Big Data and High-Performance Computing for Financial Economics, and the Colorado Finance Summit for helpful comments.

I. Introduction

Asset pricing models make a wide variety of assumptions about what investors know. Classic models assume interest in, and knowledge of, variances, covariances, and risk premia, e.g., Sharpe (1964); Merton (1987) notes limits on the number of investments that capture attention but retains focus on risk-return statistics. De Long, Shleifer, Summers, and Waldmann (1990) assume that some investors know fundamental values while others trade on noise. Hong and Stein (1999) assume that some investors follow the news and others react to price patterns. Microstructure models study other information structures. Information set concepts are also central to empirical research. Thousands of studies have examined whether prices incorporate the textbook information sets of historical prices, publicly available information, or even private information, for example.

This diversity of approaches is understandable in part because there is little direct evidence on the range of information that individual investors actually gather. In this paper, we use browser history data on an approximately representative sample of U.S. individual investors to address several basic questions: How much time do individual investors spend on stock research? Which sites do they use? Which stocks do they focus on? When do they do their research relative to their trades or corporate events? Which categories of information do they care about—and which, in contrast to themes in the literature, do they ignore?

These questions can be addressed through browser history data (also known as clickstream data) because URL addresses can be parsed to reveal details about the page visited. Also, importantly, the data provider balanced the panel against U.S. households as of the time of the sample, four months of 2007, allowing for a degree of generalization to the population of individual investors. The raw data for the households that we focus on in this paper—those that trade U.S. stocks within a brokerage account—include over 8 million clicks and 60,000 hours of Internet use. We identify 484 such households and they make 2,911 total U.S. stock trades over the four-month sample. For conciseness, we refer to household units conducting stock research and trading activity as individual investors. The data provide insight into current practice because trading through online brokerages and online research was already the norm as of 2007 and remains the norm today.[1]

Different readers may be interested in different results, but some primary conclusions include:

- Our best estimate is that the median individual investor spends seven minutes per trade research on tickers that they trade within the sample. The mean individual investor spends around 29 minutes per trade on research on traded tickers. The vast majority of stock research is done in the 24 hours before the trade.
- Most trades are preceded by the presentation of a snapshot page which includes a set of brief price and fundamental statistics and a graph of intraday prices. These are generally presented by default by the broker, and many investors don't pursue research that goes beyond a summary page.
- When individual investors do go beyond the snapshot page, they spend by far the most research time viewing price charts. In roughly declining order of interest, price charts are followed by analysts' estimates, earnings and other financials, technical price

[1] Over 80% of Schwab clients' trades were made online by 2002 (Bogan (2008)). The few investors who were still calling their broker or visiting branch offices to make trades by 2007 would have been less active less sophisticated, and economically less important than those who used the Internet. Barber and Odean (2002) find that investors trade more often after switching to online trading.

information, company websites, dividends, insider and institutional ownership and trading measures, and, lastly, traditional risk statistics such as beta or volatility.

- When choosing to perusing price charts, individual investors usually accept the default lookback period provided by the site. When they opt for a different period, they tend to choose a longer one.
- Firm size is the most important cross-sectional determinant of breadth and depth of research interest.
- Salient news, such as Apple's introduction of its iPhone, can temporarily launch a smaller stock to the uppermost ranks of aggregate attention. Individual investors also pay more research attention around earnings announcements.
- Individual investors have different research styles and focus on different categories of information. Some spend more time on research, some review more stocks, and some focus on indicators of short-term prospects versus long-term fundamentals.
- Investors' research styles connect somewhat to the characteristics of stocks that they trade. Investors who research more speculative stocks tend to focus on price charts, news, and simple snapshots as opposed to slower-moving fundamentals such as earnings and dividends.
- Google Trends volume on tickers accurately measures variation in individual investor interest as measured directly from browser data, alleviating concerns of prior research about whether it captures individual investor interest or is too noisy due to its inclusion of all searchers.

An antecedent of this paper is Lease et al. (1974), who document individual investor demographics, investment strategies, and sources of information from a survey of retail brokerage clients. Relative to the current paper, their sample was based on a nonrepresentative set of respondent clients from one brokerage; their self-reported research practices could not be observed or confirmed; and, in any event, the scope of stock research easily available to individual investors has mushroomed in the Internet era.

More recently, and closest to the data used in this paper, Gargano and Rossi (2018) study clickstream data from one online brokerage. They document investor time spent across different broad categories of pages (research, trading, balances etc.), as well as drivers of investor attention, both in the cross-section of stocks and of investors. Among other differences with our paper, Gargano and Rossi do not focus on the specific categories of information sought; the data herein reflect the full online information diet of a representative set of individual investors.

Clickstream data also provide a direct measure of investor attention and our results contribute to that literature. The work on attention often focuses on linking proxies for attention to various market outcomes. For instance, Barber and Odean (2008) show that stocks with news or extreme daily returns capture the attention of retail investors, resulting in retail traders being net buyers of attention-grabbing stocks. A related line of work in Kaniel et al. (2008) and Barber et al. (2008) finds evidence of retail trader contrarianism over short-horizon past returns. Moving past stock market-based proxies of attention, Da et al. (2015) find that aggregate internet search volume in topics related to economic downturns has predictive power over market volatility and fund flows. Subsequent work has sought to further disaggregate types of information commanding the attention of different investors. For instance, Kwan et al. (2023) study a dataset of news article reads and relate them to portfolio allocations of institutional investors.

The attention literature is a bit fragmented, however. As a crude analogy, if the information set of individual investors in aggregate was a pot of alphabet soup, a number of

papers identify the existence of individual letters—news matters, volatility attracts attention, earnings announcements affect trading, etc. Due to the representative and comprehensive nature of our sample, we can estimate the full distribution of letters in the pot as well as observe the distribution within 484 particular bowls. Relative to the literature on attention, recent studies employing ownership data of retail investors, such as Balasubramaniam et al. (2023) and Gabaix et al. (2024) are more comprehensive, but any inference on the information that individual investors inject into prices is only implicit.

The current paper is a high-level description of the types of information that individual investors seek and what they care about most. This is a first step in a longer agenda involving how the information gathered by individual investors, as evident in clickstream data, connects to how they trade, how they construct portfolios, and how they perform.

Section II describes the nature of the data and the sample of investor-household-trader units. Section III gives basic statistics on total stock research. Section IV reviews most-consulted domains, and Section V investigates which stocks are of most research interest. Section VI studies the timing of research vis-à-vis firm news and investor trading. Section VII documents which categories of information interest individual investors. Section VIII explores heterogeneity in research interests and approaches. Section IX concludes.

II. Individual investor browsing data

We begin with an introduction to the data set, including a stylized example, and discuss its unique advantages and remaining limitations. We then describe the specific household ("individual investor") sample of interest.

A. Data source

Clickstream data like ours are collected by multiple companies. Typical use cases include, for example, online retailers who use it to optimize how consumers navigate their websites. The data we use was made available by an online research company introduced in a legal context by Bakos, Marotta-Wurgler, and Trossen (2011) and Marotta-Wurgler (2011, 2012) and in a law and finance context by Laarits, Marotta-Wurgler, and Wurgler (2024). It includes the browsing behavior of tens of thousands of U.S. households across January, February, March, and June 2007. To paraphrase Bakos et al. (2011), the panel of households, recruited using random-digit-dialing distributions, installed a plug-in that collected the timing and sequence of URLs visited. All confidential or personally identifiable data, such as account numbers, addresses, passwords, and the like, have been removed by the data provider. To date, few studies have exploited detailed clickstream data in finance research; exceptions are Gargano and Rossi (2018) and Benamar, Foucault, and Vega (2021).

Inclusion in the data set is voluntary, but the data provider makes extensive efforts to reduce selection biases. Our focus in this paper is on the subset of households that trade individual stocks. Henceforth, we often refer to these household units as individual investors, following Barber and Odean (2002) and others. Subject to any remaining selection bias and sampling noise, the data offer an approximately representative picture of U.S. individual investors trading online at the time.

A stylized extract of raw data illustrates the level of detail it may include. Table 1 reports fourteen minutes of a browsing session. The session may begin at any link. This particular investor is a motorsports fan but soon switches to CNBC.com, where she clicks a link providing current data on several U.S. and international stock market indices, the US 10-year Treasury yield, and the USD-Euro exchange rate.

After a minute at that page, the investor logs into a brokerage account.[2] Within her online broker's research pages, she seeks out the day's most active names, which on that day included ImClone (IMCL). Based on our own investigation, IMCL enjoyed good news for its cancer drug's prospects. Our investor then consults the quarterly earnings performance for IMCL. From the link, we can infer that earnings were presented for the prior three years and estimates were shown for the next two years. In a separate panel, the chart showed the daily volume for IMCL over the prior three years, as well as a 13-day moving average. Further analysis of IMCL took place on the highly popular Yahoo Finance website. The investor obtains a quote on symbol ("?s=") IMCL and then looks up analysts' estimates ("ae") for same.

Following a check-in with race results, the investor returns to her brokerage's page and takes a different tactic. This time, she looks for trading ideas through a stock screener, in particular stocks with expected EPS growth of at least fifty percent over the next fiscal year. This would have yielded many results, Google (GOOG) among them. A click or two later, the investor comes to a snapshot page, which presents many types of financial and price information on the given stock in an abbreviated form. After looking at a price chart, the investor enters a market order to buy ("ordertype=1") Google shares. Then she logs into e-mail and proceeds with other activities.

While every domain has a different directory layout and requires tedious processing to extract all the available information from query strings, as described in the Data Appendix, this example illustrates how the data can provide rich detail about the research process of individual investors. This paper takes a first and high-level look at this data set. By implication, the data also shed light on the categories of information that are not interesting to individual investors despite being a mouse click away.

It is worth discussing what links do and do not reveal. We cannot directly observe what investors know or learn from other sources; in particular, we don't observe any information the investors gather offline. Although this may appear to be a severe limitation, its impact is greatly lessened by the fact that, as of 2007, the Internet was not just the easiest but often the only practical way to gather most of the data items that we track. Consensus earnings, beta, insider selling, institutional ownership, recent price trends, the day's most active stocks, forward P/E ratios—such metrics were easily accessible via free, continuously updated, one-stop-shop data feeds just a click or two away from the Buy and Sell buttons. Consider the wide range of information obtained instantly by the investor in Table 1. The Internet was—and as of this writing remains—the obvious venue for stock research for individual investors.

We also cannot observe an investor's accumulated "stock" of knowledge as opposed to the flow of information exhibited by the clickstream. In a stock-specific trading context, this is often not a major issue because many of the statistics consulted by our investor in Table 1 would be useless if stale. An investor interested in a one-week-lookback price chart cannot store such information for future trades. Dividend yield information, on the other hand, is more durable once it is observed. We can investigate these issues to see how research "ramps up" prior to the trade; if an investor was relying heavily on a stock of unobserved information, we might not expect a sudden burst of research on that stock prior to the trade (which it turns out is what we often observe).

These limitations are most significant when it comes to soft information about products, corporate reputation, and word of mouth from friends and family. Intuitions like "Apple makes cool products, and a growing number of young people seem to be using them" are abstract but

[2] The data provider scrubbed links to eliminate account numbers, personal information, email addresses, etc.

may be relevant in cases where investors are also customers or have come in contact with the product line. Informal information gathering is much less likely to be important for investments in defense contractors, oil drilling equipment manufacturers, and indeed most of the thousands of exchange-traded stocks.

Other limitations cause browser data to exaggerate, rather than understate, the scope of investor information. In particular, we cannot determine how well investors actually understand the information they access. If they struggle to interpret a detailed analyst report, the time spent perusing such data may give the impression that they value these items more than they actually do. Furthermore, without eye-tracking software, we cannot observe which specific parts of pages containing a mix of information, such as "snapshots" pages that list multiple statistics, capture attention. We take three approaches to dealing with this issue as described below. This noise may average out somewhat if different investors focus on different items of a multidimensional page, as the evidence of diverse research approaches suggests that they do.

Finally, spending only a few seconds viewing a bit of information does not necessarily point to a lack of understanding or interest. For example, investors familiar with beta will need only a moment to find that information for a given stock and, because beta is a stable stock characteristic, it will be relevant for as long as the investor can remember it. Changes in analyst opinion, on the other hand, not only take more time to find and understand, but may have a more short-term perception of investment relevance. One sure thing is that if an investor doesn't observe a type of information at all, even for a few seconds, it is unlikely to influence her decisions. For these reasons, we tabulate not just the average amount of time an investor spends on a particular type of information, but the fraction of investors interested at all.

B. Sample of individual investors

We focus on households where we observe at least one trade in a specific U.S. stock (or ADR) over the four-month sample. (Although some investors may do some activity one could regard as stock research yet not trade even a single time in four months, they are less interesting and could not be influential.) To maintain a well-defined focus on individual investors in stocks, we do not include the few investors who trade only in options, mutual funds, international stocks, bonds, or ETFs. Research on such instruments is also distinct. For example, options trading is more short-term and event-based, and fund-level investing does not involve as much company-specific research. We leave this to future work.

To identify traders, we start with a list of online brokerages operating in 2007. Six of them feature prominently in our data. In the links to pages in their domains, we search for words that suggest trading—buy, sell, ticker, order, trade, and so on—and proceed iteratively, manually inspecting each of thousands of potential trade links. Table 1 shows an example of a link that makes plain the details of a trade. See the Data Appendix for additional detail about how raw data are processed into usable data.

This process identifies 484 investors in our data, and they make 2,911 total U.S. stock and ADR trades over the four-month period. The raw data for the associated households include around 8.5 million clicks in roughly 60,000 hours of Internet use. Individual households spend between one and four months in the data. The median household spends three months in the sample while close to 40% of the households are present for the entire four months. To maintain comparability across investors who may be in the sample for a different number of months, we report statistics on household-level data, normalized either by the months they spend in the sample or the number of trades they carry out. For the same reason, we construct count variables

(such as the number of unique tickers researched) on the household-month level, and average to the household level.

Panel A of Table 2 shows basic demographic information. Head-of-household income is topcoded at $100,000/year ($155,000/year in 2025 dollars). The median is between $75,000 and $100,000. The average age of the head of household is 50 years. Households are dispersed across the U.S. and have a representative distribution of other household characteristics including family size and life stage.

Panel B reports general browsing statistics for this sample. We define a browsing session as a series of clicks with a break of no longer than 15 minutes. The mean household in this sample carries out 103 such browsing sessions per month, and the median household comes in at 92 sessions per month. During these browsing sessions, the average household visits three unique broker or other finance websites in a month. They spend over three and a half hours on brokerage sites, and another one and a half hours on other finance websites. For comparison, all other browsing time adds up to just over 37 hours per month (2239 minutes), indicating that these households spend over 10% of their total online time on finance-related sites. The time spent at brokerage and other finance sites is right-skewed. The median household spends 1.4 hours per month at brokerage sites while the 95th percentile spends close to fifteen hours.

Panel C of Table 2 summarizes trading activity in individual stocks that we were able to identify from the clickstream data; we will now switch more fully from "per household" to "per investor" terminology. The average investor in our sample trades in 1.3 browsing sessions per month; the median investor records only half a browsing session with a trade in the average month. Recall that investors are required to trade once in the four-month period to be included in our sample, hence the theoretical minimum value of per month browsing sessions with a trade is 0.25. A closely related measure is the number of stock trades. The average investor in the sample trades 2.2 times per month, while the median makes 0.67 trades per month. Trading activity is skewed: the 95th percentile household trades more than nine times per month. Trades are roughly split between buys and sells. The mean trade size is $13,540 and the median is $2,020 among the subset of 191 investors who use brokers where the URL format indicates the number of shares traded; we determine the trade price from contemporaneous CRSP and TAQ data. The act of trading itself does not take much time, as the average trading-related portion of a session with at least one trade lasts under three minutes.

III. How much time is spent on stock research?

A. Categories of content

Asset pricing research often distinguishes between cash flow news and discount rate news. In reality, investment-related information is organized differently. Brokerage sites provide a broadly similar list of company-specific data items, and investors can then drill down into these within the same domain or follow hyperlinks to other domains.[3] Given remaining heterogeneity in how websites present content, however, we must choose an appropriate level of granularity. If our content categories are too granular, we risk misinterpreting patterns if some investors, perhaps just by chance, connect to brokerages that offer a bit more or less detailed information; if the categories of interest are too coarse, we fail to exploit the full detail of the data.

Balancing these considerations led us to focus on the following content categories, which collectively represent "stock research" for purposes of the paper: (1) risk statistics, (2) earnings, (3) dividends, (4) other fundamentals (typically valuation ratios or financial statements), (5)

[3] See https://finance.yahoo.com/quote/AAPL for a current example.

analyst opinions, (6) informed ownership and trading, e.g., by insiders, funds or institutions, or short sellers, (7) price charts and price-related data, (8) technical signals, and (9) company website visits for research-relevant purposes. We classify links as "other" if the topic is determinable but of limited interest and as "indeterminate" if the link constitutes stock research but lacks the detail to further categorize. Some links are to pages that present a mixture of content, and these require special treatment as described below.

In our analysis, we focus first on aggregate totals of research time before breaking research interest down by content category. But documenting how individual investors distribute attention across these categories is clearly a unique opportunity provided by the data.

B. Content versus format

The content categories are deliberately based on content, not presentation format. The same earnings information can be conveyed through a bar chart, a news article, or a text press release. An analyst's opinion may be reported through text or summarized in a table. Past prices and returns also represent content, while a chart or a table of prices represents different formats. Although links may indicate the presentation format, and the manner in which information is presented might impact perception through salience and framing effects, we consider the content itself as the first-order input for investment decisions and the unit of increment to the information set. An analysis of presentation formats is a question for future research.

C. Total research time

One of the most basic questions the data speak to is simply *how much* stock research individual investors do. Table 3 summarizes the total online research activity conducted by investors in our sample, summing up over all the content categories. In order to maintain comparability between households that spend a differential amount of time in the sample, we construct these measures either as household-level averages of monthly values, or normalize by the number of months in sample, or normalize by the number of trades carried out by the household.

Row A. reports that individual households spend anywhere between one to four months in the sample, with three months being the median value and a quarter of the households spending at least two months. Panel B provides a sense of overall research activity, where we count as research any clicks that could be categorized the groups listed in subsection *A* above. Recall we define as a "browsing session" a sequence of clicks with a break no longer than fifteen minutes. The average household carries out close to 30 such sessions a month that have at least some research component and during these research sessions the average household sees 30 unique tickers and visits 3.3 distinct brokerage or other finance sites. We construct these count measures on the household-month level, average to the household level, and report the statistics of the resulting distribution of N=484 values. The median household carries out about 15 research sessions per month and comes across 12 unique tickers. The average browsing session with a research component includes 3.5 minutes of research. Over the course of the month, the average household carries out about 118 minutes of research, with the median at 37 minutes and the 95th percentile clocking in at close to nine hours per month. (Note that the per session averages multiplied by average sessions per month do not have to exactly equal the research per month numbers.) Finally, we report total research per trade: the average household carries out 144 minutes of research per trade with the 95th percentile coming in close to eleven hours, while the 25th and 50th percentile are at 10 and 36 minutes, respectively. We highlight this last calculation—research per trade—as the subsequent category-level analysis frequently employs this normalization to give a sense of how much research is reflected in a trade by an individual

investor. By every measure, though, research efforts are notably right skewed, with a small number of households engaging in stock research quite enthusiastically.

In Panel C we further break down the per-trade research numbers. Close to half of the 144 minutes of research per trade is matched to some sort of identifier, be it a stock ticker, an index identifier, or any other sort of information about the security in question. An average of 52 minutes per trade is matched to an explicit stock ticker, while 9 minutes are matched to funds and indices and just over a minute is matched to other categories such as REITs or currencies.

Panel D gives a sense of time spent at brokerage or other finance sites for reasons other than explicit research. Here we report the number of browsing sessions that included time at any brokerage or finance site listed in Table 4 but did not include an explicit research component. We find an average of 23 such sessions per month and the non-research time at these sites adds up to about 155 minutes per month.

IV. Which websites are used for stock research?

The next broad question is where online stock research is done. The event study literature documents that marginal investors react to information of various types, but it is often unclear where they obtain that information. Table 4 tabulates how long individual investors spend on different domains. Most domains are anonymized to protect the intellectual property of the data provider, but we can show the distribution of research time across every relevant individual domain. We break down other domains into brokerages and non-brokerage sites that include financial news or statistics. We can also identify time spent directly on Yahoo Finance, the SEC website, and the aggregate amount of time spent on the corporate websites of tickers traded within our sample.

The well-known site Yahoo Finance reports financial data, news, analyst opinions, filings, and other information in a standardized format, and it does not require an account to use. We can unmask it because it has previously been reported to be the most popular finance research site.[4] It is indeed the single most important site in our sample by a substantial margin. More than half of our investors use it at least once across the four months of the sample—in some cases, apparently through a hyperlink from a brokerage's domain, and in other cases more directly. Yahoo Finance also appears at or near the top of any Google search involving a particular stock, which in turn perpetuates its dominance. The mean investor spends 45.3 minutes per trade at the site on what we assess is stock research; the 53.5% of investors who actually use the site spend a mean of 85 minutes and a median of seven minutes per trade.

The second group of columns shows that slightly more than half of the total time on Yahoo Finance is matched to specific stock tickers. The third group includes only research links matched to tickers that were traded by the household at least once within the sample. The gap reflects some combination of time spent researching stocks already owned but not traded within the sample (such stocks, like the rest of the portfolio, are unobservable by us), time perusing information about novel stocks that were ultimately not traded, and other factors.

After Yahoo Finance, the research time across domains has a long tail. The next major sources of research are the brokerage sites themselves. We exclude the time spent trading and time on non-research activities at the same broker. Since investors typically use only their own broker for broker-domain research (few have accounts at multiple brokers), the columns for

[4] See https://ir.comscore.com/news-releases/news-release-details/yahoo-finance-ranks-top-financial-news-research-site-us-more-18 and https://www.semrush.com/website/top/united-states/finance/.

households with nonzero time spent on that broker's website are the most informative, because they exclude non-account holders who couldn't access that broker's domain anyway. From this perspective the relative dominance of Yahoo Finance again stands out.

Investors that spend any time on research at their broker's website spend an average of around an hour per trade, but the median time spent is far less. Many investors spend a minute or less on broker-site research per trade. Although nearly every investor doing at least a few seconds of broker-site research over the four-month sample, much of this may be passive in the sense that the broker may make it impossible for an investor to enter an order without being presented with a cursory quote page containing some information. This nonetheless presents relevant information, such as an intraday price chart, prior to the trade, and may still affect the decision whether and what quantity to trade.

Proceeding far down the list of importance, 6.6% of individual investors visited the SEC's domain. This overstates the interest in SEC.gov per se because visitors were typically directed by a link from another domain such as Yahoo Finance. In addition, only a fraction of these investors obtained a document that we can match to a company large enough to be in the CRSP sample. Keep in mind that we tabulate here only the time spent on the SEC website, not the time spent reading downloaded documents or documents opened in applications outside the browser. The table suggests little reason for concern about this limitation, however: Among 484 investors who perform 2,911 trades, only one investor reviewed information from the SEC website on a CRSP stock that they traded in the sample.

Another source of investment-relevant information is the websites of publicly traded U.S. corporations, specifically those matched to CRSP tickers. We group all such domains together for purposes of the table. Determining whether this browsing constitutes deliberate investment-focused research requires assumptions based on detailed examination of the clickstream, since so much internet activity occurs on domains of public corporations that is not investment-relevant (e.g., shopping on amazon.com or reviewing bank balances at citibank.com). We impose the requirement that a visit to a corporate website qualifies as "research" only when the same household clicks at least one link on another domain that implies investment-related research on the appropriate ticker. We also impose a running constraint that company website research cannot sum to more than two times the investor's overall investment-relevant research on that ticker.

Under these rules, 59.9% of investors in the sample consulted websites of publicly traded companies of demonstrated research interest to them. Among the 20% of investors that clicked on a company domain, conducted other research on the same stock, and traded the stock within our sample, the average spent about two minutes per trade on that domain and the median spent just a few seconds.

V. Which stocks attract research interest?

The most-researched stocks are listed in Table 5. At the top is Apple, Inc. (AAPL), with more than one-quarter of individual investors "researching" Apple at least once in just four sample months. This is an astounding degree of aggregate research attention in a single stock which, at the time, was not even in the top twenty in market capitalization. And, again, this is not simply counting households that peruse Apple's corporate website for new laptop specs or replacement cables; as described above, we do not include time spent at a company website as research unless the same household also studied the company from another investment-relevant perspective.

In Figure 1, we present a scatterplot of stocks that are among the top 100 stocks by market cap or the top 100 by the number of investors that engaged in investment-related research during the sample period. There is a strong correlation between breadth of interest and market cap. Microsoft (MSFT) ranks second in breadth of research interest and third in market cap, and the top five companies by market cap all fall within the top ten in terms of breadth of interest. In general, large technology companies see especially high levels of research interest.

Conversely, some stocks receive less research attention than their capitalization alone would predict. Oil stocks and financial institutions tend to attract less research attention relative to their size. Perhaps oil stocks are less interesting, or perhaps investors understand that the major driver of oil company profits is oil prices and firm-specific research is less useful. Perhaps research on financial institutions seems complex. In retrospect, AIG and many institutions involved in real estate lending could have used more scrutiny in 2007, the eve of the global financial crisis.

Table 6 conducts a regression analysis of breadth of research interest and intensity as a function of stock characteristics, the number of news articles, and other variables. The first several columns focus on the overall percentage of households with nonzero research time in that ticker-month; the last column uses the log of the total number of hours that sample households spent on research in that ticker-month. All right-side variables are standardized unless otherwise indicated.

The first takeaway is the dominant effect of firm size in explaining breadth of research interest. This reflects the pattern in Figure 1. As a single factor, log market cap explains 19% of the variation of the breadth of research interest across stocks. A one-standard deviation increase in log market cap from its mean, which corresponds to an increase in cap from about $775 million to $5.4 billion, doubles the percentage of investing households conducting some research within the sample, from an unconditional average of 1.69% up to 3.46% (=1.69%+1.77%).[5] The second specification shows that the inclusion of Fama-French 49 industry dummies increases explanatory power only marginally, to 22%.

The third specification includes two interesting endogenous factors—the retail share of trading volume, as estimated by Boehmer et al. (2021), and the percentage of individual ownership, as measured by 100% minus the percentage of institutional ownership. We confirm the expected positive relationships: Controlling for size, stocks that individuals disproportionately hold and trade also disproportionately attract their research attention.

The fourth specification considers two time-varying news variables. The first is total news coverage, not necessarily investment focused, constructed as the number of unique news items (articles and press releases) in the RavenPack database. Logs are again appropriate, as the median number of news articles per stock-month is 22, the mean is 83, and the 99th percentile is 1,061. Controlling for size, companies that are more often in the news are subject to more research attention. This specification also includes a dummy for whether the stock had an earnings announcement in the same month. Both news factors are statistically significant, but do not much increase overall explanatory power relative to size alone.

The remaining specifications add a full list of stock characteristics. A few stand out in their magnitudes. In addition to market capitalization, stock characteristics that attract research

[5] We report a linear specification for simplicity. Using logs of the percentage of households conducting research leads to very similar conclusions.

interest include low nominal share price, high volatility, no dividends, high growth (as measured by sales growth or external finance over assets), and S&P 500 Index membership.[6] Although neither share price nor S&P 500 membership is directly related to stock fundamentals, they have particularly large effects on research interest. In general, these relationships are intuitive in light of prior research on individual investors, such as Barber and Odean (2008) and Balasubramaniam et al. (2023).

The last column uses the log of the total number of research hours by investors in that stock-month as the dependent variable. As a measure of research interest, this variable can be quite noisy when, for example, a single investor focuses at great length on a particular stock, so we winsorize it at the 95th percentile. We generally see that the same factors and characteristics drive this measure of research interest as in the previous specifications.

VI. When is research done? Research around events and before trades

A. Research around events

Panel regressions from Table 6 suggest that elevated news coverage can lead to elevated research activity, but the limited length of the panel precludes the inclusion of firm fixed effects. In-depth study of research activity around specific events provides more compelling evidence into both the nature of our sample and actual investor research behavior.

In Figure 2, we analyze news-intensive periods for three companies listed at the top of Table 6, and we compare the breadth of research interest in our data to search intensity measured by Google Trends. Note that there is no mechanical relationship because Google search activity, like all search activity, is excluded from our research time tabulations.

Researchers have increasingly used Google Trends as a proxy for investor attention. Da et al. (2011) create indices of investor attention using Google search data. They find that increased search attention predicts higher prices in the subsequent weeks, eventual reversals over the long run, and IPO underperformance. Da et al. (2015) find that aggregate internet search volume in topics related to economic downturns has predictive power for market volatility and fund flows. Subsequent work on investor attention has sought to further disaggregate types of information commanding the attention of various investors. For example, Drake et al. (2012) document an increase in Google searches around earnings announcements, with elevated interest persisting long after the earnings news is released.

What does more detailed browser data, which by construction is limited to individual investors, indicate? As an illustrative example, Microsoft is the second-most researched stock in our sample. Panel A looks at a month of activity that includes a quarterly earnings announcement. We plot research behavior in the clickstream data, measured as the number of investors engaged in investment-relevant research, alongside Google Trends search volume for the ticker symbol "MSFT" as a proxy for investment-focused search activity.

There is a strong agreement between the two methodologies, supporting the approach of Drake et al. (2012) and others. The decline of interest on weekends is evident from both samples, and weekday variations roughly correlate as well. In both data sets, breadth of interest peaks around the quarterly earnings announcement. This indicates that our sample is representative of general investment-related interest despite being smaller by orders of magnitude than the Google Trends random sample of a large fraction of search activity. Also, in the other direction, our

[6] This is consistent with stock splits increasing individual investor research attention, since nominal price falls and return standard deviations rise, as argued in Shue and Townsend (2021).

sample validates that Google Trends can accurately measure time variation in individual investor interest and attention in individual tickers, which is not obvious due to Google's normalization approach.[7] This finding is of broad relevance because Google Trends data are available over the history of Internet usage and across search terms.

Apple is the most-researched company in our sample period. Panel B shows a month of activity that includes both the announcement of the iPhone and a regular quarterly earnings announcement. We again observe a local maximum in clickstream data interest around the earnings announcement and another spike around the product announcement. Here, Google Trends data have the advantage that separate search terms can be used to isolate iPhone product interest from other investment-related interest. Clickstream data, on the other hand, have the advantage that one can identify and isolate households who do other investment-related research on Apple from the mass of households who are simply interested in a novel consumer product.

Our third example, in Panel C, involves Sirius, which Table 6 shows is an outlier with far greater research interest than its capitalization would suggest. This traces to a merger announcement during the sample period. Extrapolating the number of investors in our sample who "research" Sirius to the population implies an enormous focus of investor attention on this small stock around this specific event. The popular talk show host Howard Stern moved to Sirius in January 2006, likely priming individual investor attention to corporate events. And again, investor interest based on browser data again supports the use of Google Trends.

B. Research in the run-up to a trade

The temporal pattern of overall research relative to actual trading is another basic aspect of individual investor behavior that browser data can characterize. Theory gives no general guidance, but some investors clearly do respond to news to an extent, and some types of information that investors could access quickly become stale, e.g., short-lookback price charts or the Sirius merger announcement in Figure 2. Other investment-relevant information, such as dividend policy, is more persistent and needs to be consulted less often.

Although this is conceptually straightforward exercise, there are some complexities. The simplest case is an investor that executes a single trade in a single ticker at t=0 and does no other trading throughout the other months of the sample. If so, one can simply sum up research time related to the traded ticker, as well as untraded tickers, and associate all of that research to the trade. But in the more general cases in which an investor makes multiple trades in the same ticker over time, or trades in multiple tickers in short order, one must keep in mind double-counting possibilities.

In Figure 5, we summarize all ticker-matched research in the calendar week prior to a trade. Panel A starts by simply conditioning on a buy or sell trade at t=0 (of which we have N=1,431 buys and N=1,480 sells) and looks back one calendar week to calculate the cumulative amount of research that is matched to any ticker, as a function of time remaining to the next trade, regardless of whether it is regarding the ticker traded at t=0. Each vertical bar marks a 24-hour period, and the periodic flat spots reflect afterhours pauses in research activity. We see an exponential increase in overall research time prior to a trade being executed; that is, the stock of research knowledge is rapidly augmented in the hours just prior to a trade. Roughly speaking, about half of the preceding week's ticker-matched research occurs in the 24-hour period prior to a buy, with a concentration in the minutes preceding the trade. Research prior to sells is

[7] For the same reason, it is not possible to directly use Google Trends data to estimate absolute levels of investment-related interest, i.e., the number or fraction of Googlers interested in a particular term, unlike in browser data.

somewhat more spread out. The mean cumulative research is about 20 minutes for buys and 15 minutes for sells.

The median behavior in Panel B looks different. The median household carries out no research at all until five days before a trade, and a considerable fraction of ticker-matched research that does take place occurs in the three or four minutes prior to the trade. Comparing the median and the mean behavior per trade implies that research time is skewed, and a considerable fraction of trades are preceded by at best a few minutes of ticker-specific research. Some sales may be driven by liquidity needs, but on average the cumulative pre-trade research time and pattern is similar to buys.

Next, we narrow in on the portion of this research that is attached to the specific ticker that was traded at t=0. The mean is in Panel C and median is Panel D. Given that market prices are impacted by trading, these pictures shed light on the amount and timing of stock-specific research that is injected into the market for a typical trade by an individual investor. We can again see research before a buy ramping up in the hour or two preceding a trade, with a spike in the minutes just before.

VII. What information do individual investors seek? Research by type of content

We now break down the research totals into the more granular types of underlying information. In keeping with prior, domain-level results, we report statistics on time spent in specific content categories, normalized by the number of trades that investor carries out over the course of the sample. We report means and medians of research time across all 484 investors, the percentage of investors doing any research in the category, and means and medians of research time conditional on the investor doing any research in the category. We calculate these values for all research by the household as well as separately for research that is matched to a ticker or further matched to a ticker that the same household trades during the sample period.

A. Content categories

We first tabulate time spent on pages that we can confidently assign to one of the following content categories. The numbers for some categories represent strict lower bounds because some links present multiple categories of content.

A.1. Risk statistics

The standard academic advice for individual investors is to allocate a fraction of portfolio wealth to a diversified portfolio and the rest to a riskless asset. Each investor in our sample has eschewed that advice, at least to some extent, since he or she is trading individual stocks. For such investors, academics offer the advice to at least be sure to distinguish alpha from beta. It is interesting to see the extent to which investors consult standard risk statistics, such as beta and volatility, or ignore our advice here as well.

It turns out that few investors display an explicit interest in standard risk statistics. Only seven out of 484 investors demonstrated an explicit interest in beta or volatility, and only a few drilled down into risk information on a ticker that they traded.

A.2. Dividends, earnings, and other fundamentals

Households may seek out "hard" fundamentals: dividends, earnings, valuation ratios, and the accounting statements and regulatory filings that underlie these figures. We find more interest in this type of research. Almost 40% of households seek out some sort of earnings information, with just under two minutes of research for the mean household. Dividends attract at least some direct attention by 17% of households, and those with that particular interest tend to spend

relatively more time on research. Intuition suggests that dividend information is faster to digest than earnings information.

Finally, a considerable fraction of households—53%—explicitly observe some non-earnings, non-dividend fundamental information. This category includes valuation ratios such as price to sales, annual reports, and regulatory filings. On average, we see a few minutes of such research per trade.

As discussed below, earnings and sometimes dividend information is prevalent on snapshot pages, which present a mix of content. Time spent in the specific content categories presented in Table 7 are lower bounds, and we reapportion additional time reviewing earnings in a later subsection.

A.3. Analysts

Analyst reports and estimates are of comparatively popular interest. 71% of households pursue this type of research in some measure over the course of the sample, at least for a few minutes. The mean household spends over 12 minutes per trade on analyst reports, and conditional on doing any such research they spend approximately 18 minutes. Time on analyst reports is quite skewed, with the median household with that interest carrying out only 3 minutes per trade.

A.4. Ownership

Ownership-related research includes information about short-selling, institutional or mutual fund ownership, or insider ownership or trading. This groups together market participants that an average individual investor might reasonably suspect to have comparatively superior information. The median investor who ever consults this information does so for under a minute per month; it would not take long to read the percentage of institutional ownership or short interest outstanding. Some type of ownership information is included on most major brokers' snapshot pages, although there was none on the Yahoo Finance site snapshot at the time. Overall, this category is among the least important in terms of research time share and breadth of interest.

A.5. Prices, price charts, and technical signals

By far the most prevalent types of pages the investors see are price charts and snapshots. Price chart refers to any sort of page that shows a chart of prices or reports price information in a tabular form. Snapshot refers to a combination page of price information, as well as a host of other categories such as analysts or earnings. The average household spends over 39 minutes per trade on price chart pages and another 37 minutes on snapshot pages. Nearly all households do some of this research: both of the categories see 93% of households with nonzero time.

Most snapshot pages contain a chart showing the intraday evolution of the nominal share price while most price chart pages contain a price chart showing the evolution of the nominal share price over the prior month. Only one broker and one finance website deviate from this standard, having one-year default lookbacks on the price chart pages. Theory gives no guidance as to how far back investors should be looking into past prices. Intuitively, price charts provide some notion of risk, which is useful since we have already seen that investors are not seeking out detailed risk statistics; but, for full context, a stock's price should be plotted against a market index, and that is generally not included in the default. Perhaps another plausible intuition is that investors with shorter investment horizons would be looking at shorter lookbacks.

Figure 3 shows the frequency of various lookback horizons that individual investors use. Panel A includes charts that are shown with default lookback windows and as snapshots are the predominant type of research, over 70% charts seen by investors show intraday data. The rest of chart consultations are mostly made up of default chart pages with one month lookback

windows. The bottom panel shows the frequency distribution of lookback periods when investors do opt out of the website's default. When doing so, investors tend to look further back to gather more historical information. They are most likely to opt for lookback windows between one month and one year, with only a few zooming into more recent data or zooming out beyond one year. This pattern seems consistent with the fact that most individual investors have an average holding period longer than one month, but the main takeaway is that most investors observe past prices only to the default lookback. It is intriguing to think about how the one-month default lookback could relate to the stylized facts of reversal within one month (Jegadeesh (1990)) and momentum at longer horizons (Jegadeesh and Titman (1993)).

A category closely associated with price charts is technical analysis. We find that 45% of households look at charts that include additional elements of technical analysis, such as moving averages, oscillators, Bollinger bands, or simpler return-based indicators such as 52-week highs. Overall time spent on technical analysis averages about 5 minutes per trade for the households who carry out any such research.

A.7. Company website visits

A potentially important source of stock market information is knowledge about the company's products. To capture this aspect of investor familiarity with the stock, we track time the households spend on the URLs associated with tickers that they separately research on any of the brokerage or finance sites. This means that an investor who has a Bank of America account, shops on Amazon.com, or routinely uses search websites such as Google or Yahoo does not get "research credit" for these tickers unless there is some other evidence of investment-related interest into those tickers. Specifically, we impose the constraint that the running total of company website research time can be no more than twice the running total of all other research on the ticker.

Investors in general have some exposure to the websites of companies they research. 290 households visit the website of a company that they also research in other investment-relevant respects. However, as a reflection of our 2x constraint, the overall time spent at company websites is modest, averaging only 1 minute per trade.

A.8. Other and indeterminate

The data include 13 minutes per trade of "Other," on average. These are links where we can observe or infer the type of research-related information, but it is obscure both theoretically and empirically and doesn't warrant its own primary category. We also count about 10 minutes per trade of research involving links that are truly "indeterminate," meaning we are not able to confidently ascribe any primary content category based on the information we observe.

B. Mixed content links

B.1. Snapshots

Investors frequently access detailed quote pages, often referred to as "snapshots" by their broker site or Yahoo Finance. These often serve as starting points for deeper research through other links and tabs, but sometimes they constitute the only research-related page that an investor observes. Further, in some cases, investors who trade a given ticker cannot avoid some form of fairly detailed quote based on the structure of the broker's website.

Sites have largely converged on a format for the snapshot page, but the underlying content categories represented varies. Snapshots virtually always contain current prices, daily percentage return, open/bid/ask, today's volume, 52-week range, market cap, dividend or yield, a brief earnings statistic, and a price chart. Beyond those statistics, practices vary. For example, beta was not in Yahoo Finance's snapshot page until 2012, well after our sample period. Sites

also differ in their use of lookback periods in price charts. Some include current headlines or detailed analyst forecasts, compare today's volume to a moving average, or report other fundamental data. Others are sparse or provide numerical data without context.

Absent eye-tracking software, we cannot observe interest in the different line items or charts within the snapshot page. In Table 7, we simply note the time spent on the snapshot page; at the end of this section, we use different approaches to impute the allocation of time and attention across the primary content categories displayed on the snapshot.

The average household spends 37 minutes on snapshot pages per trade. Nearly all of the households in the sample spend some time on such a page, but most do far less than this average. The median household spends only 6.5 minutes on snapshot pages per trade. Snapshot pages are usually easy to match to tickers. Since households may click on several stocks' snapshot pages before a trade, the amount of time spent on the snapshot of a stock for which we actually observe a trade in the sample is lower, coming in at 4.4 minutes per trade for the average household and 0.3 minutes per trade, or around twenty seconds, for the median household.

B.2. News and message board links

Other situations where imputations may be helpful are links to stock-related message boards and news articles for which we do not have headlines or coverage on the Internet Archive.[8] We often know that an investor is scrolling through a stock-related message board, but we do not know the content itself. The message board is simply a format that reflects a mix of underlying content, much like snapshots. News links present a similar challenge. The link may indicate that the article is about Intel but not whether it involves an earnings announcement or an analyst interview. In such cases we label the links as "Message Board" and "News – Indeterminate." See Kwan et al. (2023) for an extensive investigation of institutional investor attention to a large set of financial news articles that is matched to funds' holdings.

Of these two types of mixed-content links, news is the more important, although both are less important than snapshots in terms of time spent and breadth. News of indeterminate nature occupies a median of 2.6 minutes per trade, and message boards are used by only a fifth of households; among these, a few spend a good deal of time but even the median message-board per user spends only a minute per trade.

C. Imputing mixed-content links to content categories

We can push these allocations slightly further by redistributing the time spent on the three mixed-content categories, which are collectively significant, to specific categories of content. Under some assumptions, this allows for a more granular allocation of the average research time across types of underlying, specific content.

To impute snapshot links to specific categories of content, we obtain contemporaneous captures of the structure of each domain's snapshot using the Internet Archive and other sources. We then allocate the time spent on the domain's snapshot pages in three different ways. In the "equal weighted" imputation, we divide the time spent on the snapshot page equally among the specific categories represented. In the "visual weighted" imputation, we estimate the relative proportion of line items corresponding to each specific category represented. Finally, in the "interest weighted" imputation, we allocate snapshot time in proportion to the number of households that expressed explicit interest in that specific category in links in our data that could

[8] As a side note, Twitter—now X—had just begun operations in mid-2006. It plays no role in our analysis as the number of daily postings as of the time of our sample was miniscule. See: https://www.internetlivestats.com/twitter-statistics/. Another category of social media that may carry mixed financial content to some individual investors, blogging, was also an early stage as of the time of our sample.

be specifically allocated (as reported in the third column of Table 7). We prefer the third imputation as the most intuitive and objective measure.

For news links, we often cannot trace the underlying news article for further analysis because it was not cached in the Internet Archive. However, for two brokers we observe in the URL the article title and source (Reuters, WSJ, etc.), as well as any associated tickers. We extract the title, tickers, and the date, and match each click to an array of articles in the RavenPack data using the ticker and a date range of three days. We then rank this set of potential matches by measuring the distance between the two titles and keeping the closest, subject to a minimum distance requirement. The resulting dataset of matched articles allows us to estimate the breakdown of news article content using the detailed content classifications provided by RavenPack.

We follow a similar procedure to estimate the content breakdown of stock-related message boards, the last type of mixed-content link that we encounter. We scrape the present-day Reddit r/stocks message board and for each discussion thread we match the title, "flair" (user-supplied keywords), and original post content to our specific research content categories. We then back out content category breakdowns by ascribing equal time to each post, dividing time equally across all matched categories.

Table 8 shows how these methods ultimately assign time to primary categories. There are not dramatic differences based on the snapshot imputation approach, but one intuitive result is that the visual-weighted and interest-weighted approaches tend to assign a greater proportion of weight from visiting a snapshot page to price charts and technical information. This is because the snapshot page is often visually dominated by price information, in the case of the visual-weighted imputation, and because households display widespread interest in price charts when such activity can be directly observed, in the case of the interest-weighted imputation. Message boards and news, on the other hand, more often concern earnings.

We summarize the imputation results in Figure 4. Panel A shows mean research time per trade for all research by the all-investor sample, with red bars denoting average time per category and blue bars denoting time after imputing time from the three mixed-content categories to the remaining content categories using one of the imputation methods described in Table 8. Panel B limits the sample to the fifty investors with high ticker match rates across their research links, again tabulating all research, and Panel C limits this sample to research on the tickers that the investors trade. In each case, the red bars line up with the mean columns from Table 7 (rounded to the nearest minute), and the blue bars represent the imputed times. The bars sum to the mean total time in the appropriate sample.

These three particular panels are chosen because they represent some summary best estimates of mean time per trade allocated across categories of true content. Panel A represents our best estimate based on a putatively representative sample of hundreds of individual investors. We see again that nearly half of the overall research time is devoted to price charts and price-related information, and this becomes even more prominent when considering that such information usually represents a disproportionate share of snapshot pages. Risk statistics go from nearly ignored as a specific item of interest to a slight allocation, at least for the average household, largely because they may appear as an item on a snapshot page for a small number of domains as indicated in Table 8.

Panel B again illustrates that although the high ticker match rate sample is only a tenth as large as the all-investor sample and does not include the most skewed outliers in the all-investor sample, it displays qualitatively similar patterns in the means. This allows for more confidence

upon proceeding to Panel C, where we can apportion a high fraction of the overall research time from this sample to specific tickers that were traded by the investor, thus reflecting the research that could have been "injected" into the market by the associated trades.

Although each imputation method has its benefits and limitations, we regard the interest-weighted imputation as a compelling and objective measure. In this methodology, and with reference to Panel C, we estimate that the mean individual investor allocates her total of 29 minutes of research per trade on traded tickers as 15 minutes to price charts and related information, 2 minutes each to analysts, earnings, and other fundamentals, and the remaining time scattered across a number of other categories. While the medians are not reported in the graphs, Table 7 reminds us that the median investor in this subsample spends closer to six minutes of research per trade on tickers traded within the sample, with again the plurality of this time being devoted to price charts.

VIII. Heterogeneity in research behavior

There is clearly heterogeneity in the depth and types of research that investors carry out. We now look for patterns via principal components on investor-level research intensity in total and in different content categories.

For this analysis we restrict the data to research matched to tickers, and we create indicator variables for any investor-category that sees more than two minutes of research per month. This leads us to exclude the "risk statistics" category entirely as it is not of enough express interest to include in the estimation. The need to restrict to an investor's ticker-matched research is to be able to augment these indicator variables with a measure that captures the nature of the stocks that the particular investor focuses on. To be specific, we construct a speculative stock focus variable based the investor's propensity to research stocks across the characteristics included in Table 6; it is more positive for an investor who tilts her trading interest toward volatile, small, younger, lower nominal price "speculative" stocks and more negative for one with more interest in larger, older, dividend-paying "bond-like" stocks.[9] The inclusion of this variable into the principal components analysis is a simple way to correlate how the information of interest depends on the nature of the stock of interest.

The first principal component reported in Table 9 is a level effect. Some investors do more research of any and all categories than other investors. This is not driven by imputation because the specification includes mixed-content categories such as snapshots, message boards, and news of indeterminate nature. In light of the skewness in overall research time in the summary statistics, it is expected that this is a strong differentiator of research behavior across individual investors, and it explains about 27% of the variation. In these data, the depth of research is independent of a focus on speculative stocks.

The second principal component, which explains about 9% of the variation, is intuitive and interesting. It contrasts investors who focus on earnings, dividends, and other slow-moving fundamentals versus those who focus on news, message boards, brief summary statistics, and price charts. The latter investors also concentrate their research in speculative stocks. Speculative stocks are less likely to pay dividends, so there is less research to be done of that particular sort for such stocks. Speculative stocks are also more driven by rapidly decaying news and sentiment from message boards, and it is further intuitive that price charts are more likely to be of interest to investors who focus on speculative stocks.

[9] This is motivated by the speculative vs. bond-like stock distinction that Baker and Wurgler (2006) connect to exposure to investor sentiment.

Beyond these two components, there remains a substantial degree of unexplained heterogeneity in research behavior. This calls to mind the survey evidence that wealthy retail investors' "beliefs are mostly characterized by large and persistent individual heterogeneity" (Giglio et al. (2021), p. 1481). Table 9 provides a first step, but there is much more research to be done connecting the nature of the search for information, the formation of beliefs, the stocks of interest, and the style of trading.

IX. Conclusions

We document a number of stylized facts about the research interests and information set of individual investors. These facts help map out the nature of beliefs that individual investor trades could plausibly inject into market prices and what information guides individual investor portfolios. The main findings include:

- The median household carries out only a few minutes of stock-specific research before trading the stock, concentrated in the 24 hours before the trade.
- As a group, individual investors spend more time on price charts, followed by analyst-related information, than other categories of information. Most trades are preceded by the presentation of a snapshot of a dozen or so statistics, but these may be forced by the broker, and investors typically don't display an interest in going beyond this summary.
- Individual investors have heterogeneous research styles. Some spend more time on stock research than others; some review more stocks while others concentrate; and some focus on indicators of near-term prospects while others on long-term fundamentals. This behavior is correlated with stock characteristics in intuitive ways.

Is such behavior optimal? This paper does not attempt to answer this question. But, consider the contemporary trading environment. A more quantitative fund manager might spend years and millions of dollars refining algorithms—and then execute trades, given fresh data, within minutes or seconds. A traditional active fund manager might employ a number of industry specialists and spend days contemplating a large new position. It is optimistic to think that the typical individual investor could outperform such competitors with their own occasional minutes of research. The optimal amount of stock-specific research for the typical individual investor might be essentially zero, as advocates of indexing have long maintained. We are exploring the extent to which an individual investor's research—which is often perfunctory but occasionally in-depth—pays off in terms of better performance.

First, how does a given investor get from what she *observes* to what she *trades*? If an investor observes a particular price pattern in a chart, for example, how does that help her decide between buying, selling, or ignoring a particular holding? To the extent that information feeds into trading, browser data can shed light on how individual investors' portfolios are formed, because at a mechanical level portfolio composition is the cumulative outcome of past trading.

Second, what is the "right" amount of research for individual investors? One view is that investors should spend almost no time analyzing stocks; if they insist on trading individual stocks, they should take a dartboard approach and trust in market efficiency to ensure average returns. An alternative view, which apparently is appealing to many individual investors, is that there are returns to information gathering that justify a modicum of effort but perhaps not much more. This becomes the proverbial empirical question: How does the depth and type of *research* that investor conducts relate to the *performance* of her trades? For individual investors, is it true that "more research is needed"?

References

Arnold, Marc, Matthias Pelster, and Marti G. Subrahmanyam, 2022, Attention triggers and investors' risk-taking, *Journal of Financial Economics* 143(2), 846-875.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The Journal of Finance* 61(4), 1645-1680.

Bakos, Yannis, Florencia Marotta-Wurgler, and David R. Trossen, 2014, Does anyone read the fine print? Consumer attention to standard-form contracts, *Journal of Legal Studies* 43, 1-35.

Balasubramaniam, V., Campbell, J.Y., Ramadorai, T. and Ranish, B., 2023. Who owns what? A factor model for direct stockholding, *The Journal of Finance* 78(3), 1545-1591.

Barber, Brad M., and Terrance Odean, 2001, The Internet and the investor, *Journal of Economic Perspectives* 15, 41-54.

Barber, B.M. and Odean, T., 2002. Online investors: Do the slow die first? *The Review of Financial Studies* 15(2), 455-488.

Barber, Brad M., and Terrance Odean, 2007, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *The Review of Financial Studies* 21, 785-818.

Barber, B.M., Odean, T. and Zhu, N., 2008. Do retail trades move markets? *The Review of Financial Studies* 22(1), 151-186.

Barrot, J.N., Kaniel, R. and Sraer, D., 2016. Are retail traders compensated for providing liquidity? *Journal of Financial Economics* 120(1), 146-168.

Benamar, H., T. Foucault, and C. Vega, 2021, Demand for information, uncertainty, and the response of U.S. Treasury securities to news, *The Review of Financial Studies* 34(7), 3403-3455.

Bender, S., Choi, J.J., Dyson, D. and Robertson, A.Z., 2022. Millionaires speak: What drives their personal investment decisions? *Journal of Financial Economics* 146(1), 305-330.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *The Journal of Finance* 76, 2249-2305.

Bogan, Vicki, 2008. Stock market participation and the internet, *Journal of Financial and Quantitative Analysis* 43(1), 191-211.

Choi, J.J., Laibson, D. and Metrick, A., 2002. How does the Internet affect trading? Evidence from investor behavior in 401(k) plans. *Journal of Financial Economics* 64(3), 397-421.

Choi, J.J. and Robertson, A.Z., 2020. What matters to individual investors? Evidence from the horse's mouth. *The Journal of Finance*, 75(4), 1965-2020.

Cookson, J. Anthony, Joseph E. Engelberg, and William Mullins, 2023, Echo chambers, *The Review of Financial Studies* 36(2), 450-500.

Cookson, J.A. and Niessner, M., 2020, Why don't we agree? Evidence from a social network of investors. *The Journal of Finance* 75(1), 173-228.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *The Journal of Finance* 66, 1461-1499.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all FEARS: Investor sentiment and asset prices, *Review of Financial Studies* 28, 1-32.

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2012, Investor information demand: Evidence from Google searches around earnings announcements, *Journal of Accounting Research* 50, 1001-1040.

Fang, L. and Peress, J., 2009. Media coverage and the cross-section of stock returns. *The Journal of Finance* 64(5), 2023-2052.

Fedyk, Anastassia, 2024, Front-Page News: The Effect of News Positioning on Financial Markets, *The Journal of Finance* 79(1), 5-33.

Gargano, A. and Rossi, A.G., 2018. Does it pay to pay attention? *The Review of Financial Studies*, 31(12), 4595-4649.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, *American Economic Review* 111(5), 1481-1522.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *The Journal of Finance* 45, 881-898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65-91.

Kaniel, R., Saar, G. and Titman, S., 2008. Individual investor trading and stock returns. *The Journal of Finance* 63(1), 273-310.

Karlsson, Niklas, George Loewenstein, and Duane Seppi, 2009, The ostrich effect: Selective attention to information, *Journal of Risk and Uncertainty* 38, 95-115.

Laarits, Toomas, Florencia Marotta-Wurgler, and Jeffrey Wurgler, 2024, How much do individual investors care about SEC-mandated disclosures? Evidence from IPO and mutual fund prospectuses, New York University work in progress.

Lawrence, Alastair, James Ryans, Estelle Sun, and Nikolay Laptev, 2018, Earnings announcement promotions: A Yahoo Finance experiment, *Journal of Accounting and Economics* 66, 399-414.

Lease, Ronald C., Wilbur G. Lewellen, and Gary G. Schlarbaum, 1974, The individual investor: Attributes and Attitudes, *The Journal of Finance* 29, 413-433.

Liu, H., Peng, L. and Tang, Y., 2023. Retail attention, institutional attention. *Journal of Financial and Quantitative Analysis* 58(3), 1005-1038

Madsen, J. and Niessner, M., 2019. Is investor attention for sale? The role of advertising in financial markets. *Journal of Accounting Research* 57(3), 763-795.

Marotta-Wurgler, Florencia, 2012, Will increased disclosure help? Evaluating the recommendations of the ALI's 'Principles of the Law of Software Contracts,' *University of Chicago Law Review* 78, 165-186.

Marotta-Wurgler, Florencia, 2012, Does contract disclosure matter? *Journal of Institutional and Theoretical Economics* 168, p. 94-123.

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

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *The Journal of Finance* 32, 1151-1168.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80(3), 563-602.

Reinganum, Marc R., 1981, Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values, *Journal of Financial Economics* 9, 19-46.

Shue, K., & Townsend, R. R., 2021. Can the market multiply and divide? Non-proportional thinking in financial markets. *The Journal of Finance*, *76*(5), 2307-2357.

Sicherman, Nachum, George Loewenstein, Duane J. Seppi, and Stephen P. Utkus, 2015, Financial attention, *The Review of Financial Studies* 29, 863-897.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139-1168.

Tetlock, P.C., 2011. All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies* 24(5), 1481-1512.

Thayer, Jane, 2011, Determinants of investors' information acquisition: Credibility and confirmation, *The Accounting Review* 86, 1-22.

Umar, Tarik, 2022, Complexity aversion when seeking alpha, *Journal of Accounting and Economics* 73, 101477.

Van Nieuwerburgh, S. and Veldkamp, L., 2010. Information acquisition and under-diversification. *The Review of Economic Studies* 77(2), 779-805.

Table 1. Stylized example of browsing data. A constructed example of a household's browsing session in raw data format. The session indicates visits to non-finance websites, finance or financial news websites, and broker websites. The finance and broker websites indicate the nature of the content that is displayed, tickers involved, and some details of trades.

Row	Timestamp	Link
1	3/23/2007 2:31:11 PM	http://formula1.com/
2	3/23/2007 2:36:20 PM	http://cnbc.com/
3	3/23/2007 2:36:28 PM	http://quote.cnbc.com/quote.htm?symbols=.DJIA|.NCOMP|.SPX|GB;FTSE|DE;DAXX|FR;CAC|JP;N225|CN;SHI|US.10|$$EURUSD&requestMethod=quick&realtime=1
4	3/23/2007 2:37:30 PM	https://[broker-1].com/user/login
5	3/23/2007 2:37:37 PM	https://[broker-1].com/eresearch/markets_sectors/analysis/mostactives.jhtml
6	3/23/2007 2:38:01 PM	https://[broker-1].com/stocks/earnings/earnings.asp?symbol=IMCL
7	3/23/2007 2:38:22 PM	https://[broker-1].com/stocks/earnings/update_earningsChart.asp?showQTRChange=1&display=mountain&overlay=&indicator=volume,1,13,EMA,13&duration=1095&daysforward=545&year=2006
8	3/23/2007 2:38:51 PM	https://[broker-1].com/charts/update_chart.asp?duration=365&frequency=1day&scaling=linear&display=candlestickcolor
9	3/23/2007 2:39:14 PM	http://finance.yahoo.com/
10	3/23/2007 2:39:20 PM	http://finance.yahoo.com/q?s=IMCL
11	3/23/2007 2:39:25 PM	http://biz.yahoo.com/bw/070530/000000000.html?.v=1
12	3/23/2007 2:39:29 PM	http://finance.yahoo.com/q?s=IMCL
13	3/23/2007 2:39:32 PM	http://finance.yahoo.com/q/ao?s=IMCL
14	3/23/2007 2:40:52 PM	http://formula1.com/
15	3/23/2007 2:43:03 PM	https://[broker-1].com/tradingideas/screener/stock_screener.asp
16	3/23/2007 2:43:10 PM	https://[broker-1].com/tradingideas/screener/stock_screener_results.asp?c=IBES.EPSGrowthNFY:GEQ:50:LSS:100001
17	3/23/2007 2:43:33 PM	https://[broker-1].com/common/symbol_info/symbolInfo.asp?symbol=GOOG
18	3/23/2007 2:43:39 PM	https://[broker-1].com/stocks/snapshot/snapshot.asp?symbol=GOOG
19	3/23/2007 2:44:11 PM	https://[broker-1].com/stocks/charts/charts.asp?symbol=GOOG
20	3/23/2007 2:45:02 PM	https://[broker-1].com/invest/socreateentry?ordertype="sell"&symbol=GOOG&shares=100

21	3/23/2007 2:45:19 PM	https://[broker-1].com/user/logout
22	3/23/2007 2:45:31 PM	http://us.f518.mail.yahoo.com/ym/login?login

Table 2. Summary statistics. Full sample of 484 individual investors. All measures are constructed on the household-month level and collapsed to the household level. Income is topcoded at $100,000 per year. Age is of head of household. Trades are of US common equities and ADRs.

		Percentiles					
	Mean	5th	25th	Median	75th	95th	N
A. Demographics							
Income ($ thousands)	71.7	12.5	75.0	75.0	100.0	100.0	445
Age	50.1	29.0	39.0	50.0	59.0	74.0	467
B. Overall Browsing per Month							
N Sessions	102.8	25.0	65.8	92.2	137.0	205.7	484
Minutes per Session	25.8	10.1	16.3	22.8	32.0	51.8	484
N Unique Broker or Finance Domains	3.3	0.7	1.5	3.0	4.7	7.0	484
Minutes at Broker Domains	214.2	8.4	25.8	81.6	211.8	879.6	484
Minutes at Other Finance Domains	95.4	0	1.2	14.4	75.6	385.2	484
Minutes at Non-Finance Domains	2239.8	388.2	1102.2	1855.8	2998.8	5203.2	484
C. Overall Trading per Month							
N Sessions with a Trade	1.3	0.3	0.3	0.5	1.0	6.0	484
Minutes Trading	6.5	0.1	0.7	2.0	5.7	25.7	484
N Stock Trades	2.2	0.3	0.3	0.7	1.7	9.5	484
N Stock Buys	1.1	0	0	0.3	0.7	5.0	484
N Stock Sells	1.1	0	0	0.3	1.0	4.5	484
N Unique Tickers Traded	1.5	0.3	0.3	0.5	1.3	5.5	484
Amount Traded ($000)	13.5	0.1	0.3	2.0	8.4	52.0	191

Table 3. Aggregate research activity. Full sample of 484 individual investors. To account for differential time in the sample, measures normalized either by number of months in sample or number of trades. Count measures are constructed on the household-month level and collapsed to the household (investor) level. Research is the sum of all categories of stock research.

		Percentiles				
	Mean	5th	25th	Median	75th	95th
A. HH Months in Sample	3.0	1	2	3	4	4
B. All Research						
N Sessions Including Research	29.3	0.8	4.3	14.7	39.0	104.0
N Unique Tickers Researched	30.4	1.0	4.0	12.3	32.3	124.0
N Unique Broker/Finance Dom.	3.3	0.7	1.5	3.0	4.7	7.0
Minutes/Session Research	3.6	0.3	1.1	2.3	4.5	11.2
Minutes/Month Research	118.0	1.0	8.5	37.1	108.3	530.4
Minutes/Trade Research	144.2	1.9	10.5	36.4	112.6	659.0
C. Research Matched to any Identifier						
Minutes/Trade Research, Stocks	52.4	0.1	2.4	11.0	46.4	241.8
Minutes/Trade Research, Indices and Funds	9.2	0.0	0.0	0.2	2.6	33.1
Minutes/Trade Research, Other	1.3	0.0	0.0	0.0	0.1	3.9
D. All Other Financial Browsing						
N Sessions	23.1	1.0	5.3	14.1	35.0	72.0
Minutes/Session	5.4	0.9	2.2	4.1	6.7	14.8
Minutes/Month	154.9	4.4	18.0	58.0	152.2	657.0

Table 4. Research by domain. Total minutes of research normalized by number of trades, averaged across investors. Broker websites are those for which we observe trading activity and are anonymized. Finance sites may be specialized or general news sites; on such sites, we count time spent on financial news and financial statistic or data. Company websites are visits to the primary domain of the associated company but counted only when researched by the same investor on another finance domain.

	All Research				*Research Matched to Untraded Ticker*				*Research Matched to Traded Ticker*			
Domain	Mean All Inv	% t>0 Inv	Mean t>0 Inv	Median t>0 Inv	Mean All Inv	% t>0 Inv	Mean t>0 Inv	Median t>0 Inv	Mean All Inv	% t>0 Inv	Mean t>0 Inv	Median t>0 Inv
Yahoo Finance	45.3	53.5	84.7	6.9	21.1	44.8	47.0	4.4	5.3	27.5	19.4	1.3
Broker 1	12.3	27.3	45.3	6.2	5.3	19.2	27.5	4.0	1.4	21.5	6.4	1.5
Broker 2	10.9	18.0	60.6	10.4	3.2	15.5	20.7	3.8	0.5	13.8	3.3	0.9
Broker 3	10.8	26.0	41.6	13.5	4.5	24.6	18.2	6.0	1.3	20.9	6.4	2.1
Broker 4	9.5	16.7	56.6	5.5	1.7	9.7	17.2	5.6	0.2	7.0	2.8	1.4
Broker 5	8.7	35.3	24.6	5.8	2.9	26.7	11.0	2.4	0.8	18.6	4.3	0.7
Finance B	7.8	35.7	21.7	1.4	3.5	23.6	15.0	1.0	0.3	9.5	2.7	0.4
Finance C	7.1	34.3	20.6	1.4	4.6	22.5	20.5	2.3	0.7	13.8	4.9	0.8
Finance D	4.3	37.0	11.7	3.1	0.8	13.8	6.0	1.7	0.1	4.5	1.2	0.6
Finance E	3.4	7.6	44.2	0.6	3.0	7.2	41.5	0.4	0.3	2.1	13.1	1.2
Broker 6	3.3	3.5	94.9	1.5	2.6	3.1	83.8	1.0	0.1	1.9	6.7	0.3
Finance F	3.3	1.2	262.9	201.2	2.7	1.2	218.4	161.4	0.1	1.0	7.2	5.5
Finance G	3.1	0.4	751.5	751.5	3.1	0.4	740.8	740.8	0.0	0.4	7.7	7.7
Finance H	2.4	27.7	8.6	1.3	0.7	10.5	6.2	1.0	0.1	4.1	1.1	0.7
Finance I	2.2	11.8	18.2	2.0	0.9	10.1	9.1	1.9	0.9	6.2	13.8	1.2
Finance J	1.2	7.4	16.4	1.2	1.0	6.6	15.6	0.9	0.1	2.3	4.4	0.8
Company Website	1.2	59.9	2.0	0.1	0.8	55.0	1.4	0.0	0.4	20.0	2.1	0.1
Finance K	1.2	4.1	29.0	8.4	1.1	3.5	31.3	8.7	0.1	2.9	3.2	1.9
Finance L	1.1	6.4	17.0	3.3	0.3	4.1	6.4	1.3	0.0	2.1	0.8	0.9
Finance M	1.0	9.7	10.7	0.8	0.5	7.6	6.5	0.4	0.0	2.9	1.5	0.2
Finance N	1.0	3.3	30.2	1.0	0.1	2.3	5.4	0.8	0.0	0.2	2.5	2.5
Finance O	0.7	12.0	6.0	0.9	0.5	10.5	4.8	0.8	0.0	3.5	0.4	0.2
Finance P	0.6	9.7	5.8	1.1	0.1	3.9	2.9	1.2	0.2	2.1	8.4	0.4
Finance Q	0.5	8.1	6.3	1.6	0.2	5.2	3.6	0.9	0.0	1.4	1.0	0.3

Finance R	0.3	2.5	13.6	2.8	0.0	1.4	0.8	0.3	0.0	1.0	0.1	0.1
Finance S	0.3	9.3	2.9	0.2	0.3	6.8	3.6	0.2	0.0	4.1	0.4	0.1
SEC	0.2	6.6	3.2	0.9	0.2	3.9	4.1	0.9	0.0	0.2	0.8	0.8
Broker 7	0.1	0.4	19.6	19.6	0	0.2	2.1	2.1	0.0	0.0	0.0	0.0
Broker 8	0.0	0.8	0.3	0.4	0	0.8	0.2	0.1	0.0	0.4	0.3	0.3
NEC	0.4	18.4	2.02	0.27								
Yahoo Finance	45.3				21.1				5.3			
Finance ex-Yahoo	41.4				23.4				2.7			
Broker (Any)	55.6				20.2				4.3			
Company Website	1.2				0.8				0.4			
SEC	0.2				0.2				0.0			
Total	144.2				65.6				12.8			

Table 5. Most-researched stocks. Stocks ranked on the percentage of investing households carrying out nonzero research within the sample. Market cap as of January 2007. To eliminate shopping and other non-research activities, company website time is included only for households that also exhibit interest in other investment content categories.

			Research Excluding Company Website			*Research Including Company Website*		
Name	Ticker	Market Cap ($ billion)	% t>0	Mean t>0 Inv	Median t>0 Inv	% t>0	Mean t>0 Inv	Median t>0 Inv
APPLE INC	AAPL	74	27.3	2.6	0.8	27.3	3.0	0.8
MICROSOFT CORP	MSFT	302	22.5	4.5	0.8	22.5	5.0	0.9
CISCO SYSTEMS INC	CSCO	161	20.9	3.7	1.0	20.9	3.7	1.0
GENERAL ELECTRIC CO	GE	372	20.5	2.4	0.5	20.5	2.5	0.5
SIRIUS SATELLITE RADIO	SIRI	5	19.8	3.6	1.1	19.8	3.8	1.1
CITIGROUP INC	C	271	19	2.2	0.5	19	2.3	0.5
FORD MOTOR CO DEL	F	15	18.8	3.0	0.8	18.8	3.0	0.8
INTEL CORP	INTC	121	18	2.4	0.6	18	2.4	0.6
AT&T INC	T	236	17.4	3.7	0.8	17.4	4.0	0.8
GOOGLE INC	GOOG	114	16.3	3.7	0.7	16.3	3.8	0.7
ALTRIA GROUP INC	MO	183	15.3	4.7	1.6	15.3	5.2	1.8
WAL MART STORES INC	WMT	199	15.3	2.2	0.5	15.3	2.4	0.6
DELL INC	DELL	54	14.9	2.1	0.4	14.9	2.5	0.4
EXXON MOBIL CORP	XOM	432	14.3	2.9	0.5	14.3	2.9	0.5
GENERAL MOTORS CORP	GM	19	13.6	2.5	0.3	13.6	2.5	0.3
PFIZER INC	PFE	189	13.2	2.8	0.5	13.2	2.8	0.5
YAHOO INC	YHOO	39	13	4.4	0.4	13	5.2	0.4
HOME DEPOT INC	HD	83	13	4.2	0.6	13	4.3	0.6
GOLDMAN SACHS GROUP	GS	90	12	3.5	0.8	12	3.5	0.8
CONOCOPHILLIPS	COP	109	11.6	1.3	0.3	11.6	1.3	0.3

Table 6. Characteristics of researched stocks. Stock-month level regressions for stocks researched within the sample. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based Newey-West standard errors with two lags.

	% Investors t>0						*ln(1+Hours)*
Ln(Cap)	1.77***	1.87***	2.55***	1.66***	2.53***	2.85***	0.61***
	(16.05)	(15.81)	(15.79)	(14.26)	(11.81)	(12.87)	(20.91)
Retail Volume Share			1.02***			0.67***	0.18***
			(11.90)			(9.39)	(12.47)
Individual Ownership Share			0.42***			0.89***	0.16***
			(6.43)			(9.48)	(9.18)
ln(1 + N News Articles)				0.12**	0.10*	0.41***	0.06***
				(2.29)	(1.92)	(6.25)	(5.76)
Earnings Month (0-1)				0.62***	0.71***	0.46***	0.13***
				(7.36)	(8.82)	(5.42)	(6.36)
Ln(Age)					0.12*	0.04	0.02
					(1.84)	(0.67)	(1.63)
Ln(Price)					-1.03***	-0.61***	-0.15***
					(-5.62)	(-3.53)	(-5.73)
CAPM Beta					-0.27***	-0.14***	(0.01)
					(-5.44)	(-3.14)	(-0.73)
SD(Return)					0.46***	0.30***	0.10***
					(9.72)	(8.85)	(7.27)
Momentum					0.06*	-0.14***	0.01
					(1.66)	(-3.26)	(0.51)
Dividend>0 (0-1)					-0.39***	-0.47***	-0.16***
					(-2.60)	(-3.32)	(-5.40)
Dividend Yield					0.06	0.03	0.01
					(1.18)	(0.71)	(1.04)
Book-to-Market					0.04	0.14***	0.01
					(0.90)	(3.20)	(0.71)
Sales Growth					0.17***	0.11***	0.01
					(4.43)	(2.95)	(1.14)
External Finance					0.20***	0.13***	0.05***
					(4.17)	(2.80)	(3.38)
SP500 Member (0-1)					1.32***	0.55**	0.09*
					(5.25)	(2.38)	(1.80)
Constant	1.69***	1.10*	1.68***	1.20***	0.69**	0.24	-0.38***
	(40.30)	(1.82)	(41.65)	(8.87)	(2.06)	(0.77)	(-4.86)
FF49 Industry FE		Yes			Yes	Yes	Yes
Observations	11,280	11,280	11,280	11,280	11,280	11,280	11,280
R^2	0.19	0.22	0.24	0.19	0.29	0.33	0.22

Table 7. Minutes of research per trade. Investor-level averages of total number of minutes of research divided by total number of trades. The full sample of 484 investors is in the left set of columns. The remaining columns analyze 50 investors for which a high fraction of total stock research time can be matched to a CRSP-listed stock or ADR ticker. All Research may include a small percentage of time on options, indexes, funds, or other instruments. For each category of research, minutes per trade are calculated for the mean investor, the median investor, the percentage of investor with nonzero research time, and means and medians among investors with nonzero research time. Other denotes time on research on miscellaneous topics. Indeterminate denotes time on research of unknown type. The last two rows show the household averages of total research time per trade.

	All Investor Sample (N = 484)					*High Ticker Match Rate Sample (N = 50)*														
	All Research					All Research					Research on Untraded Tickers					Research on Traded Tickers				
	Mean All	Med. All	% t>0	Mean t>0	Med. t>0	Mean All	Med. All	% t>0	Mean t>0	Med. t>0	Mean All	Med. All	% t>0	Mean t>0	Med. t>0	Mean All	Med. All	% t>0	Mean t>0	Med. t>0
Risk Statistics	0	0	1	0.3	0.1	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0
Earnings	1.7	0	40	4.3	0.6	0.6	0	36	1.8	0.7	0.4	0	24	1.6	0.6	0.1	0	14	0.5	0.2
Dividends	0.6	0	17	3.8	0.7	0.1	0	12	1.2	0.8	0.1	0	4	1.9	1.9	0	0	2	0.2	0.2
Other Fundamentals	2.6	0	53	5	1	1.5	0	52	2.9	1.5	0.9	0	36	2.5	0.5	0.2	0	20	0.8	0.5
Analysts	12.5	0.9	71	17.5	2.5	5.1	0.4	68	7.5	1.9	2.2	0	52	4.2	1	0.6	0	40	1.6	0.3
Ownership	0.5	0	26	1.8	0.4	0.7	0	30	2.3	1.7	0.5	0	22	2.2	1.9	0	0	10	0.1	0
Price Charts	39.1	5.8	93	42.1	6.7	31.2	11.5	98	31.8	11.6	15.8	2	76	20.8	5.9	10.2	1.2	82	12.4	2
Technical	2.3	0	45	5.2	0.9	0.9	0	38	2.5	1.1	0.4	0	26	1.4	0.9	0	0	12	0.4	0.2
Company Website	1.3	0	61	2.1	0.1	1.5	0	56	2.6	1	0.9	0	46	2	0.8	0.5	0	20	2.7	0.7
Other	12.8	1.3	84	15.2	2.3	7.4	1	84	8.8	1.2	2.9	0	50	5.8	0.3	0.4	0	44	0.9	0.1
Snapshot	37	6.5	93	39.6	8.3	26.7	9.3	92	29.1	10.2	13.4	2.8	86	15.6	3.6	11.2	1.2	82	13.6	2.1
News-Indeterminate	16.2	2.6	85	18.9	4.2	7.5	2.2	84	8.9	2.7	3.6	0.3	70	5.1	1.1	1.2	0	50	2.4	1.3
Message Board	7.2	0	20	35.9	1.3	5.6	0	14	40.3	7.2	0.6	0	12	5.3	1.6	3.7	0	10	37.2	10.2
Indeterminate	10.4	1	76	13.7	2.2	7.7	0.6	66	11.6	3.5	3.2	0	56	5.6	1.9	1.1	0	46	2.3	0.3
Mean Total in Sample	144.2					96.6					44.9					29.2				
Median Total in Sample	36.4					51.3					25.8					5.7				

Table 8. Imputation of mixed-content pages to specific content categories. Time on snapshot pages is imputed to specific content categories in three ways: equal-weighted across the categories presented in the snapshot page in that domain, visually weighted according to the relative prominence of the category on the snapshot page in that domain, and interest weighted using the distribution of household interest in specific content categories in the third column of Table 7. Snapshot page structure varies across domains, so we present the min and max of the range of imputation proportions. The imputed content distribution for time on message boards is based on the distribution of content on Reddit r/stocks message board discussions. The imputed content distribution for financial news of unknown type is based on the distribution of content across RavenPack's detailed finance article content classifications. Rows that do not sum to 100% indicate remaining indeterminate content. See Data Appendix for further details.

	Proportional Imputation to Content Category								
	Risk Statistics	Earnings	Dividends	Other Fundamentals	Analysts	Ownership	Price Charts	Technical	Other
Snapshot (Equal Weighted)	0 - 0.20	0 - 0.20	0 - 0.20	0 - 0.33	0 - 0.20	0 - 0.14	0.10 - 0.33	0 - 0.33	0 - 0.22
Snapshot (Visual Weighted)	0 - 0.11	0 - 0.13	0 - 0.13	0 - 0.14	0 - 0.33	0 - 0.13	0.38 - 0.79	0 - 0.20	0 - 0.15
Snapshot (Interest Weighted)	0 - 0.01	0 - 0.19	0 - 0.08	0 - 0.28	0 - 0.23	0 - 0.09	0.18 - 0.49	0 - 0.24	0 - 0.34
Message Board - Indeterminate	0	0.27	0.02	0.07	0.02	0.01	0.26	0.02	0.33
News - Indeterminate	0.06	0.31	0.03	0.03	0.07	0.02	0.12	0.1	0.25

Table 9. Principal components of investor-level research focus. Investor-level analysis. We restrict the data to research matched to tickers, and we create indicator variables for any investor-category that sees more than two minutes of research per month. Speculative Stock Focus is a continuous variable based the investor's propensity to trade stocks across the characteristics included in Table 6; it is positive for an investor who tilts toward volatile, small, younger, lower nominal price stocks and negative for one that tilts toward larger, older, dividend-paying stocks. Too few investors are interested in risk statistics for a loading to be estimated.

	First PC	*Second PC*
Risk Statistics	.	.
Earnings	0.29	-0.35
Dividends	0.10	-0.46
Other Fundamentals	0.28	-0.29
Analysts	0.32	0.00
Ownership	0.18	-0.18
Price Chart	0.31	0.19
Technical	0.27	-0.11
Company Website	0.30	-0.03
Other	0.32	0.08
Snapshot	0.32	0.31
News - Indeterminate	0.36	0.10
Message Board – Indet.	0.16	0.45
Indeterminate	0.27	0.03
Speculative Stock Focus	0.01	0.42
Eigenvalue	3.75	1.23
Proportion Explained	26.79	8.78

Figure 1. Market cap versus number of investors conducting research. Sample limited to the top 100 stocks by market cap or top 100 by number of investors conducting any research on that ticker.

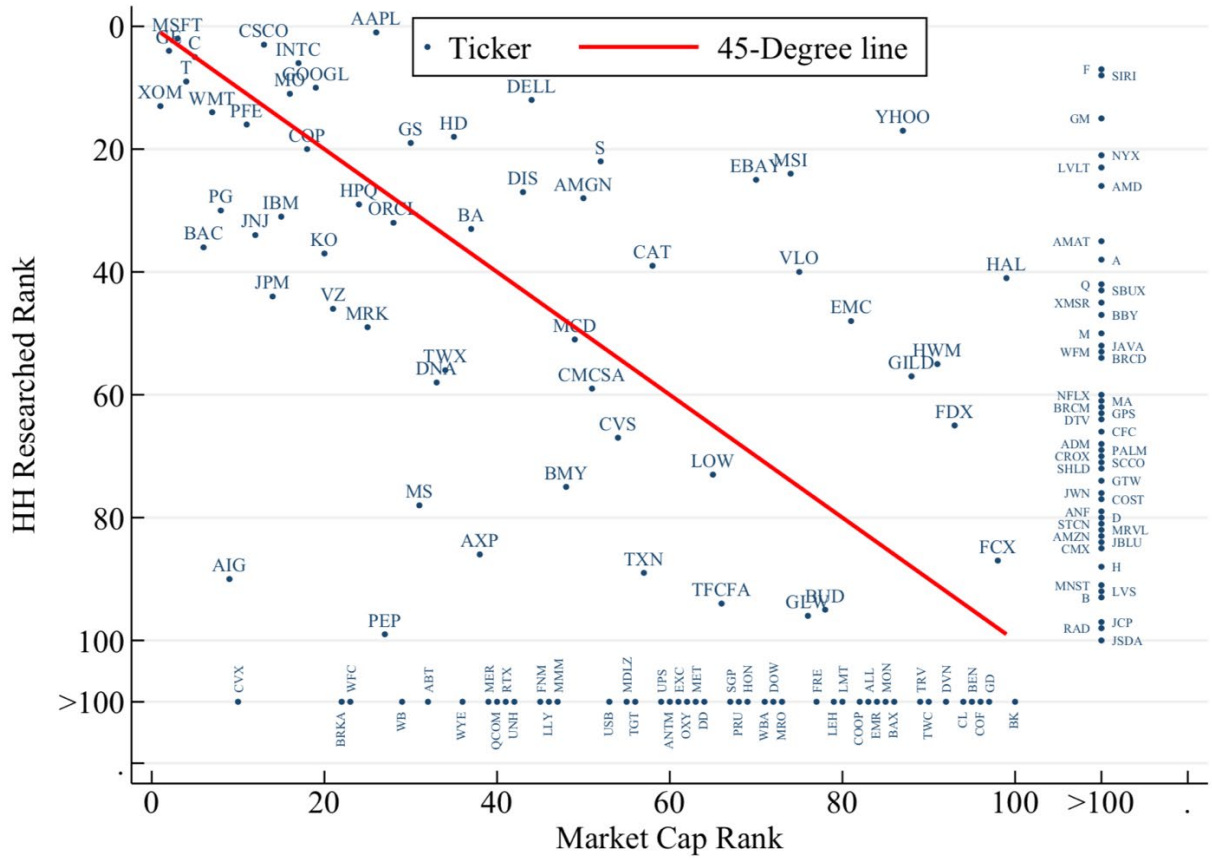


Figure 2. Research around events: Clickstream data vs. Google Trends. Panel A shows the number of investors in the full sample (N=484) with research activity on Microsoft stock in the month of its January 25, 2007 quarterly earnings announcement and compares this to the Google Trends index for interest in "MSFT". Panel B shows these two series around Apple's January 9, 2007 introduction of the iPhone and its January 17, 2007 quarterly earnings announcement, and Panel C shows these series around the February 19, 2007 announcement of a merger between Sirius and XM. Weekends are shaded.

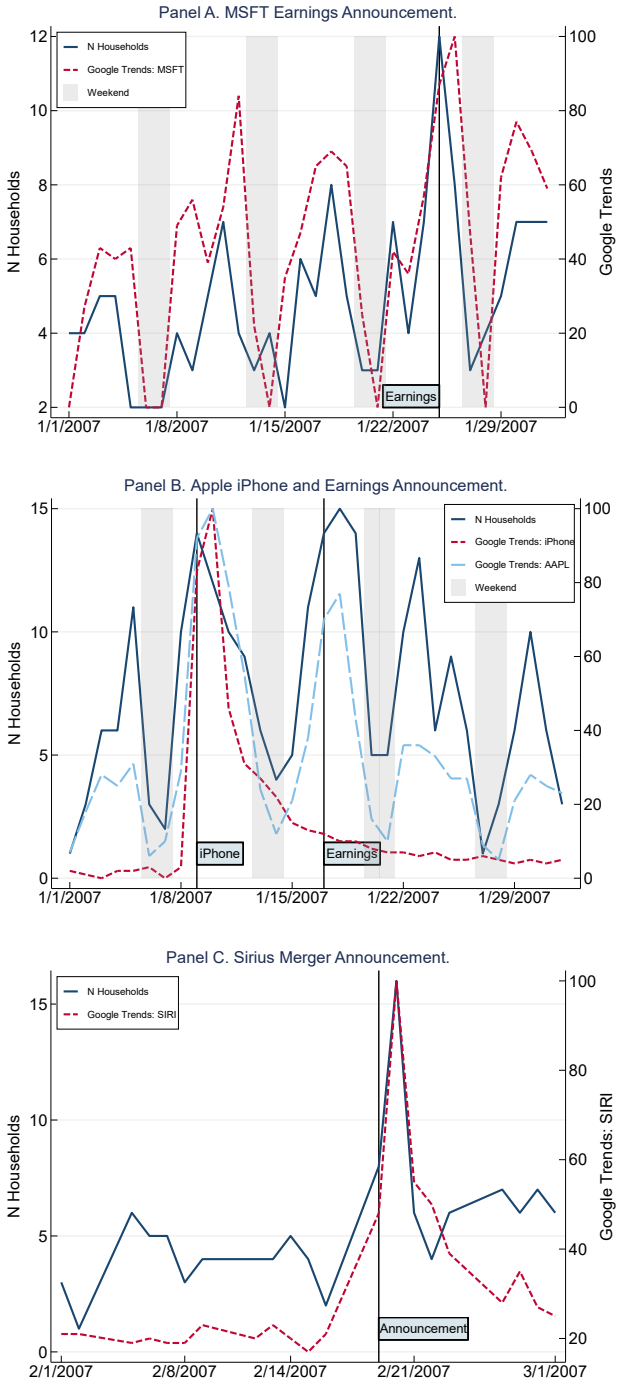


Figure 3. Length of chart lookback windows. Lookback windows viewed by investors. Panel A includes the default one-day lookback presented in the snapshot page (or intraday during trading hours). Panel B excludes such views and considers only opt-in lookback periods.

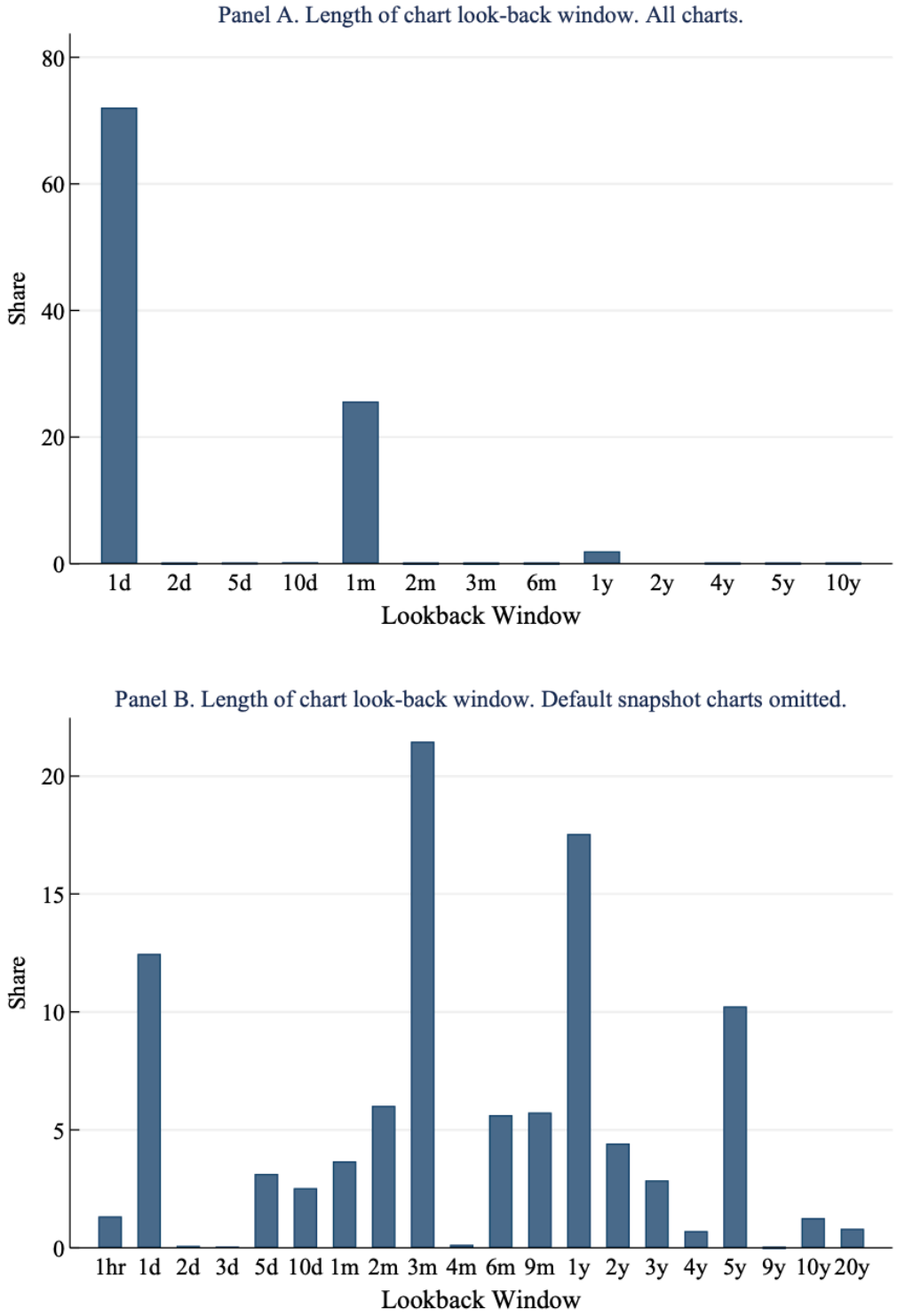
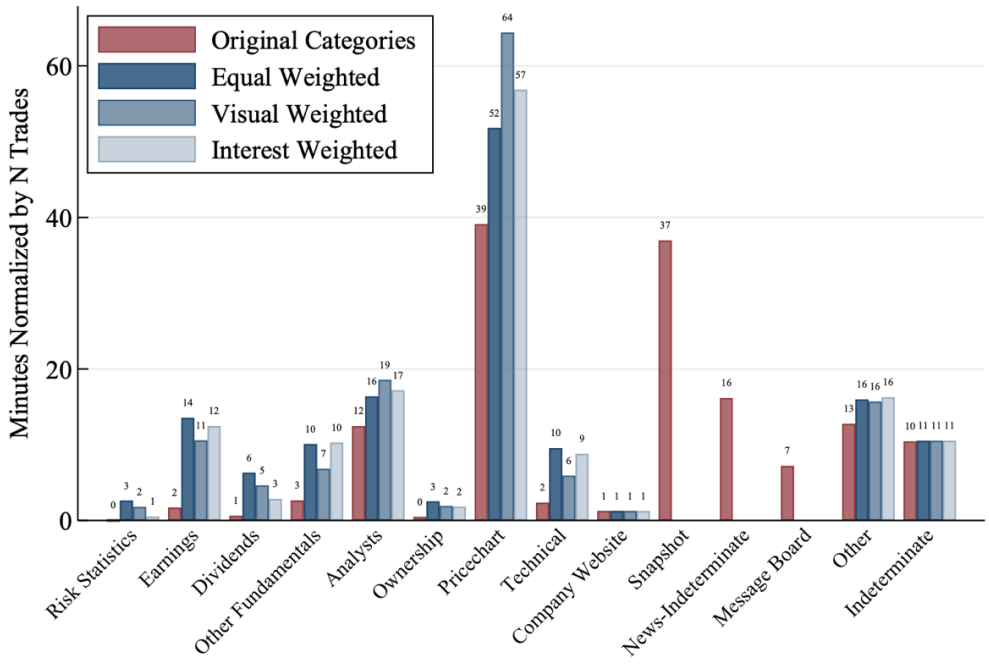
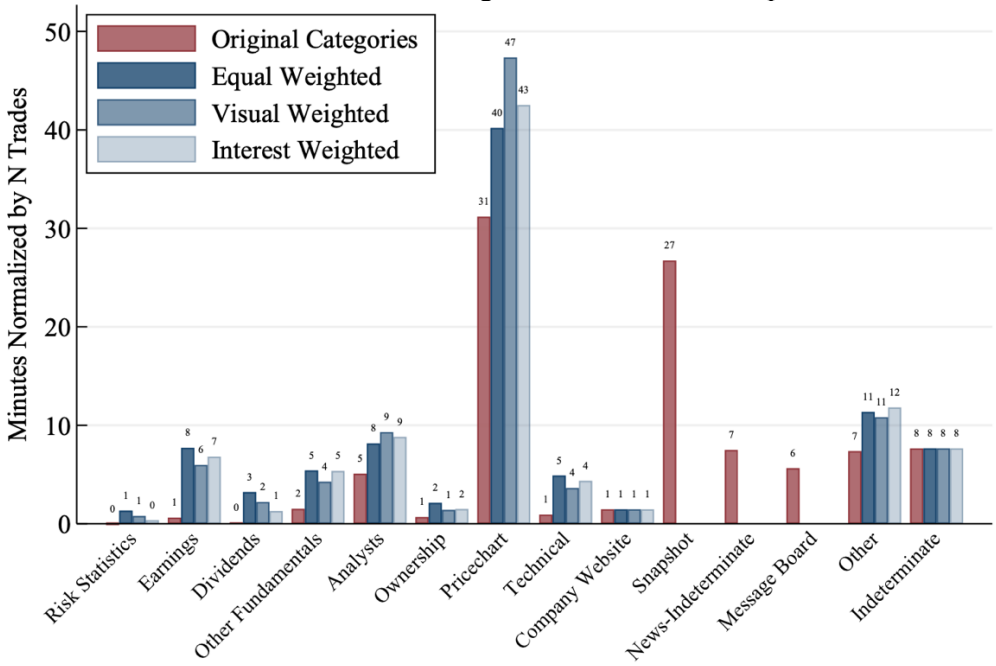


Figure 4. Total research time by category. Mean research time per trade across investors. Panel A shows mean research time per trade for the full sample (N=484), with red bars denoting average time per category and blue bars denoting time after imputing time from the three mixed-content categories to the remaining content categories using one of the imputation methods described in Table 8. Panel B limits the sample to investors with high ticker match rates in their research links (N=50), again tabulating all research, and Panel C limits this sample to research on the tickers that the investors trade.

Panel A. All research, all investor sample



Panel B. All research, high ticker match rate sample



Panel C. Research on traded tickers, high ticker match rate sample

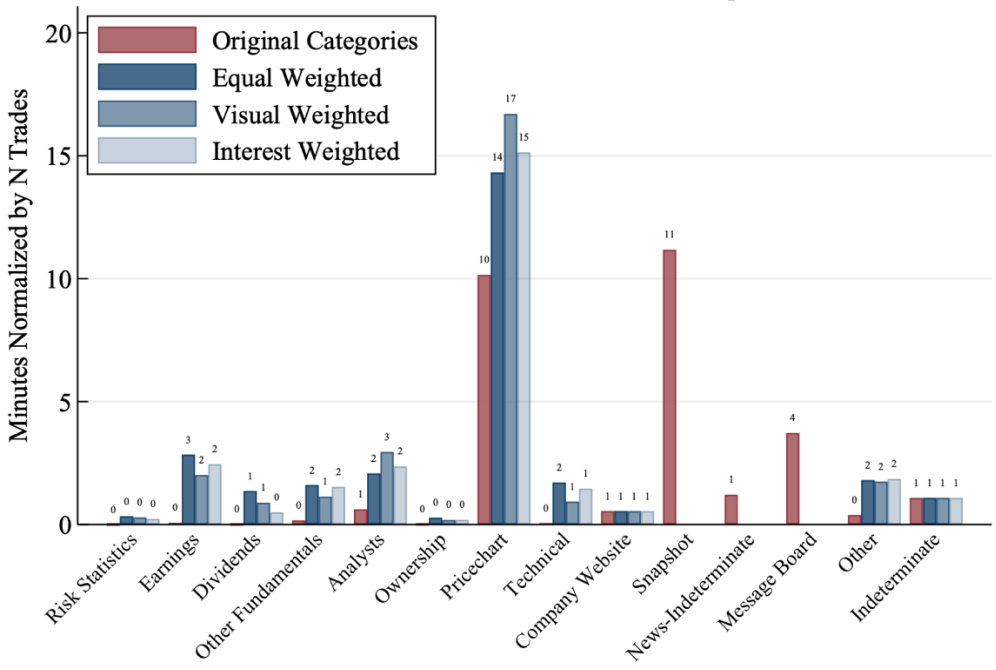
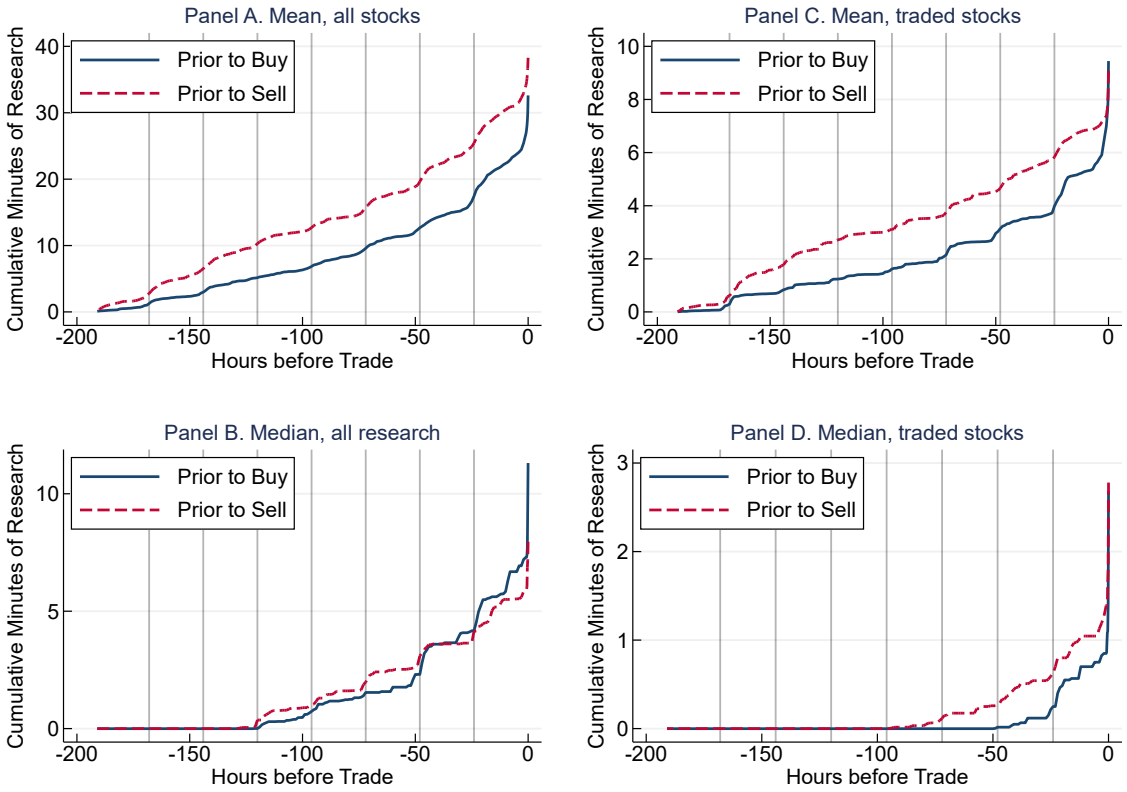


Figure 5. Research in the runup to trades.



Data Appendix

This Data Appendix describes a number of data processing steps. The starting point is clickstream data for a nationally representative sample of households for January, February, March, and June of 2007. Households move in and out of the sample and are present between one and four months. For each household we observe the timestamped sequence of all URLs ("links") that they visited, active time spent at each page, and a rough categorization of the type of content served at the specific page. The raw data prior to any exclusions or focus on investing households comprises roughly two billion links.

I. Constructing trading-household clickstream data

A. Identifying trading households (investors)

We begin with a list of 11 online brokerages operating in 2007 and the associated web domain names. Using the domain names, we read through the entire dataset and create a subset of clicks on broker websites. Within this set of clicks we flag clicks that contain terms potentially indicative of stock trades: we look for terms such as "trade", "order", "buy", "sell", "shares" and so on. For instance, the URL in row 20 of Table 1:

https://[broker1].com/invest/socreateentry?ordertype=sell&s
ymbol=GOOG&shares=100

would be flagged for the inclusion of at least two such terms. Whenever we find a link that might represent a trade, we use regular expression ("regex") matching to find all links on that broker's site that follow the same syntax. We additionally use regular expression matching to read out the stock ticker, the trade action (buy or sell), and, if available, the quantity traded and the trade price. For the above example we would use the regex pattern:

socreateentry\?ordertype=(\w+)\&symbol=(\w+)\&shares=(\d+)

to read out the action, stock ticker, and quantity traded, in that order. From the above entry we would read out "Sell", the ticker GOOG, and the quantity of 100 shares. We match the traded tickers to the CRSP header file and keep trades that that correspond to U.S. common stock and ADRs.

Having identified potential trades, we take a close look at the click sequences to confirm that they represent a stock trade. In many cases, a number of clicks occur in rapid succession pertaining to the same underling ticker. When we see a sequence of potential trade clicks relating to the same ticker and with the same trade action (buy or sell) with a gap of no longer than 30 minutes, we collapse those clicks into one trade, represented by the last click in the sequence.

With this procedure we end up with 2,911 total common stock and ADR trades by 484 distinct trading households.

B. Processing traders' browsing data

Once stock trading households are identified, we set about documenting the information they acquire before and after trading. We read through the raw data once more and keep all the clicks associated with these 484 households. We end up with over 8.5 million individual clicks corresponding to some 60,300 combined hours of browsing.

At this point we carry out two pre-processing steps. First, we check the URLs for special characters and characters that are reported in Unicode values and transform them into corresponding characters. This step makes it easier to process the URLs via regular expression matching. Second, we organize the URLs in groups corresponding to the host site. For instance, all clicks with the stems biz.yahoo.com, finance.yahoo.com, and quote.yahoo.com are assigned to the same group—Yahoo Finance—which is subsequently processed in one shared routine.

The websites we process in this manner are brokerage sites and non-broker finance sites such as Yahoo Finance, CNBC, Marketwatch, and the SEC.

II. Reading out tickers and research categories

A. Processing ticker-matched research

Our first goal is to identify clicks associated with stock tickers. We proceed in an iterative manner: we look at the URLs until we find one that represents research associated with a ticker. For instance, at Yahoo Finance we might come across a URL like in row 10 of Table 1:

http:/finance.yahoo.com/q/ao?s=IMCL

and, with the help of the Wayback Machine at

https://web.archive.org/web/20070407055358/http:/finance.yahoo.com/q/ao?s=IMCL

we learn that on the abbreviation "ao" refers to "analyst opinions". We then use a simple regex pattern to classify all such URLs as pertaining to analyst information and associate them with the corresponding ticker.

In general, in the case of non-broker finance websites such as Yahoo Finance we are often able to find a version of the page on the Wayback Machine, allowing us to directly verify the type of content. Brokerage sites, however, typically require logging in to view the research material, and hence are not captured by the Wayback Machine. However, we were in many cases able to find promotional material or video guides that illustrated the content on various pages. In some cases, however, we have to make the determination based on the URL alone. For instance, a URL might contain the term "analystresearch" and a ticker, in which case it's clear that the investor was looking up analyst reports. In instances where we can deduce nothing about the type of content, we classify the content as "indeterminate".

In the case of broker websites, we typically first allow the structure of the URLs to dictate the original classification of a given link. For instance, we might come across a URL like in row 17 of Table 1:

https://[broker1].com/stocks/snapshot/snapshot.asp?symbol=GOOG

and a regex pattern that reads out the terms separated by forward slashes identifies a slew of different research sites: stock news, fundamentals, ownership and so on. We continue processing URLs in this way until we can no longer find URLs pertaining to research that contain ticker information.

B. Special considerations

A number of clicks require further attention to fully extract the available information. The first such special case involves URLs that reference multiple tickers, such as row 3 in Table 1. In such cases we create duplicates of the underlying link and associate one copy of the link to each of the associated tickers. We then split the time spent on the original link evenly between each of the newly created copies.

Another special case involves URLs corresponding to charts. These often refer to multiple tickers as well as multiple types of information. For instance, a chart might plot the prices and volume of two stocks. In that case, we first proceed as above and split the original URL into copies corresponding to each of the associated tickers. These resulting URLs, in turn, are again split up to correspond to the different types of information. As before, we evenly split the time spent at the original site between all the resulting copies of the URL. Chart URLs often contain further information about the information the investor is seeking out, such as the lookback window and frequency of data plotted. We again use regular expression matching to extract such information. For instance, consider the URL in row 7 of Table 1:

https://[broker1].com/stocks/earnings/update_earningsChart.asp?showQTRChange=1&display=mountain&overlay=&indicator=volume,1,13,EMA,13&duration=1095&daysforward=545&year=2006

This link reveals that the investor was charting earnings over a period of 3 years (1095 days) with a period of 1.5 years (545 days) predicted into the future.

C. Cleaning tickers

Having processed the URLs and extracted ticker information we separately process the ticker information that we read out using regular expression matching. We first read out any terms appended to the ticker, such as country indicators for stocks trading abroad, ".PK" for pink sheets, and so on. We then match the resulting tickers to the CRSP header file. We inspect the unmatched tickers and match by hand, whenever possible. Reasons for unmatched tickers can be caused by stocks with multiple share classes (such as GOOG and GOOGL), or user input errors when searching for a particular entity. On occasion the ticker field instead represents the investor's input of a company name, say "WALLMART". Whenever obvious, we assign such user-inputted fields to the associated ticker, though we do not count the time from clicks involving such misspelled tickers in the research time totals.

We separately create indicators for identifiers that pertain to indexes, mutual funds, and to currencies. Tickers corresponding to indexes and currencies are often indicated by a special character on brokerage sites. For instance, in row 3 of Table 1 indexes are indicated by a leading period, while exchange rates are indicated with a double dollar sign.

D. Processing other activity at brokerage sites

Having exhausted clicks that contain ticker information, we turn to classifying other time spent at broker and finance sites. We first seek to classify research time without explicit tickers in the URL. We follow a process very similar to the one described above and use regular expression patterns to read out any information pertaining to the content of the URL.

Additionally, we classify other activity on broker sites that may reference tickers but does not constitute research. The two main categories that might contain ticker-related information are "portfolio" and "trading": clicks involving reviews of current investment positions (which we do not observe) and clicks involving inputting and confirming trades. We also classify other time spent at brokerage sites: time spent on logging in, time on the main homepage, and time on various types of non-investing activities such as educational tools, bill payments, banking, and so on.

E. Processing other research activity

As a final type of investor research activity, we keep track of time spent on websites of companies that are in the set of stocks otherwise researched by the investor. We use the COMPUSTAT header file to link CRSP Permno identifiers to company website URLs and manually augment the list of company websites the companies in the top 100 in terms of research activity.

The idea behind keeping track of this time is to provide a sense of an investor's familiarity with the products of a given company. For instance, the time an investor browses the newly released first generation iPhones on Apple.com would count as research time, if they ever research AAPL on any of the finance sites. We impose an upper bound on such research time by constraining that at each point it cannot exceed twice the total of other research on the same stock. This limit ensures that the time spent at company websites does not skew research time totals.

F. Assigning unmatched research time

As discussed above, not all research clicks reveal the underlying stock or stocks. For instance, consider the sequence of URLs in rows 10, 11 and 12 of Table 1:

http://finance.yahoo.com/q?s=IMCL

http://biz.yahoo.com/bw/070530/000000000.html?.v=1

http://finance.yahoo.com/q?s=IMCL

Here the investor is looking at the "snapshot" page for ImClone Corporation (IMCL), then reads a Business Wire press release where the ticker is not included in the URL, and then returns to the IMCL snapshot page. While we don't know the content of the press release—and that particular page is not available via the Wayback Machine—we conjecture that the press release likely pertained to IMCL as it was preceded and succeeded by research on that ticker.

Based on the prevalence of such instances, we seek to assign unmatched research clicks to tickers if the nearest ticker-matched research before and after the click in question pertains to the same underlying ticker. We apply this procedure only if the bracketing ticker-matched clicks contain exactly one ticker.

G. Revising processed data

An important final part of our procedure is an iterative process of reviewing the processed data. We do this in three main ways.

Firstly, in order to ensure that we have captured all the URLs with ticker information we search the links for popular tickers that have relatively unique letter sequences. For instance, the most researched ticker in our data is AAPL, and for each of the sites we confirm that the search of the string "AAPL" does not yield in any URLs that contain ticker information but are not yet processed with a corresponding regular expression.

Secondly, we search the links with keywords relating to the main research categories such as "snapshot", "analyst", and "earnings" to ensure that all the relevant links are classified appropriately.

Thirdly, we collapse research time by site and category and compare the percentages across different brokers and non-broker sites. If a category is conspicuously absent in a particular website, we go back and look for evidence of that type of research.

III. Recombining and harmonizing content categories

A. Original research categories

The above procedure results in close to 3,000 hours of total research time, with roughly 1,350 hours matched to any sort of identifier, be it a ticker, a company name, or an index. This data is made up of 30 distinct websites, with the most important, Yahoo Finance, contributing close to 30% percent of total research hours.

As described in Step 1 above, in processing the research clicks at the various websites we typically take the approach of classifying based on the categorizations used on the website itself. Naturally, the categories of data that we obtain in this way are not perfectly aligned across the different websites, though there are numerous broad commonalities across the websites. For instance, all of the brokerage and finance sites see more visits to a standard page best classified as a "snapshot": a ticker-specific summary page of different types of information such as price quotes, a chart of recent prices, brief data on earnings and dividends, and potentially links to news and analyst views. Another type of page present on all sites is a charting page with prices, volume, and technical indicators such as moving averages, candlesticks, and so on.

Beyond these two dominant categories, there are two types of pages that are prevalent across all broker and non-broker finance sites: analyst reports and fundamentals such as earnings

and dividends. Additionally, across different sites we may see pages on risk metrics, technical research, industry analysis, ownership, market events, or other categories.

For the less frequently clicked categories that do not amount to more than a handful distinct clicks, we code the research as "OTHER" (for not elsewhere classified), indicating that we know the type of information the investor sought out but that there is not enough investor interest to warrant inclusion in the tables. For pages where we can't tell the content we code it as "indeterminate". There are two types of pages where the content is often not obvious from the link and not captured by the Wayback Machine and hence classified as "indeterminate": news and messageboards.

B. Constructed categories

In addition to the categories suggested by the finance website, we define the following three research categories.

Company website refers to time spent at a URL associated with a company that the investor has researched in any of brokerage or finance sites and we impose a running constraint that companywebsite research can at no point add up to more than 2x of the research on the associated ticker.

Markets refers to time spent on snapshot pages where the underlying security is a stock market index, or on other pages pertaining to market-wide developments.

Businessconditions refers to time spent on pages relating to aggregate business conditions. We construct this category by looking for keywords associated with inflation, unemployment, GDP, the FOMC. We additionally add time spent at the BLS and BEA websites to this category.

C. Reassigning research time to primary categories

The approach described so far seeks to best classify each URL an investor visits. Some of these pages are amalgamations of many types of underlying primary information, while others are quite specific in their nature.

In order to attempt a full accounting of the types of information presented to investors, we construct an alternative classification where we break down all research time into primary categories. To do this we need to estimate the breakdown of underlying content categories on three main types of pages that contain a mixture of information, or where we typically don't observe the type of content: snapshots, news, and message boards.

To break down the information content of snapshots, we seek out contemporaneous screenshots of the page, or, in the case of non-broker sites, we look them up on the Wayback Machine. We then measure the relative amount of line items corresponding to each primary category. We ignore redundant information (e.g., when dividends per share are presented in addition to yields), information presented without meaningful context (e.g., daily volume presented without some notion of average volume), and other information that seems unlikely to affect individual investor behavior.

For news clicks, we typically do not observe the news article. However, for two brokers we observe in the URL the article title and source (Reuters, WSJ, etc.), as well as any associated tickers. We extract the title, tickers, and the date, and match each click to an array of articles in the RavenPack data using the ticker and a date range of three days. We then rank this set of potential matches by measuring the distance between the two titles and keeping the closest, subject to a minimum distance requirement. The resulting dataset of matched articles allows us to estimate the breakdown of news article content using the detailed content classifications provided by RavenPack.

We follow a similar procedure to estimate the content breakdown of message board discussions. We scrape the present-day Reddit r/stocks message board and for each discussion thread match the title, "flair" (user-supplied keywords), and original post content to our research content categories. We then back out content category breakdowns by ascribing equal time to each post, dividing time equally across all matched categories.