The Research Behavior of Individual Investors*

Toomas Laarits
NYU Stern School of Business

Jeffrey Wurgler
NYU Stern School of Business and NBER

December 6, 2025

Abstract

Browser data from an approximately representative sample of individual investors offers a detailed account of their search for information, including how much time they spend on stock research, which stocks they research, what types of information they seek, and when they gather information relative to events and trades. The median individual investor spends approximately six minutes on research per trade on traded tickers, mostly just before the trade; the mean spends around half an hour. Individual investors spend the most time reviewing price charts, followed by analyst opinions; they exhibit little interest in traditional risk statistics. Aggregate research interest is highly correlated with stock size, and salient news and earnings announcements draw more attention. Individual investors have different research styles, and those that focus on short-term information are more likely to trade more speculative stocks.

* We thank Valerie Baldinger, Myeongrok Doh, and Cody Wan for outstanding research assistance and Fulin Li, Terry Odean, Raghu Rau, Noah Stoffman (discussant), Paul Tetlock, Edward Tufte, and participants at NYU Stern, the University of Texas at Austin, Texas A&M University, 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 range of assumptions about what investors know or what they think they know. Classic models assume interest in, and knowledge of, variances, covariances, and risk premia, e.g., Sharpe (1964); Merton (1987) highlights limits on attention to risk-return statistics. De Long, Shleifer, Summers, and Waldmann (1990) assume that some investors have information about fundamental values while others, generally individual investors, trade on noise. Hong and Stein (1999) assume that some investors follow the news and others react to price patterns. Microstructure models and limited attention models consider still other information setups. In addition to being fundamental to modeling investor beliefs, "information set" concepts are central to empirical work: Countless studies have tested how prices reflect the textbook breakdown of information into past prices, public information, and public as well as private information.

This diversity of approaches is understandable not just because models and tests are designed to illustrate different mechanisms, but because there are relatively few comprehensive facts about the information that investors gather. In this paper, we use browser history data on an approximately representative sample of U.S. individual investors to address a set of first-order 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? And, perhaps most importantly, what types of information do individual investors care about—and what do they ignore?

These questions can be addressed through browser history data – also known as clickstream data – because URL addresses often include details about the page visited. Gargano and Rossi (2018) were the first to exploit this type of data in finance. We are able to build on and substantially extend their work because the data we use here have a number of crucial qualities. First, it contains far more detail. Second, it is comprehensive of the sample investors' browsing activity, not being limited to activity at a single brokerage domain, which allows us to observe an investors' *entire* online information diet for investors whose research is performed on any device using the household's Internet connection. Third, the sample is balanced by the data provider against U.S. households at the time of the sample, allowing for reasonable albeit (as always) tentative generalization to the population of individual investors.

The raw data for the households that we focus on—those that trade any individual U.S. stocks or ADRs within an online brokerage account—include over 8 million clicks and 60,000 hours of Internet use in four months of 2007.[1] We identify 484 household-traders and they make 2,911 stock trades over the course of the sample. For conciseness, and following prior literature, we generally refer to household units that trade stocks as individual investors.

Our main conclusions include the following:

- The subsample that provides the best estimates indicates that the median individual investor spends six minutes per trade on research specifically about the ticker(s) traded. The mean individual investor spends 29 minutes. The median investor conducts most of this stock research in the 24 hours before a trade, and most of that time in a burst immediately before the trade.
- Typical investors spend only a fraction of their research time at their broker's domain. Disparate finance news sites, especially Yahoo Finance, constitute a majority of stock-related research.

[1] The data provide insights into current practices by individual investors because online research and online trading was the norm as of this time. 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 trade by 2007 would likely have been less active, less sophisticated, and less economically important.

- Market capitalization is the most important cross-sectional determinant of individual investor research interest. Salient news, such as Apple's introduction of its iPhone, can temporarily launch a smaller stock to the highest ranks of aggregate attention. Individual investors also pay more attention around earnings announcements.
- 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. Many investors do not pursue research beyond this page.
- When individual investors do go beyond the snapshot page, they spend by far the most time on price charts and price-related information. Analysts' estimates are consulted less frequently, followed by assorted other fundamental and technical information. Risk statistics such as beta or volatility are of little apparent interest.
- Individual investors have different research styles. Some spend more time on research, and those 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. Unexplained heterogeneity is considerable.
- The data confirm that Google Trends accurately measures variation in individual investor interest. This addresses concerns in prior research that it may be too noisy due to its inclusion of all Google users; it also speaks to the representativeness of our own sample in light of the massive random sample underlying Google data.

These results provide new facts about individual investor beliefs and behavior, complementing those from, for example, Lease, Lewellen, and Scharblaum (1974), Sicherman, Loewenstein, Seppi, and Utkus (2016), and Gargano and Rossi. Lease et al. study individual investor demographics, investment strategies, and sources of information from a survey of a retail brokerage's clients. Sicherman et al. study online account logins and trading activity as a function of market conditions and investor characteristics. Gargano and Rossi's innovative paper, the closest to ours, analyzes clickstream data from 2013-2014 from an online brokerage and documents time spent across broad activities (research, trading, balances etc.) and explore drivers of investor attention. We are able to compare some of their results with our own.

Clickstream data is a direct measure of investor attention, so our results also bear on that large literature. It includes Barber and Odean (2008), who show that stocks with news or extreme daily returns appear to capture the attention of retail investors. Kaniel, Saar, and Titman (2008) and Barber, Odean, and Zhu (2008) find retail trader contrarianism against recent returns. Tetlock (2010) shows the impact of news on return reversal and volume-induced momentum. Da, Engelberg, and Gao (2015) find that aggregate Google search volume about economic downturns predicts market volatility and fund flows. Subsequent work has sought to further disaggregate types of information commanding attention, such as Kwan et al. (2025), who study online news article accesses and relate them to portfolio allocations of institutional investors.

As others have pointed out, one challenge for this literature is going beyond associations and, at an investor level, being able to observe the full line from attention to action. The literature is also piecemeal in terms of focusing on particular determinants of attention. If the aggregate information set of individual investors was a pot of alphabet soup, research on investor attention often attempts to establish the existence of particular letters in this soup. Our data allow us to estimate the full distribution of letters in the aggregate pot, directly observe the distribution within 484 particular bowls, and also directly observe trading activity. In this way the analysis supports various of the associations observed in the attention literature.

More generally, our results contribute to asset pricing modelling, especially in behavioral finance. As noted above, finance models make the full gamut of assumptions about investor information sets. The paper contributes facts that will be useful when it is important to be accurate about what information interests individual investors.

Section II describes the data and the sample of investors. 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 are of most interest. Section VIII explores heterogeneity in research approaches. Section IX concludes and comments on future directions, such as connecting research behavior to performance and portfolio formation.

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. The data was made available by an online research company and was used in an academic 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., the panel of households, recruited using random-digit-dialing distributions, installed a plug-in that collected the timing and sequence of URLs visited. All computers within the household had their browser data gathered and aggregated in this manner; the timing of the sample happened to coincide with the announcement but not the widespread use of Apple's iPhone. Confidential or personally identifiable data, such as account numbers, addresses, passwords, and the like, have been removed by the data provider. To date, only a few studies have exploited clickstream data in finance research, including 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, where the context allows, we refer to these household units as individual investors, following Barber and Odean (2002) and others. Subject to remaining selection bias and sampling noise, the data appear to offer a reasonably representative picture of U.S. individual investors trading online at the time—young and old, wealthy and less so, active and less active, and geographically balanced.

A stylized extract of raw data illustrates the level of detail that 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

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

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 the latest analysts' estimates ("ae").

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 Internet Appendix, this example illustrates how the data can provide rich detail about the research process of individual investors. This level of detail is not present in Gargano and Rossi's (2018) data, which indicates only that the investor in question is doing some research, but not what type. Accordingly, 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 click away.

It is worth discussing what links do and do not reveal. By definition, 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 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 that information for future trades. 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, but it turns out that is often what we observe.

These limitations are more significant when it comes to soft information about products, corporate reputation, and word of mouth from friends and family. "Apple makes cool products, and a growing number of young people seem to be using them" is an abstract intuition but 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. One 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 always indicate a lack of understanding or a lack of interest. Dividend information will take only a moment to find and review and, because dividends are a stable characteristic, will be relevant for some time. Changes in analyst opinion, on the other hand, take more time to find and understand and also have a short-lived investment relevance. In light of these considerations, it is useful to tabulate not just the average amount of time investors spend on an item of information, but the fraction of investors that are 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 Internet 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 are balanced to have a representative distribution of other 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. It spends 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 (2,239 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 skewed. The median household-investor spends 1.4 hours per month at brokerage sites and the 95th percentile spends close to fifteen hours.

Panel C summarizes trading activity in individual stocks that we observe; 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 investor trades more than nine times per month. Roughly speaking, this activity resembles the activity in Gargano and Rossi's sample. 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 CRSP and TAQ data). The average trading-related part 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 to a 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) analyst opinions, (6) informed ownership and trading, e.g., by insiders, funds, 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

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

but lacks the detail to further categorize. A number of links are to pages that present a mixture of content. Such links are handled separately as described below.

We focus first on aggregate totals of research time before breaking research interest down by content category. Documenting how individual investors distribute attention across these categories is a unique opportunity provided by the detail in the data.

B. Content versus format

The content categories we track are defined by content, not presentation format. Earnings information can be conveyed through a bar chart, a news article, or a text press release. Past prices and returns also represent content, while a chart or a table of prices represents different formats. Although links sometimes indicate the presentation format, and the manner in which information is presented might impact perception through behavioral salience and framing effects, in this paper we regard the content itself as the first-order input for investment decisions and the unit of increment to the information set.

C. Total research time

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

Panel A reports that investing households spend anywhere between one to four months in the sample, with three months being the median value. 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 investor carries out close to 30 such sessions a month that have at least some research component, and in the course of these research sessions the average investor sees 30 unique tickers and visits 3.3 distinct brokerage or other finance sites. We construct these count measures on the investor-month level, average to the household level, and report the statistics of the resulting distribution of N=484 values. The median investor conducts about 15 research sessions per month and comes across 12 unique tickers. The average browsing session with a research component includes 3.6 minutes of research. Over the course of the month, the average household carries out about 118 minutes of research, with the median at 37.1 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.) We did not find a strong relationship across households between the average amount traded and the average research time per trade, which argues for giving attention to median behavior.

There are a few similar estimates in the modern literature to compare to these. One is Lease et al.'s (1975) survey of "long-term" customers of a "large national retail brokerage" from 1964-1970. The response rate to their survey was 40%. The median respondent reported spending three to five hours per month "in investment analysis and decision-making for your securities portfolio" and a mean of nine hours per month. In light of the characteristics of survey respondents, one expects a bias toward active investors, so it is less surprising that this amount of time per month would put the typical survey respondents in the highest percentiles of research per month in our sample.

It is also possible to back out an estimate of average time spent on broker-site research from Gargano and Rossi (2018). We multiply their sample mean of 381 hours per weekday spent on "research" and divide by 11,000 investors in the sample, leading to a mean of 2.1 minutes per weekday of broker-site research. We estimate a mean of 1.8 minutes under the same restrictions (i.e., limiting the tabulated research to the investor's broker's website). In the next section, however, we will see that broker-site research represents less than half of the typical investor's overall research time, so these estimates of research time also represent less than half of total research activity across all web domains.

Although we focus on the activity of traders in this paper, this tabulation of the amount of research per month can also be benchmarked against investors in our data that do at least a few seconds of activity that we classify as stock-investing research but do not trade even once within the sample period. There are 9,042 such investors; they conduct a mean of 27.6 minutes of research per month (vs. 118 for traders) and a median of 3.8 minutes per month (vs. 37.1 for traders). Thus, not surprisingly, traders do several times more research than nontraders.

Finally, we report total research *per trade*: the average investor 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 dimension—research per trade—as the subsequent category-level analysis frequently employs this normalization to give a sense of how much incremental research is reflected in a typical trade by an individual investor. By every measure, though, research efforts are notably right skewed.

In Panel C we break down the per-trade research numbers further. Close to half of the 144 minutes of research per trade can be matched to some sort of identifier, be it a stock ticker, an index identifier, or any other sort of information about the security or series 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 stock 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. See Gargano and Rossi for an extensive breakdown of non-research activities at a brokerage website.

IV. Which websites are used for stock research?

The next broad question the data can answer is where online stock research is done. The event study literature documents that marginal investors react to information of various types, but it is typically unclear where they obtain that information. Gargano and Rossi's data are from a single brokerage; Table 4 tabulates how long individual investors spend on different domains, including but not limited to a brokerage. 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. We anonymize most domains to protect the intellectual property of the data provider, but we can show the distribution of research time across every relevant individual domain.

Yahoo Finance is by far the most popular finance research site. We can unmask it because it has previously been reported as such.[4] The site reports financial data, news, analyst

[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/.

opinions, filings, and other information in a standardized format, and it does not require an account to use. 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 7 minutes per trade. Attesting to the prominence of Yahoo Finance, Lawrence, Ryans, Sun, and Laptev (2018) conduct a field experiment and show that earnings announcement articles promoted to a mere one percent of Yahoo Finance users is enough to have observable impacts on abnormal returns and bid-ask spreads.

The remaining sets of columns shows that slightly more than half of the total time on Yahoo Finance is matched to specific stock tickers, and the third columns include only research links matched to tickers that were traded by that investor 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 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 investors 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 is far less. Many investors spend a minute or less time on broker-site research per trade. Although nearly every investor is doing at least a few seconds of broker-site research over the four-month sample, much of this is likely to 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 page 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 interest in SEC.gov per se because visitors were typically directed by a link from another domain, such as a brokerage or 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. Recall that we can 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 that these limitations may not be consequential: Among 484 investors who perform 2,911 trades, only a single investor reviewed information from the SEC website on a CRSP stock that he traded within 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). As described more fully in the Internet Appendix, we impose the requirement that a visit to a corporate website qualifies as "research" only when the same household-investor 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?

We enumerate the most-researched stocks in our sample in Table 5, which is inspired by a similar list in Gargano and Rossi. At the top of our list is Apple, Inc., 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. Again, this is not simply counting households that peruse Apple's corporate website for new laptop specs or replacement cables and adding that time to an investor's research activity; as described above, we do not include time spent at a company website as research unless the same household-investor also studied the company from a more clearly investment-relevant perspective.

As a crude assessment of the persistence of aggregate research attention, we can compare the list of most-researched firms in the 2007 sample with the most-researched firms in Gargano and Rossi's 2013 sample. As mentioned before, our sample is more comprehensive in terms of types of research, and we also measure research somewhat differently. In any case, six stocks appear on both lists: Apple, Microsoft, General Electric, Sirius XM, Ford, and AT&T. By contrast, 15 out of the 20 highest market cap rank stocks in January 2007 remained in the top 20 cap as of January 2013. There is measurement error in research rankings but not in cap, but nevertheless we assess that relative market cap is more stable than relative research attention.

Returning to our sample, Figure 1 shows 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 in general a strong correlation between breadth of interest and market cap. Microsoft 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, while other large stocks, including oil and financial institutions, tend to attract less research attention relative to their size. Perhaps investors consider that the major driver of oil company profits is oil prices, or perhaps research on financial institutions seems complex. In retrospect, AIG and others involved in real estate lending could have used more scrutiny at this time, the eve of the global financial crisis.

Table 6 uses regressions to relate the overall breadth of research interest and intensity as a function of stock characteristics, the volume of news articles, and other factors, which also follows Gargano and Rossi.[5] We consider only stocks that receive any research interest by any investor in the sample. (This is a relatively modest screen, since 80% of stocks in the CRSP universe are present for the regressions even upon requiring all independent variables as well.) The first several columns focus on the overall percentage of investors with nonzero research time in that ticker-month; the last column uses the log of the total number of hours that sample investors 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.[6] 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 investors conducting some research within the sample, from an unconditional average of .50% up to 1.03% (=.50%+.53%). The second specification shows that the inclusion of Fama-French 49 industry dummies increases explanatory power only marginally, to 23%.

It is not hard to imagine many reasons why large stocks receive more research attention in aggregate. For example, based on print media article counts, Fang and Peress (2009) find that "firm size has an overwhelming effect on media coverage" (p. 2030). Van Nieuwerburgh and Veldkamp (2009) predict that investors should pay more attention to higher-cap stocks. Liquidity, ETF and index inclusion, information availability, investment constraints, familiarity

[5] Balasubramaniam, Campbell, Ramadorai, and Ranish (2023) offer a detailed study of stock characteristics in Indian retail portfolios, but inferences about the information that individual investors inject into prices are only implicit; we focus on the complementary question of what characteristics attract research interest.
[6] Gargano and Rossi use a linear cap rather than log market cap value. We speculate this is why they do not find as strong a relationship with market cap.

bias, basic equilibrium considerations—indeed, it is hard to think of a mechanism that would not predict a strong positive relationship between aggregate investor attention and market cap.

The third specification includes two clearly endogenous factors—the retail share of trading volume, as estimated by Boehmer, Jones, Zhang, and Zhang (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 again seem 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. Drake, Roulstone, and Thornock (2012) observe an elevated level of Google search volume around earnings announcements. In our data, both news factors are statistically significant, but do not much increase overall explanatory power.

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 interest include low nominal share price, high volatility, no dividends, high growth (as measured by sales growth or external finance over assets), number of analysts covering the stock (as in Gargano and Rossi), and S&P 500 Index membership.[7] Although neither share price nor S&P

[7] 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).

500 membership is directly related to stock fundamentals, they have particularly strong associations with research interest. In general, these relationships are intuitive in light of prior research 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, closest to Gargano and Rossi. 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. Overall, the same factors and characteristics drive this measure of research interest as well.

VI. When is research conducted? Research around events and 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 5, 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 engine activity, is excluded from our research time tabulations.

Researchers have increasingly used Google Trends as a proxy for investor attention. Da, Engelberg, and Gao (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

Google 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. As mentioned, Drake et al. 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 at the level of individual investors tell us? 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 prior authors using Google search volume. In both data sets, breadth of interest peaks around the quarterly earnings announcement. This indicates that our sample can capture general investment-related interest despite being smaller by orders of magnitude than the Google Trends random sample of a large fraction of all search activity. Also, in the other direction, our 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.[8] This finding is relevant for other research 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

[8] For the same reason, one cannot 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.

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, while clickstream data have the advantage that one can identify and isolate investors who do other investment-related research on Apple from those who are simply interested in a new 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 fueling attention to corporate events. Here again, investor interest based on browser data supports the use of Google Trends and, in concert with this particular episode, Drake et al. find that acquisition announcements are particularly strong events in term of boosting Google search volume for a ticker symbol.

B. Research in the run-up to a trade

The temporal pattern of overall research relative to actual trading is another aspect of individual investor behavior that our 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.

Conceptually, 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 time in the sample. If so, one can simply cumulate up research time related to the traded ticker, as well as untraded tickers if desired, and report the timing of the research relative to the trade at t=0. But in more general cases in which an investor makes multiple trades in the same ticker over time, or trades in multiple tickers in short order following a common burst of research, complexities arise.

In Figure 4, we allocate research time in a traded ticker to the next trade in that stock; if no further trade occurs, we count research time as "post-trade" research. With research effort classified as a distance to next trade in the given stock, we cumulate up time for each investor-trade. In other words, cumulation of time for any given investor-trade in a specific stock starts either at the beginning of the sample, or immediately after a trade in the stock. In the cases where the investor trades multiple times over the course of our sample, we create an average cumulation of research time per trade. We plot the results for mean (and median) cumulative research time associated to any trade as well as buys and sells separately. We further limit the sample to the 50 investors whose links in the research-related clickstream can be fully or almost fully associated with specific tickers, as a result of the detail contained in the URLs of the domains they consulted (domains provide different degrees of detail in their URLs).

The left panel on Figure 4 plots the mean research time in the ticker traded at t=0 from eight days prior to the trade to two days after the trade. Each vertical bar marks a 24-hour period, and the periodic flat spots reflect afterhours pauses in research activity. Note that cumulative research at eight days prior to the trade is already different from zero, reflecting the cumulation starting at the beginning of the sample or after the last trade in the stock. We see a substantial 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 about five minutes of the total 20 minutes of pre-trade research takes place in the hours immediately prior to trade, while half of the overall pre-trade research takes place in the week prior to trade. Research in buys is more concentrated in the immediate runup while research on sells is more spread out. This may relate to the fact that research on sells involves stocks that are already in the portfolio, but that is not the case for all buys.

The median behavior in the right panel of Figure 4 is different. The median investor 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 last ten minutes prior to the trade. Comparing the median and the mean behavior per trade implies that research time is skewed, and a large number of trades are preceded by only a few minutes of ticker-specific research.

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

We next break down the research totals into types of underlying information observed, another unique opportunity afforded by the data. 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.

At this level of analysis, two samples are particularly interesting. The full sample of 484 investors give us the best estimates of total time research breakdowns across categories. We are also interested in a subsample of 50 investors for which we are able to attach a very high fraction of the investors' links to particular ticker symbols. Based on our investigation, these investors are not different from the full sample in any observable way—in particular, they trade with similar frequency, perform similar levels of overall research, and have similar age and income demographics. But their data provide the best estimates of research activity with respect to the specific stocks that the investor actually trades—in effect, a measure of the information that the average individual investor's trades "inject" into prices.

A. Mixed-content pages

The first complexity to address is that some research pages present multiple types of content. For example, investors frequently access detailed quote pages, often referred to as "snapshots" by their broker site or Yahoo Finance, which contain a variety of information—earnings, price charts, dividends, and so on. Without eye-tracking software, we cannot observe attention to the different content types within such pages. News articles of indeterminate content and time spent on message boards present similar complexities. Although snapshots, news, and message boards are objective characterizations of a dimension of research, they are not specific types of content. After reviewing time spent on such pages, we discuss a methodology for attributing time to specific content categories.

A.1. Snapshots

Snapshot pages, or detailed quote pages, are familiar. They often serve as starting points for deeper research through other links and tabs, and other times they may constitute the entirety of an investor's research time. 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.

Brokerage and finance research domains have largely converged on a format for the snapshot page, but the mix underlying content represented varies somewhat. 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 years 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.

In Table 7, which repays careful study, we report means and medians of snapshot research time across all 484 investors (in the left columns) or the high ticker match rate sample of 50 investors (in the right columns), the percentage of those investors spending any time on snapshots, and means and medians of research time conditional on the investor spending some time on snapshots.

The average investor in the full sample spends 37 minutes per trade on snapshot pages. Nearly all investors (93%) spend some time on such a page, but the median investor spends only 6.5 minutes per trade. The mean investor in the high ticker match rate sample exhibits similar behavior. Since investors may click on several stocks' snapshot pages around any given trade, the amount of time spent on the snapshot of a stock for which we actually observe a trade in the sample is lower. Referencing the last columns, our best estimate is that the average (median) investor spends 11.2 minutes per trade (1.2 minutes) on snapshots of the traded stocks.

A.2. News and message boards

We often observe that an investor is scrolling through a stock-related message board, but we do not know the content itself.[9] We treat the message board as a format that reflects a mix of underlying content, and in this way resemble snapshots. News links present a similar challenge. The link may indicate that the article is about Intel but not whether it involves an earnings

[9] Twitter—now X—had just begun operations in mid-2006. It plays no role in individual investor information during our sample as its usage was miniscule. See: https://www.internetlivestats.com/twitter-statistics/. Blogging was also in its infancy at the time of our sample. See Cookson, Mullins, and Niessner (2024) and Cookson, Lu, Mullins, and Niessner (2024) for reviews of social media outlets for financial information.

announcement or an analyst interview. In such cases we label the links as "Message Board" and "News – Indeterminate." See also Kwan et al. (2025) 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. News of indeterminate nature occupies a median of 2.6 minutes per trade, and message boards are used by only a fifth of investors; among these, a few spend a good deal of time, but even the median message-board user spends only a minute per trade. The last columns show that the majority of news pertains to stocks that are not actually traded; the average investor who trades a certain ticker spends 1.2 minutes on news about that ticker but of otherwise unknown content.

B. Imputing mixed-content links to specific content

In an attempt to provide a full, albeit somewhat noisy, accounting of the actual *content* of information presented to investors, we impute research time on these mixed-content pages to specific types of content. The method for which we report results in Table 7 is described more fully in the Internet Appendix, but the basic approach is as follows. We allocate snapshot time to specific categories of content based on investor interest in those specific categories when such attention can be observed, as reported in the third column of Table 7. We allocate news of otherwise indeterminate type to specific categories using breakdowns of present-day news articles constructed by RavenPack. And we allocate message board content using the distribution of content that we find on the present-day Reddit r/stocks message board. See Antweiler and Frank (2004) for an analysis of the content of message board postings and Neuhierl, Scherbina, and Schlusche (2013) for a detailed breakdown of the content of corporate press releases; it is not possible for us to capture this information, link-by-link, in such a granular manner.

C. Specific content

C.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).

The data indicate that few investors display special interest in standard risk statistics. Only seven out of 484 investors demonstrated specific interest in beta or volatility, and only a few drilled down into risk information on a ticker that they traded. This is not quite fair, because risk statistics are sometimes presented as line items in snapshots. Upon imputing snapshot time to specific content types, our best estimate is that the mean investor spends 1.8 of her 144.2 minutes per trade of total research on risk statistics; the median investor spends a fraction of a minute on risk statistics in total and essentially no time at all on risk statistics of the stocks that she actually trades.

C.2. Dividends, earnings, and other fundamentals

Investors 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 investors seek out some sort of earnings information, with just under two minutes of research for the mean investor. Dividends attract at least some direct attention by 17% of investors, 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 investors—53%—explicitly observe some non-earnings, non-dividend fundamental information. This disparate category includes valuation ratios such as price to sales, annual reports, cash flow statistics, 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. Upon adding this imputed time, the mean time spent on earnings, dividends, and other fundamentals rises considerably, but median behavior changes much less.

C.3. Analysts

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

C.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 the average individual investor might (or should) suspect to have comparatively superior information. The median investor who ever consults this information does so for under a minute per month. This does not necessarily indicate disinterest, since as mentioned above it would not take long to read the percentage of institutional ownership or short interest outstanding; it is more telling that only about a quarter of individual investors actively seek out such information. 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.

C.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 investor spends over 39 minutes per trade on price chart pages and another 37 minutes on snapshot pages. Nearly all investors do some of this research: both of the categories see 93% of investors 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 given that snapshots are the predominant type of research, over 70% of observed charts 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 investors 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 investors who carry out any such research.

C.6. 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 investor-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 limit the running total of company website research time to 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 investor-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 by investors at company websites is modest, averaging only 1 minute per trade.

C.7. 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.

D. Summing up

Table 7 provides new details about the interest of individual investors in different types of information. The first-order takeaways are that investors spend a disproportionate amount on price charts. Much of the remaining information comes from the contents of the snapshot page. Overall, our best estimate is that the mean (median) investor spends around 144.2 (minutes (36.4 minutes) per trade on stock research on any ticker, fund, option, or other investment instrument, and 29.2 minutes (5.7 minutes) on stock research on the particular ticker that he trades. Thus, the median trade by individual investors is associated with about six minutes of observable, ticker-specific research. As indicated in the previous section, much of this information is reviewed just prior to the trade.

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.[10] 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 8 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,

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

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.

Beyond these two components, there is a large amount of unexplained heterogeneity in research behavior. This helps to explain the survey evidence in Giglio, Maggiori, Stroebel, and Utkus (2021) that wealthy retail investors' "beliefs are mostly characterized by large and persistent individual heterogeneity" (p. 1481). Table 8 is a first step, but there is much more to be done to connect the search for information, the formation of beliefs, the stocks of interest, and the style of trading.

IX. Conclusions

This paper documents a number of facts about the research interests of individual investors. These facts help flesh 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 facts about how much stock-specific research is behind the median investor's trade, which stocks are most researched, which types of information attract most attention, and heterogeneity in research approaches. These facts inform theory and empirical research on individual investor behavior.

There are natural follow-up questions that can be answered with detailed browser data. Perhaps the most obvious is whether individual investors' research behavior is associated with their performance. The contemporary trading environment does give some cause for skepticism. A contemporaneous and certainly a modern quantitative fund manager spends years and millions of dollars refining algorithms—and then execute trades, given fresh data, within minutes or seconds. And a traditional active manager might employ a number of industry specialists and spend days contemplating a large trade. It is rather optimistic to think that the average individual investor gains an edge based on their own occasional minutes of research, especially if focused on a stock like Apple that is simultaneously being watched by hundreds of thousands if not millions of other individual investors. When it comes to the typical individual investor, it true that "more research is needed"? If so, of what type or style? Gargano and Rossi take some first steps here using their brokerage's clickstream data.

A question of at least equal importance is how an individual investor gets from what she *observes* to what she *trades*. If she sees a particular price pattern in a chart, or certain statistics in a table, how does that affect the trading decision? Since information feeds so directly into beliefs and trading, browser data can shed new light why trades are made and how the portfolios that influence investor wealth are formed.

References

Antweiler, Werner, and Murray Z,. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *The Journal of Finance* 59(3), 1259-94.

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

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.

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.

Cookson, J. Anthony, William Mullins, and Marina Niessner, 2024, Social media and finance. *Oxford Research Encyclopedia of Economics and Finance.* Forthcoming.

Cookson, J. Anthony, Runjing Lu, William Mullins, and Marina Niessner, 2024. The social signal, *Journal of Financial Economics* 158, 1038-70.

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, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *The Journal of Finance* 64, 2023-52.

Gabaix, X., Koijen, R.S.J., Mainardi, F., Oh, S., and M. Yogo, 2023, Asset demand of U.S. households, NBER working paper.

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, Ron, Saar, Gideon and Sheridan Titman, 2008. Individual investor trading and stock returns. *The Journal of Finance* 63(1), 273-310.

Kwan, Alan, Liu, Yukun, and Ben Matthies, 2025, Institutional investor attention, *The Journal of Finance* (forthcoming).

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 field 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.

Marotta-Wurgler, Florencia, 2011, 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.

Neuhierl, Andreas, Anna Scherbina, and Bernd Schlusche, 2013, Market reaction to corporate press releases, *Journal of Financial and Quantitative Analysis* 48, 1207-40.

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, 2016, Financial attention, *The Review of Financial Studies* 29, 863-897.

Tetlock, Paul C., 2010, Does public financial news resolve asymmetric information? *The Review of Financial Studies* 23, 3520-57.

Van Niewerburgh, Stijn, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle. *The Journal of Finance* 64, 1187-1215.

Table 1. Stylized example of browsing data. A constructed example of an investor'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.

<table>
<tr><th>Row</th><th>Timestamp</th><th>Link</th></tr>
<tr><td>1</td><td>3/23/2007 2:31:11 PM</td><td>http://formula1.com/puriGP</td></tr>
<tr><td>2</td><td>3/23/2007 2:36:20 PM</td><td>http://cnbc.com/</td></tr>
<tr><td>3</td><td>3/23/2007 2:36:28 PM</td><td>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</td></tr>
<tr><td>4</td><td>3/23/2007 2:37:30 PM</td><td>https://[broker-1].com/user/login</td></tr>
<tr><td>5</td><td>3/23/2007 2:37:37 PM</td><td>https://[broker-1].com/eresearch/markets_sectors/analysis/mostactives.jhtml</td></tr>
<tr><td>6</td><td>3/23/2007 2:38:01 PM</td><td>https://[broker-1].com/stocks/earnings/earnings.asp?symbol=IMCL</td></tr>
<tr><td>7</td><td>3/23/2007 2:38:22 PM</td><td>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</td></tr>
<tr><td>8</td><td>3/23/2007 2:38:51 PM</td><td>https://[broker-1].com/charts/update_chart.asp?duration=365&frequency=1day&scaling=linear&display=candlestickcolor</td></tr>
<tr><td>9</td><td>3/23/2007 2:39:14 PM</td><td>http://finance.yahoo.com/</td></tr>
<tr><td>10</td><td>3/23/2007 2:39:20 PM</td><td>http://finance.yahoo.com/q?s=IMCL</td></tr>
<tr><td>11</td><td>3/23/2007 2:39:25 PM</td><td>http://biz.yahoo.com/bw/070530/000000000.html?.v=1</td></tr>
<tr><td>12</td><td>3/23/2007 2:39:29 PM</td><td>http://finance.yahoo.com/q?s=IMCL</td></tr>
<tr><td>13</td><td>3/23/2007 2:39:32 PM</td><td>http://finance.yahoo.com/q/ao?s=IMCL</td></tr>
<tr><td>14</td><td>3/23/2007 2:40:52 PM</td><td>http://formula1.com/puriGP</td></tr>
<tr><td>15</td><td>3/23/2007 2:43:03 PM</td><td>https://[broker-1].com/tradingideas/screener/stock_screener.asp</td></tr>
<tr><td>16</td><td>3/23/2007 2:43:10 PM</td><td>https://[broker-1].com/tradingideas/screener/stock_screener_results.asp?c=IBES.EPSGrowthNFY:GEQ:50:LSS:100001</td></tr>
<tr><td>17</td><td>3/23/2007 2:43:33 PM</td><td>https://[broker-1].com/common/symbol_info/symbolInfo.asp?symbol=GOOG</td></tr>
<tr><td>18</td><td>3/23/2007 2:43:39 PM</td><td>https://[broker-1].com/stocks/snapshot/snapshot.asp?symbol=GOOG</td></tr>
<tr><td>19</td><td>3/23/2007 2:44:11 PM</td><td>https://[broker-1].com/stocks/charts/charts.asp?symbol=GOOG</td></tr>
<tr><td>20</td><td>3/23/2007 2:45:02 PM</td><td>https://[broker-1].com/invest/socreateentry?ordertype="sell"&symbol=GOOG&shares=100</td></tr>
<tr><td>21</td><td>3/23/2007 2:45:19 PM</td><td>https://[broker-1].com/user/logout</td></tr>
<tr><td>22</td><td>3/23/2007 2:45:31 PM</td><td>http://us.f518.mail.yahoo.com/ym/login?login</td></tr>
</table>

Table 2. Summary statistics. Full sample of 484 households ("investors") with at least one US common equity or ADR trade. 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. N = 484 unless otherwise indicated.

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

Table 3. Aggregate research activity. Full sample of 484 investors. To account for differential time in the sample, measures normalized either by number of trading sessions, number of months in sample, or number of trades. Count measures (indicated by "N") are constructed on the investor-month level and averaged to the investor level. Research is the total time spent on 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 or Finance Domains	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. 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	Mean if t>0	Median if t>0	Mean All Inv	% t>0	Mean if t>0	Median if t>0	Mean All Inv	% t>0	Mean if t>0	Median if t>0
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 all investors in the sample carrying out nonzero research within the sample period. Research time measured in minutes. Market cap as of January 2007.

				All Research		
Name	Ticker	Market Cap ($ billion)	Market Cap Rank	% t>0	Mean if t>0	Median if t>0
APPLE INC	AAPL	74	33	27.3	3.0	0.8
MICROSOFT CORP	MSFT	302	3	22.5	5.0	0.9
CISCO SYSTEMS INC	CSCO	161	12	20.9	3.7	1.0
GENERAL ELECTRIC CO	GE	372	2	20.5	2.5	0.5
SIRIUS SATELLITE RADIO INC	SIRI	5	600	19.8	3.8	1.1
CITIGROUP INC	C	271	4	19	2.3	0.5
FORD MOTOR CO DEL	F	15	286	18.8	3.0	0.8
INTEL CORP	INTC	121	18	18	2.4	0.6
A T & T INC	T	236	5	17.4	4.0	0.8
GOOGLE INC	GOOG	57	51	16.3	3.8	0.7
ALTRIA GROUP INC	MO	199	8	15.3	5.2	1.8
WAL MART STORES INC	WMT	43	76	15.3	2.4	0.6
DELL INC	DELL	54	56	14.9	2.5	0.4
EXXON MOBIL CORP	XOM	432	1	14.3	2.9	0.5
GENERAL MOTORS CORP	GM	19	215	13.6	2.5	0.3
PFIZER INC	PFE	189	10	13.2	2.8	0.5
YAHOO INC	YHOO	39	86	13	5.2	0.4
HOME DEPOT INC	HD	83	29	13	4.3	0.6
GOLDMAN SACHS GROUP INC	GS	90	26	12	3.5	0.8
CONOCOPHILLIPS	COP	85	28	11.6	1.3	0.3

Table 6. Characteristics of researched stocks. Stock-month level regressions for stocks researched within the sample. Left hand side variables are the share of investors with non-zero research time in a given month and natural log of (1+Minutes) where Minutes is the total number of minutes of ticker-matched research by all investors in the sample. Right hand side variables are standardized except for indicator (0-1) variables. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on Newey-West standard errors with two lags.

	% Investors t>0						Ln(1+Minutes)
Ln(Cap)	0.53***	0.56***	0.68***	0.49***	0.77***	0.81***	0.64***
	(16.28)	(15.95)	(14.22)	(14.11)	(12.07)	(13.79)	(21.37)
Retail Volume Share			0.32***			0.22***	0.20***
			(11.54)			(9.45)	(13.19)
Individual Ownership Share			0.10***			0.24***	0.15***
			(5.12)			(8.97)	(8.77)
Ln(1+ N Shareholders)			0.13***			0.12***	0.05***
			(5.02)			(4.65)	(3.17)
Ln(1 + N News Articles)				0.04***	0.04**	0.13***	0.06***
				(2.64)	(2.33)	(6.34)	(6.31)
Earnings Month (0-1)				0.14***	0.16***	0.10***	0.15***
				(5.73)	(7.02)	(4.35)	(7.76)
Ln(Age)					0.04**	(0.01)	0.02
					(2.13)	(-0.28)	(1.14)
Ln(Price)					-0.31***	-0.16***	-0.16***
					(-5.69)	(-3.26)	(-5.93)
CAPM Beta					-0.08***	-0.03**	0.00
					(-5.21)	(-2.55)	(0.09)
SD(Return)					0.13***	0.08***	0.11***
					(9.09)	(8.29)	(7.68)
Momentum					0.02**	-0.04***	0.01
					(2.01)	(-3.13)	(0.87)
Dividend>0 (0-1)					-0.13***	-0.17***	-0.18***
					(-2.95)	(-4.01)	(-5.86)
Dividend Yield					0.02	0.01	0.01
					(1.22)	(0.70)	(1.14)
Book-to-Market					0.01	0.04***	0.00
					(0.97)	(3.15)	(-0.19)

Sales Growth					0.05***	0.03***	0.01
					(4.39)	(2.87)	(1.15)
External Finance					0.06***	0.04***	0.05***
					(4.33)	(3.08)	(3.64)
SP500 Member (0-1)					0.38***	0.07	0.05
					(4.95)	(0.98)	(0.93)
Constant	0.39***	0.20	0.37***	0.25***	0.04	(0.04)	-0.50***
	(45.77)	(1.07)	(37.89)	(6.30)	(0.42)	(-0.39)	(-5.75)
FF49 Industry FE		Yes			Yes	Yes	Yes
Observations	10,683	10,683	10,683	10,683	10,683	10,683	10,683
R^2	0.191	0.23	0.265	0.198	0.302	0.351	0.257

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 and the median investor. The percentage of investors with nonzero research time, and means and medians among investors with nonzero research time are also reported. Two columns for each sample show results when time on the three mixed-content pages are allocated across specific content categories in proportion to the number of all investors with interest in that specific category per the third column; this interest-weighted imputation is described further in the Internet Appendix. Other denotes time on research on miscellaneous topics. Indeterminate denotes time on research of unknown type. The last two rows show the investor-level means and medians of research time per trade.

	All Investor Sample (N = 484)							*High Ticker Match Rate Sample (N = 50)*													
	All Research, Minutes per Trade					(With Time on Mixed-Content Pages Imputed)		All Research, Minutes per Trade					(With Time on Mixed-Content Pages Imputed)		Research on Traded Tickers, Minutes per Trade					(With Time on Mixed-Content Pages Imputed)	
	Mean	Median	% t>0	Mean if t>0	Median if t>0	Mean	Median	Mean	Median	% t>0	Mean if t>0	Median if t>0	Mean	Median	Mean	Median	% t>0	Mean if t>0	Median if t>0	Mean	Median
Snapshot	37	6.5	93	39.6	8.3			26.7	9.3	92	29.1	10.2			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			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			3.7	0	10	37.2	10.2		
Risk Statistics	0	0	1	0.3	0.1	1.8	0.2	0	0	2	0	0	0.8	0.0	0	0	2	0	0	0.3	0.0
Earnings	1.7	0	40	4.3	0.6	10.6	1.8	0.6	0	36	1.8	0.7	6.0	1.7	0.1	0	14	0.5	0.2	2.0	0.1
Dividends	0.6	0	17	3.8	0.7	4.6	0.8	0.1	0	12	1.2	0.8	2.2	0.8	0	0	2	0.2	0.2	0.9	0.1
Other Fundamentals	2.6	0	53	5	1	6.8	1.1	1.5	0	52	2.9	1.5	4.3	1.8	0.2	0	20	0.8	0.5	1.1	0.1
Analysts	12.5	0.9	71	17.5	2.5	18.6	2.1	5.1	0.4	68	7.5	1.9	9.3	2.4	0.6	0	40	1.6	0.3	3.0	0.3
Ownership	0.5	0	26	1.8	0.4	1.9	0.2	0.7	0	30	2.3	1.7	1.4	0.2	0	0	10	0.1	0	0.2	0.0
Price Charts	39.1	5.8	93	42.1	6.7	64.4	14.7	31.2	11.5	98	31.8	11.6	47.4	27.3	10.2	1.2	82	12.4	2	16.7	3.4
Technical	2.3	0	45	5.2	0.9	5.9	0.9	0.9	0	38	2.5	1.1	3.6	0.7	0	0	12	0.4	0.2	0.9	0.1
Company Website	1.3	0	61	2.1	0.1	1.3	0.0	1.5	0	56	2.6	1	1.5	0.0	0.5	0	20	2.7	0.7	0.5	0.0
Other	12.8	1.3	84	15.2	2.3	17.8	3.4	7.4	1	84	8.8	1.2	12.4	3.5	0.4	0	44	0.9	0.1	2.5	0.3
Indeterminate	10.4	1	76	13.7	2.2	10.5	1.0	7.7	0.6	66	11.6	3.5	7.7	0.7	1.1	0	46	2.3	0.3	1.1	0.0
Mean Total across Invs	144.2							96.6							29.2						
Median Total across Invs	36.4							51.3							5.7						

Table 8. Principal components of investor-level research focus. 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 observe risk statistics for a loading to be estimated.

	First PC	*Second PC*
Snapshot	0.32	0.31
News - Indeterminate	0.36	0.10
Message Board – Indet.	0.16	0.45
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
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 stock.

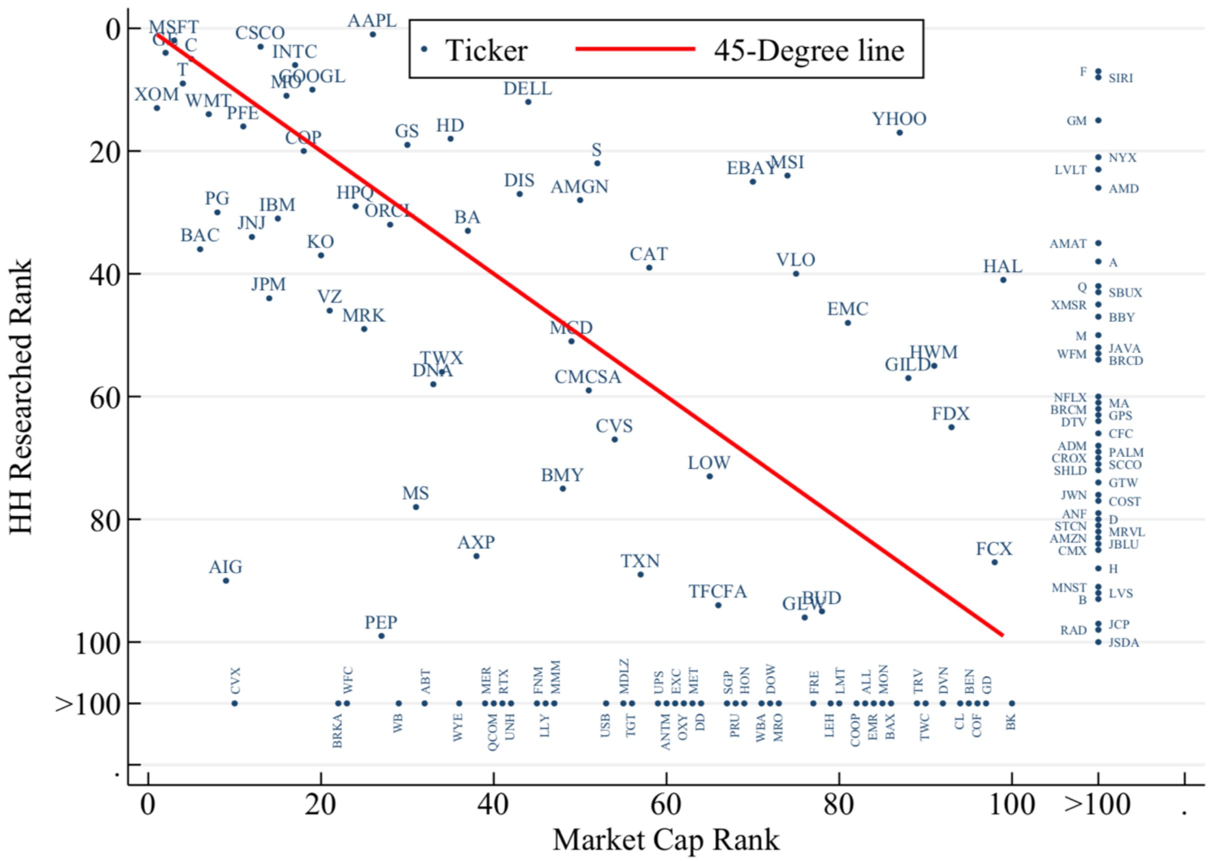


Figure 2. Research around events: Clickstream data vs. Google Trends. Panel A shows the number of investors in the full sample (N=484) with research activity on Microsoft stock in the month of its January 25, 2007 quarterly earnings announcement and compares this to the Google Trends index for interest 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.

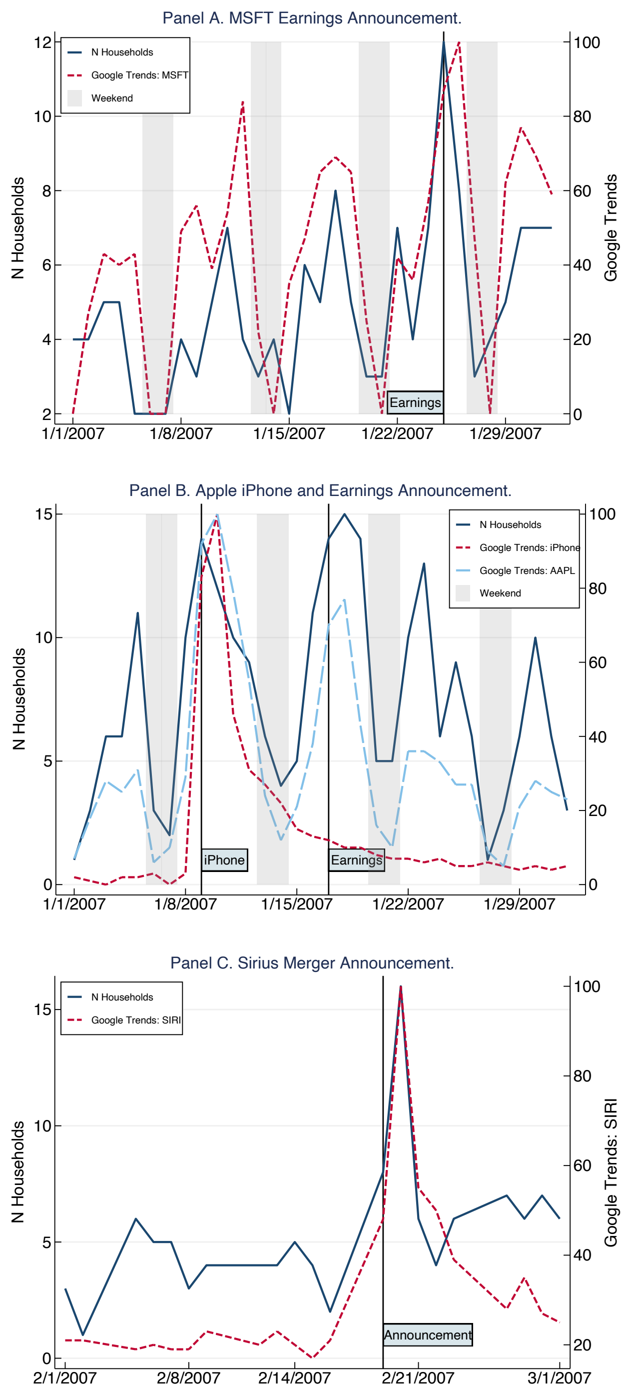
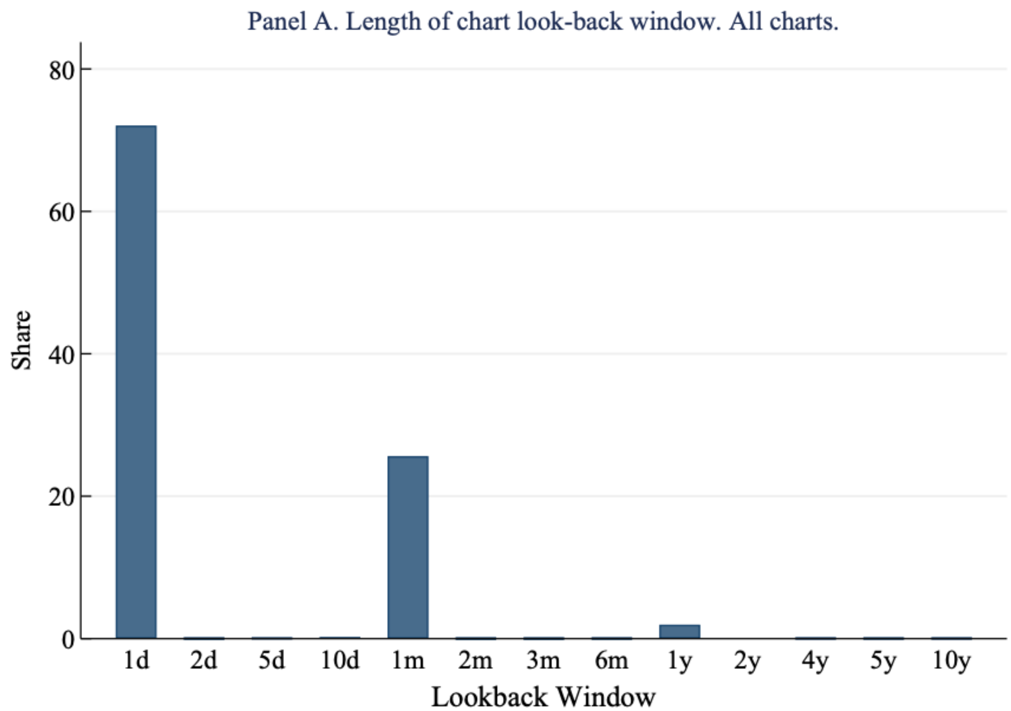


Figure 3. Length of chart lookback windows. Lookback windows viewed by investors in the full sample (N=484). Panel A includes the default lookback windows presented in the snapshot or price chart pages. Panel B excludes such views and considers only opt-in lookback periods.



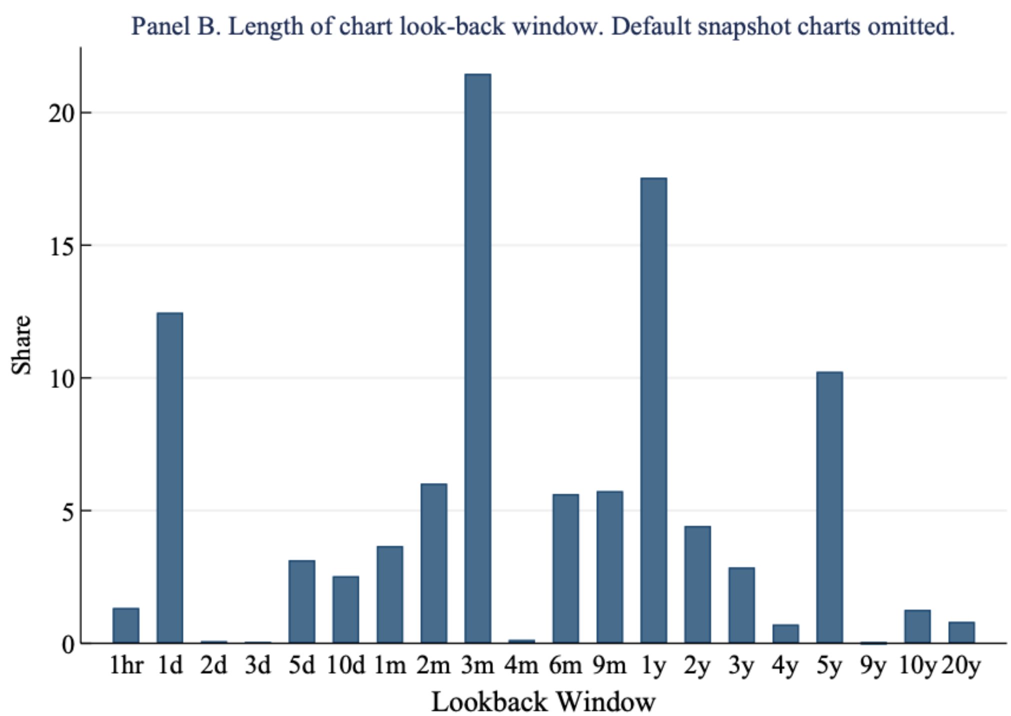
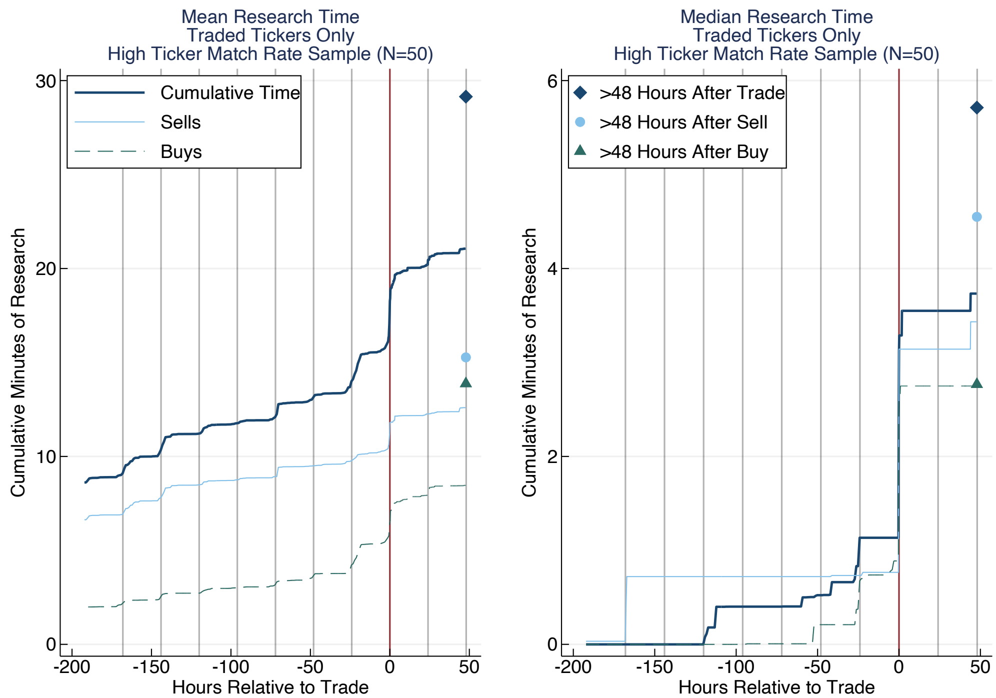


Figure 4. Research around trades. The sample is N=50 investors for which a high fraction of total stock research time can be matched to a CRSP-listed stock or ADR ticker; plots are for cumulative research time by these investors in tickers that they trade at hour 0. Figures are plotted in ten-minute increments. Small shapes on right side indicate cumulative total research time in the traded stock until the end of the investor's sample.



Internet Appendix

This Appendix describes ten key data processing steps in more detail than in the body of the paper. The starting point is clickstream data for a nationally representative sample of households for January, February, March, and June of 2007. Households are present between one and four months; the mean and median is three months. For each household we observe the timestamped sequence of all URLs (clicks/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.

Step 1. Identifying investors

We begin with a list of 11 online brokerages operating at the time of the sample and the associated domain names. We read through the entire dataset and create a subset of clicks on these brokers' websites. Within these we flag those that contain terms suggestive 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 would be flagged for the inclusion of at least two such terms:

https://[broker-1].com/invest/socreateentry?ordertype="sell"&symbol=GOOG&shares=100

Given 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. We match the traded tickers to CRSP and keep those 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 in time by the last trade in the sequence, with the notion that roughly the same "research" is associated with trades within the sequence. With this procedure we end up with 2,911 total common stock and ADR trades by 484 distinct trading households, i.e., investors.

Step 2. Processing browsing data

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

We then 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 makes it easier to process the URLs via regular expression matching. Second, we organize the URLs into 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 include brokerage sites and other non-broker finance sites such as CNBC, Marketwatch, and SEC.gov.

Step 3. Identifying tickers and classifying content

The next six steps involve reading out tickers and research categories. Our first goal is to identify clicks associated with stock tickers. We proceed in an iterative manner, inspecting the

URLs until we find one that represents research associated with a ticker. For instance, at Yahoo Finance we might come across the URL 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 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. We were in many cases able to find promotional material or video guides that illustrated the content on various pages, but in some cases, we must decide 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 the URL in row 18 of Table 1:

https://[broker-1].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 find no more URLs pertaining to research that also contain ticker information.

Step 4. 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 charts with multiple tickers as well as multiple types of information, e.g., a chart might plot the prices and volume of two stocks. In such a 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 the last three years (1095 days) and a prediction for the next one and one-half years (545 days).

Step 5. Cleaning tickers

Having processed the URLs and extracted ticker information, we separately process the ticker information 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 CRSP. We inspect the unmatched tickers and, whenever possible, match by hand. 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 ("WALLMART") which, whenever obvious, is assigned to the associated ticker. We do not count the time from clicks involving misspelled tickers in the research time totals.

We separately create indicators for identifiers of indexes, mutual funds, options, and currencies. These tickers are often indicated by a special character on brokerage sites. For example, in row 3 of Table 1, indexes are indicated by a leading period and exchange rates are indicated with a double dollar sign.

Step 6. Other activity at brokerage sites

Having exhausted clicks that contain ticker information, we classify 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 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 and is not counted as research time in this paper. The two main categories that might contain ticker-related information are "portfolio" and "trading": clicks involving reviews of current investment positions, which are generally not revealed by the URLs , and clicks involving inputting and confirming trades. We also classify other time spent at brokerage sites such as 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.

Step 7. Processing company website research

As a final type of 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 for the companies in the top 100 in terms of research activity.

The idea behind keeping track of this time, not just obviously finance research-oriented time, is to provide a sense of an investor's familiarity with the products of a given company. For that reason, the time an investor browses the newly released first generation iPhones on Apple.com would count as research time, but only if they ever research AAPL on any of the finance sites. We additionally 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.

Step 8. Processing non-ticker-matched 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 in which 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—the press release likely pertained to IMCL because it was immediately 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.

Step 9. Constructing content categories

After an iterative process of reviewing the processed data until we are confident that the URLs are properly encoded, the preceding steps lead to around 3,000 hours of total research time, with roughly 1,350 hours matched to an identifier, be it a ticker, a company name, or an index. The last two steps involve harmonizing research content categories. As described above, in processing the research clicks at the various domains we typically start with the content categorizations by the domain itself. The categories that we obtain in this way are not perfectly aligned across the different websites, but there are numerous broad commonalities. For instance, all of the brokerage and most of the other sites see many visits to a standard page best described as a "snapshot": a ticker-specific summary page of different types of information that typically includes price quotes, a chart of recent prices, brief data on earnings and dividends, and potentially links to news and analyst views. Another typical page that is available is a charts page with prices, volume, and perhaps technical indicators such as moving averages, candlesticks, and so on.

Beyond these two dominant categories, there are two other 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 more detailed pages on risk metrics, technical research, ownership, or other categories as discussed in the text. For the less-frequently clicked categories that amount to no more than a handful distinct clicks, we code the research as “Other”, 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. The “Other” category includes disparate information such as time spent on snapshot pages where the underlying security is a stock market index, or on other pages pertaining to industry- or market-wide developments. It also includes time spent on pages relating to aggregate business conditions, which we construct based on keywords associated with inflation, unemployment, GDP, and the FOMC. We additionally add time spent at the BLS and BEA websites to this category. By definition, these research activities are not matched to tickers.

For pages where it is not possible to infer any aspect of the content the URL being highly suggestive that it constitutes stock research, we code it as “indeterminate”.

Step 10. Imputing mixed-content links to content categories

Finally, three types of pages either contain a mixture of content or do not indicate the specific category of content: snapshots, news, and message boards. In order to attempt a full accounting of the content of information presented to investors, we construct an alternative classification in which we assign research time on these pages into primary categories.

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 particular category in links in our data that could be specifically allocated (as reported in the third column of Table 7). The third imputation is an appealing and objective measure, and it is what we use in the construction of relevant columns of Table 7, but no results in Table 7 change meaningfully under the two alternative approaches (results available on request).

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, *Wall Street Journal*, 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 IA.1 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 investors 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 focus on earnings.

Table IA.1. 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 investor interest in specific content categories in the third column of Table 7 (and which is used in Table 7's imputed time columns). 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 unobservable type is based on the distribution of content across RavenPack's detailed finance article content classifications. Rows that do not quite sum to 100% indicate remaining indeterminate content.

	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	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