NBER WORKING PAPER SERIES

IQ FROM IP: SIMPLIFYING SEARCH IN PORTFOLIO CHOICE

Huaizhi Chen
Lauren Cohen
Umit Gurun
Dong Lou
Christopher Malloy

Working Paper 24801
http://www.nber.org/papers/w24801

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2018, Revised April 2019

We would like to thank Anup Agrawal, Kenneth Ahern, Gurdip Bakshi, Gennaro Bernile, Bhagwan Chowdry, Eugene Fama, Fangjian Fu, William Goetzmann, Zhiguo He, Petri Jylha, Tobias Moskowitz, Jun Pan, Christo Pirinsky, Andrew Weinstock, Russell Wermers, and seminar participants at the Chicago Booth School of Business, Chinese University of Hong Kong, King's College London, Renmin University, Shanghai Advanced Institute for Finance, University of International Business and Economics, University of Massachusetts at Amherst, Singapore Management University, University of Southern California, Oppenheimer Funds, PanAgora Asset Management, Arrowstreet Capital, the 2017 University of Miami Behavioral Finance Conference, the 2018 American Finance Association Meetings in Philadelphia, the 2018 UCLA Anderson Finance Conference, the 2018 QuantCon Conference in New York City, the 6th Luxembourg Asset Management Summit, and the 2018 Taiwan Finance Association Meetings in Taipei. We are grateful for funding from the National Science Foundation and the Paul Woolley Center at the London School of Economics. Please send correspondence to: Lauren Cohen, Harvard Business School, Baker Library 279, Soldiers Field, Boston, MA 02163, phone: 1-617-495-3888, email: lcohen@hbs.edu. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

IQ from IP: Simplifying Search in Portfolio Choice
Huaizhi Chen, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher Malloy
NBER Working Paper No. 24801
July 2018, Revised April 2019
JEL No. G02,G11,G14,G23

ABSTRACT

Using a novel database that tracks web traffic on the SEC's EDGAR server between 2004 and 2015, we show that institutional investors gather information on a very particular subset of firms and insiders, and their surveillance is very persistent over time. This tracking behavior has powerful implications for their portfolio choice, and its information content. An institution that downloaded an insider-trading filing 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. Moreover, the average tracked stock that an institution buys generates annualized alphas of over 12% relative to the purchase of an average non-tracked stock. We find that institutional managers tend to track members of the top management teams of firms (CEOs, CFOs, Presidents, and Board Chairs), and tend to share educational and location-based commonalities with the specific insiders they choose to follow. Collectively, our 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.

Huaizhi Chen
University of Notre Dame
238 Mendoza
College of Business
hchen11@nd.edu

Lauren Cohen
Harvard Business School
Baker Library 273
Soldiers Field
Boston, MA 02163
and NBER
lcohen@hbs.edu

Umit Gurun
University of Texas at Dallas
School of Management
800 W Campbell Rd. SM41 75080
Richardson, TX
umit.gurun@utdallas.edu

Dong Lou
Department of Finance
London School of Economics
Houghton Street
London, WC2A 2AE
United Kingdom
d.lou@lse.ac.uk

Christopher Malloy
Harvard Business School
Baker Library 277
Soldiers Field
Boston, MA 02163
and NBER
cmalloy@hbs.edu

I. Introduction

There is a fundamental search problem inherent in portfolio choice. Moreover, in light of 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. These forces create 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?

Exactly how investors approach this foundational problem has remained largely a black box to researchers. We view our paper as an important step in the direction of peaking behind this curtain, looking under the hood of this process, and providing a novel window into - as well as micro-level foundations with regard to - exactly what large portfolio managers do in the search process; we explore how it is conducted, and what impact this has on their observable portfolio choices.

In particular, 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 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 (i.e., firms where the fund manager downloads a key filing) 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 institutional investors (through their download behavior on the SEC's website—which we are able to map to the IP addresses of institutional investment firms) 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. 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 acquisition 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 the explicit transmission of material non-public information, it might be difficult (or prohibitively costly in any scalable manner) for an institution to gain a comparative advantage in interpreting 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 could feasibly 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 related 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.

To better understand our approach, consider the following example firm and investment manager tracking and trading. AOK Inc. is a large, publicly traded firm on the NYSE, which in 2018 operated in all 50 states, with over 1,000 stores throughout North America and Asia - shipping to over 80 countries worldwide. The firm was widely held in large positions by a number of institutional investors. Moreover, the firm had a number of insiders. Two of these insiders, Mr. Sampson and Dr. Jenkins (both independent board members), were actively trading over their times at AOK Inc.

In 2013, Mr. Sampson – who lived close to the firm's Headquarters – made an unusually large dump of AOK Inc. stock. This trade was reported to the SEC, with the trade showing up on the SEC's site on July 30, 2013. The next day (July 31), mutual fund manager Mr. Thompkins – manager of a large active equity fund located close to Mr. Sampson – checked Mr. Sampson's trade, downloading the file from the SEC's site. This was usual behavior for Mr. Thompkins, who had made a habit of keeping particularly close tabs on the trades of Mr. Sampson. Mr. Thompkins, followed this tracking the very next day (August 1) with an unusually large sale of his position in AOK that he held in his fund for three years. This turned out to be a very savvy move. Following Mr. Sampson's insider sale – echoed by Mr. Thompkins dumping – AOK plummeted in the following months. It dropped 8% in the month following, and burned down nearly 28% in the quarter following the trade. AOK was downgraded by two large banks covering the firm, then on its next earnings announcement missed on profits, revenues, and same store sales, along with guiding downward.

Now consider the trading of Dr. Jenkins. Dr. Jenkins was also an independent director at AOK, a few years following Mr. Sampson. In March of 2015, she made an untimely purchase of AOK on March 14. Quickly thereafter, a fund manager, Mr. Rothman, tracked this trade and made an outsized purchase in the same direction. Now, not only were Mr. Rothman and Dr. Jenkins located closely to each other, they also shared the same alma mater, Harvard Business School. In the month following Dr. Jenkins trade, and the paired tracked buy by Mr. Rothman, AOK had a large number of positive return realizations. It marched up over 7% the first month, and climbed over 41% in the quarter following the buys. Moreover, it upped guidance, picked up initiation by a new investment bank, and beat on both earnings and same store sales at its next earnings announcement.

Had one taken the simple strategy of replicating Mr. Rothman's and Mr. Thompkins' tracked trades, (which we term as "tracked buys" and "tracked sells" throughout the paper), one would have made abnormal returns of over 7% in the following month, in each case, and even larger returns accruing in the subsequent months. Moreover, these returns did not reverse, as - in sharp contrast to any overreaction pattern - they represented fundamental information that was revealed and incorporated into prices (e.g., real quantities of same store sales and profit margins).

In this paper we demonstrate that the above example of AOK Inc. – and its tracking fund managers - represents a much more systematic pattern across the entire universe of investment manager behavior and informed trade profitability. We are able to systematically and predictably identify institutional profitable trading on their "tracked" firms and insiders throughout our sample period (2004-2015), using novel data from the Securities and Exchange Commission. Further, our classification scheme is able to richly identify informed trading even out of seeming identical behavior (i.e., informed buying and selling vs. all other buying and selling). These abnormal returns exist through the present day, and are even slightly *stronger* in point estimate in the most recent period.

In order to do this for the full sample, we first document that institutional managers have a very specific set of firms (and insiders) that they track. We then find that their tracking and monitoring activities, again measured using their downloading behavior, have powerful implications for their portfolio choices and future returns. For instance, the fact that an institution downloaded an insider-trading filing 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 = 14.37) 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 similarly strong persistence at the individual insider-level tracking. For instance, we observe an 18.7% (t = 9.20) increase in the probability of downloading, say, Jamie Dimon's insider trading filing if the manager downloaded the same insider's filing the prior quarter.

We show that these "tracked" insider trades have predictability for future firm operations and returns. In particular, when an institution buys a stock following a tracked insider buy, this stock outperforms other institutional purchases by a risk-adjusted 4-factor alpha of 300 basis points per quarter on average (t=2.10), or 12% annualized abnormal return per year. Similarly, when an institution sells a stock following a tracked insider sell, this portfolio underperforms other institutional sales by a risk-adjusted 4-factor alpha of 190 basis points per quarter (t=3.48), or over 7% annualized abnormal return per year. These returns are unaffected by known risk determinants or factor model adjustment chosen (i.e. 4-factor and DGTW).

In addition, if the results we find reflect institutional managers exhibiting a true comparative advantage in their tracked stocks, we might expect these managers to know when *not* to "follow" the tracked insiders' behavior. 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 exactly this pattern in the data. Firms in which institutions buy alongside tracked insider purchases tend to outperform those in which tracked insiders buy but the institutions choose *not* to buy alongside the insiders; we find analogous results on the sell side as well. A final counterfactual we explore is the returns to following all insider buys and sells; we show that the returns we document are not simply earned by following all insider buys (or sells), but rather that institutions' tracking behavior appears to allow them to engage in valuable stock selection above and beyond mimicking all insider trades.

There are a number of characteristics that have been shown in the literature to predict more profitable insider trades. It could be that our results on tracked insider trades are simply repackaging these past results. In order to test against this, we split the universe of trades by these firm and insider characteristics. We show that institutions' ability to outperform on their tracked insider trades holds across: large and small firms, value and growth firms, high and low past return firms, and high and low turnover firms. We then turn to the trades of CEOs and other top officers along with opportunistic insider trades – both of which have been shown to be especially informative. We show that managers do have the ability to track and trade even within these especially profitable subsets of trades. In particular, they appear able to choose precisely *which* CEO, CFO, top officer and opportunistic trade are most profitable relative to all other CEO, CFO, top officer and opportunistic trades.

Importantly, we show that the outperformance that we document on these tracked trades persists. The abnormal returns do not immediately reverse, 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.

Further, since our institutional holdings data is primarily quarterly in nature – as opposed to the SEC download files and insider transaction data which both contain precise timestamps – we also use more granular data from ANcerno on the daily trades of institutional investors to directly map the timing of institutional trades to the timing of insider transactions. We show that institutions that trade in the direction of recent insider transactions do so relatively quickly, with a large proportion of these trades (almost 50%) occurring within 30 days after the insider trade; far less occurs before the insider trades, or after this 30-day window. Moreover, a sizable percentage of the return occurs in the few days/weeks following the tracked trade, underscoring the importance of real-time tracking in this linked relationship between the institution and a given insider.[1]

We explore the mechanism at work behind our results in a variety of ways. First we show that institutional managers tend to track members of the top management teams of firms (CEOs, CFOs, Presidents, and Board Chairs) and accountants, and shy away from tracking outside directors and insiders with PhDs. Next we take our institutional holdings data and isolate the fund managers within each institutional investment firm whose holdings correlate most with the insider trades tracked by their fund company, and explore these "cherry picking" fund managers in greater depth. We find that these cherry picking fund managers are more likely to have an educational or location-based link to the insider in question. These findings are consistent with the idea that fund managers choose to track and mimic the trades of the specific insiders who not only possess the

[1] We also explore alternate (conservative) timing conventions to show the robustness to these alternative conventions.

most valuable information, but also those with whom they have lower-cost channels for obtaining private information.

Our finding of strong outperformance by both tracked insider buys and insider sales relates to the larger literature on insider trading. The literature has found a few systematic empirical facts regarding insider trading profitability. Insider buys are followed by, on average, systematic positive abnormal returns of roughly 50-100 basis points in the following month (this literature dates back to Lorie and Niederhoffer (1968) and Jaffe (1974)). However, the literature has generally found little predictability associated with the average insider sale (Jeng et al. (2003) and Lakonishok and Lee (2001)). Reasons for this result include potential liquidity, diversification, and other motives that could the information content of insider selling behavior. The reason this dynamic is critical is that insider sales – with their average statistically zero return – make up over 80% of all insider trades (Seyhun (1988) and Cohen et al. (2012)). Thus, in only finding robust predictability of insider buys, this comprises less than 20% of all insider trading activity. Importantly, given that we find evidence in the paper of profitable tracking and trading of both insider buying and insider selling, it suggests that fund managers appear able to exploit the rich information in the entirety of insider trading (and not solely the less than 20% linked to buys).

II. Literature Review

Our work relates to 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.[2] 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 (2008), 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, Coles, and Lemmon (2000), Lakonishok and Lee (2001), and Marin and Olivier (2008)) focus on the 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 (2012) 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

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

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 (2007), 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." 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. Bozanic, Hooppes, Thornock, and Williams (2017) focus on IRS use of the EDGAR data; Gibbons, Iliev, and Kalodimos (2018) explore sells-side analysts' downloading activity; Iliev, Kalodimos, and Lowry (2018) examine mutual fund family downloads of proxy filings and the governance implications of this type of monitoring. While all of these papers explore the SEC logfile data in various ways, none of them explicitly link the users of the data to individual institutional investors (i.e., the 13F filers that feature in our analysis) in order to explore the investment implications of SEC searches in a direct and granular way.

III. Data and the Setting

We combine data from a variety of sources in order to execute the empirical tests in this paper. We use CRSP to obtain stock related information, the Thompson Reuters Insiders database to obtain insider transactions, and Thompson Reuters Ownership data to obtain stock holdings of fund families. We also construct three unique and novel datasets using various sources. In this section, we describe in detail how we construct these three datasets, which entail: (1) matching IP addresses to 13-F organizations, (2) matching 13-F organization names to daily trades reported in the ANcerno database (described further below), and (3) collecting biographical background information both for corporate insiders and mutual fund managers.

A. Matching IP Addresses to 13-F Organizations

To match IP address to 13-F organizations, we follow a four-step procedure. First, we obtain the data on SEC 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 onwards, and are posted by the SEC on a quarterly basis with a 6-months delay. We use mainly the data logs from 2004 to 2015 due to the limitation of the IP geolocation data.

The IP addresses in the dataset are partially anonymized using a static cipher. The data describe the access of filings by different IP addresses. A standard IP address would have the form ###.###.###.###. Each octet ### separated by periods has a numerical value between 0 and 255. Instead, an IP address in the EDGAR data has the form ###.###.###.&&&. The first 9 digits of the EDGAR server IP correspond to the actual IP that visited the website (024.145.236). The last 3 digits correspond to a cipher (*jcf*). 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).

The static ciphers in the EDGAR log are 3-digit alphabetical codes that uniquely represent octets between 0 and 255 throughout the data. This is apparent as the same IP with the same cipher would visit similar forms for the same firm in multiple days and months; and ciphers with the lowest frequency tend to be the same two: *ghf* and *ghg* throughout the data. We ended up matching *ghf* and *ghg* to 0 and 255 respectively. IP addresses ending with 0 and 255 tend to be reserved for network administrative purposes, and are generally less frequent than other addresses.

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 which 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* will have one match to 111. Since there are multiple matches between a cipher and a set of octets, we compose a scoring system that counts the number of times each cipher is matched to each octet. The most frequent match is our candidate for the cipher link. The link between a cipher and its most frequently matched octet is distinct for the vast majority of the ciphers (236 out of 256 ciphers). For example, *aba*'s most frequent octet match is 009, no other cipher matches to 009 as their most frequent match.

The rest of the matching, where multiple ciphers are matched to a single octet, is done through a process of elimination. For example, if both *aaa* and *aab* score 001 as their most frequent match, we distinguish the two by examining which of the two ciphers are more frequently linked to 001. If 001 has many more matches with *aaa* than *aab*, then *aaa* is matched to 001. In this case 001 is removed as a potential choice for *aab*'s matching. *aab* is then linked to its next most frequent octet. We only had to iterate this process once in order to match the remaining 20 pairs.

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. In Table 1, we illustrate the cipher table we used in order to de-anonymize the IP addresses of the 13-F organizations in our sample.

In the second step, we hand-match names of the 13-F organizations to the list of potential organizations from MaxMind by research assistants. We start off with two thousand 13-F organizations with the largest average AUM between Q1 2004 and Q4 2015. 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 the third step, we identify the link between each 13-F filings institution and their IP addresses, and 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 construct a sample for 779 unique 13-F filing institutions and the filings they track in the EDGAR.

In the final step, we scrape the insider trading filings from the SEC website in order 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 4 (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 Q1 2004 to Q4 2015.

B. Matching 13-F Organizations to Daily Trades Reported in ANcerno

In this section we describe how we match the institution identifier from the ANcerno database (managercode) to the 13-F institution identifier (mgrno), which we use to aggregate the tracking of insiders by individual IP addresses. The primary data source consists of detailed institutional stock transactions from ANcerno (formerly Abel Noser), a leading consulting firm that works with institutional investors to monitor and optimize their equity trading costs. ANcerno's clients include major pension plan sponsors United Airlines, mutual fund families (i.e., money managers), such as Fidelity Investments and Putman Investments, and a small number of brokerage firms. This dataset is also used in other studies, such as Goldstein, Irvine, Kandel, and Wiener (2009), Puckett and Yan (2011), Hu, Mclean, Pontiff, and Wang (2013), and Cohen, Lou, and Malloy (2016).

Our sample period for the ANcerno data is Q1 2004 to Q3 2011. In this period, we start off with 11,649 *managercode-year-quarter* observations in the ANcerno database. This reports the name of the institutional manager for each of its client portfolios. Using this information, we are able to manually match the trading by institutions from the ANcerno database to the available institutions gathered from EDGAR's SEC data. As a result of this procedure, we are able to identify 91 of the 779 mgrnos from the 13-F data in ANcerno.

C. Collecting Biographical Information for Corporate Insiders and Fund Managers

To collect biographical information on corporate insiders, we use the BoardEx database, which collects information about individuals who have been on the board of or assumed an executive manager role at a publicly traded firm or a major private firm. The set of personal information includes academic qualifications, current and past job positions, and memberships in professional and other groups. To collect the biographical backgrounds of mutual fund managers, we use the Morningstar database, which contains fund-level performance measures as well as manager profiles. We manually parse through the education information of fund managers after searching for each fund name in the database, and pay special attention to make sure that we capture the fund manager who is in charge of the fund at the time of the trade. In total, we have collected biographical information on 53,744 corporate insiders and 225 mutual fund managers associated with our sample of insiders and institutional portfolios.

Table 2 reports the summary statistics of the forms downloaded by matched financial institutions, and the frequency of these forms in the EDGAR database. The first three columns report the top accessed forms, their respective number of downloads by all matched institutional IP addresses, and the relative frequency of these downloads. The next three columns remove Mass Download IP addresses- IP addresses that access more than 3000 filings per day- and report the same three statistics. The last three columns are the top most filed forms, their respective number of filings, and the relative frequency of these forms in the EDGAR Database.

There were over 400 million form files requested by institutional IPs over the sample period. The two most requested filing types were corporate 8-Ks and insider trading filings (Form 4s). 8-Ks are required to be filed by firms to notify shareholders of material events transpiring at the firm. After removing mass downloads, Form 4 downloads still represent over 7% of the total downloads. Form 4s also represent over 36% of all forms filed in the EDGAR Database.

Panel B of Table 2 reports the correlation between the holding firm checked by institutions through 10-K forms, 8-K forms, and Insider Trading Forms (4, 5, 6, and their amendments). This panel shows that there is a positive correlation between an institution's tendency to check the insider filings along with the 10-K and 8-K forms. In Panel C of Table 2, we report summary statistics on the number of IPs, institutions, and the percent of their portfolio represented by the stocks they track each quarter.

Table 3 lists the top 30 active institutions (in terms of total value of the active portfolio last observed) that we are able to link to an IP address. Many of the largest fundamental stock picking investment firms are represented in this list.

IV. Empirical Results

A. Persistence in Tracking Behavior

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 3, 4, and 5) on the EDGAR server. We conduct a panel OLS regression 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 filing by a given firm in quarter t-1 increases her likelihood of downloading an insider-trading filing from the same firm by more than 41.3% (t = 14.37) 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 fixed effects including portfolio fixed effects and year-stock fixed effects. Our results remain economically large. For example, in Column 3 (with the full set of year-stock and portfolio fixed effects), the coefficient on lagged search dummy is 0.255 (t = 11.86)) — 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 = 9.20) higher likelihood of downloading an insider trading filing by the same executive in the following quarter. Again, including portfolio and year-insider fixed effects has little impact on our results. For example, in Column 6, the fully specified regression, the coefficient on lagged search behavior drops only slightly to 13.3% (t = 6.31).

B. Contemporaneous Trading of Fund Managers and Corporate Insiders

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 and examine whether institutions trade in the same direction as the insiders that they follow.

In Table 5 we report the results of panel regressions of the direction of contemporaneous trading by institutions on the direction of trading by the insiders checked by these institutions. The dependent variable, FundDirection, is equal to -1 if the institution sells the stock, 1 if the institution buys the stock, and 0 otherwise. CheckedInsiderDirection is set equal to -1 if the insiders that were checked by the institution sold in net, and 1 if the insiders bought in net. Similarly, the variable CheckedInsiderBuy is equal to 1 if the insiders checked bought in net, and CheckedInsiderSell is equal to 1 if the insiders checked sold in net. Meanwhile AllInsiderDirection is set to -1 if *all* the insiders from the firm sold in net during the quarter, and 1 if all the insiders from the firm bought in net. And again, AllInsiderBuy is set equal to 1 if all the insiders from the firm bought in net during the quarter; and AllInsiderSell is set to 1 if all the insiders from the firm sold in net during the quarter. The panel is weighted by the inverse of the number of positions in each portfolio (so that a portfolio with many positions does not dominate the regression).

As shown in the first row of Table 5, our institutions' trading behavior is highly correlated with the trading of the insiders they actively track. The coefficient on CheckedInsiderDirection is large and positive (=0.077, t-stat=4.43) on the direction of trades for existing positions. This coefficient is derived on a bidirectional basis, and implies that if the checked insider buys the stock, the probability that an institution buys this stock increases by 15% (from the unconditional likelihood of buying additional shares of an existing position of 51%); additionally, if the checked insider sells the stock, the probability that an institution sells this stock increases by 16% (from the unconditional likelihood of selling a share of an existing position 47%). From Columns 2 and 3, we see trading responses of fund managers to tracked insider buys and tracked insider sales separately from EDGAR. We control for period and fund fixed effects in these regressions. Moreover, in Columns 4-6 we control for the overall trading of the fund per quarter with quarter x institution fixed effects. The tracked insider trades remain large and significant predictors of fund manager behavior even controlling for these fixed effects. These results highlight the particularly important information that appears to be embedded in tracked insider trades for institutional trading behavior.

C. Portfolio Returns to Active Insider Tracking

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 correctly choose which firms/managers' insider trades 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 equal-weighted calendar-time portfolios over our 135-month sample period using various sorting criteria, and compute the return differentials across these portfolios.[3] The timing and the construction of these tests is as follows: a) in quarter *t*-1 we observe the trading behavior of institutions at the quarterly frequency (note that later in the paper we explore daily trading records for a subset of our sample to pinpoint the timing even more directly), as well as the exact insider trading date and corresponding download date by these institutional investors, and b) in quarter *t* we

[3] We discuss various weighting schemes, including value- and trade-weighting below.

compute the quarterly returns to the separate parts of these investors' portfolios weighed by their total net assets in quarter *t*-1. We illustrate this timing using a hypothetical example in Figure 1.

In the daily trading tests that we present later in the paper, we demonstrate that the vast majority of fund manager trades that appear around insider trades occur *after* the reported insider trade date, so we do not believe our results are driven by investors trading before the insider trades, and before they download the SEC file. However, to further allay this concern we also employ an alternate timing convention in Appendix Table A1 where we force the insider trade and download dates to occur in the quarter *prior* to any subsequent trades by the fund manager, and then explore returns in the quarter following those fund manager trades. Given that the insider trading literature has documented that the returns to following insider trades primarily accrue in the 30 days immediately after the reported trade dates (a fact we replicate in our sample), the idea of forcing our portfolio tests to wait three full months after the insider trade to measure any return benefits of insider following seems overly conservative. However, as shown in the Appendix Table A1, even using this conservative timing convention we still find evidence of abnormal returns for fund managers who trade in the same direction as their tracked insider trades.

In Table 6 Panel A we report the raw, DGTW-, and 4-factor-adjusted returns of portfolios relating to insiders who bought stocks followed through the EDGAR system. We begin by computing the average stock returns held in the institution portfolios weighed by their total net assets (TNA). As shown in the first row of Table 6, during our sample period, the average stock held within an institutional portfolio earns a marginally significant and economically small 4 Factors-adjusted return of 34bps (t = 2.16) per quarter.

We next construct a variety of counterfactuals and benchmark portfolios to isolate the information content of fund managers' purchases following insider trades that they actively follow. For instance, we divide each institution's entire portfolio into several sub-components: the first sub-portfolio – shown in row 2 – includes all institution's positions except for firms whose checked insiders bought. Row 3 then computes a long-short portfolio, which is held for one quarter, where the long portfolio consists of stocks where a fund manager downloads an insider purchase filing, and the short portfolio consists of the portfolio in row 2. This long-short portfolio, while non-trivial in magnitude (around 1.87% per quarter), is nonetheless insignificantly different from zero. It is worth noting however that this portfolio mixes both stocks that are downloaded and bought by the manager, and stocks that are simply downloaded.

However, in row 4 we now construct a long-short portfolio where we solely long stocks where an investment manager checks the insiders' buy transaction and *also purchases*, while keeping the same short portfolio in row 2. This long-short portfolio earns a statistically and economically significant alpha of 3% in the following quarter ($t = 2.08$). Columns 4-7 then break down this L-S portfolio into its L and S legs, as well as give information regarding the size of both legs. Throughout Table 6, it is clear from Columns 4-7 that the bulk of the alpha from the L-S portfolio is coming from the "tracked-trade" leg of the portfolio (L) – based on the active tracking and trading by the institutional manager.

Next, in row 5 we change the benchmark and compare the returns of a portfolio for which the fund manager checks an insider purchase and also buys the stock (the long portfolio) to the returns to all other stocks that fund managers also buy in that same quarter (the short portfolio). Note that here the benchmark consists of all other stock purchases by active fund managers. This long-short portfolio again earns 3% alpha in the following quarter ($t = 2.10$). The results here indicate that these tracked insider purchases contain information beyond the content of institutional purchases alone.

In rows 6 and 7 of Table 6 Panel A, we construct an alternate benchmark that consists of all stocks where a fund manager downloads the filing after an insider purchase, but chooses *not* to buy the stock in that quarter. As noted earlier, if our results truly reflect institutional managers exhibiting a comparative advantage in information processing, we expect these managers to know when *not* to "follow" the tracked insiders' trades. For example, if the institution can decipher that the given trade was for personal reasons (as opposed to information-based), the manager would not want to follow that trade of the tracked insider. Indeed, we find that these insider trades have less predictive ability for future returns. This shown in row 6, when we use this counterfactual as our short portfolio and use the same long portfolio as in rows 4 and 5, the long-short spread of 3.25% per quarter is economically large (and statistically significant at the 90% confidence). In row 7, if we further restrict the long portfolio to only situations where the fund manager makes a large increase in the stock position from a zero or insignificant initial position (less than 20 basis points of total portfolio weight), the long-short spread increases to 3.93% alpha per quarter.

Our final counterfactual/benchmark portfolio involves examining the returns to all stocks that are not checked by an institution but where both the insiders and the institution purchased shares. This benchmark portfolio allows us to compare the coordinated trading by institutions and their checked insiders against the potentially coincidental simultaneous trading of unchecked insiders and institutions. The portfolio of tracked insider buys with a corresponding fund manager purchase again outperforms this benchmark of all other insider purchases by up to 2.97% alpha per quarter (t=2.09). Collectively, our findings point to large and significant return predictability that stems from fund managers' active tracking of select insiders, beyond just the interaction of insider and institutional trades.

Table 6 Panel B presents the analogous results around insider sales, as opposed to insider purchases. The results are remarkably consistent with the purchase results shown in Panel A, albeit at somewhat smaller magnitudes for the L/S portfolios in rows 5 through 8—more in the 6% to 8.5% annualized range, as opposed to the 9% to 13% annualized range for the purchase results. For instance, consider row 5 where the long side requires both the institution and its tracked insiders to sell the security in question, and where the benchmark portfolio includes all holdings which are sold by the institution but not by tracked insiders; in this specification tracked-sales underperform the benchmark by 190bps (t = 3.48) in the following quarter. In row 6, we define the benchmark as including all holdings that are not sold by the institution but are sold by tracked insiders. Similar to the result in Panel A, the return differential for the tracked stocks that are sold relative to tracked stocks that institutions choose *not* to sell is economically and statistically significant. Finally, in row 8, if the counterfactual benchmark portfolio is instead defined as the set of insider sells which are not downloaded by the institution, the L/S spread is again significant at 183bps (t = 3.41).

Collectively, the results in Table 6 indicate that institutions are able to identify the most informative trades by closely tracking the trades of specific corporate insiders.

D. Portfolio Returns across Firm and Insider Characteristics

While the results in Table 6 provide evidence that institutions are able to identify, track, and trade on especially informative insider trades, the rich literature on insider trading has also established a number of characteristics that have been shown to also predict more profitable insider trades. For instance, Lakonishok and Lee (2001) show that insiders' trades appear to be more profitable in smaller firms (vs. larger firms). It could be that our results on tracked insider trades are simply repackaging these past results. In order to test against this, we split the trading universe of institutions by various stock characteristics, and explore the ability of the institutions – with their tracking of insider trades – to generate outperformance *within* each of these subsamples. We test fund managers' ability to profitably track trades across: a.) large and small firms, b.) value and growth firms, c.) high and low past return firms, d.) high and low turnover firms, and e.) high and low institutional ownership firms.

The results are shown in Table 7. Each panel represents independent double-sorts with the respective firm characteristic being considered, with the cross-section of stocks first split upon the median of each characteristic per quarter. In particular, they show the L-S return spread of (buying-selling of tracked insider trades) – (buying-selling of all other non-tracked trades) by the same managers in the given subset of the universe (similar to row 5 in Table 6). In this way, the results aim to identify solely the tracking premium amongst the given subset of firms (e.g., large firms separately from small firms).

The results in Table 7 Panels A show that fund managers' ability to outperform on their tracked insider trades broadly holds across: large and small firms, value and growth firms, high and low past return firms, high and low turnover firms.[4] The slight exception is that of institutional ownership (IO) – where the profits from tracked insider trades, though positive in both, seem to be somewhat larger in low IO firms – though this tendency is only statistically significant when the Fama-French four-factor adjustment is

[4] The *t*-stats in these double-sorted subset universes are slightly weaker than what we have in the original full-sample test, largely driven by the reduced sample size of each smaller sub-sample.

taken into account. Stepping back, the results across all five sorts bolster the idea that institutions' ability to profit on tracked insider trades broadly extends across firm characteristics.

We next move on to explore fund managers' outperformance on tracked insider trades within and across characteristics of the insiders themselves. In particular, Ravina and Sapienza (2010) find evidence of stronger outperformance by top executives (e.g., CEOs, as opposed to other insider), while Cohen, Malloy, and Pomorski (2012) show that much of the outperformance by insider trading is concentrated amongst opportunistic insider trades (as opposed to routine trades). We thus examine whether tracked insider trades can still outperform non-tracked insider trades within each of these subsets.

To test this, similar to above, we measure different types of insider trades and split the universe of measured trades into: f.) routine vs. opportunistic trades, and g.) trades by top executives (e.g., CEO, chairman) vs. trades of other insiders, and again run independent tests of managers' ability to outperform on their tracked trades within each of these universes. Here, we focus particularly on the relevance of information production by the fund managers in these cases and consequently examine the performance of L-S return spread of (buying-selling of tracked insider trades) – (buying-selling of all other non-tracked trades that coincide with untracked insiders).

Panel B of Table 7 displays our results. It is clear that managers do have the ability to track and act on the especially profitable trades of top executives and opportunistic trades. Importantly, the fund managers appear able to choose precisely which CEO, CFO, top officer and opportunistic trades are most profitable relative to all other CEO, CFO, top officer and opportunistic trades.

V. Mechanism

In this section we explore the mechanism behind our key results in greater depth. We examine the characteristics of the insiders who are tracked, the characteristics of the fund manager doing the tracking, and the characteristics of the matched insider-institution pairs, in order to investigate the drivers of the return predictability we document above.

A. Characteristics of Insiders Who Are Tracked By Institutions

First, we examine the profiles of the specific insiders who are being searched in our sample. The first column of Table 8 contains the names of the characteristics associated with the firm insiders. The second column reports the percentage of each type of profile across all BoardEx reported insiders. The third column reports the percentage of each type for profiles that are actively checked by institutional investors, and computes differences between the percentage of checked insiders relative to all insider profiles.

Table 8 indicates that institutional managers tend to track members of the top management teams of firms (CEOs, CFOs, Presidents, and Board Chairs) and accountants, and shy away from tracking outside directors and insiders with PhDs. For instance, CEO-related insider trade reports account for 28.7% of the total downloads by institutions, even though they make up only 9.8% of all insider transactions, a difference of 18.8%. We find similarly large differences for Presidents, Board Chairs, and CFOs, suggesting that institutional managers perceive these filings to be more value-relevant than ordinary director filings (which they download at a significantly lower rate than their overall incidence in the population of filings).

B. Characteristics of Fund Managers Who Track and Mimic Insider Trades

Next, we take our institutional holdings data and attempt to isolate the fund managers within each institutional investment firm whose holdings correlate most with the insider trades tracked/downloaded by their institution. We label these highly correlated fund managers as "cherry picking" fund managers. Specifically, for each portfolio *j* in the 13-F family *k*, we run the following regression:

$$FundDirection_{j,i,t} = \alpha_{j,t} + \beta_{j,t} \cdot InsiderDirection_{k,i,t} + \epsilon_{j,i,t}$$

- $FundDirection_{j,i,t}$ is in the set (-1, 0, 1) which indicates the direction of the trading by the portfolio *j* on stock *i* in quarter *t*.
- $InsiderDirection_{k,i,t}$ is in the set (-1, 0, 1) which indicates the direction of an indicator for whether the insiders at firm *i* observed by the 13-F family *k* had in net sold shares at *t*.
- If $\beta_{j,t}$ is > 90% significance level, we call the fund *j* at *t* a Cherry Picker.

We then examine the geographic and educational characteristics of these fund managers in Table 9. Panel A of Table 9 reports the distribution of top locations (by state) of all matched mutual fund managers, and specifically the fund managers we dub to be cherry pickers. Panel A shows that cherry pickers are more likely to reside in Massachusetts, which has a high concentration of fundamental stock pickers in the Boston area (e.g., Fidelity, Wellington, etc.), and also Maryland (where T. Rowe Price and many other institutional stock selection houses are located). Panel B then records the distribution of education backgrounds of matched mutual fund managers relative to the cherry pickers. The universities with the highest absolute differences between manager and cherry picker distributions are reported.

C. Exploring the Source of Fund Manager - Insider Links

In Table 10 we try to pinpoint the mechanism more cleanly by focusing on the manager-insider links that we observe in the data. Specifically, we try to better understand why certain fund managers choose to follow certain corporate insiders. To do so, we explore the connection links between a cherry picking mutual fund manager and an insider, based on commonalities in location and educational backgrounds. Again, we define cherry pickers as fund managers whose trades each quarter correlate with net checked insider trading at the 90% significance level. The Match Indicator is a dummy variable for whether the cherry picker had searched for the insider within his portfolio. The sample includes all potential insiders linked by each fund manager's portfolio holdings.

Table 10 shows that while there is still much unexplained variation in the match between institutions and tracked insiders, educational background and location are significant predictors of this likelihood of a tracking match, even controlling for past matching, as well as time fixed effects, education fixed effects, and geography fixed effects. Collectively, the findings in Tables 9 and 10 are consistent with the idea that fund managers choose to track and mimic the trades of the specific insiders who not only possess the most valuable information, but also those with whom they have lower-cost channels for obtaining private information.

VI. Alternate Timing Conventions, Daily Trading Records, and Robustness

We also explore a variety of alternate timing conventions, as well as novel data on the daily trading behavior of institutions in order to verify the robustness of our findings, and solidify the interpretation of our results.

First, in Appendix Table 1, we re-run the return predictability analysis from Table 6, but impose a longer lag structure in the return tests. In Table A1 we require the insider trade and download dates to occur in the quarter *prior* to any subsequent trades by the fund manager, and then explore returns in the quarter following those fund manager trades. We note that this builds in a minimum of 90 days from the insider trade date to the beginning of the measured future return period. As noted earlier, given that the insider trading literature has documented that the returns to following insider trades primarily accrue in the 30 days immediately after the reported insider trade dates, the idea of forcing our portfolio tests to wait a minimum of 3 full months after the insider trade dates to measure any return benefits of insider following is a conservative approach.

Appendix Tables A2 and A3 show two alternative weighting schemes to the equal-weighting we include in the main analysis. In particular, we show the results using value-weighting (using the holding weights of the tracked insider positions) in Appendix Table A2, and trade-weighting (putting larger weights on larger trades by institutions following these tracked insider trades relative to their initial values) in Appendix Table A3. The results reported in Appendix Tables A2 and A3 are broadly consistent with the evidence we presented in Table 6.[5]

To further motivate the timing convention used in Table 6, we collect an alternate source of institutional holdings at the *daily* level from the ANcerno database described earlier. This data contains daily stock-level holdings and trades at the institution level for a subset of our overall sample (the description of our matching of this institutional data to the precise fund-manager level is contained in Section III). Using this data, we can granularly test how quickly institutions trade following their tracking of insider trades. Figure 2 tabulates the proportion of gross purchases of stock by institutions over a number of trading days after checking that insiders had purchased a stock in net. Figure 3 presents

[5] In Appendix Table A4 we exclude the last two years of our sample (post-2013) to address the concern that there is staleness in the Thomson reported holdings data (https://wrds-www.wharton.upenn.edu/pages/support/research-wrds/researchguides/research-note-regarding-thomson-reuters-ownership-data-issues/). The portfolio results remain very similar (large and significant).

the analogous results for sales of stocks, as opposed to purchases. Both Figures 2 and 3 paint a similar picture: most institutional trading that follows in the direction of insider trading happens quickly, within 30 days, and far less happens right before or right after that 30-day window post-insider trade. These results suggest that our return predictability results in Table 6 are highly unlikely to be an artifact of institutions trading before the insider trade filing that they subsequently download.

Moreover, we use the ANcerno sample to explore the return results of tracked trades in a setting where we can measure the tracking and precise trade timing of institutional managers. In particular, we see the exact date on which the trade is executed by the manager following the tracking of the insider trade. Although this is a much smaller sample – both in the cross-section and time-series – and so the power is lower, we still see patterns of future returns for these tracked trades that are consistent with the evidence we document using the 13-F filings. If an ANcerno institution downloads an insider form and trades in the same direction as this insider over the next 15 trading days, we classify these trades as following the insider. We plot the subsequent cumulative DGTW returns of these following trades in Figure 4.

In Figure 5 we also investigate the long-run returns to tracked insider trades by fund managers. Specifically, we show in this figure that the outperformance that we document on these tracked trades continues for a number of months following the tracked insider and institution trading in the same direction. Importantly, the cumulative returns do not immediately reverse, 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.

VII. Conclusion

With the proliferation of information signals in both quantity and dimensionality in recent decades, investors face an increasingly complex portfolio choice problem. How does an investor reduce the dimensionality of the investment problem sufficiently to know which signals to track and collect? In this paper, we provide novel evidence to shed light on this investment process – giving micro-foundations for the information collection and utilization process of large delegated portfolio managers. Specifically, using web traffic on the SEC's EDGAR server between 2004 and 2015, we find that mutual fund managers track a very particular subset of firms and insiders. In addition, fund manager tracking activity not only remains persistent over time, but also has powerful implications for their portfolio choice.

Moreover, managers seem able to know precisely *which* of their tracked insiders' trades (e.g., Jamie Dimon's trades) to follow, and which not to, as the trades they track and choose to act upon significantly outperform those that they track and choose not to trade along with. This tracked-trade outperformance holds broadly across firm characteristics (e.g., small and large firms, value and growth firms, high and low past return firms); additionally, it holds even considering only the most information-rich subsets of insider trades (e.g., opportunistic insider trades and trades by top executives). Lastly, the abnormal returns that following these tracked trades continue to accrue for the quarters following the trades and do not reverse, suggesting that the information contained in the trades is important for fundamental firm value, and is revealed and incorporated into firm value only following the information-rich tracked trades. While a number of factors likely determine the precise choice of such information flow, we propose and verify two significant predictors: close proximities and alumni-network connections between fund managers and firm insiders.

Stepping back, as the costs of producing, disseminating, collecting, and processing information continue to fall, signal proliferation will only accelerate. This will make dimensionality reduction a growing problem facing investors for the foreseeable future. We believe that our study – using novel, rich, detailed data on fund manager tracking and trading behavior – is a first step to micro-founding and understanding successful attempts to do precisely this. Future research should push ahead even further to establish alternate ways that investors can solve this problem and engage as important information collectors and price-setters in modern capital markets.

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.

Bozanic, Z., Hoopes, J. L., Thornock, J. R. and Williams, B. M. (2017), IRS Attention. Journal of Accounting Research, 55: 79-114.

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, D. Lou, and C. Malloy, 2016, Cloaked trading, Journal of Investment Consulting 17, 207–248.

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.

Gibbons, Brian and Iliev, Peter and Kalodimos, Jonathan, Analyst Information Acquisition via EDGAR. Penn State University Working Paper.

Goldstein, M.A., Irvine, P., Kandel, E. and Wiener, Z., 2009. Brokerage commissions and institutional trading patterns. The Review of Financial Studies, 22(12), pp.5175-5212.

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.

Hirshleifer, D. and Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. Journal of accounting and economics, 36(1-3), pp.337-386.

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.

Hu, G., McLean, R.D., Pontiff, J. and Wang, Q., 2013. The year-end trading activities of institutional investors: Evidence from daily trades. The Review of Financial Studies, 27(5), pp.1593-1614.

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.

Iliev, Peter and Kalodimos, Jonathan and Lowry, Michelle, Investors Attention to Corporate Governance. Penn State University Working Paper.

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.

Massimo Massa, Andrei Simonov; Is College a Focal Point of Investor Life?, Review of Finance, Volume 15, Issue 4, 2011, Pages 757–797.

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

Puckett, A. and Yan, X., 2011. The interim trading skills of institutional investors. The Journal of Finance, 66(2), pp.601-633.

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

Ravina, Enrichetta and Paola Sapienza, 2010, What Do Independent Directors Know?, Review of Financial Studies 23, 962-1003.

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, University of Colorado Working Paper.

Table 1. Cipher Table (Pre-Publication)

This table reports the mapping of IP addresses' hidden octet (code) to actual octet. The procedure we follow to identify these one-to-one mapping is as follows: We match the first 9 digits (the unciphered portion) of each IP addresses on the Edgar Server to the first 9 digits of IP addresses on a separate web trafficked server. Each hidden octet (code) is matched to each actual octet a number of times. We choose the most frequent matching actual octet as the deciphering octet. We mask the octets below, planning to make the entire table of ciphers available to all researchers post peer-review.

Code	Octet	Code	Octet	Code	Octet	Code	Octet	Code	Octet	Code	Octet	Code	Octet	Code	Octet
ghf	0	---	32	---	64	---	96	---	128	---	160	---	192	---	224
jbj	1	---	33	---	65	---	97	---	129	---	161	---	193	---	225
jdd	2	---	34	---	66	---	98	---	130	---	162	---	194	---	226
ggf	3	---	35	---	67	---	99	---	131	---	163	---	195	---	227
---	4	---	36	---	68	---	100	---	132	---	164	---	196	---	228
---	5	---	37	---	69	---	101	---	133	---	165	---	197	---	229
---	6	---	38	---	70	---	102	---	134	---	166	---	198	---	230
---	7	---	39	---	71	---	103	---	135	---	167	---	199	---	231
---	8	---	40	---	72	---	104	---	136	---	168	---	200	---	232
---	9	---	41	---	73	---	105	---	137	---	169	---	201	---	233
---	10	---	42	---	74	---	106	---	138	---	170	---	202	---	234
---	11	---	43	---	75	---	107	---	139	---	171	---	203	---	235
---	12	---	44	---	76	---	108	---	140	---	172	---	204	---	236
---	13	---	45	---	77	---	109	---	141	---	173	---	205	---	237
---	14	---	46	---	78	---	110	---	142	---	174	---	206	---	238
---	15	---	47	---	79	---	111	---	143	---	175	---	207	---	239
---	16	---	48	---	80	---	112	---	144	---	176	---	208	---	240
---	17	---	49	---	81	---	113	---	145	---	177	---	209	---	241
---	18	---	50	---	82	---	114	---	146	---	178	---	210	---	242
---	19	---	51	---	83	---	115	---	147	---	179	---	211	---	243
---	20	---	52	---	84	---	116	---	148	---	180	---	212	---	244
---	21	---	53	---	85	---	117	---	149	---	181	---	213	---	245
---	22	---	54	---	86	---	118	---	150	---	182	---	214	---	246
---	23	---	55	---	87	---	119	---	151	---	183	---	215	---	247
---	24	---	56	---	88	---	120	---	152	---	184	---	216	---	248
---	25	---	57	---	89	---	121	---	153	---	185	---	217	---	249
---	26	---	58	---	90	---	122	---	154	---	186	---	218	---	250
---	27	---	59	---	91	---	123	---	155	---	187	---	219	---	251
---	28	---	60	---	92	---	124	---	156	---	188	---	220	---	252
---	29	---	61	---	93	---	125	---	157	---	189	---	221	---	253
---	30	---	62	---	94	---	126	---	158	---	190	---	222	---	254
---	31	---	63	---	95	---	127	---	159	---	191	---	223	---	255

Table 2. Summary Statistics

This table reports summary statistics of our sample. In Panel A, we report summary statistics of the forms downloaded by institutions, along with the proportions of forms in the Edgar database. The first three columns report the most accessed forms, their respective numbers of downloads by all matched institutions, and the relative frequencies of these downloads. The next three columns remove mass-download IP addresses – i.e., IP addresses that access more than 3000 filings per day – and report the same three statistics. The last three columns report the most filed forms, their respective numbers of filings, and their relative frequencies in the Edgar Database. Panel B reports the correlations between firms checked by institutions through tracking 10-K, 8-K, and Insider Trading Forms (3, 4, 5, and their amendments). Finally, Panel C reports the distribution of the number of matched institutions as well as their search behavior in each quarter.

Panel A. Summary statistics on the download frequency by form type

Edgar Downloads by Institutions			After Removing Mass Downloads			All Edgar Forms		
Form	#Downloads	Frequency	Form	#Downloads	Frequency	Form	# of Forms	Frequency
4	291,592,283	55.6%	8-K	19,871,337	17.8%	4	6,006,779	36.6%
8-K	68,471,817	13.0%	10-K	19,523,529	17.5%	8-K	1,413,387	8.6%
10-Q	29,551,127	5.6%	10-Q	19,468,598	17.4%	SC 13G/A	591,842	3.6%
10-K	24,935,559	4.8%	4	8,043,938	7.2%	3	552,843	3.4%
13F-HR	13,889,856	2.6%	6-K	2,563,774	2.3%	10-Q	536,219	3.3%
4/A	9,268,341	1.8%	DEF 14A	2,120,153	1.9%	497	375,547	2.3%
8-K/A	7,874,539	1.5%	424B2	2,073,561	1.9%	SC 13G	347,617	2.1%
SC 13G	6,357,204	1.2%	424B5	1,943,211	1.7%	6-K	336,734	2.1%
SC 13G/A	5,706,116	1.1%	S-1/A	1,891,509	1.7%	424B3	258,821	1.6%
6-K	3,984,856	0.8%	424B3	1,862,154	1.7%	SC 13D/A	206,492	1.3%
DEF 14A	2,847,715	0.5%	13F-HR	1,832,043	1.6%	13F-HR	202,615	1.2%

Panel B. 8K-10K Correlation

	10K	8K	Insiders
10K	1.00		
8K	0.28	1.00	
Insiders	0.11	0.17	1.00

Panel C. Number of Matched Institutions per Quarter

	Mean	Std	Min	Q1	Median	Q3	Max
Number of IPs Per Qtr	208.47	47.65	95	177	209	246	280
Number of Institution Per Qtr	69.70	12.21	41	63	70	77	92
Weight of Checked Stocks Per Mgrno/Qtr	8.98%	17.50%	0.00%	0.35%	2.00%	7.47%	100.00%

Table 3. Top 30 Linked Institutions

This table reports the top 30 matched institutions (largest in terms of portfolio value).

	Mgrno	Institution Name	Portfolio Value
1	12740	CAPITAL RESEARCH & MGMT	$644,502,498,763
2	27800	FIDELITY MGMT & RESEARCH (US)	$623,465,559,994
3	55390	MELLON BANK NA	$479,115,002,653
4	71110	T. ROWE PRICE ASSOCIATES, INC.	$435,541,810,456
5	62890	BANK OF AMERICA CORPORATION	$377,795,325,262
6	91910	WELLINGTON MGMT CO, L.L.P.	$357,154,732,394
7	11836	CAPITAL WORLD INVESTORS	$352,914,042,331
8	25610	AXA FINANCIAL, INC.	$352,084,782,849
9	90457	VANGUARD GROUP, INC.	$351,551,062,880
10	58950	MSDW & COMPANY	$316,603,792,017
11	58835	JPMORGAN CHASE & COMPANY	$316,062,544,205
12	65260	NORTHERN TRUST GLOBAL INVTS	$305,692,675,586
13	11835	CAPITAL RESEARCH GBL INVESTORS	$299,354,230,407
14	72400	PUTNAM INVESTMENT MGMT, LLC	$285,307,501,875
15	41260	GOLDMAN SACHS & COMPANY	$245,836,983,966
16	10586	AMVESCAP PLC LONDON	$237,493,511,896
17	7800	DEUTSCHE BK AKTIENGESELLSCHAFT	$226,994,949,895
18	48170	JANUS CAPITAL MANAGEMENT LLC	$225,037,238,864
19	50160	LEGG MASON INC	$215,319,086,243
20	84900	CITIGROUP INVESTMENTS INC.	$200,525,553,607
21	39300	FRANKLIN RESOURCES, INC.	$196,253,825,506
22	11371	NORGES BK INVT MGMT (NBIM)	$188,412,117,446
23	54600	MASSACHUSETTS FINL SERVICES CO	$179,313,992,099
24	10039	GEODE CAPITAL MGMT, L.L.C.	$169,824,996,700
25	8350	BERKSHIRE HATHAWAY	$153,892,447,818
26	18265	COLLEGE RETIRE EQUITIES	$145,381,192,291
27	45639	COLUMBIA MGMT INV ADVISERS LLC	$145,135,260,505
28	37700	WACHOVIA CORPORATION	$141,242,468,650
29	56780	MERRILL LYNCH CAPITAL MARKETS	$139,563,896,834
30	23270	DODGE & COX, INC.	$135,031,475,226

Table 4. Persistence of Insider Tracking by Institutions

This table examines the persistence in insider tracking behavior of institutions. In Panel A, the dependent variable in Columns 1-3 is a dummy variable (CheckedFirm at t+1) that equals one if an institution tracks insider filings from a particular firm in quarter t+1. The main independent variable of interest is CheckedFirm at t that equals one if the same institution tracks insider filings from the same firm in quarter t. In Columns 4-6, the dependent variable is a dummy (CheckedInsider at t+1) that equals one if an institution tracks filings by a particular insider in quarter t+1. The main independent variable of interest is CheckedInsider at t that equals one if the same institution tracks the same insider's filings in quarter t. We further control for portfolio, year x stock and year x stock x insider fixed effects. Panel B reports the unconditional probabilities of an institution downloading insider filings from a given firm or by a given insider. T-statistics, reported in parenthesis, are based on standard errors double clustered by quarter and firm. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Persistence of Tracking Behavior

	Checked Firm at t+1			Checked Insider at t+1		
Checked Firm at t	0.413***	0.259***	0.255***			
	(14.37)	(12.20)	(11.86)			
Checked Insider at t				0.187***	0.142***	0.133***
				(9.20)	(6.58)	(6.31)
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
Adj. R^2	0.1947	0.2696	0.2833	0.0387	0.0842	0.0958
No.	1,338,919	1,338,919	1,338,919	11,190,087	11,190,087	11,190,087

Panel B. Unconditional Tracking

	Checked Firm	Checked Insider
Unconditional Probability	4.80%	1.13%

Table 5. Contemporaneous Trading

This table reports panel regressions of the direction of contemporaneous trading by an institution on the direction of trading by the insiders tracked by the institution. The dependent variable, FundDirection, is equal to -1 if the institution sells the stock, 1 if the institution buys the stock, and 0 otherwise. The main independent variable of interest is CheckedInsiderDirection, which equals -1 if the insiders checked by the institution sell in aggregate, 1 if the insiders buy in aggregate, and 0 otherwise. CheckedInsiderBuy is equal to 1 if the insiders checked by the institution buy in aggregate and 0 otherwise. CheckedInsiderSell is equal to 1 if the insiders checked by the institution sell in aggregate and 0 otherwise. AllInsiderDirection is equal to -1 if all insiders of the firm (checked and unchecked) sell the stock in aggregate, 1 if all insiders buy in aggregate, and 0 otherwise. AllInsiderBuy is equal to 1 if all the insiders of the firm buy in aggregate and 0 otherwise. AllInsiderSell is equal to 1 if all the insiders of the firm sell in aggregate and 0 otherwise. Observations are weighted by the inverse of the number of positions held by each institution (so that a portfolio with many positions do not dominate the analysis). T-statistics, reported below the coefficients, are based on standard errors double clustered by quarter and firm. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	FundDirection (-1 for sell, 1 for buy)					
CheckedInsiderDirection	0.077***			0.071***		
	(4.43)			(4.18)		
CheckedInsiderBuy		0.159***			0.160***	
		(6.54)			(5.83)	
CheckedInsiderSell			-0.084***			-0.076***
			(-3.43)			(-3.21)
AllInsiderDirection	0.019***			0.020***		
	(3.94)			(4.21)		
AllInsiderBuy		0.035***			0.034***	
		(3.58)			(3.71)	
AllInsiderSell			-0.042***			-0.044***
			(-4.37)			(-4.66)
Time FE	Yes	Yes	Yes	No	No	No
Institution FE	Yes	Yes	Yes	No	No	No
Time x Institution FE	No	No	No	Yes	Yes	Yes
Adj. R^2	0.0265	0.0266	0.0264	0.0908	0.0909	0.0907
No.	2,469,803	2,469,803	2,469,803	2,469,792	2,469,792	2,469,792

Table 6. Portfolio Returns

This table reports calendar-time portfolio returns of various types of institutional holdings and trading. Panel A corresponds to tracked insider buys and Panel B corresponds to tracked insider sells. In terms of timing, we require that institutions both view the insider trading record and potentially trade the underlying stock in quarter t. We then analyze the equal weighted returns to those trades in quarter t+1. Row 1 shows the returns of the average stock held within an institutional portfolio. Row 2 includes all holdings where the institution does not check/download a Form 4 filing. Row 3 then computes a long-short portfolio, where the long portfolio consists of stocks of which a fund manager downloads an insider trading filing, and the short portfolio consists of the portfolio in row 2. Row 4 computes a long-short portfolio where the long portfolio consists of all stocks of which a fund manager checks the insiders' transaction and also trade shares in the same direction during the same quarter, while keeping the same short portfolio represented in row 2. Row 5 constructs a long-short portfolio, where the long portfolio includes all stocks of which the fund manager checks insider trades and also trade in the same direction as the insider, the short portfolio includes all other stocks that the fund manager buys/sells in the same quarter. In rows 6 and 7, we then construct an alternate benchmark that consists of all stocks where a fund manager downloads the filing after an insider trades, but chooses not to trade in the same direction in that quarter. In Row 8, the benchmark portfolio includes all stocks that are not checked by a fund manager but are traded by insiders in the same direction as the fund manager. Reported below are the quarterly excess, DGTW, and 4-Factor adjusted returns of these aforementioned portfolios (as well as the long and short sides). T-statistics, reported in parenthesis, are based on Newey-West standard errors.

Panel A: Tracked Insider Buys	Excess Returns	DGTW	4-Factor Alpha	L % of Assets	L 4F Alpha	S % of Assets	S 4F Alpha
1) All Positions	2.78%	0.22%	0.34%	100%			
	(1.92)	(2.02)	(2.16)				
2) All Positions Except Checked Insider Buying	2.76%	0.22%	0.33%	98.7%			
	(1.93)	(1.93)	(2.14)				
3) Checked Insider Buying vs 2)	1.87%	2.02%	1.77%	1.28%	2.10%	98.7%	0.33%
	(1.56)	(1.84)	(1.57)		(1.84)		(2.14)
4) Checked Insider Buying and Bought vs 2)	2.96%	2.90%	3.00%	0.68%	3.33%	98.7%	0.33%
	(2.02)	(2.12)	(2.08)		(2.30)		(2.14)
5) Checked Insider Buying and Bought vs.	2.97%	2.90%	3.00%	0.68%	3.33%	49.8%	0.33%
Rest Bought	(2.03)	(2.13)	(2.10)		(2.30)		(1.93)
6) Checked Insider Buying and Bought vs.	3.68%	2.55%	3.25%	0.68%	3.33%	0.61%	0.08%
Checked and Not Bought	(2.13)	(1.63)	(1.91)		(2.30)		(0.08)
7) Checked Insider Buying and Bought vs.	4.28%	2.90%	3.93%	0.15%	3.80%	0.15%	-0.14%
Checked and Not Bought (Zero Initial Positions)	(2.39)	(1.79)	(2.23)		(2.46)		(-0.13)
8) Checked Insider Buying and Bought vs.	3.02%	2.97%	2.97%	0.68%	3.33%	2.65%	0.35%
Not Checked Insider Buying and Bought	(2.17)	(2.22)	(2.09)		(2.30)		(0.95)

Panel B: Tracked Insider Sales	Excess Returns	DGTW	4-Factor Alpha	L % of Assets	L 4F Alpha	S % of Assets	L 4F Alpha
1) All Positions	2.77%	0.23%	0.34%	100%			
	(1.93)	(2.22)	(2.25)				
2) All Positions Except Checked Insider Selling	2.78%	0.24%	0.36%	91.9%			
	(1.93)	(2.36)	(2.32)				
3) Checked Insider Selling vs 2)	-1.09%	-1.12%	-1.14%	8.01%	-0.78%	91.9%	0.36%
	(-2.56)	(-2.64)	(-2.71)		(-1.85)		(2.32)
4) Checked Insider Selling and Sold vs 2)	-1.83%	-1.62%	-1.89%	4.46%	-1.54%	91.9%	0.36%
	(-3.40)	(-3.07)	(-3.50)		(-2.83)		(2.32)
5) Checked Insider Selling and Sold vs.	-1.85%	-1.63%	-1.90%	4.46%	-1.54%	48.0%	0.36%
Rest Sold	(-3.42)	(-3.07)	(-3.48)		(-2.83)		(2.41)
6) Checked Insider Selling and Sold vs.	-1.06%	-1.12%	-1.49%	4.46%	-1.54%	3.55%	-0.05%
Checked and not Sold	(-1.51)	(-1.78)	(-2.15)		(-2.83)		(-0.12)
7) Checked Insider Selling and Sold vs.	-1.64%	-1.71%	-2.10%	0.79%	-1.77%	0.70%	0.34%
Checked and not Sold (Small Initial Positions)	(-1.95)	(-2.23)	(-2.53)		(-2.71)		(0.65)
8) Checked Insider Selling and Sold vs.	-1.78%	-1.62%	-1.83%	4.46%	-1.54%	45.3%	0.29%
Not Checked Insider Selling and Sold	(-3.41)	(-3.07)	(-3.41)		(-2.83)		(1.98)

Table 7. Portfolio Returns across Firm and Insider Characteristics

This table reports calendar-time portfolio returns of various types of institutional holdings and trading using different subsample. Panel A classifies all stocks into two halves based on whether they are above or below the median lagged market value, book-to-market ratio, past one-year return, share turnover, or institutional ownership respectively. We report the long-short spread, which is (buying-selling of tracked insider trades) – (buying-selling of all other non-tracked trades) by institutions (i.e. Row 5 of Panel A minus Row 5 of Panel B of Table 6). Panel B classifies all insider trades to routine vs. opportunistic (non-routine) trades (following Cohen, Malloy, and Pomorski, 2012), and into trades by top executives (e.g., CEO, chairman) vs. trades by other insiders. We report (buying-selling of tracked insider trades) – (buying-selling of all other non-tracked trades that coincide with untracked insiders of the same insider group) by institutions (i.e. Row 8 of Panel A minus Row 8 of Panel B of Table 6). Reported below are the quarterly excess, DGTW, and 4 Factor adjusted returns of these aforementioned L-S spreads. T-statistics, reported in parenthesis, are based on Newey-West standard errors.

Panel A. Portfolio Returns across Firm Characteristics

a) Small vs. Large Firm Size (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
Large Size	3.14	2.76	2.82
	(2.01)	(1.98)	(1.96)
Small Size	3.26	3.45	3.42
	(1.31)	(1.48)	(1.39)

b) Low vs. High BM (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
High BM	3.10	2.99	3.28
	(2.03)	(2.11)	(2.31)
Low BM	4.88	4.38	4.85
	(2.06)	(1.96)	(2.13)

c) Low vs. High Past Returns (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
High MOM	3.93	4.01	3.68
	(2.08)	(2.24)	(1.91)
Low MOM	4.21	3.26	4.47
	(1.94)	(1.59)	(2.24)

d) Low vs. High Turnover (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
High Turnover	4.65	4.17	4.72
	(2.66)	(2.60)	(2.69)
Low Turnover	2.89	2.60	2.58
	(1.12)	(1.14)	(1.13)

e) Low vs. High Institutional Ownership (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
High Inst	1.41	1.64	0.74
	(0.77)	(0.98)	(0.44)
Low Inst	4.65	4.30	4.88
	(2.32)	(2.27)	(2.56)

Panel B. Portfolio Returns across Trading and Insider Characteristics

f) Routine Vs Non-Routine (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
Routine	3.46	3.89	4.19
	(1.09)	(1.45)	(1.40)
Opportunistic	4.54	4.55	4.56
	(2.90)	(3.03)	(2.84)

g) Insiders Types (Buying Minus Selling)

	Excess Returns	DGTW	4-Factor Alpha
All Top Officers	3.96	4.09	4.10
	(2.84)	(3.21)	(2.97)
CEO & Chairman	3.14	3.31	3.33
	(1.64)	(1.79)	(1.71)
Rest of Top Officers	3.52	2.75	3.63
	(2.55)	(2.02)	(2.66)
Rest of Insiders	2.05	1.56	1.83
	(1.21)	(0.95)	(1.07)

Table 8. Characteristics of Checked Insider Profiles

This table reports the characteristics of the insiders being tracked in our sample. These characteristics include the insider's highest education degree (MBA, PhD, MD), position in the firm (accountant, director, CEO, president, chairman, CFO), and graduating institution (Yale, Harvard, Princeton). The first column reports the percentage of each type of insiders for the entire BoardEx sample. The second column reports the percentage of each type of insiders being checked by at least one institution. The differences between the checked sample and the BoardEx sample are reported in the third column.

	All Profile	Checked Profile	Difference
Accountant	11.25%	15.55%	4.30%
MBA	34.53%	34.57%	0.04%
PhD	7.19%	6.46%	-0.73%
MD	2.13%	1.92%	-0.21%
Director	70.23%	54.58%	-15.65%
CEO	9.83%	28.71%	18.88%
President	19.91%	33.69%	13.78%
Chairman	9.49%	27.28%	17.80%
CFO	9.95%	16.18%	6.24%
Harvard	9.92%	11.65%	1.73%
Princeton	1.53%	1.91%	0.38%
Yale	1.80%	2.01%	0.21%

Table 9. Characteristics of Cherry-Pickers

This table reports the characteristics of fund managers within the matched 13-F mutual fund families. Cherry Pickers are fund managers whose trades correlate positively with the insider trades tracked by the 13-F mutual fund family at or above the 10% significance level. Panel A reports the geographic distribution of all mutual fund managers, and that of the cherry pickers, as well as the difference between the two. Panel B reports the education background distribution of all mutual fund managers, and that of the cherry pickers, as well as the difference between the two.

Panel A. Geographic Distribution of Cherry-Pickers

State	Distribution of Managers	Distribution of Cherry Pickers	Difference
AZ	0.89%	1.48%	0.59%
CA	8.67%	5.90%	-2.77%
CO	0.53%	0.37%	-0.16%
CT	2.30%	0.37%	-1.93%
FL	0.53%	0.37%	-0.16%
GA	1.06%	0.74%	-0.32%
IL	3.89%	4.43%	0.53%
KS	1.24%	0.74%	-0.50%
MA	22.66%	30.63%	7.97%
MD	7.43%	11.44%	4.01%
MO	3.19%	2.58%	-0.60%
NC	2.66%	1.11%	-1.55%
NJ	3.36%	4.06%	0.70%
NY	20.89%	21.03%	0.15%
OH	5.66%	3.69%	-1.97%
OK	0.18%	0.37%	0.19%
PA	6.37%	7.38%	1.01%
TX	6.20%	1.85%	-4.35%
WA	0.18%	0.00%	-0.18%
WI	2.12%	1.48%	-0.65%

Panel B. Education Background Distribution of Cherry-Pickers

School	Distribution of Managers	Distribution of Cherry Pickers	Diff
University of Wisconsin	5.79%	7.27%	1.48%
SUNY	2.79%	3.50%	0.71%
Harvard University	5.07%	5.39%	0.32%
University of Virginia	1.60%	1.75%	0.15%
Brown University	1.09%	1.23%	0.15%
Princeton University	1.14%	1.23%	0.10%
Yale University	1.29%	1.36%	0.07%
University of Michigan	1.03%	1.10%	0.07%
California State University	3.10%	3.11%	0.01%
Stanford University	2.28%	2.27%	0.00%
New York University	3.52%	3.50%	-0.01%
Columbia University	3.15%	3.11%	-0.04%
Boston College	1.76%	1.69%	-0.07%
Northwestern University	1.24%	1.17%	-0.07%
University of Minnesota	1.24%	1.17%	-0.07%
University of Pennsylvania	5.48%	5.39%	-0.09%
Dartmouth College	1.81%	1.56%	-0.25%
University of Illinois	2.28%	1.95%	-0.33%
University of Chicago	5.02%	4.67%	-0.34%
University of California	9.31%	7.14%	-2.17%

Table 10. Match of Insiders and Cherry-Pickers

This table explores potential links between cherry-picking fund managers and the insiders they track. We examine common links through locations and education backgrounds. Cherry Pickers are fund managers whose trades correlate positively with the insider trades tracked by the 13-F institution at or above the 10% significance level. The dependent variable is a match indicator that equals one if the cherry picker searches for a given insider and zero otherwise. The main independent variables include an education-link dummy that equals one if the cherry picker and the insider graduate from the same school, and a location-link dummy that equals one if the cherry picker and the insider are based in the state. The sample includes all potential insiders-fund manager pairs based on fund holdings. We also control for quarter, insider-education, manager-education, insider-location, and manager-location fixed effects. T-statistics, reported below the coefficients, are based on standard errors clustered by quarter. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Match Indicator					
Education Link	0.328%**	0.349%***	0.325%***	0.370%***	0.668%***	0.391%***
	(2.64)	(2.76)	(2.75)	(3.85)	(6.20)	(3.91)
Location Link	1.42%***	1.40%***	1.31%***	0.955%***	0.501%**	0.502%**
	(4.84)	(4.98)	(4.77)	(3.92)	(2.15)	(2.20)
Lag Match Indicator			21.9%***	18.9%***	20.04%***	17.91%***
			(12.32)	(11.86)	(11.39)	(11.26)
Quarterly FE	No	Yes	Yes	Yes	Yes	Yes
Insider Education FE	No	No	No	Yes	No	Yes
FM Education FE	No	No	No	Yes	No	Yes
Insider Location FE	No	No	No	No	Yes	Yes
FM Location FE	No	No	No	No	Yes	Yes
Adj-R^2	0.0005	0.0141	0.0303	0.0533	0.0455	0.0618
No. Obs.	1,748,892	1,748,892	1,748,892	1,748,892	1,705,546	1,705,546

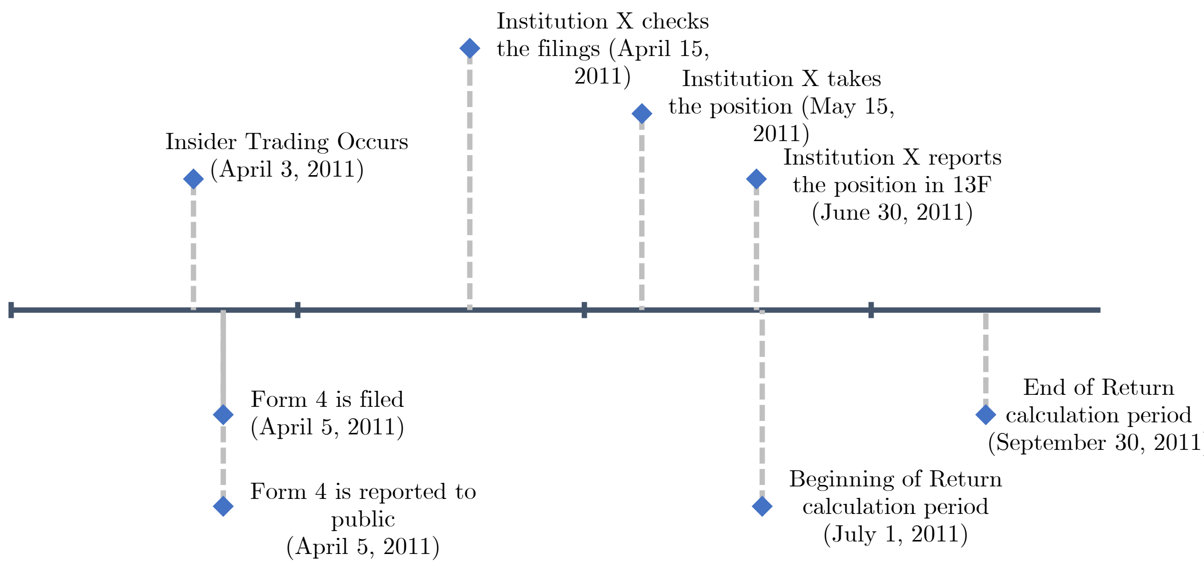

Figure 1. This figure illustrates the timing of our portfolio return test with a hypothetical example.

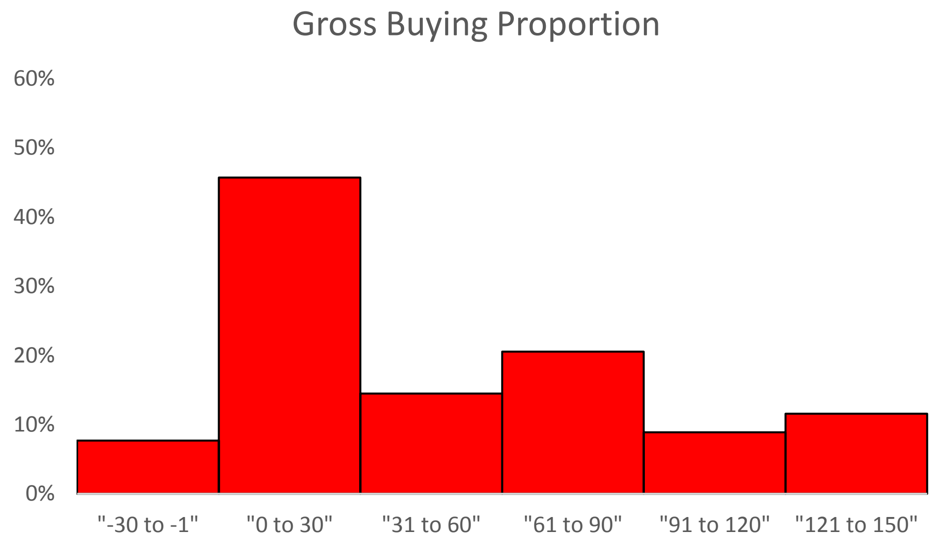

Figure 2. This figure shows the proportion of gross buying of a stock by institutions over the next X (ranging from 1 to 150) number of trading days after checking insider filings.

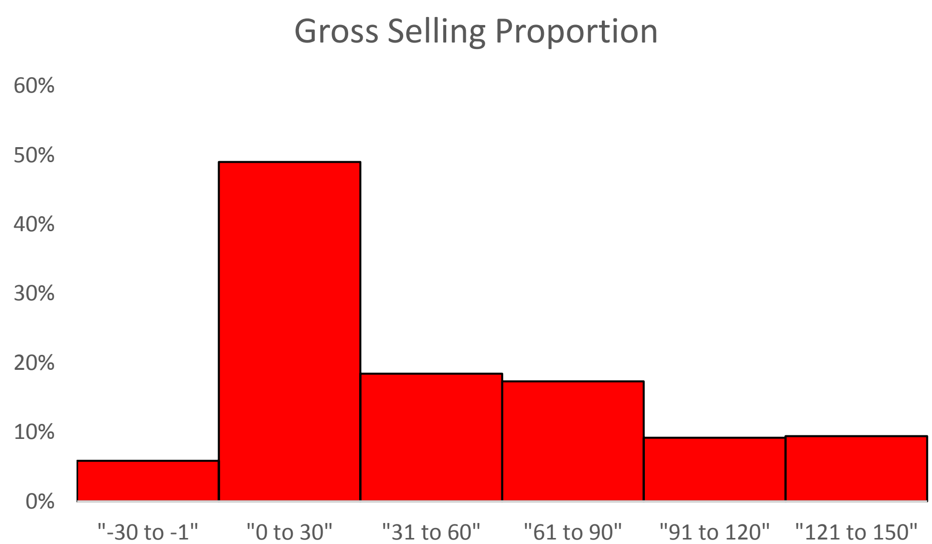

Figure 3. This figure shows the proportion of gross selling of a stock by institutions over the next X (ranging from 1 to 150) number of trading days after checking insider filings.

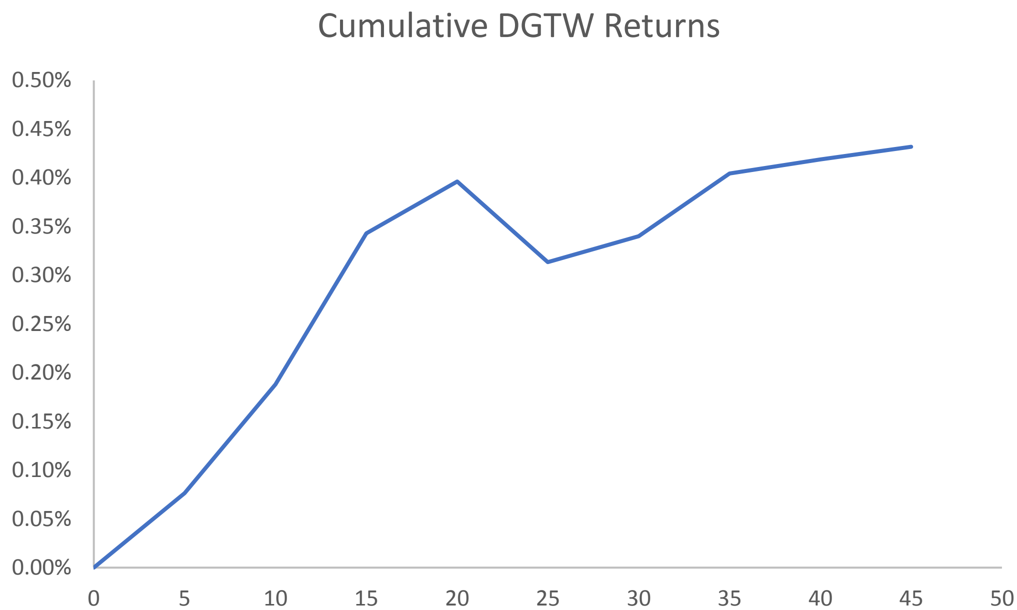

Figure 4. We plot the subsequent cumulative DGTW returns of tracked-insider trades. Specifically, If an Ancerno institution downloads an insider form and trades in the same direction as the insider over the next 15 trading days, we classify these trades as tracked-insider trades.

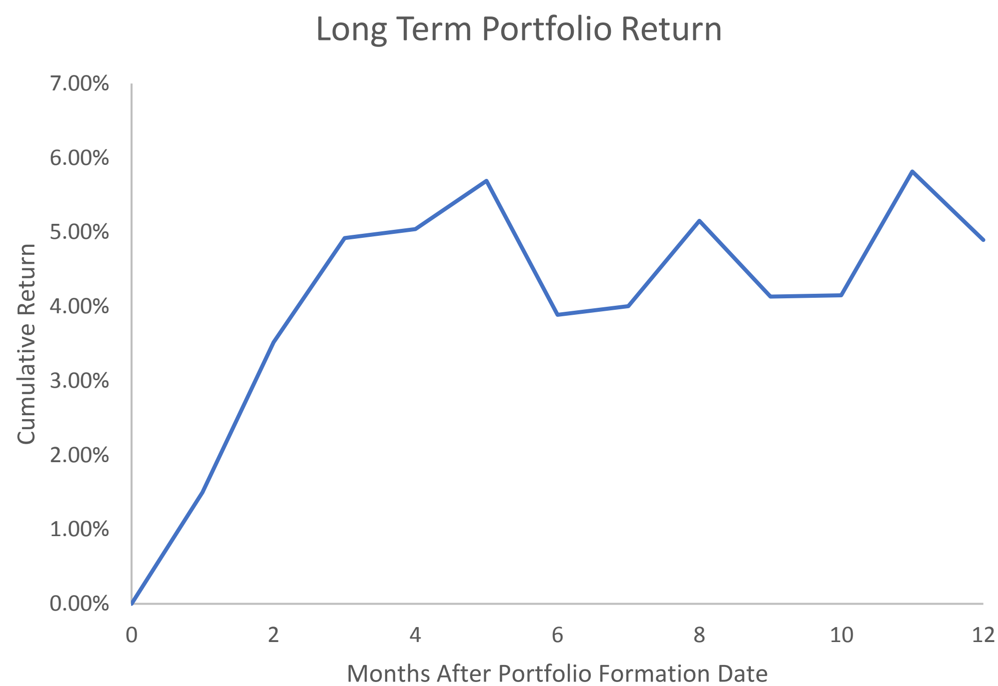

Figure 5. This figure shows the long-run returns to tracked-insider trades by fund managers. Specifically, it is the difference between the portfolio described in row 5 of Table 6 Panel A, and that in row 5 of Table 6 Panel B over the next 12 months.

Appendix Table A1. Alternative Timing

This table reports calendar-time portfolio returns of various types of institutional holdings and trading. Panel A corresponds to tracked insider buys and Panel B corresponds to tracked insider sells. In terms of timing, we require that institutions both view the insider trading record (which takes place in quarter t) and trade the underlying stock in quarter t+1. We then analyze the equal-weighted returns to those trades in quarter t+2. Row 5 constructs a long-short portfolio, where the long portfolio includes all stocks of which the fund manager checks insider trades and also trade in the same direction, the short portfolio includes all other stocks that the fund manager buys/sells in the same quarter. In Row 8, the benchmark portfolio includes all stocks that are not checked by a fund manager but are traded by insiders in the same direction as the fund manager. Reported below are the quarterly Raw, DGTW, and 4 Factor adjusted returns of these aforementioned portfolios (as well as the long and short sides). T-statistics, reported in parenthesis, are based on Newey-West standard errors. 5% statistical significance is indicated in bold.

Panel A: Tracked Insider Buys	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Buying and Bought vs.	1.20%	0.86%	1.11%
Rest Bought	(1.05)	(0.76)	(1.01)
8) Checked Insider Buying and Bought vs.	1.38%	1.00%	1.18%
Not Checked and Bought	(1.30)	(0.90)	(1.09)

Panel B: Tracked Insider Sales	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Selling and Sold vs.	**-1.22%**	**-1.16%**	**-1.42%**
Rest Sold	**(-2.22)**	**(-2.16)**	**(-2.67)**
8) Checked Insider Selling and Sold vs.	**-1.07%**	**-1.10%**	**-1.26%**
Not Checked and Sold	**(-2.05)**	**(-2.08)**	**(-2.42)**

Appendix Table A2. Value-Weighted Portfolio Returns

This table reports calendar-time portfolio returns of various types of institutional holdings and trading. Panel A corresponds to tracked insider buys and Panel B corresponds to tracked insider sells. In terms of timing, we require that institutions both view the insider trading record and potentially trade the underlying stock in quarter *t*. We then analyze the holding-value weighted returns to those trades in quarter *t*+1. Row 5 constructs a long-short portfolio, where the long portfolio includes all stocks of which the fund manager checks insider trades and also trade in the same direction, the short portfolio includes all other stocks that the fund manager buys/sells in the same quarter. In Row 8, the benchmark portfolio includes all stocks that are not checked by a fund manager but are traded by insiders in the same direction as the fund manager. Reported below are the quarterly Raw, DGTW, and 4 Factor adjusted returns of these aforementioned portfolios (as well as the long and short sides). T-statistics, reported in parenthesis, are based on Newey-West standard errors. 5% statistical significance is indicated in bold.

Panel A: Tracked Insider Buys	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Buying and Bought vs.	2.18%	2.27%	1.75%
Rest Bought	(1.50)	(1.70)	(1.22)
8) Checked Insider Buying and Bought vs.	2.63%	**2.71%**	2.18%
Not Checked and Bought	(1.77)	**(1.96)**	(1.45)

Panel B: Tracked Insider Sales	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Selling and Sold vs.	**-1.70%**	**-1.33%**	**-1.85%**
Rest Sold	**(-2.80)**	**(-2.39)**	**(-3.02)**
8) Checked Insider Selling and Sold vs.	**-1.68%**	**-1.35%**	**-1.84%**
Not Checked and Sold	**(-2.77)**	**(-2.42)**	**(-3.00)**

Appendix Table A3: Trade-Size-Weighted Portfolio Returns

This table reports calendar-time portfolio returns of various types of institutional holdings and trading. Panel A corresponds to tracked insider buys and Panel B corresponds to tracked insider sells. In terms of timing, we require that institutions both view the insider trading record and trade the underlying stock in quarter t. We then analyze the trade-size weighted returns to those trades in quarter t+1. A stock's trade-size weight is the increase (decrease) in portfolio weight of the tracked buy (the tracked sale) relative to the stock's final weight (initial weight). Row 5 constructs a long-short portfolio, where the long portfolio includes all stocks of which the fund manager checks insider trades and also trade in the same direction, the short portfolio includes all other stocks that the fund manager buys/sells in the same quarter. In Row 8, the benchmark portfolio includes all stocks that are not checked by a fund manager but are traded by insiders in the same direction as the fund manager. Reported below are the quarterly Raw, DGTW, and 4 Factor adjusted returns of these aforementioned portfolios (as well as the long and short sides). T-statistics, reported in parenthesis, are based on Newey-West standard errors. 5% statistical significance is indicated in bold.

Panel A: Tracked Insider Buys	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Buying and Bought vs.	**3.92%**	**3.95%**	**4.05%**
Rest Bought	**(2.09)**	**(2.22)**	**(2.28)**
8) Checked Insider Buying and Bought vs.	**3.75%**	**3.65%**	3.63%
Not Checked and Bought	**(2.00)**	**(2.04)**	(1.96)

Panel B: Tracked Insider Sales	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Selling and Sold vs.	**-2.89%**	**-2.64%**	**-3.11%**
Rest Sold	**(-3.37)**	**(-3.29)**	**(-3.81)**
8) Checked Insider Selling and Sold vs.	**-3.31%**	**-2.96%**	**-3.61%**
Not Checked and Sold	**(-3.78)**	**(-3.38)**	**(-4.25)**

Appendix Table A4: Exclude the Post-2013 Period from Our Sample

This table reports calendar-time portfolio returns of various types of institutional holdings and trading excluding the post 2013 sample period. Panel A corresponds to tracked insider buys and Panel B corresponds to tracked insider sells. In terms of timing, we require that institutions both view the insider trading record and trade the underlying stock in quarter t, and analyze the equal-weighted returns to those trades in quarter t+1. Row 5 constructs a long-short portfolio, where the long portfolio includes all stocks of which the fund manager checks insider trades and also trade in the same direction, the short portfolio includes all other stocks that the fund manager buys/sells in the same quarter. In Row 8, the benchmark portfolio includes all stocks that are not checked by a fund manager but are traded by insiders in the same direction as the fund manager. Reported below are the quarterly Raw, DGTW, and 4 Factor adjusted returns of these aforementioned portfolios (as well as the long and short sides). T-statistics, reported in parenthesis, are based on Newey-West standard errors. 5% statistical significance is indicated in bold.

Panel A: Tracked Insider Buys	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Buying and Bought vs.	1.58%	1.27%	1.47%
Rest Bought	(1.02)	(0.92)	(1.00)
8) Checked Insider Buying and Bought vs.	1.70%	1.48%	1.64%
Not Checked and Bought	(1.17)	(1.09)	(1.11)

Panel B: Tracked Insider Sales	Excess Returns	DGTW	4-Factor Alpha
5) Checked Insider Selling and Sold vs.	**-1.88%**	**-1.51%**	**-1.85%**
Rest Sold	**(-3.02)**	**(-2.44)**	**(-2.96)**
8) Checked Insider Selling and Sold vs.	**-1.82%**	**-1.52%**	**-1.81%**
Not Checked and Sold	**(-3.02)**	**(-2.46)**	**(-2.93)**

Appendix A5: Description of the Deciphering Procedure

The deciphering process decrypts the IP addresses from the SEC Edgar log files in order to match them to a commercial IP geolocation/organizational database. The IP addresses in the dataset are partially anonymized using a static cipher. A standard IP address have the form ###.###.###.###. Each octet ### separated by periods has a numerical value between 0 and 255. An IP address in the Edgar data has the form ###.###.###.&&&. The first 9 digits of the Edgar server IP correspond to the actual IP that visited the website (024.145.236). The last 3 digits correspond to a hidden octet (*jcf*). 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).

The static ciphers in the Edgar log are 3-digit alphabetical codes that *uniquely* represent octets between 0 and 255 throughout the data. It is very apparent that this cipher is static in nature.

1) The same cipher would visit similar forms in multiple days and months. For example, the same IP address, with *aca* as the cipher, would visit AIG filings in January 2009, February 2009, and so on.
2) We find that ciphers with the lowest frequency tend to be the same two: *ghf* and *ghg* throughout the data. We ended up matching *ghf* and *ghg* to 0 and 255 respectively. IP addresses ending with 0 and 255 tend to be reserved for network administrative purposes, and are generally less frequent than other addresses.

We obtain another dataset of server logs from a private but well-trafficked website in order to decrypt the EDGAR IP address. Specifically, we identify all instances where the first 9 digits of an IP address in the Edgar files match an IP address from this private server.

If 1.1.1.aaa visited the SEC server in its history and 1.1.1.111 visited this website in 2016, then the cipher *aaa* will have one match to 111. Since there are multiple matches between a cipher and a set of octets, we compose a scoring system that counts the number of times each cipher is matched to each octet. The most frequent match is our candidate for the cipher link. The link between a cipher and its most frequently matched octet is distinct for the vast majority of the ciphers (236 out of 256 ciphers). For example, *aba*'s most frequent octet match is 009, no other cipher matches to 009 as their most frequent match. In this set of 236 matches, the identified octet is on average 96% more frequent than the next most frequent match. I.e. *aej* is matched to its more frequent octet, 71, 298 times. The next most frequent octet, 135, has about 139 matches to *aej*.

The remainder of the matching - where multiple ciphers are matched to a single octet - is done through a process of elimination. For example, if both *aaa* and *aab* score 001 as their most frequent match, we distinguish the two by examining which of the two ciphers are more frequently linked to 001. If 001 has many more matches with *aaa* than *aab*, then *aaa* is matched to 001. In this case 001 is removed as a potential choice for *aab*'s matching. *aab* is then linked to its next most frequent octet. We only had to iterate this process once in order to match the remaining 20 pairs. When we examine the frequency of octet to cipher matches in this subset, the most frequent cipher for each of these octet accounts for significantly more matches than the next most frequent cipher (116% more frequent).

At the end of our decryption, we obtain 256 pairs of unique matchings between the hidden octet and the actual octet for the EDGAR database. We present our results in table 1.