IQ from IP: Simplifying Search in Portfolio Choice*

Huaizhi Chen
Harvard Business School

Lauren Cohen
Harvard Business School and NBER

Umit G. Gurun
University of Texas at Dallas and NBER

Dong Lou
London School of Economics and CEPR

Christopher J. Malloy
Harvard Business School and NBER

*Contact Information: Huaizhi Chen, Harvard Business School, Baker Library 273, Soldiers Field, Boston, MA 02163, email: hchen@hbs.edu; Lauren Cohen, Harvard Business School, Baker Library 273, Soldiers Field, Boston, MA 02163, email: lcohen@hbs.edu, Phone: 1-617-495-3888; Umit G. Gurun, University of Texas at Dallas, Jindal School of Management, Richardson, TX 75093, email: umit.gurun@utdallas.edu, Phone: 1-972-883-5917; Dong Lou, London School of Economics, email: d.lou@lse.ac.uk, Phone: +442071075360; Christopher Malloy, Harvard Business School, Baker Library 277, Soldiers Field, Boston, MA 02163, email: cmalloy@hbs.edu , Phone: 1-617-495-4383.

IQ from IP: Simplifying Search in Portfolio Choice

ABSTRACT

Using a novel database that tracks web traffic on the SEC's EDGAR servers between 2003 and 2016, we show that mutual funds exert effort to reduce the dimensionality of their portfolio selection problem. Specifically, we show that mutual fund managers' gather information on a very particular subset of firms and insiders, and their surveillance stays largely unchanged over time. This tracking has powerful implications for their portfolio choice, and its information content. An institution that downloaded an insider-trading filling by a given firm last quarter increases its likelihood of downloading an insider-trading filing on the same firm by more than 41.3 % this quarter, which is 8 times larger than the unconditional probability of an institution downloading at least one insider trading filing in a quarter from any firm in her existing portfolio (4.8%). Moreover, the average tracked stock that an institution sells generates 7.5% annualized DGTW-adjusted alpha, whereas the sale of an average non-tracked stock has close to zero DGTW adjusted alpha. The outperformance of tracked trades continues for a number of quarters following the tracked insider/institution sale and does not reverse within the sample period. Collectively, these results suggest that the information in tracked trades is important for fundamental firm value, and is only revealed following the information-rich dual trading by insiders and linked institutions.

Keywords: Tracked trades, return predictability, institutional trading, insider trading

JEL Classification: G11, G14, G23

I. Introduction

There is a fundamental search problem inherent in all portfolio choice. In fact, with the decreasing cost of creating, processing, and transmitting information, the proliferation of information signals has increased greatly in both quantity and dimensionality in recent decades. Of course this creates a classic signal-noise problem, in which an agent must search ever larger matrices to decipher and create profitable signals. In a Grossman-Stiglitz world, an agent will be happy to collect information up to their private marginal value of expected return from that activity. However, with hundreds of thousands of information signals being produced in any given day, how does an investor reduce the dimensionality of the investment problem sufficiently to know even which subset (or class) of signals have the *potential* to be informative and provide this return in expectation?

In this paper, we propose that this reduction in dimensionality is a critical, and yet understudied, step in the investment process. Using rich, proprietary data provided by the Securities and Exchange Commission (SEC) on every document downloaded from their online site—including the exact timing and the IP address of the agent downloading--we provide new evidence on the search process in delegated portfolio management. In particular, we show that fund managers follow, and download, information on a very particular subset of firms, and that this set of firms stays highly constant over time. Further, their trades on these "tracked" firms are significantly more informative for future operations and future firm performance, relative to their other trades.

The key innovation in our paper is that we are able to explicitly link the monitoring behavior of individual fund managers (through their download behavior—which we are able to map to the IP addresses of institutional investors, and hence identify them) to specific events on the stocks in their own portfolios. No prior study has been able to examine search behavior at the level of a specific institutional investor. In particular, we focus on how institutional fund managers track the trades of corporate insiders in the stocks they own.

We examine this laboratory for a number of reasons. First, compensation and hiring vs. firing decisions of fund managers – in addition to external human capital valuation such as possible hedge fund transitions -- are often determined by managers' performance relative to their peers. In fact many of the industries' highest profile rankings (e.g., Morningstar, Kiplinger, Barron's, etc.) are relative rankings amongst fund managers competing within a mandate. Thus, career concerns give fund managers a highly incented framework in which to care about the maximization of relative performance. Given this tournament-setup, a natural argument in a fund manager's maximization-function would be to find a signal (or set of signals) on which they have a comparative advantage relative to other managers. This begins to put some structure on the information dimensionality and resultant tracking problem that managers face.

Turning to insider trades, these are a potentially attractive candidate for relative comparative advantage signals for mutual fund managers. First, insiders are – by definition – a class of agents with privileged access and private information regarding their firms. Second, of all the factors of production – and all the information signals produced on a firm – insider trades are likely amongst the most valuable for unlocking a powerful (and legal) comparative advantage for a given fund manager. For instance, if a firm announces a new product launch, outside of explicit transmitting of material non-public information, it might be difficult (or prohibitively costly in any scalable manner) for an institution to gain a comparative advantage on this signal relative to other institutions.

However, contrast this with an insider trade within the same firm. The trade itself is public information – a sell, for instance. However, following the publicly disclosed sell, an institutional fund manager who owns the stock and hence has a connection to that firm can contact someone at the firm and inquire whether the sell was for personal liquidity reasons; for instance, to purchase a vacation home. Once determining that the sell was unrelated to personal liquidity needs – information which the insider is perfectly legally free to tell her connection (i.e., it is not considered material non-public information to speak about vacation home purchases) – the fund manager can more accurately interpret this public signal of the tracked stock and trade accordingly. Importantly, this is an advantage of a connected manager – in that her competitor funds without a connection to the given insider may have a more costly process in gathering the same private information. This results in the tracking of connected stocks – and in particular signals generated by a manager's precise connection at that stock, such as insider trades – being a potentially natural way to reduce the dimensionality of the competitive portfolio choice problem that delegated portfolio management faces.

We document that mutual fund managers have a very specific set of firms (and insiders) that they track. Moreover, their tracking activities have powerful implications for their portfolio choice, and its information content. For instance, the fact that an institution downloaded an insider-trading filling by a given firm last quarter increases her likelihood of downloading an insider-trading filing from the same firm by more than 41.3% (t = 30.92) this quarter. For reference, the unconditional probability of an institution downloading at least one insider trading filing in a quarter from any firm in her existing portfolio is 4.8%. In other words, our persistence result – an 8 times increase in probability - is not only statistically significant, but also economically important. We find this is driven by persistence at the individual insider-level tracking. For instance, an increase in the probability of 18.7% (t = 24.72) of downloading Jamie Dimon's insider trading filing if the manager downloaded the same filing the prior quarter.

Importantly, the behavior of these "tracked" insiders' behavior is closely linked to the portfolio choice decisions of fund managers themselves. For instance, the probability that an institution sells a given stock in its portfolio increases by 20% if one of the "tracked" firms in its portfolio has insider selling in that quarter. As a placebo, if another firm in the same institutions' portfolio has identical insider selling as the tracked firm, but is not tracked, the institution is no more likely than random to sell the given stock. This shows that the selling behavior has nothing to do with insider trading itself, but is instead linked to the tracking behavior of the fund manager.

Next we show that these "tracked" insider trades have predictability for future firm operations and returns. In particular, the average non-tracked stock that an institution sells has an insignificant DGTW-adjusted alpha of close to zero (-11 basis points per quarter (t=0.23)). However, when an institution performs this same action – selling a stock - when a tracked insider sells along *with* the tracking institution in that quarter, this portfolio underperforms the non-tracked sell by a highly significant DGTW-adjusted return of 185 basis points per quarter (t=2.77), or roughly 7.5% annualized abnormal return.

If the results we find reflect institutional managers exhibiting a true comparative advantage in their tracked stocks, we might expect the managers to know when not to "follow" the tracked insiders' behavior, as well. For example, if the institution can decipher that the given trade was for liquidity reasons (as opposed to information-based), the manager would not want to mimic that trade of the tracked insider. This implies that when we observe institutions choosing not to follow the trades of their tracked insiders, these insider trades should have less predictive ability for future returns. We find this pattern in the data. In particular, firms in which institutions sell alongside tracked insider sales tend to underperform those in which tracked insiders sell but the institutions choose *not* to sell alongside the insiders. The magnitude of the difference is large and highly significant, at 120 basis points per quarter in abnormal return (t=3.80).

Lastly, we show that the outperformance that we document on these tracked trades continues for a number of quarters following the tracked insider (and institution) sell. Importantly, it never reverses, suggesting that the information in tracked trades is important for fundamental firm value, and is only revealed following the information-rich dual trading by insider and linked institution.

Moreover, a sizable percentage of the return occurs in the direct proximity following the insider trade itself, underscoring the importance of the real-time tracking in this linked relationship between the institution and a given insider.

The paper proceeds as follows. Section II lays out the background for the setting we examine in the paper. Section III presents our data collection procedures, and summary statistics. Section IV provides our main results on the robust behavior of institutions tracking particular insiders and the return predictability pattern of institutions tracked insider trades. Section V concludes.

II. Literature Review

Our work relates to a several strands of the literature, including papers analyzing the investment performance of mutual fund managers, articles exploring the characteristics and profitability of insider trading, and a slew of studies documenting gradual information diffusion and limited attention in the stock market.

The area of the mutual fund literature most closely related to our paper is the collection of work examining whether mutual fund managers possess stock-picking ability. This remains an open question, because while many papers (Jensen (1968), Malkiel (1995), Gruber (1996), and Carhart (1997)) find that active managers fail to outperform passive benchmark portfolios (even before expenses), several others (Grinblatt and Titman (1989, 1993), Grinblatt, Titman, and Wermers (1995), Daniel et al. (1997), and Wermers (1997)) find that active managers do exhibit stock-picking ability. In terms of specific characteristics known to correlate with superior performance, Chevalier and Ellison (1999) use biographical data on managers to show that fund managers from undergraduate institutions with higher average SAT scores earn abnormal returns.[1] Other evidence shows that fund managers tend to overweight nearby companies (Coval and Moskowitz (1999), and earn higher returns on their local holdings (Coval and Moskowitz (2001)). Closest to this paper is Cohen, Frazzini, and Malloy (2006), who find that fund managers place bigger bets on firms they are connected to through their education network, and perform significantly better on these holdings relative to their non-connected holdings. Hong, Kubik, and Stein (2005) also document word-of-mouth effects between same-city mutual fund managers with respect to their portfolio choices. We add to this list by exploring to what extent mutual funds actively investigate the insider trades on stocks within their own portfolios. Our approach highlights another channel through which fund managers earn abnormal returns.

Our paper is also closely related to a large literature examining the behavior of corporate insiders. Many of these papers study the cross-sectional return forecasting ability of insider trades aggregated at the firm level. Numerous papers (see, for example Lorie and Niederhoffer (1968), Jaffe (1974), Seyhun (1986, 1998), Rozeff and Zaman (1988), Lin and Howe (1990), Bettis, Vickery, and Vickery (1997), Lakonishok and Lee (2001), and Marin and Olivier (2008)) focus on the

[1] Massa and Simonov (2005) also document a relation between the portfolio choices of individual investors and their past educational backgrounds.

abnormal returns to firms in relation to various metrics of firm-level insider trading. Seyhun (1998) summarizes this evidence and reports that several different trading rules lead to abnormal returns. In addition, Jeng et al. (2003) show that insider purchases earn abnormal returns of more than 6% per year, while insider sales fail to earn significant abnormal returns.

Several papers take a more granular approach and examine individual insider-level data in order to identify which insiders are truly informed. For example, Cohen, Malloy, and Pomorski (2010) show that the past trading records of insiders can be used to identify which insiders are likely to be trading on information and which are not. In addition, Piotroski and Roulstone (2005) demonstrate that insider trades reflect both contrarian beliefs as well as private information about future cash flows, and Ke, Huddard and Petroni (2003) demonstrate that insiders trade before significant accounting disclosures. Kahle (2000) shows that long-run stock returns associated with seasoned equity offerings (SEOs) are significantly related to measures of insider trading, and Clarke, Dunbar, and Kahle (2001) provide evidence consistent with insiders exploiting windows of opportunity by trying to issue overvalued stock. Finally, Jagolinzer (2009) presents more evidence of strategic trading by insiders by focusing on a subset of insiders who publicly disclose 10b5-1 plans; he shows that insiders initiate sales plans before negative returns and terminate sales plans before positive returns.

Our paper can also be situated within the large and growing literature on limited attention, and the slow diffusion of information in the stock market. Many of these papers argue that if investors have limited resources and capacity to collect, interpret, and finally trade on value-relevant information, we should expect stock prices to incorporate information only gradually. For instance, because of gradual information diffusion (Hong and Stein, 2007) and/or gradual capital diffusion (slow moving capital (Duffie, 2010)), this information may be impounded into stock prices slowly. Meanwhile, there is a substantial literature studying investors' limited attention to information. Theoretical papers such as Merton (1987), Hong and Stein (1999), and Hirshleifer and Teoh (2003), argue that with investors subject to binding attention and resource constraints, delayed information flows can lead to expected returns that are not explained by traditional asset pricing models. Numerous empirical studies find supporting evidence for these models. For example, Huberman and Regev (2001), Barber and Odean (2006), DellaVigna and Pollet (2006), Hou (2006), Hong, Torous, and Valkanov (2007), Cohen and Frazzini (2008), and Cohen and Lou (2011) find that investors respond quickly to salient, eye-catching information, but tend to ignore information that is less obvious yet nonetheless essential to firm value.

Since our work utilizes the log file from the SEC Edgar database, our paper is also related to a few recent papers that use this data to explore different, but related issues in corporate finance and asset pricing. For example, Loughran and McDonald (2016) provide a first descriptive analysis of this dataset and show that--after sifting out robot requests--the average publicly-traded firm has their annual report requested only 28.4 total times by investors immediately after the 10 K-filing; they conclude that the "lack of annual report requests suggests that investors generally are not doing fundamental research on stocks." Meanwhile Lee, Ma, and

Wang (2016) apply a "co-search" algorithm to the SEC log file in order to identify economically-related peer firms; they show that firms appearing in chronologically adjacent searches by the same individual are fundamentally similar on multiple dimensions. Finally, Drake, Roulstone, and Thornock (2016) show that EDGAR activity is positively related with corporate events (particularly restatements, earnings announcements, and acquisition announcements), poor stock performance, and the strength of a firm's information environment; EDGAR activity is also related to, but distinct from, other proxies of investor interest such as trading volume, business press articles, and Google searches. While all of these papers explore the SEC logfile data in various ways, none of them are able to explicitly link the users of the data to specific firms (i.e., the 13F filers that feature in our analysis); by doing so we are able to explore the investment implications of SEC searches in a more direct and granular way.

III. Data and the Setting

In this section, we describe how we handled the data for the insider monitoring analysis. First, we obtain the data of filings and the IP addresses of their viewers from the SEC at the log file website (https://www.sec.gov/data/edgar-log-file-data-set). The log files are available from 2003 to 2016, and are posted by the SEC on a quarterly basis with a 6-months delay.

Table 1 provides summary statistics on our sample. There were roughly 38 million form file requests over the 2003-2016 sample period. The two most requested filing types were corporate 8-Ks and insider trading filings. 8-Ks are required to be filed by firms to notify shareholders of material events transpiring at the firm.[2] There were over 6 million requests for 8-Ks. Insider filings make up the next most requested documents, at 5.88 million form requests. 10-K (annual) and 10-Q (quarterly) reports - which are less frequently filed by firms - both had roughly 5 million form requests over the sample.

The IP addresses in the dataset are partially anonymized using a static cipher. The data describe the access of fillings by different IP addresses. That is, each row of the data corresponds to a certain IP address (24.145.236.jcf) viewing a specific filing coded by an accession number (0000891020-04-000160) at a specific time and date (12:00 am on April 31, 2004).

In order to match the data to organization level information, we first de-anonymized the ciphers. This is done by using another set of server logs from a private but well-trafficked website. Assuming that the intersection of IP addresses for existing IP are similar between two servers, we are able to deduce which cipher, say aaa, corresponds to what number between 0 and 255 using the frequency of potential matches.

For instance, if 1.1.1.aaa visited the SEC server and 1.1.1.111 visited the private website in 2016, then the cipher aaa to 111 has 1 additional match. The number of matches between a cipher and its most frequently matched last IP octet is distinct for the vast majority of the ciphers. Out of the 256 ciphers, over 230

[2] More detail is given here as to when an 8-K is required to be filed, along with usual format and content contained there-in: https://www.sec.gov/fast-answers/answersform8khtm.html.

have a most frequently matched octet that does not intersect with any other potential pairing. The last digit octet is the most frequent for only 1 cipher. The rest of the pairings are matched using a process of elimination. For example, if 001 is the most frequent octet for both aaa and aab, but 001 has many more matches with aaa than aab, then aaa is matched to 001. In this case, aab is then linked to its next most frequent octet until all 256 pairs are matched.

Once the IP data is deciphered, we connect the specific filing IP address to a set of organizations using a dataset of organization IP addresses from MaxMind. The IP of organizations data is released on a periodic basis. The IP address linked to each viewing from Edgar is matched with the last available organization data for that IP address at the time of the viewing.

After this step, we hand match names of the 13F organizations to the list of potential organizations from MaxMind by Research Assistants. We start off with two thousand 13F organizations with the largest average AUM between 2004 and 2016. Since IPs are non-static and the MaxMind data also changes from period to period, certain institutions appear more frequently and longer than others. In Table 2 we report the statistics on the number of 13F reporting organizations that are linked in the final data, how frequently they appear, and their respect final appearance AUMs.

On average, we observe an IP matched is matched to an institution for 3.87 quarters, out of possible 30 quarters, whereas the average number of quarters that a 13F institution is observed 5.46. When we investigate the number of unique securities filings visited by each 13F manager per quarter, we find a wide range, from 1 to 2,682, with a mean of 104.15 visits. Compared with the number of unique securities filings visited by each IP per quarter, which includes both professional money managers and also other organizations, we observe significantly fewer visits, 41.95. On average, we observe 208 unique IPs per quarter, of which 69 of them are 13 Filing institutions. In Table 3, we report the top 30 institutions (in terms of total value of the portfolio last observed) that we were able to link to an IP address.

After we identify the link between each 13F filings institution and their IP addresses, we focus on the documents accessed in the EDGAR system. We use the WRDS accession filing database to link each IP viewing to a specific filing. This specific accessing filing contains a mapping of each EDGAR document to a COMPUSTAT firm. After this step, we are able to observe which institution tracked which filing in the EDGAR, i.e. Fidelity at xxx.xxx.xxx.xxx viewed a specific Form 3 of Apple on a particular date.

In the final step, we scrape the insider trading filings from SEC website to obtain the datacodes recorded in each form. This datacode in each insider form allow us to obtain the accession numbers necessary to match to the Thompson Insider database. For example, an insider trade Form 3 (0000891020-04-000160) represents Tim Cook's unloading of shares. After following these steps, we are able to observe which of the identified IP addresses of 13-F organizations accessed which particular insider trading forms on the Edgar server from 2004 to 2016.

IV. Empirical Analysis

The main thesis of our paper is that investors, considering their resource constraints, should optimally choose to focus their information gathering efforts on a subset of the firms and a subset of the signals where they have a comparative advantage in terms of collecting and interpreting the information. To illustrate, if investor A has a comparative edge in interpreting information from the healthcare industry (due to, for example, her prior work experience), we expect the investor to focus her research activity, and consequently her portfolio holdings, in this industry.

Moreover, since comparative advantages in information processing are accumulated (developed) through years of experience and interactions with other economic agents, and are thus unlikely to change rapidly over time, we expect persistent patterns in investors' information-gathering activity. We start our empirical analysis by examining the following question: conditional on investor A searching for regulatory filings by company X in one period, do we see searches by the same investor on the same firm in the *next* period?

Table 4 reports the persistence in institutions' search behavior for insider trading filings (Forms 4, 5, and 6) on the EDGAR server. We conduct a panel OLS where the dependent variable is a dummy that equals one if an institution downloads at least one insider-trading filing by a given firm in quarter t. The main independent variable of interest is a similar dummy defined in quarter t-1. As can be seen from Column 1 of Panel A, there is substantial persistence in institutions' search behavior. The fact that an institution downloaded an insider-trading filling by a given firm in quarter t-1 increases her likelihood of downloading an insider-trading filing from the same firm by more than 41% (t = 30.92) the next period. For reference, the unconditional probability of an institution downloading at least one insider trading filing in a quarter from any firm in her existing portfolio is

4.8%. In other words, our persistence result is not only statistically significant, but also economically important.

In Columns 2 and 3, we further include a host of control variables, as well as portfolio fixed effects and year-stock fixed effects. Our results remain economically large. For example, in Column 3 (with the full set of controls and fixed effects), the coefficient on lagged search dummy is 0.255 ($t = 20.75$)) — i.e., an institution that downloaded insider filings of a given firm in the prior quarter has a 25.5% higher chance of downloading insider filings by the same firm again in the following quarter.

In the next three columns of Table 4, we narrow in on the specific insiders. In other words, we track not only institutions' searching for insider filings by Apple, but also the specific filings by Tim Cook. The results are consistent with those shown in the first three columns. As can be seen from Column 4, an institution that downloaded an insider trading filing by a given executive in a quarter has an 18.7% ($t = 24.72$) higher likelihood of downloading an insider trading filing by the same executive in the following quarter. Again, including portfolio and year-times-insider fixed effects has little impact on our results. For example, in Column 6, where we include the full set of controls and fixed effects in our regression, the coefficient on lagged search behavior drops only slightly to 13.3% ($t = 21.48$).

After establishing that institutions' search behavior on EDGAR is highly persistent (that is, each institution tends to follow the same group of firms and insiders over time), we next turn to institutions' trading decisions. In particular, we examine whether institutions trade in the same direction as the insiders that they follow. To this end, we classify trading in each stock by each institution as either a buy or a sell, based on changes in the number of shares. (Our results are largely unchanged if we instead define trading using changes in portfolio weights.) We then compute the fraction of institutions that buy or sell a particular stock in a given quarter, conditioning on their contemporaneous search activity on the EDGAR server.

As shown in the first two rows of Table 5, in each quarter and on a TNA-weighted basis, the unconditional fraction of institutions buying a given stock is 53%, while that of selling a given stock is 40%. (The two figures do not sum up to one because a portion of the institutions does not trade the stock in each quarter.) The asymmetry between buys and sells is consistent with the fact that institutions are growing rapidly in our sample period (from holding around 60% of the entire market to over 85%).

In the next two rows, we examine trading behavior by institutions whose "tracked insiders" (i.e., insiders whose filings are downloaded by the institutions) had sell transactions in the same quarter. Focusing solely on this subset, we find that the fraction of institutions (again on a TNA-weighted basis) that sells the stock rises to 47%. The difference between the 47% in this setting and the 40% in the unconditional test is highly statistically significant. This suggests that institutions indeed act upon the cues they get from their tracked insiders.

In the last two rows of the table, we conduct a placebo test, drawing on the idea that institutions should only follow the insiders that they actively search for, and not the ones for which they do not search. Consequently, we examine the trading behavior of institutions whose tracked insiders do *not* have sell transactions but their untracked insiders do. To illustrate, imagine manager X that has a sell transaction. This placebo setting focuses on institutions that have not downloaded manager X's filings on EDGAR. Not surprisingly, compared to the unconditional case, we see very little change in institutions' trading behavior (41% vs. 40% in terms of selling, 53% vs. 53% in terms of buying).

Given that institutions tend to trade in the same direction as the subset of managers they follow, a natural question to ask is whether institutions earn abnormal returns from these trades. If institutions optimally choose which firms/managers to follow based on their comparative advantages to process/interpret information, we should expect these trades to generate positive abnormal returns.

To test this idea, we form calendar-time portfolios using various sorting criteria, and compute the return differentials among these portfolios. To start, we compute the total portfolio return of an average institution in our sample. As shown in the first row of Table 6, during our sample period, the average institution earns a statistically insignificant DGTW-adjusted quarterly return of -11bps (t = 0.45). This is consistent with prior findings that institutions as a whole are unable to outperform the market.

We then divide each institution's entire portfolio into two sub-components: the first sub-portfolio includes all holdings where the institution's tracked insiders have sell transactions in the same quarter; the second sub-portfolio, which serves as a benchmark, includes all other holdings (i.e., where the institutions' tracked insiders do not have reported sell transactions). As shown in the third row of the table, the former underperforms the latter by 118bps (t = 3.59) on a DGTW-adjusted basis in the subsequent quarter. This is consistent with insider selling being informative for future stock returns.

In the fourth row (our main analysis), we construct our primary portfolio requiring that both the institution and its tracked insiders sell the security in question. This portfolio underperforms the benchmark by 185bps ($t = 2.73$) on a DGTW-adjusted basis in the following quarter. In the next two rows, we employ two alternative benchmark portfolios. In row five, the benchmark portfolio includes all holdings which are sold by the institution but not by tracked insiders. We obtain very similar a return differential between our primary portfolio and this new benchmark, 185bps ($t = 2.77$), in the following quarter. In row six, we define the benchmark as including all holdings that are not sold by the institution but are sold by tracked insiders. Again, the return differential remains economically large and statistically significant. These results suggest that institutions are able to identify the most informative trades by their tracked insiders.

V. Conclusion

With the proliferation of information signals in both quantity and dimensionality in recent decades, investors face an increasingly complex portfolio choice problem. Most investors simply do not have enough resources and time to comb through all the information available to them. With hundreds of thousands of information signals being produced in any given day, how does an investor reduce the dimensionality of the investment problem sufficiently to know which signals to track and collect? In this paper, we show evidence that sophisticated investors such as mutual fund managers try to reduce dimensionality problem in a particular way. Using web traffic on the SEC's EDGAR servers between 2003 and 2016, we show that mutual fund managers' follow information on a very particular subset of firms and insiders. Mutual funds tracking activity not only stays highly unchanged over time but also has powerful implications for their portfolio choice and performance. The outperformance of tracked trades continues for a number of quarters following the tracked insider (and institutional) sale and does not reverse within the sample period, suggesting that the information in tracked trades is important for fundamental firm value, and is only revealed following the information-rich dual trading by insider and linked institution.

References:

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

Bettis, Carr, Jeffrey Coles, and Michael Lemmon, 2000, Corporate policies restricting trading by insiders, Journal of Financial Economics 57, 191-220.

Carhart, Mark, 1997, On persistence in mutual fund performance, Journal of Finance 52, 57-82.

Chevalier, Judith, and Glenn Ellison, 1999, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, Journal of Finance 54, 875-899.

Clarke, Jonathan, Craig Dunbar, and Kathleen M. Kahle, 2001, Long-run performance and insider trading in completed and canceled seasoned equity offerings, Journal of Financial and Quantitative Analysis 36, 415-430.

Cohen, L. and D. Lou, 2012, Complicated firms, Journal of Financial Economics, 83, 383-400.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. Journal of Finance 63, 1977–2011.

Cohen, L., Frazzini, A., and Malloy, C., 2008, The Small World of Investing: Board Connections and Mutual Fund Returns, Journal of Political Economy 116, 951-979.

Cohen, Lauren, Christopher Malloy, and Lukasz Pomorski. "Decoding inside information." The Journal of Finance 67.3 (2012): 1009-1043.

Coval, Joshua, and Tobias Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, Journal of Finance 54, 2045-2074.

Coval, Joshua, and Tobias Moskowitz, 2001, The geography of investment: Informed trading and asset prices, Journal of Political Economy 109, 811- 841.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, Journal of Finance 52, 394-415.

DellaVigna, S., Pollet, J., 2006. Investor inattention, firm reaction, and Friday earnings announcements. Journal of Finance 64, 709–749.

Drake, Michael S, Darren T. Roulstone, and Jacob R. Thornock, The Determinants and Consequences of Information Acquisition via EDGAR, 2016, Contemporary Accounting Research, Forthcoming

Duffie, D., 2010. Asset price dynamics with slow-moving capital. Journal of Finance 65, 1237–1267.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, Journal of Business 62, 394-415.

Grinblatt, Mark, and Sheridan Titman, 1993, Performance measurement without benchmarks: An examination of mutual fund returns, Journal of Business 66, 47-68.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, American Economic Review 85, 1088-1105.

Gruber, Martin J., 1996, Another puzzle: The growth in actively managed mutual funds, Journal of Finance 51, 783-810.

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

Hong, H., Stein, J., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Hong, H., Tourus,W., Valkanov, R., 2007. Do industries lead the stock market? Journal of Financial Economics 83, 367–396.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, Journal of Finance 60, 2801-2824.

Hou, K., 2007, Industry information diffusion and the lead-lag effect in stock returns, Review of Financial Studies 20, 1113–1138.

Huberman, G., Regev, T., 2001. Contagious speculation and a cure for cancer: a nonevent that made stock prices soar. Journal of Finance 56, 387–396.

Jaffe, Jeffrey, 1974, Special information and insider trading, Journal of Business 47, 410-428

Jagolinzer, Alan, 2009, SEC Rule 10b5-1 and insiders' strategic trade, Management Science 55, 224-239

Jeng, Leslie, Andrew Metrick, and Richard Zeckhauser, 2003, Estimating the returns to insider trading: a performance-evaluation perspective, The Review of Economics and Statistic 85, 453-471.

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945-1964, Journal of Finance 23, 389-416.

Kahle, Kathleen, 2000, Insider trading and the long-run performance of new security issues, Journal of Corporate Finance 6, 25-53.

Ke, Bin, Steven Huddard, and Kathy Petroni, 2003, What insiders know about future earnings and how they use it: Evidence from insider trades, Journal of Accounting and Economics 35, 315-346.

Lakonishok, Josef, and Inmoo Lee, 2001, Are insiders' trades more informative?, Review of Financial Studies 14, 79-111.

Lee, Charles M. C., Paul Ma, and Charles Wang, "Search-Based Peer Firms: Aggregating Investor Perception through Internet Co-Searches,"Journal of Financial Economics 2015, Vol. 116, Issue 2, Pages 410-431.

Lin, Ji-Chai, and John Howe, 1990, Insider trading in the OTC market, Journal of Finance 45, 1273-1284.

Lorie, James, and Victor Niederhoffer, 1968, Predictive and statistical properties of insider trading, Journal of Law and Economics 11, 35-53.

Loughran, Tim and McDonald, Bill, The Use of EDGAR Filings by Investors (December 3, 2016), Journal of Behavioral Finance forthcoming.

Malkiel, Burton G., 1995, Returns from investing in equity mutual funds 1971-1991, Journal of Finance 50, 549-572.

Marin, Jose, and Jacques Olivier, 2008, The dog that did not bark: Insider trading and crashes, Journal of Finance 63, 2429-2476.

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

Piotroski, Joseph, and Darren Roulstone, 2005, Do insider trades reflect contrarian beliefs and superior knowledge about cash flow realizations?, Journal of Accounting and Economics 39, 55-81

Rozeff, Michael, and Mir Zaman, 1988, Market efficiency and insider trading: New evidence, Journal of Business 61, 25-44

Seyhun, H. Nejat, 1986, Insiders' profits, costs of trading, and market efficiency, Journal of Financial Economics 16, 189-212.

Seyhun, H. Nejat, 1988, The information content of aggregate insider trading, Journal of Business 61, 1-24.

Wermers, Russ, 1997, Momentum investment strategies of mutual funds, performance persistence, and survivorship bias, Working paper, University of Colorado.

Table 1. Summary Statistics on the EDGAR web traffic

This table presents the sample coverage used in our primary analyses. The first column of this table provides information on the user request. We report the total number of user requests as well as requests for the top 5 forms. The second and third column reports the total number of unique CIKs and Compustat firms that received a user request, where CIKs are the primary unique identifier for the EDGAR. Insider trading forms include form 3, 4, and 5 and are required to be filed by every director, officer or owner of more than ten percent of a class of equity securities. Form 3 is the initial filing form. Form 4 reports the changes in holdings. Form 5 reports the annual statement of insider holdings. 8-Ks are used to disclose information to the public about financial and operating conditions that change in between the periodic material information events. These events include change in management (e.g. director, executive turnover) and other important events such as acquisition, bankruptcy, and labor layoffs. 10-K forms are audited annual report of a firm's annual business operations. 10-Qs are unaudited financial statements. Form 6K are used by foreign private issuers.

	Form Access (in millions)	Unique CIK Codes	Unique Compustat Firms
Total	37.65	395,811	24,891
Insider Filings	5.88	184,793	13,461
8-K	6.28	32,380	17,273
10-K	4.51	27,313	15,071
10-Q	4.99	22,678	16,202
6-K	0.88	2,474	1,862

Table 2. Summary Statistics
This table reports the summary of the matching procedure outcome.

	Mean	Std	Min	Median	Max	N
Observed Quarters Per IP	3.87	5.18	1	2	30	1,618
Observed Quarters Per Mgrno	5.46	6.83	1	3	30	383
Number of Securities Tracked Per IP/QTR	41.95	223.15	1	3	2,463	6,254
Number of Securities Tracked Per MGRNO/QTR	104.15	385.05	1	5	2,682	2,091
Number of Ips Per Qtr	208.47	47.65	95	209	280	30
Number of Mgrnos Per Qtr	69.70	12.21	41	70	92	30
Weight of Tracked Sells Per Mgrno/Qtr	8.98%	17.50%	0.00%	2.00%	1.00%	2091
Number of Insiders Per Security/MGRNO/QTR Observation	2.01	1.68	1	1	32	217,780

Table 3. Top 30 Linked Institutions

This table reports the top 30 institutions (in terms of total value of the portfolio last observed) that we are able to link to an IP address.

	Mgrno	Institution Name	Total Value of Portfolio
1	90457	VANGUARD GROUP, INC.	$1,344,548,411,848.90
2	27800	FIDELITY MANAGEMENT	$725,200,575,428.07
3	71110	T. ROWE PRICE ASSOCIATES, INC.	$453,427,103,470.14
4	91910	WELLINGTON MANAGEMENT CO	$358,216,083,562.33
5	55390	MELLON BANK NA	$356,260,626,015.85
6	58835	J.P MORGAN CHASE & CO.	$318,920,960,270.88
7	10586	AMVESCAP PLC LONDON	$243,324,228,800.61
8	39300	FRANKLIN RESOURCES INC	$202,455,950,199.15
9	41260	GOLDMAN SACHS & COMPANY	$195,875,918,775.99
10	58950	MSDW & COMPANY	$195,645,104,691.13
11	62890	BANK OF AMERICA CORPORATION	$162,832,627,544.76
12	23000	DIMENSIONAL FD ADVISORS, INC.	$147,787,218,282.19
13	25610	AXA FINANCIAL, INC.	$137,179,724,006.76
14	65850	WELLS FARGO & (NORWEST CORP)	$103,778,616,033.34
15	48360	JENNISON ASSOCIATES LLC	$103,020,767,464.88
16	71200	PRIMECAP MANAGEMENT COMPANY	$94,650,141,887.34
17	91430	BLACKROCK INVT MGMT (UK) LTD.	$93,855,928,757.16
18	63050	NEUBERGER BERMAN, LLC	$91,974,624,138.91
19	11386	BLACKROCK ADVISORS, LLC	$88,448,401,719.39
20	47833	AMERICAN CENT INVESTMENT	$84,105,516,452.61
21	8100	MEWHINNEY&STRAUS	$71,881,728,789.38
22	12000	CALIFORNIA PUBLIC EMP RET SYS	$71,194,008,483.36
23	67470	OPPENHEIMERFUNDS, INC.	$69,518,405,083.30
24	91100	WADDELL & REED INVT MGMT CO	$69,494,354,719.60
25	81120	STATE FARM MUT AUTOMOBILE INS	$68,727,063,162.71
26	76760	CHARLES SCHWAB INVT MGMT, INC.	$64,691,445,195.45
27	43485	HARRIS ASSOCIATES L.P.	$60,199,891,873.38
28	89310	U.S. TRUST COMPANY N.A.	$59,680,142,694.96
29	67600	PNC FINL SERVICES GROUP INC	$57,082,906,339.17
30	37700	WACHOVIA CORPORATION	$56,417,822,808.96

Table 4. Persistence of Insider Tracking by Institutions

In Panel B of this table, we report the persistence of institutions monitoring of insider activities. We regress a dummy (Tracked Firm at t+1) that indicates whether the institutions tracked the insider filing on its lagged value and a set of fixed effects, such as portfolio, year-stock and year x stock x insider fixed effects. We require the institutions to be matched in consecutive quarters. These tests use 28 quarters of data following second quarter of 2006, due to the missing data between Q4/2005 and Q3/2006. T-statistics, clustered by quarters, is reported in parenthesis. Panel B reports the unconditional tracking probabilities calculated as the probability of an institution tracking the insider filings of any firm in its portfolio.

Panel A.

	Tracked Firm at t+1			Tracked Insider at t+1		
Tracked Firm at t	0.413	0.259	0.255			
	(30.92)	(20.90)	(20.75)			
Tracked Insider at t				0.187	0.142	0.133
				(24.72)	(23.52)	(21.48)
Portfolio Fixed Effect	No	Yes	Yes	No	Yes	Yes
Year x Stock Fixed Effect	No	No	Yes			
Year x Stock x Insider Fixed Effect				No	No	Yes
R2	0.19	0.272	0.28	0.04	0.08	0.10
N	1,338,919	1,338,919	1,338,919	11,190,087	11,190,087	11,190,087

Panel B.

Unconditional Tracking	4.80%	1.13%

Table 5. Trading Direction

In Panel A of this table, we report the conditional and unconditional probability of fund buying/selling events. The unit of observation is a position per fund quarter observation weighted by position size. The sample includes 149183 observations for 350 funds, between Q1 2005 and Q1 2013. In Panel B, we report the estimates of an ordered logit model where we regress direction of the trade on *Inside Sell*, *All Inside Sell* and *Weight*. *Inside Sell* is an indicator variable for when the insiders tracked by the fund sells shares in aggregate. *All Inside Sell* is an indicator variable for when the insiders of a firm sells in aggregate over the past year, and *Weight* is the weight of the position.

Panel A.

	TNA Weighted Average	Weight Weighted
P(Fund Sells stock)	40%	56.7%
P (Fund Buys stock)	53%	35.5%
P (Fund Sells stock| Sell by Tracked insiders)	47%	57.4%
P (Fund Buy stock| Sell by Tracked insiders)	52%	28.4%
P (Fund Sells stock | Sell by Untracked insiders)	41%	57.0%
P (Fund Buys stock | Sell by Untracked insiders)	53%	35.6%

Panel B.

	Direction
Inside sell	-0.291
	(-5.29)
All inside sell	0.021
	(0.57)
Weight	-2.463
	(-4.65)
Year Qtr FE	Yes
R2	0.001
N	149,183

Table 6. Returns

Panel B of this table reports the raw, DGTW adjusted and 4-factor adjusted returns of various portfolios. Panel A describes the timeline of procedure we follow to form these portfolios. T-statistics, clustered at the quarterly level, is reported in parenthesis.

Panel A.

Timing Horizon:	(t-2, t-1)	(t-1, t)	(t, t+1)
Event:	Viewing Record	Trading	Returns

Panel B.

Equal Weighted Portfolios Averaged by TNA	% of assets	Raw Return	DGTW-Adjusted	(Quarterly Adjusted) 4 Factors Adjusted
All Portfolios	100%	2.64%	-0.11%	0.30%
		(1.36)	(0.45)	(1.94)
Untracked Portfolios	93%	2.66%	-0.12%	0.28%
		(1.35)	(0.46)	(1.58)
Tracked insider selling vs. everything else	7%	-1.28%	-1.18%	-1.25%
		(2.91)	(3.59)	(3.49)
Tracked and sold vs. everything else	5%	-1.70%	-1.85%	-1.72%
		(2.37)	(2.73)	(2.59)
Tracked and sold vs. Not Tracked and sold	5%	-1.73%	-1.85%	-1.73%
		(2.46)	(2.77)	(2.69)
Tracked and sold vs. Tracked and not sold				
Tracked insider selling vs. Not Tracked Insider selling	5%	-1.23%	-1.20%	-1.17%
		(3.32)	(3.80)	(3.43)

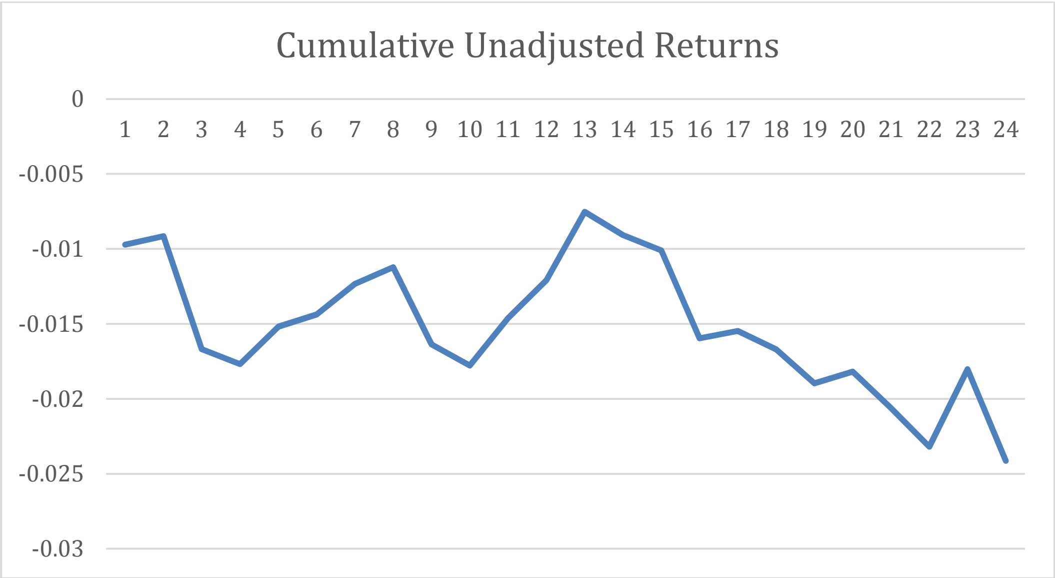

Figure 1. This figure illustrates the cumulative unadjusted returns.

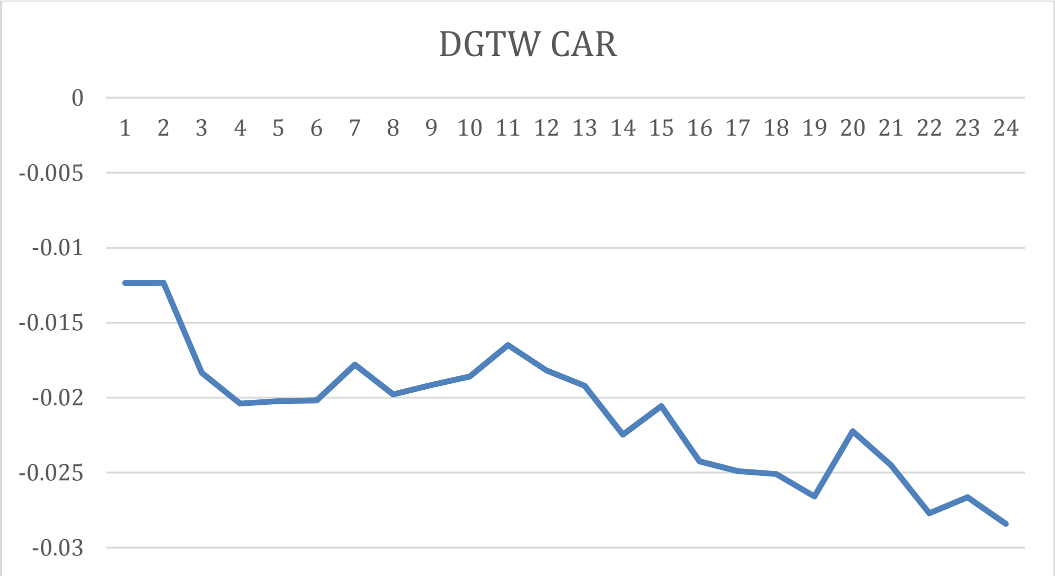

Figure 2. This figure illustrates the DGTW cumulative abnormal return.