Informed Trading under the Microscope: Evidence from 30 Years of Daily Hedge Fund Trades

JinGi Ha, Jianfeng Hu, and Yuehua Tang *

December 4, 2024

ABSTRACT

We develop a new measure of daily aggregate hedge fund trades in individual U.S. stocks over 30 years, from 1993 to 2022. Leveraging this high-frequency measure, we find that hedge fund pre-announcement trading significantly predicts abnormal returns around corporate events (e.g., earnings announcements) and various types of public news arrivals. Furthermore, we uncover that hedge funds increasingly utilize alternative data, such as satellite images, to generate informed trading signals. We do not find similar results for trades of non-hedge fund institutional investors. In addition, daily aggregate hedge fund trading positively predicts future stock returns in the cross-section without subsequent reversals, especially for stocks with weak informational environments and high arbitrage costs. Finally, we find that hedge fund trading enhances price efficiency as it reduces variance ratios of stock returns and mitigates price response to earnings announcements.

Keywords: Hedge Funds, Informed Trading, Market Efficiency, Return Predictability, Alternative Data
JEL Classification: G12, G14, G23, D82

*We would like to thank and the seminar participants at Aarhus University and Singapore Management University for comments. All remaining errors are ours. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant and IREC, The Institute of Finance and Banking, and Seoul National University. Jianfeng Hu also acknowledges financial support from the Lee Kong Chian Fund for Excellence. Hu is at Singapore Management University (jianfenghu@smu.edu.sg); Ha is at Soongsil University (jgha@ssu.ac.kr); Tang is at the University of Florida (yuehua.tang@warrington.ufl.edu).

I. Introduction

The ability of hedge funds to engage in informed trading and generate superior returns has long intrigued both academics and practitioners. Numerous researchers have examined hedge fund strategies and performance using low-frequency data such as quarterly 13F holdings or monthly fund returns from commercial hedge fund databases.[1] While these datasets offer valuable insights, they often fail to capture the nuances of hedge fund trading strategies, particularly with respect to the precise timing of trades, leaving significant gaps in understanding how hedge funds exploit short-lived information or contribute to price discovery. Recent research using higher frequency data has yielded limited insights, relying on proprietary datasets covering only a small and selective sample of funds.[2] As a result, a broader understanding of hedge fund trading behavior remains elusive, particularly in comprehensively examining how hedge funds execute trades in response to time-sensitive signals. High-frequency data, capturing daily or intraday trading patterns, can unlock new insights into how hedge funds explore short-term inefficiencies and seize investment opportunities.

This paper seeks to fill this gap by developing a novel dataset of daily aggregate hedge fund trades in all U.S. stocks, spanning 30 years from 1993 to 2022. Through this comprehensive and granular dataset, we aim to address several unresolved questions: Can hedge funds successfully trade on short-lived information? If so, what are their information sources, and how do these sources evolve over time? What role do hedge funds play in price discovery, and how does their activity compare to other institutional investors? Our analysis of daily hedge fund trades allows for a detailed exploration of their trades' timing, magnitude, and market

[1]For studies on hedge fund 13F holdings, see, among others, Brunnermeier and Nagel (2004), Griffin and Xu (2009), Aragon and Martin (2012), Agarwal, Jiang, Tang, and Yang (2013), Jiao, Massa, and Zhang (2016), Agarwal, Ruenzi, and Weigert (2017), Cao, Liang, Lo, and Petrasek (2018), Chen, Da, and Huang (2019), and Kumar, Mullally, Ray, and Tang (2020). For analyses of hedge fund skills using monthly returns, see, among others, Fung and Hsieh (2004), Aragon (2007), Kosowski, Naik, and Teo (2007), Jagannathan, Malakhov, and Novikov (2010), Sun, Wang, and Zheng (2012), Cao, Chen, Liang, and Lo (2013), Gao, Gao, and Song (2018), and Agarwal, Ruenzi, and Weigert (2024). See Agarwal, Mullally, and Naik (2015) for a comprehensive review.

[2]See Jame (2018) and Çötelioğlu, Franzoni, and Plazzi (2021) for studies on hedge funds' liquidity provision, using transaction data from the brokerage firm Abel Noser, which covers a highly limited sample of fewer than 100 hedge fund firms.

impact, offering new insights into the mechanisms behind hedge fund trading strategies and their influence on market efficiency.

Early studies on institutional trading estimate institutional order flow using high-frequency TAQ data, assuming that institutions are more likely to place large orders.[3] Later research casts doubt on the usefulness of a simple trade size cutoff rule.[4] A pivotal study by Campbell, Ramadorai, and Schwartz (CRS, 2009) addresses these limitations by assigning propensities to trades with different sizes using a regression approach calibrated to quarterly changes in institutional ownership from 1993 to 2000, and shows that institutional order flows are informed about earnings news.

In this paper, we build on the CRS method to estimate daily hedge fund and non-hedge fund trades across a large cross-section of stocks from 1993 to 2022 by combining TAQ data with hedge fund (HF) and non-hedge fund (NHF) holding changes in 13F filings. Using proprietary institutional transaction-level data from Abel Noser covering a sample of both HFs and NHFs, we uncover distinct trade size distributions between HFs and NHFs. Hedge funds rely more on medium-sized trades, whereas non-hedge funds split orders into smaller trades but also execute a substantial volume through very large trades. This trade size variation allows us to disentangle HF and NHF trades when calibrating TAQ order flows in multiple trade size bins to quarterly holdings change in hedge funds and non-hedge funds separately. Our extension of the CRS model thus enhances the precision of institutional trade estimation, enabling a more detailed analysis of hedge fund return predictability in the cross-section of stocks, as well as around corporate events and news arrival. This level of granularity was previously unavailable due to data limitations, and it contributes to a clearer understanding of hedge fund trading dynamics relative to other institutional investors.

Leveraging our high-frequency hedge fund trading data, we first investigate the potential informational sources hedge funds may possess. We start with examining how hedge

[3]See, e.g., Lee (1992), Lee and Radhakrishna (2000), Battalio and Mendenhall (2005), Malmendier and Shanthikumar (2007), and Hvidkjaer (2008).

[4]See, e.g., Cready, Kumas, and Subasi (2014).

fund and non-hedge fund trades respond to salient corporate events, including earnings announcements, analyst rating updates, and price jumps. Hedge fund trades prior to the event positively correlate with cumulative abnormal returns around all types of corporate events. This indicates that hedge funds are informed about firm fundamentals such as earnings and value-relevant events and can reap the benefits from their information advantage through trading ahead of such events. Non-hedge fund institutions' trades, however, tend to be negatively associated with the abnormal returns around these events, highlighting their relative disadvantage in processing information around corporate disclosures.

Furthermore, hedge funds can have additional information advantages from early access to news (e.g., Bolandnazar, Jackson, Jiang, and Mitts (2020)) or by strategically collaborating with news agencies to release their private information (Ljungqvist and Qian (2016)). Taking advantage of the flexibility of our estimate of hedge fund order flow in a large cross-section over three decades, we examine hedge funds' trading behavior around a comprehensive sample of firm-specific news from RavenPack, including both anticipated and unanticipated news. We find that hedge funds have an edge in both fundamental news, such as mergers and acquisitions, as well as non-fundamental news, such as insider trades and investor relation matters. They engage in informed trading before the arrival of value-relevant information. In contrast, we do not find the same pattern for non-hedge fund institutions' trades, consistent with the literature (e.g., Huang, Tan, and Wermers (2020)).

Importantly, we uncover that the use of alternative data sources, such as satellite image data, significantly enhances the predictive power of hedge fund trades for stock returns. For instance, hedge fund trades show a stronger positive impact on stock prices when trading on information derived from satellite data related to retail traffic. This reflects the capacity of hedge funds to integrate innovative sources such as big data into their investment processes, a capability less prevalent among non-hedge funds, thus sustaining their role as informed traders. Thus, big data has become a valuable resource for hedge funds to maintain their information edge.

In the next set of analyses, we evaluate whether hedge funds' ability to trade ahead of short-lived information generalizes to return predictability in the broader sample. If yes, it can speak directly to how pervasive hedge funds' informed trading is. We do so by studying the return predictability of hedge fund versus non-hedge fund trades using Fama-MacBeth regressions. We find that daily aggregate HF trades significantly predict future stock returns over the next five trading days, with consistently positive coefficients across various time lags, indicating a permanent price impact. The economic impact is sizable, with a one standard deviation increase in HF net purchase corresponding to a 3.7 basis point (bps) increase in the next day's stock return. In contrast, although NHF trades are relatively highly correlated with HF trades (with a correlation of 0.67) and exhibit strong positive contemporaneous price impact, NHF trades display significantly negative coefficients when predicting future stock returns, suggesting their trading generates only a transitory price impact. This disparity highlights the informational advantage of hedge funds, as their trades integrate information into prices more effectively than non-hedge funds, with no subsequent reversals in their predictive power.

Reinforcing our interpretation of the return predictability as evidence of informed trading, the predictive power of hedge fund trades varies across several stock characteristics. Specifically, hedge funds exhibit stronger return predictive ability in small, illiquid stocks, where trading costs are high, and in stocks with low analyst coverage and institutional ownership, where information asymmetry is more pronounced. The cross-sectional variation in return predictability suggests that hedge funds are particularly effective in trading environments with higher limits to arbitrage, where their ability to process information provides a competitive advantage over other investors.

In the final set of analyses, we explore the broader implications of hedge fund trading on market efficiency. By acting as informed traders, hedge funds can embed their private information into the stock price and improve market efficiency. Focusing on one of the most critical corporate events for information updates, earnings announcements, we find that hedge fund pre-announcement trades significantly reduce post-earnings announcement drift (PEAD) as well as the initial price reaction to earnings surprise, consistent with our earlier findings that hedge funds' pre-announcement trades incorporate information into the stock prices. Moreover, we also find evidence that hedge fund trading facilitates price discovery because the stock variance ratio decreases in the intensity of hedge fund low both in the cross-section and post-earnings announcements. Hedge funds play a pivotal role in adjusting stock prices to reflect new information, leading to a quicker and more accurate incorporation of information into stock prices. In all these tests we also include the non-hedge fund institutional trades but do not find consistent results of their effects on price efficiency. These results underscore hedge funds' unique role in enhancing financial markets' informational efficiency.

Overall, our evidence shows that hedge funds act as informed traders whose trades predict future stock returns without reversals, a characteristic not observed in non-hedge funds' trades. In addition, hedge funds engage in informed trading ahead of both fundamental news, such as earnings announcements, and non-fundamental news. They increasingly utilize alternative data sources like satellite images in their investment process. This informed trading by hedge funds contributes to market efficiency, as evidenced by smaller stock price responses to earnings surprises and reduced variance ratios in stocks with higher hedge fund trading intensity.

Our study makes several contributions to the literature. First, to the best of our knowledge, we are the first to develop a daily aggregate hedge fund trade dataset that spans a large cross-section of U.S. stocks over 30 years. Most prior research on hedge fund trading behavior relies on low-frequency, quarterly holdings data from 13F filings, leaving gaps in our understanding of hedge fund activity at higher frequencies. A few exceptions exist, such as studies focusing on hedge funds as liquidity providers using transaction data from the brokerage firm Abel Noser, but these cover a highly selective sample of fewer than 100 hedge fund firms (e.g., Jame (2018) and Çötelioğlu, Franzoni, and Plazzi (2021)). Our approach advances the literature by offering high-frequency measures of hedge fund trading across a comprehensive sample of stocks, helping to shed light on their trading behavior in a more granular manner. In particular, our focus on hedge fund trading at the daily frequency differs from and complements earlier studies on arbitrage or informed trading at the quarterly frequency by combining 13F holdings and short interest data (e.g., Jiao, Massa, and Zhang (2016) and Chen, Da, and Huang (2019)).[5]

Second, our study provides new insights into hedge funds as informed traders. While previous research has explored the informational content of hedge fund strategies using low-frequency data (see footnote 1), there is very limited evidence on hedge fund skills at higher frequencies, largely due to data constraints. Leveraging the new measures we develop, we comprehensively analyze hedge funds' informational sources and how they have evolved over time. One of our key contributions is documenting that hedge funds make informed trades ahead of corporate news. This evidence contrasts with earlier findings based on high-frequency data that non-hedge fund institutions primarily trade after corporate news releases (e.g., Huang, Tan, and Wermers (2020)). Our paper adds to earlier studies by Campbell, Ramadorai, and Schwartz (2009) and Hendershott, Livdan, and Schurhoff (2015) by documenting that hedge funds are the significantly more informed group among institutional investors regarding corporate events and news.

Third, our study is also related to the nascent literature on big data in finance. While a few recent studies have examined the impact of alternative data on stock prices, firm managers, and mutual fund managers (e.g., Zhu (2019), Katona, Painter, Patatoukas, and Zeng (2024), Dessaint, Foucault, and Fresard (2024), and Bonelli and Foucault (2024)), there is little direct evidence on which group of investors benefit from the availability of alternative data. Our study is the first to document evidence of hedge funds' adoption of alternative data in recent years to maintain their edge as informed traders.

[5]We note that Chen, Da, and Huang (2019) have some analysis on anomaly trading at the daily frequency by combining daily security lending data with daily trading of a highly selective sample of hedge funds from Abel Noser. Importantly, their focus is on anomaly mispricing. Our paper differs by focusing on informed trading via fundamental information and big data using a more comprehensive sample of hedge funds.

Lastly, we shed new light on hedge funds' impact on the informational efficiency of stock prices. In particular, we offer direct, daily-level evidence of hedge funds' pivotal role in enhancing price efficiency, both broadly and around key corporate events. This complements earlier findings based on lower-frequency data (e.g., Boehmer and Kelley (2009) and Cao, Liang, Lo, and Petrasek (2018)).

The rest of the paper is organized as follows: Section II describes our estimation method of aggregate hedge fund order flow and summary statistics. Section III uses the daily order flow data to understand hedge funds' role as informed traders. Section V investigates the impact of hedge fund trading on market efficiency. Finally, Section VI presents our conclusions.

II. Estimate daily hedge fund trades

A. Data and sample selection

We construct our sample using common stocks listed on the NYSE, AMEX, and Nasdaq that are available in the CRSP, TAQ, and Thomson Reuters 13F Ownership databases spanning the years 1993 to 2022. We apply a series of filters to refine the sample for analysis. First, we exclude stocks with prices below $5 to avoid including low-priced stocks, which are typically more volatile and may not reflect the institutional trading patterns we aim to study. Additionally, we exclude stocks with percentage bid-ask spreads above 50% to eliminate illiquid stocks that are unlikely to attract hedge fund interest and may distort the analysis due to wide bid-ask spreads. To ensure the sample comprises stocks with sufficient market capitalization, we further restrict the sample to include only stocks with a market capitalization above the 5% breakpoint of NYSE-listed stocks each quarter. This restriction helps focus our analysis on more liquid and widely held stocks, which are more relevant to institutional investors. After applying these criteria, our final sample contains a total of 13,924 unique stocks and 22,696,498 stock-day observations, providing a comprehensive dataset for estimating daily hedge fund and non-hedge fund order flows with high granularity.

We identify hedge fund firms as 13F-filing institutions whose primary business is sponsoring or managing hedge funds in Thomson Reuters 13F Ownership data following Agarwal, Jiang, Tang, and Yang (2013). Hedge fund firms are classified based on a range of sources, including institutions' own websites, industry directories and publications, news article searches, and Form ADVs filed by investment companies. Out of the full list of 13F institutions, 1,854 are identified as hedge fund firms over our sample period, representing one of the most comprehensive samples of hedge fund firms used in the literature. In part of our analysis, we also use transaction data of institutional investors from Abel Noser and identify a list of 82 hedge fund managers in the Abel Noser data following Jame (2018).

B. *Difference in hedge fund and non-hedge fund trading*

Although researchers can infer institutional investors' trading demand from their quarterly disclosure of stock holdings, it is not possible to directly observe their activities at a higher frequency, limiting opportunities for more granular research. Campbell, Ramadorai, and Schwartz (CRS, 2009) develop a method to estimate daily aggregate institutional investors' net order flow. This approach follows the conventional wisdom that large trades are more indicative of institutional activity. Unlike other studies using single trade-size cutoffs, however, they use a regression approach to leverage information across the full spectrum of transaction sizes. Specifically, they first use quarterly observations to estimate the relation between institutional ownership changes in 13F filings and order imbalances across different trade size bins derived from the TAQ data. Then they extrapolate the quarterly relation to daily observations to estimate daily institutional trades using observable order imbalance data from the TAQ.

Our methodology extends the CRS framework by distinguishing between hedge fund and non-hedge fund institutions, leveraging differences in their trade size distributions. We use a proprietary institutional trading dataset from Abel Noser covering a sample of both HFs and NHFs to verify this distinction. Specifically, we identify hedge fund trades in Abel Noser following the approach of Jame (2018). We then plot the distribution of aggregate hedge fund and non-hedge fund trades across trade size bins. Results are reported for the number of trades, trade volume, buy volume, and sell volume within each trade size category as a percentage of total activity across all size bins. Following CRS, the trade size bins have lower limit points of \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1,000,000.

As shown in Figure 1, hedge funds place the majority of their orders in medium trade sizes, particularly within trade size bins 7 to 15. In contrast, non-hedge funds appear to focus more on splitting their orders, employing the largest proportion (27%) of smallest trades (below \$2,000) across the distribution. This divergence highlights a fundamental difference in the trading patterns of the two types of institutions, with NHFs more inclined toward extensive order fragmentation.

[Place Figure 1 about here]

More importantly, significant differences also emerge in the volume distribution across trade size bins between HFs and NHFs, which is crucial for estimating the relation between order flow and ownership change in 13F filings. While both HFs and NHFs conduct the bulk of their trading volume in the largest trade size bins—aligning with the institutional characteristic of using large trades—hedge funds maintain a greater presence in medium-sized trades. Although NHFs split orders to a great extent, they fill their positions mostly through the largest orders (above \$1,000,000) as almost 60% of their trading volume comes from this group of trades. In contrast, the same group of orders accounts for about only 35% of HF trading volume. HFs accumulate substantial trading volume (46%) in bins 13 to 17 with trade sizes ranging from \$100,000 and \$700,000. In the same region along the trade volume distribution, NHFs fill only 30% of their total volume. This same pattern is evident for both buying and selling, as illustrated in the bottom panels. The reliance on medium trades suggests that hedge funds may use these sizes tactically, potentially as part of a strategy to optimize transaction costs, manage liquidity needs, and maintain order anonymity.

The results in Figure 1 indicate that non-hedge funds tend to prioritize extensive order slicing and dicing yet still execute a large proportion of their total volume through very large trades. In contrast, hedge funds more frequently employ medium-sized orders. One possible explanation for this difference is that hedge funds, if more informed than other institutions, may prioritize immediacy to capitalize on short-term information without fully splitting their orders. At the same time, informed traders might avoid very large trades to limit their price impact, using medium trades as a compromise between rapid execution and minimized market impact. Notably, these differences are persistent during the sample period, as shown in Figure A1 in our internet appendix to the paper, which reinforces the validity of using the whole spectrum of trade sizes to distinguish hedge fund and non-hedge fund trades.

By incorporating these distinctions, our extension of the CRS method enables a more nuanced estimation of daily institutional flows, specifically distinguishing hedge fund trading behavior from that of other institutions. This approach not only allows for the estimation of hedge fund order flow but also deepens our understanding of how different types of institutions interact with the market on a daily basis, shedding light on the varying strategies and preferences that define institutional trading.

C. *Estimate daily hedge fund and non-hedge fund trades*

Similar to the CRS method, our estimation is based on the following equation:

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \sum_{Z=1}^{19} \beta(Z,v) F_{Z,i,q} + \epsilon_{i,q}, \tag{1}$$

where for a stock i in a quarter q, α is a set of four quarter dummies, Y is either aggregate hedge fund or non-hedge fund ownership (in separate estimations) from 13F, F_Z is aggregate order imbalance based on the Lee and Ready (1991) algorithm scaled by shares outstanding in a trade-size bin Z, and U is aggregate unclassified trades scaled by shares outstanding for which the Lee and Ready (1991) algorithm cannot determine the direction.[6] Hedge fund and non-hedge fund ownership is identified following the methodology of Agarwal, Jiang, Tang, and Yang (2013) in the Thomson Reuters 13F Ownership data. Following CRS, we assign trades into nineteen size bins whose lower limit points are \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. To smooth out the coefficient variation across transaction size and mitigate estimation errors in certain bins (e.g., very large trades for small stocks, which are rare), CRS apply a yield curve function from Nelson and Siegel (1987) to model the structure of β across trade-size bins:

$$\beta(Z, v) = b_{01} + b_{02}v + (b_{11} + b_{12}v + b_{21} + b_{22}v)[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22}v)e^{-Z/\tau}, \qquad (2)$$

where τ is a constant to estimate and v is set to the lagged level of hedge fund or non-hedge fund ownership ($Y_{i,q-1}$) as in CRS. Following CRS, we use non-linear least squares to estimate the coefficients in Equation (1) for each firm size quintile based on NYSE breakpoints of market capitalization at the start of each quarter. Concerning that both types of institutional investors may change their trading styles in the relatively long sample period of 30 years, we estimate Equation (1) in three decade-long subperiods separately.

The estimated coefficients of the CRS model are not immediately intuitive due to the complexities of the Nelson and Siegel yield curve approach. As a result, we report these estimates in Table A1 of the Internet Appendix. The estimated coefficients are highly sig-

[6] We restrict TAQ observations to regular transactions between 9:30:00 to 15:59:59 EST. We exclude trades under the sale condition of the Opened Last ('O'), Sold Sale ('Z'), Bounced ('B'), Pre- and Post-Market Close Trades ('T'), Sold Last ('L'), Bunched Sold ('G'), Average Price Trades ('W'), Rule 127 Trade ('J'), and Rule 151 Trade ('K').

nificant for both hedge funds and non-hedge funds, encompassing all firm size quintiles and all subperiods examined. To enhance interpretability, and in line with CRS, we compute the trade-size coefficients implied by the estimated coefficients in Table A1 of the Internet Appendix. First, we set the lagged level of quarterly institutional ownership to its in-sample mean. Second, we standardize the net flow coefficients by subtracting their mean and dividing by their standard deviation. Then, we illustrate the implied trade-size coefficients in Figure 2.

Figure 2 reveals important distinctions in trading behavior between hedge funds and non-hedge funds over the three subperiods. While there are some similarities, such as general patterns of increasing coefficients for large trades and decreasing coefficients for smaller trades, key differences emerge, especially when examining specific stock size groups and time periods. First, in the period of 1993-2002, hedge funds' ownership sensitivities to size-dependent order flow diverge significantly across small, medium, and large stocks, as shown in Panel (a). Hedge funds mainly take liquidity in small trades (bins 1 to 5) for small-cap stocks. They provide liquidity for relatively large trades (bins 8 to 19). For medium-cap stocks, hedge funds display a strong positive sensitivity to medium-sized trades in bins 8 to 13, and a strong negative sensitivity to large trades in bins 16 to 19. For large-cap stocks, the curve follows a different trajectory. Hedge funds seem to provide liquidity in small trades as the estimated sensitivities are negative in bins 1 to 3. The sensitivities turn positive as trade size increases and peak at bin 7 before gradually diminishing to zero. This variation suggests that hedge funds tailor their trading strategies depending on the size of the stock, employing complex and adaptive approaches to optimize their trades, manage market impact, or exploit specific informational advantages. In contrast, non-hedge funds exhibit far less variation across firm size quintiles, as depicted in Panel (d). The sensitivity curves for medium and large stocks are nearly identical, indicating that non-hedge funds trade these stocks in a largely similar manner. NHFs seem to provide liquidity in small trades in bins 1 to 6, and mainly take liquidity from trades in bins 8 to 14. When it comes to small-cap stocks, NHFs' sensitivity curve is similar to HFs' in this subperiod.

Second, moving to the 2003-2012 period, the differences between hedge funds and non-hedge funds persist but exhibit a shift with the most distinct sensitivity curves coming from large-cap stocks. Hedge funds have moderate negative sensitivities to small and medium trades in size bins 1 to 10, and strong positive sensitivities to large trades in size bins 14 to 19, as illustrated in Panel (b). In comparison, when trading large-cap stocks, non-hedge funds show stronger negative sensitivities to small trades in bins 1 to 6 than hedge funds in Panel (e). The sensitivities turn positive after bin 8 but flattens to the end of the size spectrum. The distinct sensitivity curves indicate that hedge funds can be more aggressive in taking bulky liquidity for large-cap stocks than non-hedge funds in this period, possibly due to informational reasons. The sensitivity curves for small and medium stocks are more similar between the two types of funds although the curves still show different levels of curvature and peaks.

In the final subperiod from 2013 to 2022, while the overall patterns of hedge funds and non-hedge funds appear more similar than in earlier periods, key differences in their sensitivities remain evident. For medium-cap stocks, hedge funds exhibit small negative sensitivities to small trades (bins 1 to 3) and have the largest positive sensitivities to trade-size bins 7 to 12, as illustrated in Panel (c). In comparison, Panel (f) shows that for the same group of stocks, non-hedge funds have negative sensitivities for bins 1 to 6, and the positive sensitivities concentrate in bins 9 to 14. Conversely, for large stocks, hedge funds show positive sensitivities to trade size bins 10 to 18 but exhibit negative sensitivities to small trades in bins 1 to 4. Non-hedge funds again exhibit preference over relatively smaller trades than hedge funds when trading large-cap stocks. The positive sensitivities come from bins 9 to 15. And for the largest trades in bin 19, non-hedge funds have a strong negative sensitivity which is absent in hedge funds' behavior.

Our estimation using quarterly data shows clear difference in trading styles between the two types of institutional investors. The contrasting patterns indicate that hedge funds tend to utilize a more dynamic strategy, adapting their trade sizes and execution methods based on stock characteristics and market conditions. They often prefer medium-sized trades, particularly when trading small and medium stocks. Non-hedge funds, meanwhile, appear to follow a broader but potentially less adaptive approach.

[Place Figure 2 about here]

In our last step of estimating daily hedge fund and non-hedge fund trades, we calculate the expected daily change of ownership separately for hedge funds and non-hedge funds, $\mathrm{E}[\Delta Y_{i,d}]$, taking the estimates in Equation (1):

$$\Delta Y_{i,d} = \alpha_d + \rho \Delta Y_{i,d-1} + \phi Y_{i,d-1} + \beta^U U_{i,d} + \sum_{Z=1}^{19} \beta(Z, v) F_{Z,i,d} + \epsilon_{i,d}, \tag{3}$$

where d indexes a day. The order flow variables can be calculated every day using the TAQ data. We set to zero unobservable daily variables of aggregate institutional ownership such as $\Delta Y_{i,d-1}$ and $Y_{i,d-1}$ as well as a set of daily dummies, α_d following CRS (2009). The frequency conversion is possible under an exogeneity assumption that the error terms, $\epsilon_{i,d}$, are not correlated with all of its leads and lags within a quarter. Our extension of the original CRS method allows us to estimate daily HF and NHF trades based on their distinct sensitivities to trade-size dependent order flows.

D. Summary statistics

Table I provides summary statistics, presenting the time-series averages of cross-sectional statistics for hedge fund order flow (HF), non-hedge fund order flow (NHF), total order flow (TOF), and risk-adjusted mid-quote return (Return) across four sample periods: 1993–2022 in Panel A, 1993–2002 in Panel B, 2003–2012 in Panel C, and 2013–2022 in Panel D. HF and NHF are estimated as described in Section II.C. TOF is calculated as buyer-initiated volume minus seller-initiated volume scaled by the number of shares outstanding with the trade direction classified based on the Lee and Ready (1991) algorithm. Return is a risk-adjusted mid-quote return with respect to the Fama−French (1993) factors and Carhart (1997) momentum factor. All the order flow variables are scaled by the number of shares outstanding and can be interpreted as the proportion of shares traded.

Over the full sample period, HF has an average value of 0.009, with means ranging from 0.005 to 0.010 across the other sample periods. NHF shows a higher mean than HF, averaging 0.024 in 1993–2022 and ranging between 0.018 and 0.032 in the subperiods. The standard deviations of both HF and NHF are approximately three times larger than their means, indicating substantial variability. The mean of TOF is 0.036 in 1993–2022, with values ranging from 0.022 to 0.056 across the subperiods. Notably, HF, NHF, and TOF all exhibit positive skewness. The average of Return is consistently 0.000 across all sample periods, with a slight positive skewness.

Table I also reports the time-series averages of the cross-sectional Spearman correlations among the four variables. HF and NHF are strongly correlated, with a correlation coefficient of 0.671 over the full sample period (1993–2022). Notably, the correlation between HF and NHF has increased over time: 0.547 in 1993–2002, 0.668 in 2003–2012, and 0.798 in 2013–2022. This evolving relationship in HF and NHF trading behavior aligns with the patterns observed in Figure 1. Additionally, HF shows a lower correlation with TOF than NHF, indicating that hedge fund trades are less aligned with overall market order flow compared to non-hedge fund trades. For example, in the 1993–2022 period, HF's correlation with TOF is 0.177, while NHF's correlation with TOF is higher at 0.320. Similarly, HF has a weaker correlation with Return than NHF does, with HF's correlation with Return at 0.070 in 1993–2022, compared to NHF's correlation of 0.097. This suggests that while both HF and NHF positively impact contemporaneous prices, HF's influence on current stock prices is weaker than NHF's. However, in the most recent sample period (2013–2022), the correlations of HF with TOF and Return become close to those of NHF, reflecting the similarity of HF and NHF trade behaviors in 2013-2022.

We formally examine the influence of HF on contemporaneous prices in Table A2 of the internet appendix. The results show that HF and NHF exert positive and statistically significant contemporaneous price pressure, with NHF trades having a stronger economic impact than HF trades, though this effect diminishes over time.

[Place Table I about here]

E. *Determinants of hedge fund and non-hedge fund trading*

To better understand the factors influencing the trading behavior of hedge funds and non-hedge funds, we employ a Fama−MacBeth (1973) regression model that includes key firm characteristics: market beta (MktBeta), market capitalization (MktCap), book-to-market ratio (BM), short-term reversal (REV), return momentum (MOM), relative bid-ask spread (SPRD), share volume turnover ratio (TURN), idiosyncratic risk (IdioRisk), and the mispricing index (MISP) proposed by Stambaugh, Yu, and Yuan (2012).[7] We use momentum as a separate characteristic together with other price and volume based characteristics, and exclude the momentum factor from the original mispricing index because all the other anomalies included in MISP are calculated using fundamental information obtained from financial statements. These firm characteristics are well known for their relations to future stock returns and we are interested to see how hedge funds and non-hedge funds react to them.

Because the firm characteristics are typically measured at the monthly frequency, we aggregate daily HF and NHF into monthly observations. The analysis is based on a total of 1,161,084 stock-month observations. Table II reports the time-series averages of cross-section estimated coefficients for the following model:

$$\begin{aligned}\mathrm{OF}_{i,m} &= \alpha_m + \beta_m^B \mathrm{MktBeta}_{i,m} + \beta_m^C \mathrm{MktCap}_{i,m} + \beta_m^{MB} \mathrm{MB}_{i,m} + \beta_m^R \mathrm{REV}_{i,m} + \beta_m^M \mathrm{MOM}_{i,m} \\ &\quad + \beta_m^S \mathrm{SPRD}_{i,m} + \beta_m^T \mathrm{TURN}_{i,m} + \beta_m^I \mathrm{IdioRisk}_{i,m} + \beta_m^{MISP} \mathrm{MISP\ excl.\ MOM}_{i,m} + \epsilon_{i,m}.\end{aligned}$$

[7] We document detailed calculation methods for these firm characteristics in the Appendix.

To account for serial correlations, we use Newey and West (1987) standard errors with five lags to calculate the t-statistics.

Our analysis reveals that the differences between hedge funds and non-hedge funds in their responses to well-known return predictors at the monthly frequency are less pronounced than one might anticipate. Both types of funds demonstrate a propensity to buy high-beta and large-cap stocks, trade contrarian to both short-term and long-term past returns, and favor liquid stocks characterized by low bid-ask spreads and high turnover. This behavior deviates from established return predictability based on monthly data. Specifically, their preference for large-cap stocks runs counter to the expectations set by the size premium, which favors smaller firms. Similarly, they trade against well-documented winning strategies that bet against beta (Frazzini and Pedersen (2014), investment in stock return momentum (Jegadeesh and Titman (1993)), and collect illiquidity premium on stocks with wide bid-ask spreads (Amihud and Mendelson (1986)). Thus, at daily frequency, both types of institutional investors prioritize transaction cost management and trade execution by tilting towards larger and more liquid stocks.

However, important distinctions emerge between the two groups, reflecting their divergent strategies and market roles. Hedge funds display a clear preference for value stocks, aligning with the value premium, and they actively engage in trading based on mispricing derived from fundamental information. This suggests a nuanced approach where hedge funds seek out undervalued opportunities to generate returns. On the other hand, hedge funds also show a strong inclination toward stocks with high idiosyncratic volatility, a characteristic indicating greater information asymmetry or higher rewards for being informed. In contrast, NHF exhibits no significant response to these characteristics.

In summary, our low-frequency analysis presents some evidence of how hedge funds create an edge over non-hedge funds by trading stocks with certain characteristics. While hedge funds favors larger and more liquid stocks likely for transaction cost reasons, they also exhibit a preference for value stocks and actively trade on mispricing derived from fundamental information and stocks displaying high idiosyncratic volatility where rewards are high for being informed. To better understand the information edge hedge funds might possess, we turn to higher frequency analyses to examine their role in price discovery and trading around news events, capturing how they leverage short-term information.

[Place Table II about here]

III. Hedge funds as informed traders

In this section, we test the hypothesis that hedge fund order flow is more informative about future stock prices than non-hedge fund order flow. Prior studies using lower-frequency data provided suggestive evidence for this conjecture. For instance, Kosowski, Naik, and Teo (2007) and Jagannathan, Malakhov, and Novikov (2010) document persistent superior performance in hedge funds. Agarwal, Jiang, Tang, and Yang (2013) find positive and significant abnormal returns associated with hedge funds' confidential holdings in 13F filings. However, research using high-frequency hedge fund order flow across a broad sample is limited, and we still do not have an in-depth understanding of how hedge funds exploit short-lived information and what contributes to their information edge. To address this gap, we examine hedge fund and non-hedge fund order flow through three sets of event studies focusing on information shocks, including: (1) scheduled and unscheduled corporate events, (2) arrival of news articles about both fundamental and non-fundamental information, and (3) initial coverage of alternative data.

A. Corporate event studies

In this subsection, we examine the reaction of hedge fund and non-hedge fund trades to salient corporate events. Our sample includes quarterly earnings announcements, analyst rating updates to capture unscheduled corporate announcements, and permanent price jumps not related to earnings news, which reflect other material information. Quarterly earnings announcements and analyst rating updates are sourced from the I/B/E/S database, while permanent price jumps are defined as two standard-deviation shocks that do not fully reverse within the following ten days. After merging the event data with our trading sample, we have $1,291,649$ stock-day observations for these corporate events.

Table III presents the estimated coefficients from the following model, using ordinary least squares regressions with firm and year fixed effects:

$$\begin{aligned}\mathrm{CAR}_{i,t-1,t+1} \;&=\; \alpha_i + \alpha_y + \sum_{k=2}^{6} \beta_k^{HF}\mathrm{HF}_{i,t-k} + \sum_{k=2}^{6} \beta_k^{NHF}\mathrm{NHF}_{i,t-k} + \sum_{k=2}^{6} \gamma_k^{T}\mathrm{TOF}_{i,t-k} \\ &\quad + \sum_{k=2}^{6} \gamma_k^{R}\mathrm{Return}_{i,t-k} + \gamma^{S}\mathrm{SPRD}_{i,t-2} + \gamma^{S}\mathrm{AMI}_{i,t-2} + \epsilon_{i,t},\end{aligned}$$

where for each event i on day t, CAR is the cumulative abnormal return from days $t-1$ to $t+1$, and all the explanatory variables are the same as defined in Table VI with event subscription i instead of firm subscription. We cluster the standard errors around firm and year in calculating the t-statistics. We report results for the main variables in Table III and exclude the control variables for brevity.

We first examine these three event types separately over the full sample period (1993–2022).The results are reported in the first column for each event type in Table III. Consistent across all event samples, we find that the estimated coefficients for HF are always positive and significant on day $d-2$, just before the event window, with the positive effects persisting through the subsequent days ($d-3$ to $d-6$). Specifically, HF shows coefficients of 8.091 (t-stat = 5.16) for earnings announcements, 1.870 (t-stat = 2.73) for analyst rating updates, and 11.950 (t-stat = 5.82) for permanent price jumps. In contrast, NHF is either negatively or insignificantly associated with cumulative abnormal returns around corporate events, both on day $d-2$ and in the following days ($d-3$ to $d-6$). These results suggest that the return predictability of hedge fund trades is closely tied to fundamental information flow surrounding significant corporate events.

We then analyze the three event types across two subperiods, separated by the enactment of Regulation Fair Disclosure (RegFD): 1993–1999 and 2000–2022. Table III present the results for these subperiods. Consistent with the full sample results, HF shows positive and significant coefficients on day $d-2$ across all event types, while NHF displays negative coefficients. However, both the statistical and economic significance of HF and NHF coefficients weaken in the 2000–2022 subperiod compared to 1993–1999. The enactment of RegFD, which mandates timely disclosure of material information by firms, potentially reduces the informational advantage of hedge funds by narrowing the gap between informed and uninformed investors. As a result, the price impact of both hedge fund and non-hedge fund trades around corporate events has diminished in the more recent period.

[**Place Table III about here**]

B. HF trading around news

In this subsection, we examine the reaction of hedge fund and non-hedge fund trades to a comprehensive collection of firm-specific news stories. Our news data come from RavenPack between 2000 and 2022, sourced from the Dow Jones Newswire. To ensure that each news story is specifically about a given firm, we include only those news with a relevance score of 100.[8] Additionally, we require the "event relevance" score, also provided by RavenPack, to be 100, ensuring that we capture only news articles that mention a company in the headline. To mitigate concerns about the look-ahead bias, we exclude events for a given stock that occur within 30 calendar days of another news event related to the same stock. Furthermore, to accurately compute the predictive power of hedge fund and non-hedge fund trades in pre-event periods, we adjust the news event date to the next trading day if the news article is issued after 4:00 pm. Lastly, we include all news "groups" within the business "topic" in the

[8] According to the RavenPack User Guide 1.0 (2020), the "relevance score" ranges from 0 to 100, indicating how closely a news item pertains to a particular company: 0 meaning the company is only passively mentioned and 100 indicating the company is predominantly featured.

RavenPack database that cover more than 5,000 firms to ensure representation of smaller firms (Kolasinski and Yang (2018)).[9] After merging with our trading sample, we obtain $1,956,008$ stock-day observations for these RavenPack news events.

We present the results from RavenPack news event studies in Table IV. Panel A focuses on news groups related to firm fundamentals, including earnings, analysts, mergers and acquisitions, assets, credit, dividends, equity actions, labor issues, products and services, revenues, and partnerships. Across all these news groups, hedge fund trades consistently show positive and significant return predictability on firm-specific news within the five days from $d-2$ to $d-6$. By comparison, non-hedge fund trades during the same period exhibit a negative price impact on the day of the news event across all fundamental news groups. Panel B examines non-fundamental news groups, including insider information, investor relations, marketing, order imbalance, price targets, and stock prices. Even in these non-fundamental news categories, the results remain qualitatively similar: hedge fund trades demonstrate positive coefficients, while non-hedge fund trades show negative coefficients in the lagged period.

[**Place Table IV about here**]

Taken the results of Tables III and IV together, the findings from these event studies suggest that the return predictability of hedge fund trades is closely tied to the flow of both fundamental and non-fundamental information in financial markets.

C. Alternative data

This subsection investigates the impact of alternative data, such as satellite imagery from RS Metrics, on the predictive power of hedge fund trades for stock returns on subsequent dates. Satellite data provides valuable insights into forthcoming quarterly reports (e.g.,

[9] We combine three news groups—bankruptcy, credit-ratings, and credit—into a single group named 'credit' due to the limited coverage of firms in the bankruptcy and credit-rating groups and exclude a few news groups with very limited observations.

Kang, Lorien, and Wong (2021); Bonelli and Foucault (2024); Katona, Painter, Patatoukas, and Zeng (2024)). With access to this data, hedge funds may have more accurate information regarding firm performance. For this reason, we expect hedge fund trades to demonstrate stronger predictive power for the stock returns of major retailers covered by satellite data after its release.

We acquire RS Metrics data between May 2009 and September 2018 because we require a three-year window post coverage in this event study. We begin by identifying the exact date when RS Metrics commenced the provision of state-level parking lot traffic data, extracted from satellite imagery, in the U.S. market.[10] Next, to address concerns that stocks covered by RS Metrics might systematically differ from other stocks, we match the 48 covered stocks with three control stocks that (1) are not covered by RS Metrics in our sample period, (2) belong to the same broad industry (i.e., Wholesale Trade or Retail Trade sectors), and (3) have the closest market capitalization to the corresponding covered stocks. The sample period spans three years before and three years after RS Metrics' coverage initiation for each covered stock. Combined with our trading data, this yields $186,956$ stock-day observations.

Table V presents the estimated coefficients from the following model, using a difference-in-differences approach with firm and year fixed effects:

$$\begin{aligned}
\text{Return}_{i,d} &= \alpha_i + \alpha_y + \beta_1 \text{TREAT} \times \text{POST} \times \text{HF}_{d-1} + \beta_2 \text{TREAT} \times \text{POST} \times \text{NHF}_{d-1} \\
&+ \beta_3 \text{TREAT} \times \text{POST} \times \text{TOF}_{d-1} + \beta_4 \text{TREAT} \times \text{HF}_{d-1} + \beta_5 \text{TREAT} \times \text{NHF}_{d-1} \\
&+ \beta_6 \text{TREAT} \times \text{TOF}_{d-1} + \beta_7 \text{POST} \times \text{HF}_{d-1} + \beta_8 \text{POST} \times \text{NHF}_{d-1} + \beta_9 \text{POST} \times \text{IOF}_{d-1} \\
&+ \beta_{10} \text{TREAT} \times \text{POST} + \beta_{11} \text{TREAT} + \beta_{12} \text{POST} + \beta_{13} \text{HF}_{d-1} + \beta_{14} \text{NHF}_{d-1} + \beta_{15} \text{TOF}_{d-1} \\
&+ \gamma \text{Controls}_{d-1} + \epsilon_{i,d},
\end{aligned}$$

where, for stock i on day d, TREAT is a dummy equal to one if the stock is covered by RS Metrics, and POST is a dummy equal to one after RS Metrics initiates coverage of

[10] RS Metrics is recognized as the first vendor to introduce nearly real-time daily data feeds, enabling sophisticated investors, such as hedge funds, to incorporate this data into their trading strategies (see Katona, Painter, Patatoukas, and Zeng (2024) for more details).

the stock. We include market capitalization (MktCap) as a control variable because large retailer stocks are significantly more likely to be selected for satellite coverage (Bonelli and Foucault (2024)), along with lagged stock returns and illiquidity measures. Standard errors are clustered by firm and year.

In Table V, we find a positive and significant coefficient for β_1 in both columns with and without control variables, suggesting that hedge fund trades exhibit stronger predictive power for stock returns on the subsequent day after the introduction of satellite data on major retailers in U.S. markets. The economic impact is notable: a one-standard-deviation increase in hedge fund trades is associated with a 1.09% increase in mid-quote return on the subsequent day for treated stocks after the introduction of satellite data. In contrast, the estimated coefficient for β_2 is negative in both columns, indicating that non-hedge funds have relatively weaker return predictability for the retailers' stock returns on the next day. The contrast clearly shows that hedge funds gain an informational advantage compared to non-hedge funds. This interpretation aligns with Katona, Painter, Patatoukas, and Zeng (2024), who argue that the substantial costs associated with acquiring and processing satellite data mean that only a select group of informed traders like hedge funds, with advanced analytical capabilities, typically use satellite data vendors.

[Place Table V about here]

We also perform a back-of-the-envelope calculation for the potential profit of hedge funds derived from acquiring the alternative data. Specifically, we evaluate the profit of hedge fund order flow. Our estimate of HF reveals the hedge fund trading direction and quantity through aggressive orders. However, we do not observe their trades directly. Instead, we assume hedge funds build their positions at an average cost the same as the close price every trading day and hold the position for one day only. In this way, we can compute the holding period profits of hedge funds for all treated stocks after the coverage of RS Metrics. Over a three-year post-coverage window, the total profits for all hedge funds on these 48 treated stocks amount to \$2.87 billion. The sizable trading profit clearly shows the value of acquiring private information by taking advantage of alternative data. Overall, the active exploration of alternative data appears to provide an important edge for hedge funds in generating superior performance.

IV. Return predictability in the cross-section

Our investigation so far provides strong and clear evidence of the informational advantage of hedge funds over their peer institutions in a comprehensive set of events relevant to firm values. We show that hedge funds trade prior to the arrival of such news consistent with the subsequent market reaction to the event. In this section, we evaluate whether hedge funds' ability to trade ahead of short-lived information generalizes to return predictability in the broader sample. This analysis speaks directly to how pervasive informed trading is by hedge funds by testing the predictive power of hedge fund trading over the entire sample period (not just around informational events as in the analysis earlier).

A. *Baseline predictability*

Note that HF and NHF order flows are highly correlated (0.671) and that both HF and NHF order flows exert positive and statistically significant contemporaneous price pressure. To examine potential return predictive ability of HF and NHF, we run Fama and Macbeth (1973) regressions in the full cross-section of stocks in our sample. Table VI reports the estimated coefficients of the following model:

$$\begin{aligned}\text{Return}_{i,d} \;=\; & \alpha_d + \sum_{k=1}^{5} \beta_{d,k}^{HF}\text{HF}_{i,d-k} + \sum_{k=1}^{5} \beta_{d,k}^{NHF}\text{NHF}_{i,d-k} \\ & + \sum_{k=1}^{5} \gamma_{d,k}^{T}\text{TOF}_{i,d-k} + \sum_{k=1}^{5} \gamma_{d,k}^{R}\text{Return}_{i,d-k} + \gamma_d^{B}\text{SPRD}_{i,d-1} + \gamma_d^{A}\text{AMI}_{i,d-1} + \epsilon_{i,d}.\end{aligned}$$

where for stock i on day d, SPRD is relative bid-ask spread, AMI is Amihud (2002) illiquidity measure, and the other variables are the same as defined in Table I. To account for serial correlations, we use Newey−West (1987) standard errors with ten lags to calculate the t-statistics. For brevity, Table VI reports the coefficient estimates of HF and NHF only, while the regressions always include the full set of control variables.

For the full sample period spanning 1993 to 2022, the estimated coefficient of HF is positive and statistically significant at the first lag, with a value of 1.540 and a t-statistic of 17.10. Economically, this translates into a notable effect: a one standard-deviation increase in HF corresponds to a 3.7 bp increase in the next day's stock return. This positive effect persists across multiple lags, with coefficients of 0.047 (t-stat = 0.75), 0.144 (t-stat = 2.28), 0.108 (t-stat = 1.70), and 0.155 (t-stat = 2.44) at the second, third, fourth, and fifth lags, respectively. In contrast, non-hedge fund (NHF) trades display a negative and significant coefficient of −0.283 at the first lag (t-stat = −12.23), with this negative impact persisting over subsequent lags. Together with our evidence on contemporaneous price impact, these findings imply that hedge fund trades exert a lasting impact on stock prices. In contrast, non-hedge fund trades tend to create only temporary price pressure that reverses later. This supports the notion that hedge funds hold an informational advantage over other institutional investors.

Turning to the subperiod analysis, we observe that the persistent price impact of hedge fund trades remains robust but diminishes over time. During the 1993–2002 period, the coefficient for the first lag of HF is 2.883 (t-stat = 15.27), declining to 1.298 (t-stat = 11.86) in the 2003–2012 period, and further tapering to 0.443 (t-stat = 4.84) in the 2013–2022 period. Similarly, NHF exhibits reversals across all subperiods, with coefficients of −0.572 (t-stat = −12.56) in 1993–2002, −0.202 (t-stat = −7.65) in 2003–2012, and −0.077 (t-stat = −2.11) in 2013–2022. This attenuation in both statistical and economic significance over time may reflect the evolution of a more transparent and efficient market. In Table A3 of the internet appendix, we also explore the predictability at the monthly horizon and find consistent results that HF significantly and positively predicts future returns in the cross-section while NHF has a negative predictive ability only. Collectively, these results reinforce the idea that hedge funds contribute to price discovery by incorporating private information through their trades, while the influence of non-hedge fund trades remains more transient, consistent with the notion that hedge funds possess a unique informational edge in financial markets.

[**Place Table VI about here**]

B. Cross-sectional variation in the predictability

Next, we investigate the predictive power of HF and NHF within subsamples of stocks to explore cross-sectional variation in predictability. In Table VII, Panel A, we begin by conditioning on liquidity factors, including firm size, spread, and Amihud illiquidity. To accomplish this, we sort all stocks into two groups based on the daily cross-sectional median of each liquidity proxy. Table VII Panel A reports the estimated coefficients of HF and NHF from the regression model in Table VI for each subsample. The results can be summarized as follows: First, focusing on the subsequent day's pricing effect, HF exhibits positive and significant coefficient estimates across all the liquidity-based subsamples, while NHF shows negative and significant estimates. Second, HF's pricing effect appears permanent across all the liquidity subsamples, with no significant reversals observed. Third, HF's pricing effects on the subsequent day vary inversely with liquidity measures in terms of statistical significance. Specifically, HF's first-lag coefficients are 3.066 (t-stat = 17.64) for small stocks and 1.176 (t-stat = 12.52) for large stocks; 1.116 (t-stat = 12.09) for stocks with narrow spreads and 2.324 (t-stat = 15.80) for stocks with wide spreads; 0.761 (t-stat = 8.78) for liquid stocks with low Amihud measures and 4.492 (t-stat = 19.78) for illiquid stocks with high Amihud measures. A stronger pricing effect in illiquid stocks is consistent with the notion that illiquid stocks have higher information asymmetry and incur higher arbitrage costs; therefore, rewards are higher for informed investors in these stocks.

In Table 5, Panel B, we examine subsamples of stocks based on the information environment, measured by the number of sell-side analysts (Analyst Coverage) and institutional ownership (Ownership). As in Panel A, we separate all stocks into two groups based on the daily cross-sectional median of each proxy for the information environment, and then replicate the regression model from Table VI. The results show that HF has positive and significant predictive power in all Analyst and Ownership subsamples, while the coefficient of NHF is negative and significant in all the subsamples. HF's pricing effect is notably stronger for stocks with low analyst coverage or low institutional ownership. Specifically, HF has a first-lag coefficient of 2.389 (t-stat = 16.61) for stocks with low analyst coverage and 0.758 (t-stat = 8.80) for stocks with high analyst coverage. Similarly, HF's first-lag coefficient is 2.475 (t-stat = 15.15) for stocks with low institutional ownership and 1.070 (t-stat = 13.14) for stocks with high institutional ownership. These findings support the notion that more opaque stocks incur higher arbitrage risks and information asymmetry, and these are stocks that hedge funds have a unique advantage.

In summary, the results presented in this subsection demonstrate that HF's performance remains robust across various stock subsamples, with particularly strong performance in stocks facing higher arbitrage costs and those with less favorable information environments.

[Place Table VII about here]

V. HF trading and market efficiency

After documenting high-frequency evidence of informed trading by hedge funds, we move on to test the implication on market efficiency as a result of hedge fund trades. Prior research indicates that hedge funds, as sophisticated investors, play a crucial role in reducing mispricing and improving market efficiency (Kokkonen and Suominen (2015); Rösch, Subrahmanyam, and Dijk (2017); Cao, Liang, Lo, and Petrasek (2018); Chen, Kelly, and Wu

(2020)). For example, Kokkonen and Suominen (2015) show that hedge funds actively correct misvaluations by trading both undervalued and overvalued stocks, thereby contributing to market efficiency. Similarly, Cao, Liang, Lo, and Petrasek (2018) provide evidence that stocks held by hedge funds experience substantial improvements in informational efficiency, with hedge fund ownership playing a more significant role in price efficiency than other institutional investors. However, there remains limited research regarding the impact of hedge fund order flow at high frequency on price efficiency across a broad set of stocks. To address this gap, we examine the relationship between hedge fund trading and price efficiency at a daily frequency. Our analysis focuses on two key areas: 1) the effect of hedge fund trades on price efficiency around earnings announcements, and 2) the overall impact of hedge fund trades on price efficiency in the broader market.

A. Earnings announcements

We first examine the relationship between hedge fund trading and price efficiency around earnings announcements. Earnings announcements provide an ideal setting for investigating the impact of hedge fund trades on price efficiency for two key reasons. First, earnings announcements are arguably the most important value-relevant events, where firms disclose their past profitability and offer projections for future performance (Beyer, Cohen, Lys, and Walther (2010)). Second, informed traders, such as hedge funds, are particularly active prior to earnings announcements (Weller (2018)). To measure price efficiency, we use three distinct metrics: (i) post-earnings announcement drift (PEAD), which captures slow diffusion of information after the announcement; (ii) Weller's (2018) jump ratio, which measures immediate stock price reaction to newly announced information; and (iii) the change in the variance ratio, which captures the overall shift in price efficiency before and after earnings announcements. If hedge funds engage in informed trading, their trades can incorporate private information into the stock prices and reduce the magnitude of all three measures of price efficiency. Earnings announcement dates are sourced from Compustat, I/B/E/S, and

RavenPack.

Table VIII presents results from the following least square regression model with firm and quarter fixed effects:

$$\text{y}_{i,d} = \alpha_i + \alpha_q + \beta^{HF}\text{HF}_{i,d} + \beta^{NHF}\text{NHF}_{i,d} + \gamma\text{ControlVariables}_{i,d} + \epsilon_{i,d},$$

where, for stock i on earnings announcement day d in quarter q, the dependent variables are $\text{CAR}_{d+1,d+61}$, Jump Ratio Rank, and ΔVR. $\text{CAR}_{d+1,d+61}$ represents the cumulative abnormal return, compounded over the 60-day post-announcement period covering days 1 to 61 following the earnings announcement. The Jump Ratio is calculated as the ratio of the abnormal return on the earnings announcement day (AR_d) to the cumulative abnormal return over the pre-announcement period of days -21 to 0 ($\text{CAR}_{d-21,d}$). To mitigate the influence of extreme values, Jump Ratio is ranked cross-sectionally, and Jump Ratio Rank is defined as a categorical variable ranging from 0 (for the lowest decile) to 9 (for the highest decile). ΔVR represents the difference between the variance ratio averaged over days +21 to +1 and the variance ratio averaged over days -21 to -1. The variance ratio itself is the absolute difference between the 15-to-60 second stock return variance and 1. Because both the Jump Ratio Rank and ΔVR are unsigned, we use the absolute values of HF and NHF summed over days -21 to -1 in those tests. The control variables include institutional ownership (InstOwn), the number of analysts covering the stock (NumAnalyst), market capitalization, relative bid-ask spread, Amihud illiquidity, cumulative abnormal return, and the standard deviation of abnormal returns.

Column (1) of Table VIII presents the results for post-earnings announcement drift (PEAD). Hedge fund trades on the day of the earnings announcement (HF_d) exhibit a negative and significant coefficient of -8.069, with a t-statistic of -2.19. This negative relationship suggests that hedge fund trades facilitate the rapid incorporation of earnings information into stock prices, thereby reducing the size of PEAD and improving price efficiency. In contrast, non-hedge fund trades (NHF_d) show a positive and significant coefficient of 3.168, with a t-statistic of 2.02, indicating that non-hedge funds may impede the timely incorporation of earnings information, ultimately amplifying PEAD.

Column (2) examines Weller's (2018) jump ratio, which measures the market response to newly announced information. The results show that both hedge fund and non-hedge fund trades, averaged over the 21 days prior to earnings announcements, have negative and significant coefficients: -1.038 (t-stat = -10.10) for hedge fund trades and -0.653 (t-stat = -13.35) for non-hedge fund trades. These negative coefficients suggest that active trading by both hedge funds and other institutions prior to earnings announcements incorporates private information into stock prices and reduces the informational value of the announcement.

Finally, in Column (3), the change in the variance ratio (ΔVR) is used to assess overall price efficiency before and after earnings announcements. The absolute value of hedge fund trades on the announcement day ($|\text{HF}|_d$) is negatively and significantly associated with ΔVR, with a coefficient of -2.628 and a t-statistic of -2.41. This indicates that hedge fund trades contribute to a reduction of market risk through their trading on the earnings announcement date. Although non-hedge fund trades ($|\text{NHF}|_d$) also display a negative coefficient, it is not statistically significant, suggesting that their effect on price efficiency is weaker and less consistent.

Overall, the results across all columns demonstrate that hedge fund trades play a significant role in enhancing price efficiency around earnings announcements, while the evidence regarding non-hedge funds is mixed and inconclusive.

[Place Table VIII about here]

B. *Price efficiency in the cross-section*

In this subsection, we analyze the role of daily hedge fund trading in price efficiency in the full cross-section of stocks using the following Fama-MacBeth regression model:

$$\Delta\text{VR}_{i,d} = \alpha_d + \beta_d^{HF}|\text{HF}|_{i,d-1} + \beta_d^{NHF}|\text{NHF}|_{i,d-1} + \gamma\text{ControlVariables}_{i,d-3} + \epsilon_{i,d},$$

where, for stock i on day d, ΔVR represents the two-day change in the variance ratio: ΔVR $= VR_{i,d} - VR_{i,d-2}$. Other variables are the same as those defined in Table VIII.

Table IX reports the estimated coefficients. We find that, across all subperiods, hedge fund trades consistently show negative estimated coefficients, indicating that hedge fund trades reduce the variance ratio and enhance price efficiency throughout the sample period. In contrast, the effect of non-hedge fund trades on price efficiency is time-varying: non-hedge fund trades significantly impair price efficiency in the 1993-2002 period but improve it after 2002. Notably, both hedge fund and non-hedge fund trades exhibit negative estimated coefficients in the 2013-2022 period, suggesting that both types of institutions contributed to improved price efficiency in the most recent sample period. This finding aligns with our earlier results in Figure 2 and Tables I and A2, supporting the interpretation that non-hedge fund trading behavior has become increasingly similar to that of hedge funds over time.

[Place Table IX about here]

VI. Conclusion

In this paper, we study daily aggregate hedge fund trades in individual U.S. stocks to shed light on their role as informed traders over the past 30 years. By constructing a measure of daily hedge fund order flow as an extension of Campbell, Ramadorai, and Schwartz (2009)'s method, we provide compelling evidence that hedge funds are informed traders and contribute significantly to price discovery in equity markets. Our findings show that hedge fund order flow not only predicts future stock returns without reversal but also plays a more critical role in price formation than non-hedge fund order flow.

Through various tests, we confirm that hedge fund trades have superior predictive power for stock returns, particularly in stocks with weaker informational environments or higher arbitrage costs. Unlike non-hedge fund institutions, hedge funds exhibit a permanent impact on stock prices, suggesting that their trades are based on superior information. Additionally, we demonstrate that hedge funds trade more efficiently around key information events, such as earnings announcements and major corporate news, integrating information quickly into prices. One key contribution is to be the first to uncover that hedge funds make informed trades ahead of corporate events and new arrivals. Our analysis of hedge fund trading extends to the utilization of alternative data, such as satellite imagery, further reinforcing the notion that hedge funds possess advanced capabilities for extracting and acting on new, less conventional forms of information. We find that hedge funds are better positioned to process and capitalize on such data, further enhancing their role in price efficiency.

This research contributes to the literature in two important ways. First, we estimate daily hedge fund order flow that can be used to analyze hedge fund and non-hedge fund trading behavior at higher frequencies. Second, our study highlights the unique contribution of hedge funds to market efficiency, showing that their trades are more informative and effective compared to those of other institutional investors. The implications of these findings are broad, opening avenues for future research. Further exploration into how hedge funds leverage big data and artificial intelligence technologies and how their trading behavior might evolve in increasingly efficient markets would provide valuable insights for understanding their continued role in global financial markets.

Appendix. Firm Characteristics Related to Institutional Trading

We construct a monthly sample of common stocks listed on the NYSE, AMEX, and Nasdaq that are available in the CRSP, Compustat, TAQ, and Thomson Reuters 13F Ownership databases from 1993 to 2022. We exclude stocks priced below $5 to avoid market microstructure effects. we also exclude stocks with month-end market capitalization below the 5th percentile of NYSE breakpoints, focusing on more liquid and widely held stocks. For determinant variables requiring firm-level data from Compustat, we use annual financial statements, ensuring that the Compustat reporting date (item RDQ) precedes the end of the month. For variables based on stock data from CRSP, we rely on information recorded for the given month or earlier, as reported by CRSP. Our final monthly sample comprises 1,161,084 stock-month observations, merged with a trading dataset containing monthly aggregated hedge fund and non-hedge fund order flows.

The determinant variables are defined as follows:

1. MISP excl. MOM: The mispricing index for a stock is calculated as the average decile rank values across 10 anomaly variables, measured at the end of each month, with values ranging from 10 to 100. For each anomaly, stocks are ranked based on their values for that anomaly variable, with the highest rank assigned to the value associated with the highest average abnormal return. A higher rank suggests a greater degree of underpricing relative to that anomaly. Thus, stocks with the highest MISP values are considered the most underpriced, while those with the lowest values are considered the most overpriced. The 10 anomaly variables used in the calculation are listed below:

   - Net stock issues (Ritter (1991); Loughram and Ritter (1995))
   - Composite equity issues (Daniel and Titman (2006))
   - Failure probability (Campbell, Hilscher, and Szilagyi (2008))
   - O-Score (Ohlson (1980))

- Total accruals (Sloan (1996))
- Net operating assets (Hirshleifer, Hou, Teoh, and Zhang (2004))
- Gross profitability (Novy-Marx (2013))
- Asset growth (Cooper, Gulen, and Schill (2008))
- Return on assets (Fama and French (2006); Chen, Novy-Marx, and Zhang (2011))
- Investment-to-asset (Titman, Wei, and Xie (2004))

2. MktBeta: Market beta, estimated over the past 36 months using the Fama–French three-factor model.
3. MktCap: Market capitalization, defined as the product of the stock's price and the number of shares outstanding at the end of each month.
4. MB: Market-to-book ratio, calculated as the ratio of market capitalization to book equity value at the end of each month.
5. REV: Short-term reversal, defined as the one-month-ahead return at the end of each month.
6. MOM: Return momentum, calculated as the cumulative return over the period from month $m-2$ and $m-6$.
7. SPRD: Relative bid-ask spread, calculated as the bid-ask spread divided by the average of the bid and ask prices at the end of each month.
8. TURN: Share volume turnover ratio, defined as the monthly trading volume divided by the total number of shares outstanding.
9. IdioRisk: Idiosyncratic risk, measured as the standard deviation of the residuals from the Fama–French three-factor model over the past 36 months.

References

[1] Agarwal, Vikas, Jiang, Wei, Tang, Yuehua, and Yang, Baozhong, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, The Journal of Finance **68**, pp. 739–783.

[2] Agarwal, Vikas, Mullally, K. A., and Naik, N. Y., 2015, The economics and finance of hedge funds: A review of the academic literature, Foundations and Trends in Finance **10**, pp. 1–107.

[3] Agarwal, Vikas, Ruenzi, Stefan, and Weigert, Florian, 2017, Tail risk in hedge funds: A unique view from portfolio holdings, Journal of Financial Economics **125**, pp. 610–636.

[4] Agarwal, Vikas, Ruenzi, Stefan, and Weigert, Florian, 2024, Unobserved performance of hedge funds, The Journal of Finance **79**, pp. 3203–3259.

[5] Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, Journal of Financial Markets **5**, pp. 31–56.

[6] Amihud, Yakov and Mendelson, Haim, 1986, Asset pricing and the bid-ask spread, Journal of Financial Economics **17**, pp. 223–249.

[7] Aragon, G. O., 2007, Share restrictions and asset pricing: Evidence from the hedge fund industry, Journal of Financial Economics **83**, pp. 33–58.

[8] Aragon, G. O. and Martin, J. S., 2012, A unique view of hedge fund derivatives usage: Safeguard or speculation?, Journal of Financial Economics **105**, pp. 436–456.

[9] Battalio, R. H. and Mendenhall, R. R., 2005, Earnings expectations, investor trade size, and anomalous returns around earnings announcements, Journal of Financial Economics **77**, pp. 289–319.

[10] Beyer, A., Cohen, D., Lys, T., and Walther, B., 2010, The financial reporting environment: Review of the recent literature, Journal of Accounting and Economics **50**, pp. 296–343.

[11] Boehmer, E. and Kelley, E. K., 2009, Institutional investors and the informational efficiency of prices, Review of Financial Studies **22**, pp. 3563–3594.

[12] Bolandnazar, Mohammadreza, Jackson Robert J., Jr., Jiang, Wei, and Mitts, Joshua, 2020, Trading against the random expiration of private information: A natural experiment, The Journal of Finance **75**, pp. 5–44.

[13] Bonelli, M. and Foucault, T. *Displaced by big data: Evidence from active fund managers*. 2024.

[14] Brunnermeier, M. K. and Nagel, S., 2004, Hedge funds and the technology bubble, Journal of Finance **59**, pp. 2013–2040.

[15] Campbell, John, Ramadorai, Tarun, and Schwartz, Allie, 2009, Caught on tape, Journal of Financial Economics **92**, p. 26.

[16] Campbell, John Y., Hilscher, Jens, and Szilagyi, Jan, 2008, In search of distress risk, The Journal of Finance **63**, pp. 2899–2939.

[17] Cao, Charles, Chen, Yong, Liang, Bing, and Lo, Andrew W., 2013, Can hedge funds time market liquidity?, Journal of Financial Economics **109**, pp. 493–516.

[18] Cao, Charles, Liang, Bing, Lo, Andrew W., and Petrasek, Lubomir, 2018, Hedge fund holdings and stock market efficiency, The Review of Asset Pricing Studies, rax015–rax015.

[19] Carhart, M. M., 1997, On persistence in mutual fund performance, The Journal of Finance **52**, pp. 57–82.

[20] Chen, Long, Novy-Marx, Robert, and Zhang, Lu, 2011, An alternative three-factor model, Working Paper,

[21] Chen, Yong, Da, Zhi, and Huang, Dayong, 2019, Arbitrage trading: The long and the short of it, The Review of Financial Studies **32**, pp. 1608–1646.

[22] Chen, Yong, Kelly, Bryan, and Wu, Wei, 2020, Sophisticated investors and market efficiency: Evidence from a natural experiment, Journal of Financial Economics **138**, pp. 316–341.

[23] Cooper, Michael J., Gulen, Huseyin, and Schill, Michael J., 2008, Asset growth and the cross-section of stock returns, The Journal of Finance **63**, pp. 1609–1651.

[24] Çötelioğlu, Efe, Franzoni, Francesco, and Plazzi, Alberto, 2021, What constrains liquidity provision? Evidence from institutional trades, Review of Finance **25**, pp. 485–517.

[25] Cready, William, Kumas, Abdullah, and Subasi, Musa, 2014, Are trade size-based inferences about traders reliable? Evidence from institutional earnings-related trading, Journal of Accounting Research **52**, pp. 877–909.

[26] Daniel, Kent and Titman, Sheridan, 2006, Market reactions to tangible and intangible information, The Journal of Finance **61**, pp. 1605–1643.

[27] Dessaint, Olivier, Foucault, Thierry, and Fresard, Laurent, 2024, Does alternative data improve financial forecasting? The horizon effect, The Journal of Finance **79**, pp. 2237–2287.

[28] Fama, Eugene F. and French, Kenneth R., 2006, Profitability, investment, and average returns, Journal of Financial Economics **82**, pp. 491–518.

[29] Fama, Eugene F. and Macbeth, James D., 1973, Risk, return, and equilibrium: Empirical tests, Journal of Political Economy **81**, pp. 607–636.

[30] Frazzini, Andrea and Pedersen, Lasse Heje, 2014, Betting against beta, Journal of Financial Economics **111**, pp. 1–25.

[31] Fung, William and Hsieh, David A., 2004, Hedge fund benchmarks: A risk-based approach, Financial Analysts Journal **60**, pp. 65–80.

[32] Gao, P. George, Gao, Pengjie, and Song, Zhaogang, 2018, Do hedge funds exploit rare disaster concerns?, The Review of Financial Studies **31**, pp. 2650–2692.

[33] Griffin, John M. and Xu, Jin, 2009, How smart are the smart guys? A unique view from hedge fund stock holdings, The Review of Financial Studies **22**, pp. 2531–2570.

[34] Hendershott, T., Livdan, D., and Schurhoff, N., 2015, Are institutions informed about news?, Journal of Financial Economics **117**, pp. 249–287.

[35] Hirshleifer, David, Hou, Kewei, Teoh, Siew Hong, and Zhang, Yinglei, 2004, Do investors overvalue firms with bloated balance sheets?, Journal of Accounting and Economics **38**, pp. 297–331.

[36] Huang, Alan Guoming, Tan, Hongping, and Wermers, Russ, 2020, Institutional trading around corporate news: Evidence from textual analysis, The Review of Financial Studies **33**, pp. 4627–4675.

[37] Hvidkjaer, S., 2008, Small trades and the cross-section of stock returns, Review of Financial Studies **21**, pp. 1123–1151.

[38] Jagannathan, Ravi, Malakhov, Alexey, and Novikov, Dmitry, 2010, Do hot hands exist among hedge fund managers? An empirical evaluation, The Journal of Finance **65**, pp. 217–255.

[39] Jame, Russell, 2018, Liquidity provision and the cross-section of hedge fund returns, Management Science **64**, pp. 3288–3312.

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

[41] Jiao, Yawen, Massa, Massimo, and Zhang, Hong, 2016, Short selling meets hedge fund 13f: An anatomy of informed demand, Journal of Financial Economics **122**, pp. 544–567.

[42] Kang, J. K., Lorien, S., and Wong, Forester, 2021, The firm next door: Using satellite images to study local information advantage, Journal of Accounting Research **59**, pp. 713–750.

[43] Katona, Zsolt, Painter, Marcus O., Patatoukas, Panos N., and Zeng, Jean, 2024, On the capital market consequences of big data: Evidence from outer space, Journal of Financial and Quantitative Analysis, pp. 1–29.

[44] Kokkonen, Joni and Suominen, Matti, 2015, Hedge funds and stock market efficiency, Management Science **61**, pp. 2890–2904.

[45] Kolasinski, Adam C. and Yang, Nan, 2018, Managerial myopia and the mortgage meltdown, Journal of Financial Economics **128**, pp. 466–485.

[46] Kosowski, R., Naik, N. Y., and Teo, M., 2007, Do hedge funds deliver alpha? A bayesian and bootstrap analysis, Journal of Financial Economics **84**, pp. 229–264.

[47] Kumar, Nitish, Mullally, Kevin, Ray, Sugata, and Tang, Yuehua, 2020, Prime (information) brokerage, Journal of Financial Economics **137**, pp. 371–391.

[48] Lee, C. M. C., 1992, Earnings news and small trades - an intraday analysis, Journal of Accounting and Economics **15**, pp. 265–302.

[49] Lee, Charles M. C. and Radhakrishna, Balkrishna, 2000, Inferring investor behavior - evidence from torq data, Journal of Financial Markets **3**, p. 29.

[50] Lee, Charles M. C. and Ready, Mark J., 1991, Inferring trade direction from intraday data, The Journal of Finance **46**, pp. 733–746.

[51] Ljungqvist, Alexander and Qian, Wenlan, 2016, How constraining are limits to arbitrage?, The Review of Financial Studies **29**, pp. 1975–2028.

[52] Loughram, Tim and Ritter, Jay R., 1995, The new issues puzzle, The Journal of Finance **50**, pp. 23–51.

[53] Malmendier, U. and Shanthikumar, D., 2007, Are small investors naive about incentives?, Journal of Financial Economics **85**, pp. 457–489.

[54] Nelson, C. R. and Siegel, A. F., 1987, Parsimonious modeling of yield curves, Journal of Business **60**, pp. 473–489.

[55] Newey, Whitney K. and West, Kenneth D., 1987, A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance-matrix, Econometrica **55**, pp. 703–708.

[56] Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, Journal of Financial Economics **108**, pp. 1–28.

[57] Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research **18**, pp. 109–131.

[58] Ritter, Jay R., 1991, The long-run performance of initial public offerings, The Journal of Finance **46**, pp. 3–27.

[59] Rösch, Dominik M., Subrahmanyam, Avanidhar, and Dijk, Mathijs A. van, 2017, The dynamics of market efficiency, The Review of Financial Studies **30**, pp. 1151–1187.

[60] Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, The Accounting Review **71**, pp. 289–315.

[61] Stambaugh, Robert F., Yu, Jianfeng, and Yuan, Yu, 2012, The short of it: Investor sentiment and anomalies, Journal of Financial Economics **104**, pp. 288–302.

[62] Sun, Z., Wang, A., and Zheng, L., 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, Review of Financial Studies **25**, pp. 96–143.

[63] Titman, Sheridan, Wei, K. C. John, and Xie, Feixue, 2004, Capital investments and stock returns, Journal of Financial and Quantitative Analysis **39**, pp. 677–700.

[64] Weller, Brian M., 2018, Measuring tail risks at high frequency, The Review of Financial Studies **32**, pp. 3571–3616.

[65] Zhu, Christina, 2019, Big data as a governance mechanism, The Review of Financial Studies **32**, pp. 2021–2061.

Figure 1. This figure shows the distribution of trading activity in the Abel Noser institutional trade database across trade size bins. We report results for the number of trades, trade volume, buy volume, and sell volume within each trade size category as a percentage of total activity of all size bins. The trade size bins have lower limit points of \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million.

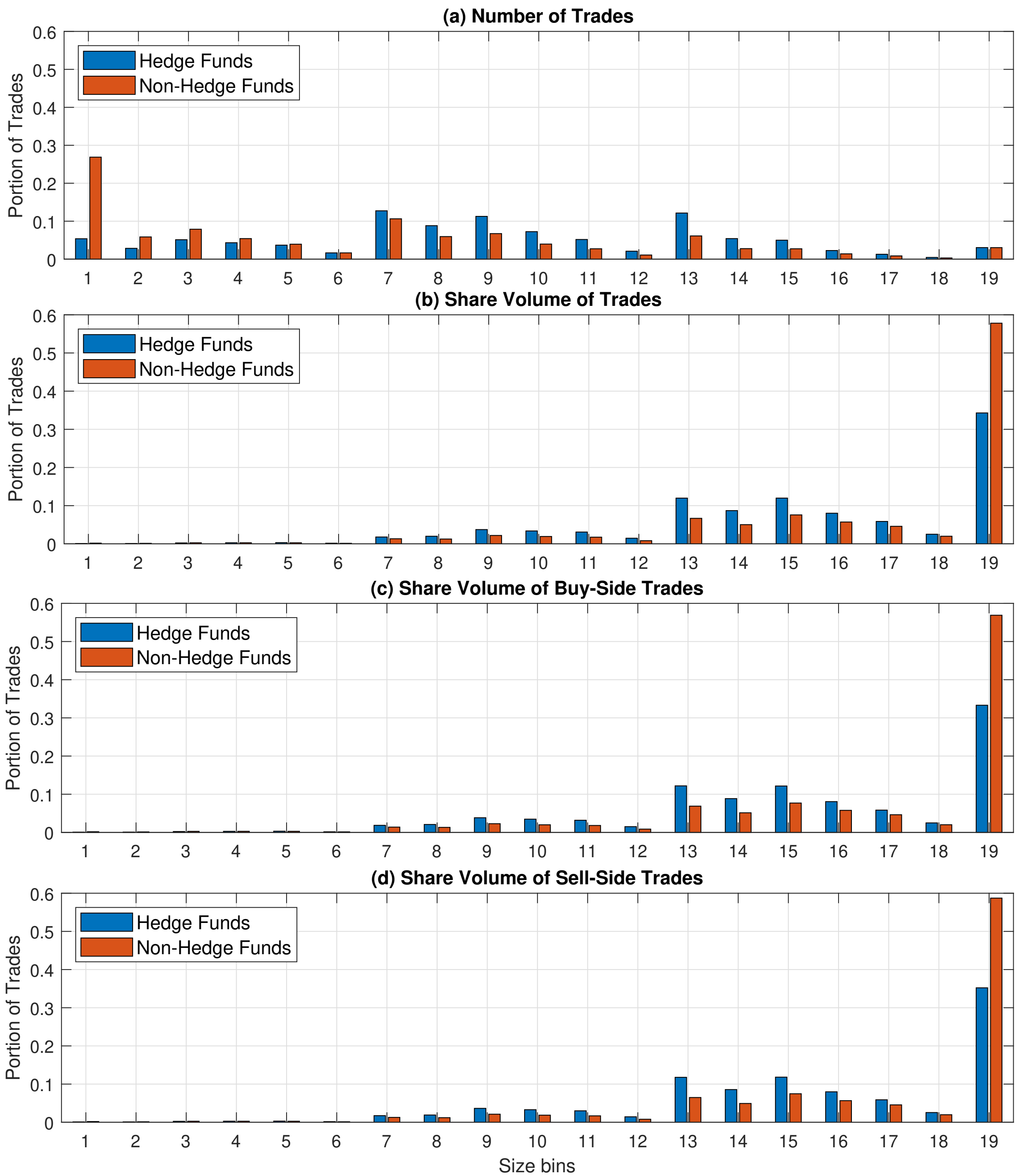


Figure 2. This figure shows the variation in coefficients reflecting the sensitivities of the publicly observable Lee and Ready (1991) order imbalance to hedge fund or non-hedge fund ownership across trade size bins and stock sizes, based on the results in Table A1 of the Internet Appendix. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients and dividing by the cross-sectional standard deviation of bin coefficients. The trade size bins have lower limit points of \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq that have information in TAQ, CRSP, and Thomson Reuters's 13F data.

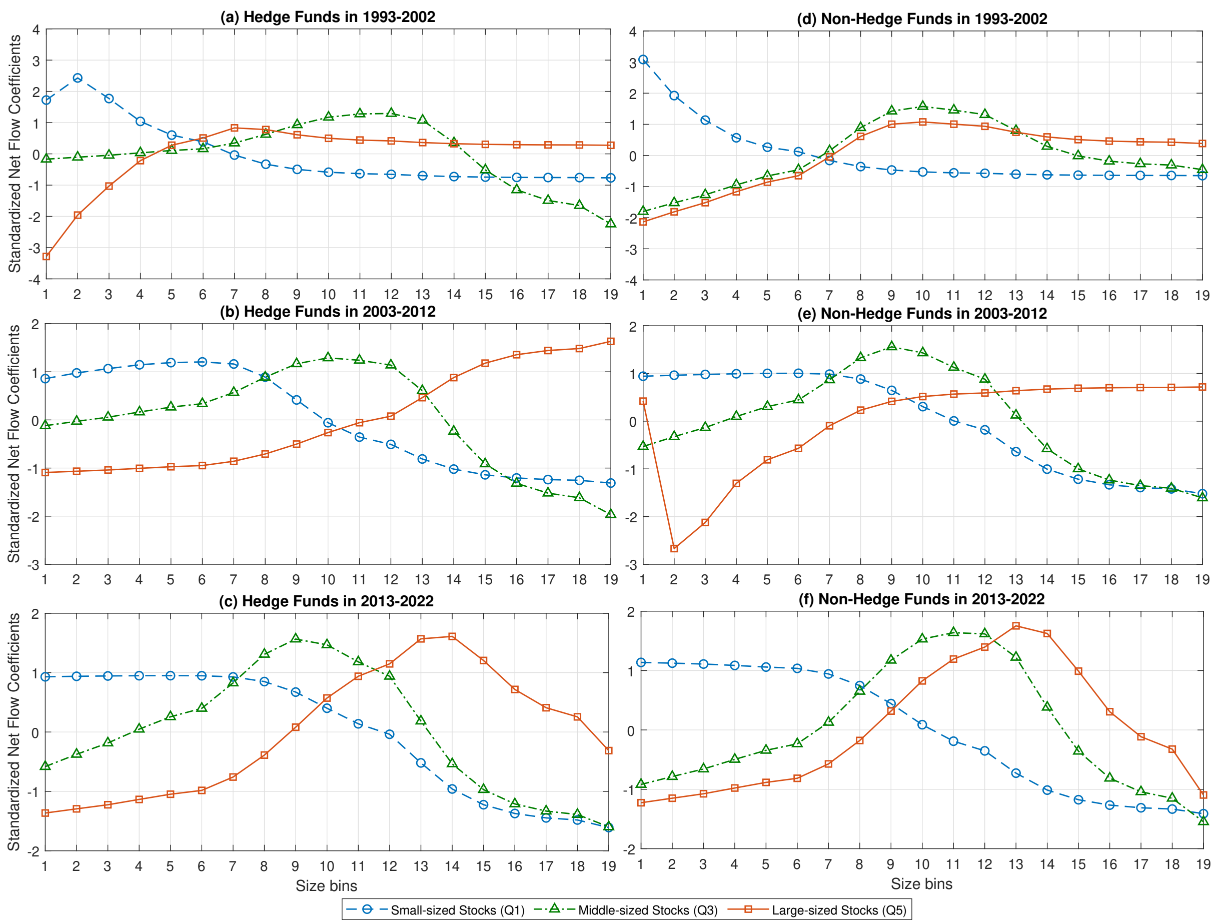


Table I. Summary Statistics

This table shows the time-series averages of cross-sectional statistics. HF and NHF are daily hedge fund and non-hedge fund order flow, respectively, estimated following Campbell, Ramadorai, and Schwartz (2009): taking the estimated coefficients in Table A1 of the Internet Appendix, HF and NHF are calculated as the expected change of daily hedge fund and non-hedge fund ownership, $E[\Delta Y_{i,d}]$, respectively.

$$\Delta Y_{i,d} = \beta^U U_{i,d} + \beta^{UY} Y_{i,q-1} \times U_{i,d} + \sum_{Z=1}^{19} \beta(Z, Y_{i,q-1}) F_{Z,i,d} + \epsilon_{i,d},$$

where for a stock i in a day d in a quarter q, ΔY_d is the change of daily hedge fund or non-hedge fund ownership, Y_q is aggregate hedge fund or non-hedge fund ownership in 13F, F_Z is aggregate Lee and Ready (1991) order imbalance scaled by shares outstanding in a trade-size bin Z, $\beta(Z, Y_{i,q-1})$ is defined as in Table A1 of the Internet Appendix. TOF is daily total order flow in TAQ based on Lee and Ready (1991) algorithm. Return is daily risk-adjusted mid-quote stock return with respect to the Fama−French−Carhart four factors. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data. Panels A, B, C and D report statistics in 1993-2022, 1993-2002, 2003-2012, and 2013-2022, respectively.

						Spearman Rank Correlation			
	Mean	Stdev	Min	Median	Max	HF	NHF	TOF	Return
Panel A. Full sample period (1993-2022)									
HF	0.009	0.025	-0.063	0.003	0.134	1.000			
NHF	0.024	0.072	-0.159	0.006	0.353	0.671	1.000		
TOF	0.036	0.170	-0.641	0.016	0.818	0.177	0.320	1.000	
Return	0.000	0.028	-0.251	-0.001	0.422	0.070	0.097	0.192	1.000
Panel B. 1993-2002									
HF	0.005	0.015	-0.026	0.001	0.077	1.000			
NHF	0.023	0.063	-0.079	0.002	0.315	0.547	1.000		
TOF	0.022	0.195	-0.722	0.003	0.842	0.183	0.277	1.000	
Return	0.000	0.031	-0.290	-0.001	0.421	0.073	0.101	0.300	1.000
Panel C. 2003-2012									
HF	0.012	0.025	-0.041	0.005	0.132	1.000			
NHF	0.032	0.079	-0.163	0.011	0.364	0.668	1.000		
TOF	0.029	0.160	-0.585	0.015	0.679	0.152	0.489	1.000	
Return	0.000	0.024	-0.215	-0.001	0.336	0.056	0.110	0.115	1.000
Panel D. 2013-2022									
HF	0.010	0.036	-0.123	0.003	0.193	1.000			
NHF	0.018	0.074	-0.235	0.004	0.381	0.798	1.000		
TOF	0.056	0.156	-0.616	0.029	0.934	0.196	0.194	1.000	
Return	0.000	0.027	-0.248	0.000	0.510	0.081	0.081	0.161	1.000

Table II. Firm Characteristics and Institutional Trading

This table presents time-series averages of coefficient estimates from cross-sectional regressions of the following equation,

$$
\begin{aligned}
\text{OF}_{i,m} \;=\;& \alpha_m + \beta_m^B \text{MktBeta}_{i,m} + \beta_m^C \text{MktCap}_{i,m} + \beta_m^{MB} \text{MB}_{i,m} + \beta_m^R \text{REV}_{i,m} + \beta_m^M \text{MOM}_{i,m} \\
&+ \beta_m^S \text{SPRD}_{i,m} + \beta_m^T \text{TURN}_{i,m} + \beta_m^I \text{IdioRisk}_{i,m} + \beta_m^{MISP} \text{MISP excl. MOM}_{i,m} + \epsilon_{i,m},
\end{aligned}
$$

where for stock i in a month m, OF is a hedge fund or non-hedge fund order flow estimate described in Section II.C, MktBeta is market beta from a Fama–French three-factor model over the past three years; MB is a market-to-book ratio; REV is short-term reversal; MOM is intermediate-term momentum; SPRD is relative bid-ask spread; TURN is a share volume turnover ratio; IdioRisk is idiosyncratic risk from a Fama–French three-factor model over the past three years; and MISP excl. MOM is a mispricing index proposed by Stambaugh, Yu, and Yuan (2012) that excludes a momentum factor. The definitions of these determinants are detailed in the Appendix. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, Compustat, and Thomson Reuters's 13F data. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

	HF_m	NHF_m
Intercept	-0.044	-1.080***
	(-1.37)	(-7.27)
MktBeta_m	0.020***	0.048***
	(7.81)	(9.18)
MktCap_m	0.006***	0.094***
	(2.71)	(8.01)
MB_m	-0.067***	-0.007
	(-2.60)	(-0.09)
REV_m	-0.062***	-0.063***
	(-7.03)	(-3.30)
MOM_m	-0.036***	-0.051***
	(-6.39)	(-4.74)
SPRD_m	-3.128***	-6.265***
	(-5.11)	(-4.61)
TURN_m	0.079***	0.250***
	(22.11)	(11.43)
IdioRisk_m	0.180***	-0.080
	(5.14)	(-1.29)
MISP excl. MOM_m	0.016**	-0.017
	(2.29)	(-0.73)
Adjusted R^2	0.342	0.305
Number of Stocks	2,455	2,455
Number of Months	360	360

Table III. Predicting CAR around Corporate Events

This table presents ordinary least squares regression results for the following equation,

$$\mathrm{CAR}_{i,t-1,t+1} = \alpha_i + \alpha_y + \sum_{k=2}^{6} \beta_k^{HF}\mathrm{HF}_{i,t-k} + \sum_{k=2}^{6} \beta_k^{NHF}\mathrm{NHF}_{i,t-k} + \sum_{k=2}^{6} \gamma_k^{T}\mathrm{TOF}_{i,t-k} + \sum_{k=2}^{6} \gamma_k^{R}\mathrm{Return}_{i,t-k} + \gamma^{S}\mathrm{SPRD}_{i,t-2} + \gamma^{S}\mathrm{AMI}_{i,t-2} + \epsilon_{i,t},$$

where for each corporate event i announced on day t in a year y, CAR is cumulative abnormal return from day $t-1$ to $t+1$ and all the explanatory variables are the same as defined in Tables I and II with event subscription i instead of firm subscription. For brevity, we only report the coefficient estimates of the hedge/non-hedge fund order flow estimates with the full set of control variables in the regressions. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data from 1993 to 2022. The corporate events we study are quarterly earnings announcements (Earnings Announcement) from I/B/E/S, analysts rating updates (Analysts Rating Update) from I/B/E/S, and extreme price movement (Extreme Price Change) exceeding two standard deviations of daily returns and not followed by return reversal for at least ten days from CRSP. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on firm and year clustered standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

	Earnings Announcement			Analysts Rating Update			Extreme Price Change		
	1993-2022	1993-1999	2000-2022	1993-2022	1993-1999	2000-2022	1993-2022	1993-1999	2000-2022
HF_{d-2}	8.091***	11.351***	7.156***	1.870***	5.259***	1.360*	11.950***	28.451***	10.042***
	(5.16)	(2.71)	(4.28)	(2.73)	(2.62)	(1.94)	(5.82)	(4.40)	(4.70)
HF_{d-3}	2.789*	3.186*	2.840	0.879	4.888**	0.372	6.319***	17.895***	4.836***
	(1.94)	(1.66)	(1.63)	(1.30)	(2.21)	(0.53)	(4.97)	(4.89)	(4.09)
HF_{d-4}	5.589***	6.960*	5.588***	2.410***	2.839***	2.324***	7.818***	12.729***	7.420***
	(5.47)	(1.73)	(4.96)	(4.14)	(2.90)	(3.53)	(4.63)	(4.71)	(3.70)
HF_{d-5}	1.839	-1.307	2.454	2.985***	5.419***	2.649***	3.811**	14.207***	2.788
	(1.00)	(-0.44)	(1.12)	(3.62)	(3.14)	(3.16)	(2.28)	(4.98)	(1.55)
HF_{d-6}	1.791	-1.553	2.029	1.087	2.743	1.081	5.109***	15.376**	3.760***
	(1.40)	(-0.70)	(1.34)	(1.56)	(0.88)	(1.52)	(3.57)	(2.27)	(2.98)
NHF_{d-2}	-1.636***	-2.510***	-1.264***	-0.610**	-1.142**	-0.473	-3.079***	-4.598***	-2.825***
	(-4.36)	(-4.64)	(-2.93)	(-2.27)	(-2.54)	(-1.49)	(-4.69)	(-5.01)	(-3.45)
NHF_{d-3}	-1.079*	-0.936***	-1.131	-0.451*	-1.503***	-0.213	-1.915***	-4.321***	-1.399***
	(-1.83)	(-2.64)	(-1.45)	(-1.85)	(-3.50)	(-0.78)	(-3.89)	(-3.07)	(-3.00)
NHF_{d-4}	-1.404***	-1.416**	-1.474**	-0.928***	-1.585***	-0.749***	-2.056***	-2.843***	-1.937***
	(-3.01)	(-2.33)	(-2.45)	(-4.41)	(-2.98)	(-3.49)	(-4.21)	(-4.28)	(-3.15)
NHF_{d-5}	-0.829	-0.470	-0.937	-0.912***	-1.497	-0.778***	-1.645***	-3.173***	-1.379**
	(-1.30)	(-0.49)	(-1.17)	(-2.95)	(-1.41)	(-3.14)	(-3.03)	(-3.83)	(-2.14)
NHF_{d-6}	-0.391	-0.585	-0.239	-0.633***	-0.389	-0.751***	-1.978***	-3.543***	-1.649***
	(-0.72)	(-0.78)	(-0.35)	(-3.49)	(-0.95)	(-3.51)	(-4.89)	(-5.36)	(-3.46)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.006	0.013	0.004	0.004	0.004	0.004	0.008	0.011	0.006
Observation	310,841	92,092	218,749	530,651	144,384	386,267	539,764	157,236	382,528

Table IV. Predicting CAR around RavenPack News

This table reproduces Table III to examine the predictive ability of hedge fund and non-hedge fund order flow estimates for CAR around corporate news events that RavenPack designates as indicated on the top of each column. This table presents regression results in all RavenPack company news groups with covering more than 5,000 firms in 2000-2022. Credit news group represents three credit-related RavenPack news groups of bankruptcy, credit, and credit-ratings. News groups of exploration, indexes, industrial-accidents, regulatory, stock-picks are excluded. For brevity, we only report the coefficient estimates of hedge fund order flow (HF) and non-hedge fund order flow (NHF), while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on firm and year clustered standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

Panel A. Fundamental news

	Earnings	Analysts	M&A	Assets	Credit	Dividends	Equity Actions	Labor Issues	Products Services	Revenues	Partnership
HF_{d-2}	5.446***	1.939	8.182***	2.446	-1.148	4.894***	-5.122	3.932**	1.872	7.559***	-0.983
	(3.56)	(1.50)	(3.07)	(1.33)	(-0.63)	(3.07)	(-1.23)	(2.34)	(0.67)	(3.31)	(-0.34)
HF_{d-3}	3.950***	3.126*	3.275	2.020	0.724	0.101	1.264	2.522**	1.276	2.036	-2.288
	(3.87)	(1.86)	(1.10)	(0.80)	(0.30)	(0.04)	(0.35)	(2.19)	(0.55)	(1.23)	(-1.00)
HF_{d-4}	3.596***	2.170	6.708***	9.955***	3.450**	-0.751	0.820	1.711	4.546	5.086***	3.545
	(3.19)	(1.31)	(3.76)	(3.36)	(2.13)	(-0.22)	(0.21)	(0.86)	(1.36)	(3.28)	(1.10)
HF_{d-5}	1.242	5.192***	10.711***	1.141	0.941	3.318**	12.453**	1.418	-1.447	1.932	4.203*
	(1.02)	(4.06)	(3.08)	(0.32)	(0.59)	(2.31)	(2.46)	(0.65)	(-0.76)	(1.02)	(1.72)
HF_{d-6}	-1.275	4.822***	0.756	-3.284	1.721*	2.552	1.566	1.065	3.508*	-0.374	-0.348
	(-0.76)	(3.14)	(0.30)	(-0.91)	(1.85)	(1.23)	(0.42)	(0.92)	(1.90)	(-0.20)	(-0.12)
NHF_{d-2}	-1.597***	-0.815***	-1.539	-0.318	0.752	-0.769	0.016	-0.948	0.395	-1.334**	0.673
	(-2.89)	(-3.14)	(-1.36)	(-0.32)	(1.33)	(-1.60)	(0.02)	(-1.19)	(0.49)	(-2.20)	(1.06)
NHF_{d-3}	-0.593	-0.275	-0.901	-2.035*	-0.476	-0.136	-0.462	-0.461	0.053	-0.693	0.436
	(-1.60)	(-0.58)	(-1.15)	(-1.71)	(-0.60)	(-0.21)	(-0.36)	(-0.65)	(0.10)	(-1.20)	(0.69)
NHF_{d-4}	-1.121**	-0.579	-0.228	-1.925**	-0.109	-0.149	0.767	-0.877*	0.150	-1.111*	-0.424
	(-2.00)	(-1.24)	(-0.35)	(-2.38)	(-0.15)	(-0.16)	(1.08)	(-1.83)	(0.19)	(-1.93)	(-0.56)
NHF_{d-5}	-0.220	-0.795**	-2.169**	0.472	0.056	0.242	-2.800**	-0.420	-0.191	-1.066*	-1.478*
	(-0.53)	(-2.08)	(-2.31)	(0.38)	(0.07)	(0.53)	(-2.22)	(-0.57)	(-0.40)	(-1.86)	(-1.90)
NHF_{d-6}	0.081	-1.734***	-0.276	1.899*	-0.946**	-1.246*	-2.347**	-0.019	-1.044	-0.360	-0.009
	(0.19)	(-3.96)	(-0.25)	(1.74)	(-2.31)	(-1.76)	(-2.26)	(-0.04)	(-1.42)	(-0.57)	(-0.01)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.004	0.004	0.016	0.009	0.006	0.003	0.030	0.004	0.005	0.002	0.011
Observation	273,626	172,905	73,523	25,189	78,067	80,236	92,084	118,375	114,744	193,390	33,841

(Continued)

Table IV – *Continued*

Panel B. Non-fundamental news

	Insider	Investor Relation	Marketing	Order Imbalance	Price Target	Stock Price
HF$_{d-2}$	1.747**	1.951**	1.230	2.891	1.292	3.342
	(2.11)	(2.43)	(0.98)	(1.25)	(0.48)	(0.93)
HF$_{d-3}$	1.598**	0.293	0.557	-2.712	1.555	4.511*
	(2.34)	(0.26)	(0.37)	(-0.95)	(1.27)	(1.74)
HF$_{d-4}$	1.096***	2.630**	1.822	2.135	0.996	-0.708
	(3.23)	(2.15)	(1.45)	(1.06)	(1.02)	(-0.33)
HF$_{d-5}$	0.570	1.400	-0.884	3.363*	2.094	6.023*
	(1.59)	(0.87)	(-0.65)	(1.67)	(0.97)	(1.69)
HF$_{d-6}$	0.119	0.412	3.512***	0.125	1.240	0.441
	(0.21)	(0.55)	(4.00)	(0.06)	(0.72)	(0.11)
NHF$_{d-2}$	-0.175	-0.273	-0.312	-0.978	-1.321	-1.831
	(-0.82)	(-0.74)	(-0.59)	(-1.54)	(-1.03)	(-0.99)
NHF$_{d-3}$	-0.590***	0.001	-0.427	-0.037	0.206	-2.235
	(-3.33)	(0.00)	(-1.07)	(-0.09)	(0.18)	(-1.43)
NHF$_{d-4}$	-0.172	-0.997***	-1.122***	0.672	-0.087	0.767
	(-1.09)	(-2.60)	(-2.91)	(1.23)	(-0.10)	(0.95)
NHF$_{d-5}$	-0.321**	-0.400	0.736	-0.365	-1.146	-2.189*
	(-2.45)	(-1.06)	(1.34)	(-0.85)	(-1.14)	(-1.94)
NHF$_{d-6}$	-0.167	-0.162	-1.265***	-0.118	-0.075	-1.816**
	(-0.72)	(-0.50)	(-4.27)	(-0.26)	(-0.09)	(-2.20)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.004	0.001	0.002	0.002	0.008	0.017
Observation	543,153	251,852	99,373	69,153	68,378	100,439

Table V. Heterogeneous Effects Conditioning on RS Metrics Coverage Initiation

This table presents ordinary least squares regression results for the following equation,

$$
\begin{aligned}
\text{Return}_{i,d} = {} & \alpha_i + \alpha_y + \beta_1 \text{TREAT} \times \text{POST} \times \text{HF}_{d-1} + \beta_2 \text{TREAT} \times \text{POST} \times \text{NHF}_{d-1} \\
& + \beta_3 \text{TREAT} \times \text{POST} \times \text{TOF}_{d-1} + \beta_4 \text{TREAT} \times \text{HF}_{d-1} + \beta_5 \text{TREAT} \times \text{NHF}_{d-1} \\
& + \beta_6 \text{TREAT} \times \text{TOF}_{d-1} + \beta_7 \text{POST} \times \text{HF}_{d-1} + \beta_8 \text{POST} \times \text{NHF}_{d-1} + \beta_9 \text{POST} \times \text{IOF}_{d-1} \\
& + \beta_{10} \text{TREAT} \times \text{POST} + \beta_{11} \text{TREAT} + \beta_{12} \text{POST} + \beta_{13} \text{HF}_{d-1} + \beta_{14} \text{NHF}_{d-1} + \beta_{15} \text{IOF}_{d-1} \\
& + \beta_{16} \text{MktCap}_{d-1} + \beta_{17} \text{Return}_{d-1} + \beta_{18} \text{AMI}_{d-1} + \beta_{19} \text{SPRD}_{d-1} + \epsilon_{i,d},
\end{aligned}
$$

where for stock i on day d, TREAT is a dummy equal to one if the stock is covered by RS Metrics and POST is a dummy equal to one after RS Metrics initiates coverage of the stock. To address concerns on stocks covered by RS Metrics are different from other stocks, we match 48 treated stocks with three stocks that do not experience coverage by RS Metrics. We select the three stocks with the closet market capitalization in the same industry (first digit of SICCD=5) to that of the corresponding treated stock. Sample period is from three years before to three years after RS Metrics coverage initiation for each treated stock. All variables are the same as defined in Tables I and IX. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on firm and year clustered standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

	(1)	(2)
TREAT × POST × HF_{d-1}	0.319***	0.305***
	(2.95)	(3.50)
TREAT × POST × NHF_{d-1}	-0.532	-0.573
	(-1.45)	(-1.50)
TREAT × POST × TOF_{d-1}	-0.151	-0.188
	(-0.69)	(-0.84)
TREAT × HF_{d-1}	-0.027	-0.062
	(-0.05)	(-0.13)
TREAT × NHF_{d-1}	0.249	0.295
	(1.09)	(1.19)
TREAT × TOF_{d-1}	0.019	0.029
	(0.18)	(0.28)
POST × HF_{d-1}	-0.540	-0.576
	(-0.71)	(-0.75)
POST × NHF_{d-1}	0.396	0.450
	(1.14)	(1.28)
POST × TOF_{d-1}	0.115	0.135
	(0.76)	(0.84)
TREAT × POST	-0.004	-0.003
	(-0.90)	(-0.72)
TREAT	0.003	0.003
	(1.44)	(1.17)
POST	-0.003**	-0.002
	(-2.00)	(-1.39)
HF_{d-1}	0.459	0.520
	(0.69)	(0.78)
NHF_{d-1}	-0.183	-0.228
	(-0.68)	(-0.83)
IOF_{d-1}	0.013	0.007
	(0.21)	(0.10)
MktCap_{d-1}		-0.033***
		(-3.61)
Return_{d-1}		-1.602***
		(-2.97)
AMI_{d-1}		19.025***
		(3.32)
SPRD_{d-1}		0.728
		(0.03)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Adjusted R^2	0.004	0.007
Observation	186,956	186,879

Table VI. Return Predictability

This table presents time-series averages of coefficient estimates from cross-sectional regressions of the following equation,

$$\begin{aligned}\text{Return}_{i,d} &= \alpha_d + \sum_{k=1}^{5} \beta_{d,k}^{HF}\text{HF}_{i,d-k} + \sum_{k=1}^{5} \beta_{d,k}^{NHF}\text{NHF}_{i,d-k} \\ &+ \sum_{k=1}^{5} \gamma_{d,k}^{T}\text{TOF}_{i,d-k} + \sum_{k=1}^{5} \gamma_{d,k}^{R}\text{Return}_{i,d-k} + \gamma_d^B\text{SPRD}_{i,d-1} + \gamma_d^A\text{AMI}_{i,d-1} + \epsilon_{i,d}.\end{aligned}$$

For brevity, we only report the coefficient estimates of hedge fund order flow (HF) and non-hedge fund order flow (NHF), while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data from 1993 to 2022. All variables are the same as defined in Tables I and II. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

	1993-2022	1993-2002	2003-2012	2013-2022
HF_{d-1}	1.540***	2.883***	1.298***	0.443***
	(17.16)	(15.27)	(11.86)	(4.84)
HF_{d-2}	0.047	0.082	0.196**	-0.137*
	(0.75)	(0.60)	(2.01)	(-1.67)
HF_{d-3}	0.144**	0.378***	0.078	-0.024
	(2.28)	(2.69)	(0.83)	(-0.29)
HF_{d-4}	0.108*	0.289**	0.125	-0.091
	(1.70)	(2.12)	(1.22)	(-1.13)
HF_{d-5}	0.155**	0.126	0.233**	0.106
	(2.44)	(0.90)	(2.52)	(1.26)
NHF_{d-1}	-0.283***	-0.572***	-0.202***	-0.077**
	(-12.23)	(-12.56)	(-7.65)	(-2.11)
NHF_{d-2}	-0.043**	-0.048	-0.124***	0.044
	(-2.17)	(-1.34)	(-4.50)	(1.20)
NHF_{d-3}	-0.041**	-0.037	-0.078***	-0.007
	(-2.07)	(-1.02)	(-2.79)	(-0.19)
NHF_{d-4}	-0.046**	-0.082**	-0.066***	0.009
	(-2.47)	(-2.37)	(-2.62)	(0.26)
NHF_{d-5}	-0.088***	-0.078**	-0.097***	-0.090**
	(-4.62)	(-2.16)	(-3.94)	(-2.48)
Adjusted R^2	0.031	0.022	0.028	0.042
Number of Stocks	2,973.1	3,704.3	2,600.6	2,616.9
Number of Days	7,544.0	2,509.0	2,517.0	2,518.0

Table VII. Heterogeneous Effects

This table examines the predictive ability of hedge fund and non-hedge fund order flow estimates in subsamples based on firm characteristics. We separate our sample into two groups using the daily cross-sectional median of a proxy for liquidity and information environment. Then, within each subsample, we replicate the regression anaysis in Table VI. The liquidity proxies are market capitalization (Size), relative bid-ask spread (Spread), and Amihud illiquidity (Amihud). The information proxy is analyst coverage (Analyst) and institutional ownership (Ownership). For brevity, we only report the coefficient estimates of hedge fund order flow (HF) and non-hedge fund order flow (NHF), while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data from 1993 to 2022. All variables are the same as defined in Table I. Panal A is for liquidity proxies and Panel B is for information proxies. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

Panel A. Liquidity	Size		Spread		Amihud	
	Small	Large	Narrow	Wide	Liquid	Illiquid
HF_{d-1}	3.066***	1.176***	1.116***	2.324***	0.761***	4.492***
	(17.64)	(12.52)	(12.09)	(15.80)	(8.78)	(19.78)
HF_{d-2}	0.046	0.238***	0.104	-0.051	-0.028	0.471***
	(0.37)	(3.49)	(1.40)	(-0.46)	(-0.42)	(3.23)
HF_{d-3}	0.308**	0.139**	0.176**	0.050	-0.009	0.778***
	(2.28)	(1.97)	(2.50)	(0.46)	(-0.14)	(5.42)
HF_{d-4}	0.081	0.205***	0.089	0.084	0.043	0.404***
	(0.68)	(2.73)	(1.28)	(0.75)	(0.61)	(2.95)
HF_{d-5}	0.243**	0.223***	0.113	0.292***	-0.003	0.606***
	(2.01)	(3.04)	(1.56)	(2.74)	(-0.04)	(4.23)
NHF_{d-1}	-0.490***	-0.179***	-0.192***	-0.416***	-0.134***	-0.368***
	(-8.69)	(-7.42)	(-7.99)	(-9.92)	(-5.75)	(-5.82)
NHF_{d-2}	-0.202***	-0.041**	-0.031	-0.076**	0.019	-0.090*
	(-4.23)	(-2.01)	(-1.43)	(-2.03)	(0.96)	(-1.68)
NHF_{d-3}	-0.280***	-0.019	-0.027	-0.075**	0.027	-0.164***
	(-5.78)	(-0.98)	(-1.31)	(-1.99)	(1.41)	(-3.09)
NHF_{d-4}	-0.205***	-0.038*	-0.004	-0.148***	0.001	-0.043
	(-4.64)	(-1.95)	(-0.23)	(-4.03)	(0.07)	(-0.85)
NHF_{d-5}	-0.361***	-0.030	-0.047**	-0.193***	0.006	-0.194***
	(-8.08)	(-1.51)	(-2.26)	(-5.35)	(0.29)	(-3.84)
Adjusted R^2	0.034	0.046	0.042	0.035	0.046	0.034
Number of Stocks	1,482.9	1,490.2	1,494.8	1,478.3	1,490.3	1,482.7
Number of Days	7,544.0	7,544.0	7,544.0	7,544.0	7,544.0	7,544.0

(Continued)

Table VII – *Continued*

Panel B. Information environment

	Analyst Coverage		Ownership	
	Low	High	Low	High
HF_{d-1}	2.389***	0.758***	2.475***	1.070***
	(16.61)	(8.80)	(15.15)	(13.14)
HF_{d-2}	0.052	0.049	-0.126	0.180***
	(0.46)	(0.64)	(-1.04)	(2.70)
HF_{d-3}	0.112	0.139*	0.297**	0.097
	(1.02)	(1.77)	(2.50)	(1.42)
HF_{d-4}	0.346***	0.017	0.049	0.141**
	(3.26)	(0.22)	(0.42)	(2.08)
HF_{d-5}	0.266**	0.111	0.134	0.178**
	(2.51)	(1.42)	(1.17)	(2.57)
NHF_{d-1}	-0.365***	-0.108***	-0.345***	-0.186***
	(-8.96)	(-4.48)	(-7.88)	(-8.06)
NHF_{d-2}	-0.057	0.017	-0.068*	-0.049**
	(-1.41)	(0.75)	(-1.76)	(-2.32)
NHF_{d-3}	-0.068*	-0.006	-0.136***	0.001
	(-1.83)	(-0.28)	(-3.33)	(0.05)
NHF_{d-4}	-0.083**	-0.006	-0.069*	-0.032
	(-2.31)	(-0.30)	(-1.95)	(-1.63)
NHF_{d-5}	-0.154***	-0.017	-0.164***	-0.051***
	(-4.17)	(-0.76)	(-4.30)	(-2.63)
Adjusted R^2	0.033	0.045	0.042	0.035
Number of Stocks	1,282.9	1,275.2	1,423.5	1,427.2
Number of Days	7,544.0	7,544.0	7,544.0	7,544.0

Table VIII. Price Efficiency around Earnings Announcement Dates

This table presents results from a regression of price efficiency proxies around earnings announcements on hedge fund and non-hedge fund order flow estimates:

$$y_{i,d} = \alpha_i + \alpha_q + \beta^{HF}\mathrm{HF}_{i,d} + \beta^{NHF}\mathrm{NHF}_{i,d} + \gamma\mathrm{ControlVariables}_{i,d} + \epsilon_{i,d},$$

where for stock i on earnings announcement day d in quarter q, the dependent variables are $\mathrm{CAR}_{d+1,d+61}$, Jump Ratio Rank, and ΔVR in columns 1 to 3. We use Compustat, I/B/E/S, and Raven Pack to identify earnings announcement dates for all common stocks on NYSE, AMEX, and Nasdaq with market capitalization above the 5% NYSE breakpoints from 1993 to 2022. $\mathrm{CAR}_{d+1,d+61}$ is the cumulative abnormal return compounded over the 60-day post-announcement period over days (1, 61) following the earnings announcement. Jump Ratio is the ratio of cumulative abnormal return on the earnings announcement (AR_d) divided by cumulative abnormal return over days (-21, 0) relative to earnings announcement ($\mathrm{CAR}_{d-21,d}$). To reduce the influence of extreme values, we rank Jump Ratio in cross section and define Jump Ratio Rank as a categorical value between 0 (for a stock with the lowest decile of Jump Ratio) and 9 (for a stock with the highest decile of Jump Ratio). ΔVR is the difference between variance ratio averaged over days +21 to +1 and variance ratio averaged over days -21 and -1. Variance ratio is the absolute value of the difference between the ratio of 15-to-60 second stock return variance and one. $|\mathrm{OF}|_{d-21,d-1}$ is the absolute value of order flow summed over days -21 to -1. Control variabales include insitutional ownership (InstOwn), the number of analysts covering the stock (NumAnalyst), market capitalization averaged over days $-p$ to $-q$ ($\mathrm{MktCap}_{d-p,d-q}$), relataive bid-ask spread averaged over days $-p$ to $-q$ ($\mathrm{SPRD}_{d-p,d-q}$), Amihud illiquidity averaged over days $-p$ to $-q$ ($\mathrm{AMI}_{d-p,d-q}$), cumulative abnormal return over days $-p$ to $-q$ ($\mathrm{CReturn}_{d-p,d-q}$), and standard deviation of abnormal returns over days $-p$ to $-q$ ($\mathrm{SReturn}_{d-p,d-q}$). All coefficient estimates in columns (1) and (3) are multiplied by 100. Corresponding t-statistics based on firm and quarter clustered standard errors are reported in parentheses. Superscripts ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

(1) y = $\mathrm{CAR}_{d+1,d+61}$		(2) y = Jump Ratio Rank		(3) y = ΔVR	
HF_d	-8.069**	$\lvert\mathrm{HF}\rvert_{d-21,d-1}$	-1.038***	$\lvert\mathrm{HF}\rvert_d$	-2.628**
	(-2.19)		(-10.10)		(-2.41)
NHF_d	3.168**	$\lvert\mathrm{NHF}\rvert_{d-21,d-1}$	-0.653***	$\lvert\mathrm{NHF}\rvert_d$	-0.566
	(2.02)		(-13.35)		(-1.17)
TOF_d	-0.511	$\lvert\mathrm{TOF}\rvert_{d-21,d-1}$	1.306***	$\lvert\mathrm{TOF}\rvert_d$	-0.423**
	(-0.77)		(34.20)		(-2.55)
				$\mathrm{VR}_{d-42,d-22}$	-7.157***
					(-8.80)
$\mathrm{InstOwn}_{d-1}$	2.316***	$\mathrm{InstOwn}_{d-22}$	2.744***	$\mathrm{InstOwn}_{d-22}$	0.798***
	(4.19)		(27.17)		(4.16)
$\mathrm{NumAnalyst}_{d-1}$	-0.489**	$\mathrm{NumAnalyst}_{d-22}$	1.018***	$\mathrm{NumAnalyst}_{d-22}$	0.240***
	(-2.23)		(36.20)		(3.46)
$\mathrm{MktCap}_{d-21,d-1}$	-0.111	$\mathrm{MktCap}_{d-42,d-22}$	-0.093***	$\mathrm{MktCap}_{d-42,d-22}$	-0.091*
	(-0.52)		(-4.74)		(-1.69)
$\mathrm{SPRD}_{d-21,d-1}$	-11.725	$\mathrm{SPRD}_{d-42,d-22}$	38.831***	$\mathrm{SPRD}_{d-42,d-22}$	6.979
	(-0.42)		(21.43)		(1.03)
$\mathrm{AMI}_{d-21,d-1}$	0.597	$\mathrm{AMI}_{d-42,d-22}$	0.921***	$\mathrm{AMI}_{d-42,d-22}$	0.631**
	(0.63)		(7.85)		(2.01)
$\mathrm{CReturn}_{d-21,d-1}$	-2.957*	$\mathrm{CReturn}_{d-42,d-22}$	0.182**	$\mathrm{CReturn}_{d-42,d-22}$	0.217
	(-1.73)		(2.39)		(1.26)
$\mathrm{SReturn}_{d-21,d-1}$	30.600***	$\mathrm{SReturn}_{d-42,d-22}$	5.428***	$\mathrm{SReturn}_{d-42,d-22}$	9.718***
	(2.68)		(3.98)		(3.27)
Firm FE	Yes	Firm FE	Yes	Firm FE	Yes
Quarter FE	Yes	Quarter FE	Yes	Quarter FE	Yes
Adjusted R^2	0.002	Adjusted R^2	0.686	Adjusted R^2	0.024
Observation	148,846	Observation	148,135	Observation	146,020

Table IX. Price Efficiency Using Fama–MacBeth (1973) Regressions

This table presents time-series averages of coefficient estimates from cross-sectional regressions of the following equation,

$$\Delta \mathrm{VR}_{i,d} = \alpha_d + \beta_d^{HF}|\mathrm{HF}|_{i,d-1} + \beta_d^{NHF}|\mathrm{NHF}|_{i,d-1} + \gamma \mathrm{ControlVariables}_{i,d-3} + \epsilon_{i,d},$$

where for stock i on day d, ΔVR is a two-day change in variance ratio: $VR_{i,d} - VR_{i,d-2}$. Control variabales include insitutional ownership (InstOwn), the number of analysts covering the stock (NumAnalyst), market capitalization (MktCap), relataive bid-ask spread (SPRD), Amihud illiquidity (AMI), and Return is daily risk-adjusted mid-quote stock return with respect to Carhart (1997) four factors. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data from 1993 to 2022. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on Newey–West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

	Entire (1993-2022)	First (1993-2002)	Second (2003-2012)	Third (2013-2022)
$\lvert\mathrm{HF}\rvert_{d-1}$	-3.237***	-5.640***	-1.509***	-2.569***
	(-5.65)	(-3.55)	(-3.21)	(-6.00)
$\lvert\mathrm{NHF}\rvert_{d-1}$	-0.105	0.661**	-0.302	-0.672***
	(-0.73)	(2.00)	(-1.56)	(-3.45)
$\lvert\mathrm{TOF}\rvert_{d-1}$	-0.562***	0.039	-0.258***	-1.465***
	(-8.43)	(0.26)	(-2.96)	(-19.91)
VR_{d-3}	-2.547***	-1.976***	-2.357***	-3.306***
	(-42.46)	(-15.95)	(-28.56)	(-43.15)
$\mathrm{InstOwn}_{d-3}$	-0.008	0.073	-0.003	-0.094***
	(-0.32)	(1.09)	(-0.12)	(-2.79)
$\mathrm{NumAnalyst}_{d-3}$	-0.173***	-0.069***	-0.097***	-0.354***
	(-13.94)	(-3.07)	(-5.84)	(-16.55)
MktCap_{d-3}	-0.820***	-0.848	-0.536*	-1.075***
	(-3.35)	(-1.23)	(-1.79)	(-7.12)
SPRD_{d-3}	31.595***	-2.786	34.193***	63.233***
	(9.50)	(-1.29)	(5.04)	(10.39)
AMI_{d-3}	0.248	-5.779**	2.618***	3.883***
	(0.27)	(-2.30)	(3.76)	(5.45)
Return_{d-3}	1.709***	-0.820	1.807***	4.130***
	(5.55)	(-1.21)	(4.29)	(10.91)
Intercept	1.008***	0.488***	0.686***	1.846***
	(21.22)	(6.04)	(10.86)	(24.79)
Adjusted R^2	0.007	0.006	0.005	0.010
Number of Stocks	1,738.5	1,023.7	2,069.2	2,120.0
Number of Days	7,539.0	2,507.0	2,514.0	2,518.0

Internet Appendix

Informed Trading under the Microscope: Evidence from 30 Years of Daily Hedge Fund Trades

In the paper titled *"Informed Trading under the Microscope: Evidence from 30 Years of Daily Hedge Fund Trades,"* we develop a novel measure of daily aggregate hedge fund trades in individual U.S. stocks. Due to space constraints, we could not include all the empirical findings discussed throughout our research. This Internet Appendix complements the main paper by presenting three additional sets of results to provide readers with comprehensive insights. We are happy to provide further supporting evidence upon request.

This Internet Appendix is organized as follows. Section A explores the differences in trading behavior between hedge funds and non-hedge funds, utilizing the Ancerno institutional trade database. Also, it provides the complete set of estimated coefficients from the regression model developed by Campbell, Ramadorai, and Schwartz (2009). Section B evaluates the predictive power of hedge fund trades for stock returns over subsequent months.

A. Estimate daily hedge fund trades

A. Difference in hedge fund and non-hedge fund trading

Replicating Figure 1 from our main paper, we use the Abel Noser dataset to examine whether the distinctions between hedge fund (HF) and non-hedge fund (NHF) trading behaviors persist across three sub-periods: 1999–2003, 2004–2007, and 2008–2012. Hedge fund trades are identified in Abel Noser following the methodology of Jame (2018). We plot the distribution of aggregate HF and NHF trades across trade size bins, presenting the number of trades, trade volume, buy volume, and sell volume within each trade size category as percentages of total activity across all size bins. Following CRS, trade size bins have lower limit points of \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1,000,000.

As shown in Figure A1, while HF and NHF both conduct the majority of their trading volume in the largest trade size bins, HF consistently exhibit a stronger presence in medium-sized trades across all three sub-periods. This persistence is particularly evident in the share volume of trades, highlighting the stability of their trading strategies over time. By contrast, NHFs consistently rely predominantly on the largest orders (above \$1,000,000), which account for nearly 60% of their trading volume in each sub-period. For HFs, this same group of trades contributes to approximately 35% of their total volume, with little variation over time. Hedge funds also consistently allocate substantial trading activity (46%) to trade size bins 13 to 17, ranging from \$100,000 to \$700,000. In comparison, NHFs allocate only 30% of their total trading volume to these bins, a pattern that remains stable across sub-periods. This reliance on medium-sized trades is a distinguishing feature of HF trading behavior, suggesting a tactical approach to optimizing transaction costs, managing liquidity needs, and maintaining order anonymity. However, the persistence across sub-periods is not observed in the number of trades, buy volume, or sell volume distributions, further emphasizing that this stability is uniquely reflected in the share volume of trades.

When examining the number of trades, the distinctive patterns of HFs and NHFs exhibit less persistence compared to trade volume. In 1999–2003, both hedge funds and non-hedge funds place the majority of their orders within medium trade sizes, particularly in trade size bins 7 to 15. However, in the subsequent sub-periods (2004–2007 and 2008–2012), non-hedge funds shift their focus toward splitting orders, with the smallest trades (below \$2,000) accounting for the largest proportion of their activity. In contrast, hedge funds maintain a consistent distribution of trades across the three sub-periods, showing little variation in the number of trades. This divergence underscores a fundamental difference in trading strategies between the two types of institutions, with NHFs demonstrating a stronger tendency toward extensive order fragmentation over time.

[Place Figure A1 about here]

A.2. Estimate daily hedge fund and non-hedge fund trades

Similar to the CRS method, our estimation is based on the following equation:

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \sum_{Z=1}^{19} \beta(Z, v) F_{Z,i,q} + \epsilon_{i,q}, \tag{1}$$

where for a stock i in a quarter q, α is a set of four quarter dummies, Y is either aggregate hedge fund or non-hedge fund ownership (in separate estimations) from 13F, F_Z is aggregate order imbalance based on the Lee and Ready (1991) algorithm scaled by shares outstanding in a trade-size bin Z, and U is aggregate unclassified trades scaled by shares outstanding for which the Lee and Ready (1991) algorithm cannot determine the direction.[1] Hedge fund and non-hedge fund ownership is identified following the methodology of Agarwal, Jiang, Tang, and Yang (2013) in the Thomson Reuters 13F Ownership data. Following CRS, we assign trades into nineteen size bins whose lower limit points are \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. To smooth out the coefficient variation across transaction size and mitigate estimation errors in certain bins (e.g., very large trades for small stocks, which are rare), CRS apply a yield curve function from Nelson and Siegel (1987) to model the structure of β across trade-size bins:

$$\beta(Z, v) = b_{01} + b_{02}v + (b_{11} + b_{12}v + b_{21} + b_{22}v)[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22}v)e^{-Z/\tau}, \tag{2}$$

[1] We restrict TAQ observations to regular transactions between 9:30:00 to 15:59:59 EST. We exclude trades under the sale condition of the Opened Last ('O'), Sold Sale ('Z'), Bounced ('B'), Pre- and Post-Market Close Trades ('T'), Sold Last ('L'), Bunched Sold ('G'), Average Price Trades ('W'), Rule 127 Trade ('J'), and Rule 151 Trade ('K').

where τ is a constant to estimate and v is set to the lagged level of hedge fund or non-hedge fund ownership ($Y_{i,q-1}$) as in CRS. Following CRS, we use non-linear least squares to estimate the coefficients in Equation (1) for each firm size quintile based on NYSE breakpoints of market capitalization at the start of each quarter. Concerning that both types of institutional investors may change their trading styles in the relatively long sample period of 30 years, we estimate Equation (1) in three decade-long subperiods separately.

The estimated coefficients of the CRS model are reported in Table A1, shown separately for hedge funds and non-hedge funds. The estimated coefficients are highly significant across the board for both hedge funds and non-hedge funds encompassing all firm size quintiles and across all subperiods.

[Place Table A1 about here]

A.3. Contemporaneous price impact

We compare the contemporaneous price impact of hedge fund order flow (HF) versus non-hedge fund order flow (NHF) to evaluate their skills in trade execution. Table A2 presents the estimated coefficients of the following model:

$$\begin{aligned}\text{Return}_{i,d} \;=\;& \alpha_d + \beta_d^{HF}\text{HF}_{i,d} + \beta_d^{NHF}\text{NHF}_{i,d} \\ &+ \sum_{k=1}^{5}\gamma_{d,k}^{T}\text{TOF}_{i,d-k} + \sum_{k=1}^{5}\gamma_{d,k}^{R}\text{Return}_{i,d-k} + \gamma_d^{B}\text{SPRD}_{i,d-1} + \gamma_d^{A}\text{AMI}_{i,d-1} + \epsilon_{i,d},\end{aligned}$$

where for stock i in a day d, HF and NHF represent monthly aggregated order flow from hedge funds and non-hedge funds, respectively, as estimated in Section A. TOF is the monthly aggregated total order flow in TAQ, calculated using the Lee and Ready (1991) algorithm. Return is a mid-quote return with respect to Fama-French-Carhart four factors. SPRD is relative bid-ask spread and AMI is Amihud (2002) illiquidity measure. For brevity, we only report the estimated coefficients for HF and NHF in Table A2. The t-statistics in Table A2 are calculated by Newey−West (1987) standard errors with ten lags in the consideration of serial correlations.

During the full sample period (1993–2022), the estimated coefficients for hedge funds (HF) and non-hedge funds (NHF) are 3.131 (with a t-statistic of 10.46) and 2.795 (with a t-statistic of 33.02), respectively. This indicates that both hedge funds and non-hedge funds exert positive and statistically significant contemporaneous price pressure. Although HF's estimated coefficient is higher than NHF's, its economic significance is lower due to HF's smaller standard deviation. Specifically, a one standard-deviation increase in HF is associated with a 7.6 basis points (bp) increase in contemporaneous return, whereas NHF's price impact reaches 19.0 bp, underscoring a stronger economic influence.

We further examine the time-series dynamics of price impact across three subperiods, as shown in Table A2. The results indicate that HF's price impact is consistently lower than NHF's across all subperiods. In the 1993-2002 period, a one standard-deviation increase in HF is associated with an 11.1 bp increase in contemporaneous return, compared to NHF's 22.6 bp. From 2003-2012, HF's impact turns negative, while NHF's impact rises to 26.3 bp. In the most recent subperiod (2013-2022), a one standard-deviation increase in HF leads to a 6.8 bp increase, while NHF's impact moderates to 11.0 bp. Despite a general reduction in price impact over time, HF trading consistently generates less contemporaneous price pressure than NHF trading, suggesting a potentially more nuanced or less disruptive approach to market interactions by hedge funds.

[Place Table A2 about here]

B. Hedge funds as informed traders

A.1. Return predictability at monthly frequency

We construct a monthly sample of common stocks listed on the NYSE, AMEX, and Nasdaq that are available in the CRSP, Compustat, TAQ, and Thomson Reuters 13F Ownership databases from 1993 to 2022. To avoid market microstructure effects, we exclude stocks priced below $5. Additionally, we exclude stocks with month-end market capitalizations below the 10th percentile of NYSE breakpoints, focusing on more liquid and widely held stocks. For determinant variables requiring firm-level data from Compustat, we use annual financial statements, ensuring that the Compustat reporting date (item RDQ) precedes the end of the month. For variables based on stock data from CRSP, we rely on information recorded during the given month or earlier, as reported by CRSP. Our final monthly sample comprises 1,161,084 stock-month observations, merged with a trading dataset containing monthly aggregated hedge fund and non-hedge fund order flows.

Using the monthly sample, we examine the return predictive ability of HF and NHF with Fama and Macbeth (1973) regressions. Table A3 reports the estimated coefficients of the following model:

$$\begin{aligned}\text{Return}_{i,m} \;=\;& \alpha_m + \beta_m^{HF}\text{HF}_{i,m-1} + \beta_m^{NHF}\text{NHF}_{i,m-1} + \gamma_m^{T}\text{TOF}_{i,m-1} + \gamma_m^{R}\text{Return}_{i,m-1}\\ &+\gamma_m^{B}\text{SPRD}_{i,m-1} + \gamma_m^{A}\text{AMI}_{i,m-1} + \gamma_m^{M}\text{MISP}_{i,m-1} + \gamma_m^{I}\text{IdioRisk}_{i,m-1} + \epsilon_{i,m}.\end{aligned}$$

where for stock i in a month m, HF and NHF represent monthly aggregated order flow from hedge funds and non-hedge funds, respectively, as estimated in Section A. TOF is the monthly aggregated total order flow in TAQ, calculated using the Lee and Ready (1991) algorithm. Return is a mid-quote return with respect to Fama-French-Carhart four factors. SPRD is the relative spread at the end of the month, and AMI is the Amihud (2002) illiquidity measure over the month. IdioRisk refers to idiosyncratic risk from a Fama–French three-factor model over the past three years. MISP is a mispricing index proposed by Stambaugh, Yu, and Yuan (2012), while MISP excl. MOM is a mispricing index that excludes the momentum factor. To account for serial correlations, we use Newey and West (1987) standard errors with five lags to calculate the t-statistics.

In Table A3, hedge fund trades (HF) exhibit a positive and statistically significant coefficient of 0.525 with a t-statistic of 2.71 across two columns, regardless of the mispricing index used. This translates into an economic effect where a one standard-deviation increase in HF is associated with a 1.4 bp increase in the subsequent month's stock return. Conversely, non-hedge fund trades (NHF) display a negative and statistically significant coefficient of -0.251 (t-stat $= -4.16$), indicating that NHF trades tend to exert downward pressure on stock prices. These results suggest that hedge fund trades have a durable impact on stock prices even at a monthly frequency, while non-hedge fund trades primarily create temporary price distortions. This reinforces the view that hedge funds possess a distinct informational advantage over other institutional investors.

[Place Table A3 about here]

References

[1] Agarwal, Vikas, Jiang, Wei, Tang, Yuehua, and Yang, Baozhong, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, The Journal of Finance **68**, pp. 739–783.

[2] Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, Journal of Financial Markets **5**, pp. 31–56.

[3] Campbell, John, Ramadorai, Tarun, and Schwartz, Allie, 2009, Caught on tape, Journal of Financial Economics **92**, p. 26.

[4] Fama, Eugene F. and Macbeth, James D., 1973, Risk, return, and equilibrium: Empirical tests, Journal of Political Economy **81**, pp. 607–636.

[5] Jame, Russell, 2018, Liquidity provision and the cross-section of hedge fund returns, Management Science **64**, pp. 3288–3312.

[6] Lee, Charles M. C. and Ready, Mark J., 1991, Inferring trade direction from intraday data, The Journal of Finance **46**, pp. 733–746.

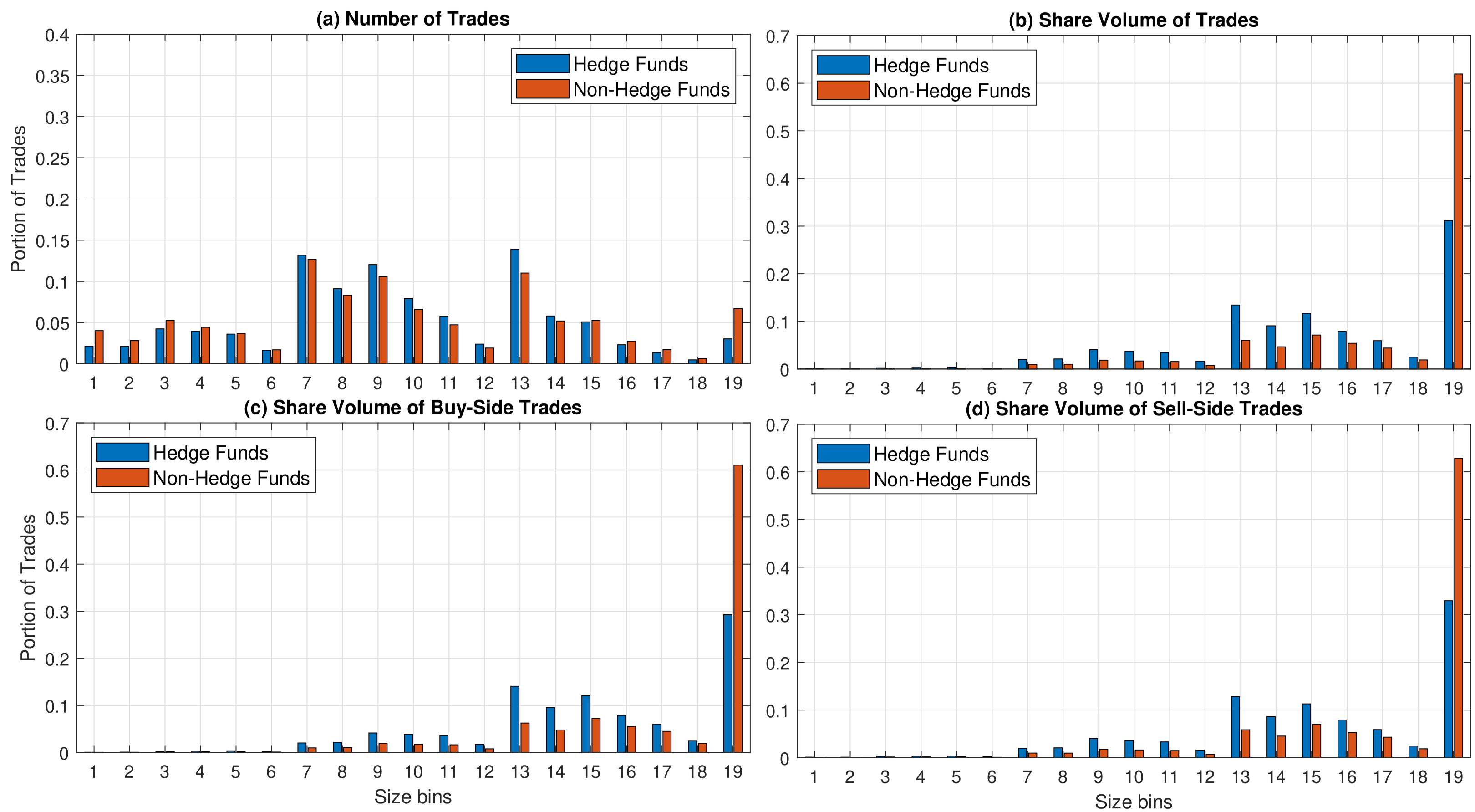
[7] Nelson, C. R. and Siegel, A. F., 1987, Parsimonious modeling of yield curves, Journal of Business **60**, pp. 473–489.

[8] Newey, Whitney K. and West, Kenneth D., 1987, A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance-matrix, Econometrica **55**, pp. 703–708.

[9] Stambaugh, Robert F., Yu, Jianfeng, and Yuan, Yu, 2012, The short of it: Investor sentiment and anomalies, Journal of Financial Economics **104**, pp. 288–302.

Figure A1. This figure shows the distribution of trading activity in the Abel Noser institutional trade database across trade size bins. We report results for the number of trades, trade volume, buy volume, and sell volume within each trade size category as a percentage of total activity of all size bins. The trade size bins have lower limit points of \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. Panels A, B, and C report the distributions in 1999-2003, 2004-2007, and 2008-2012, respectively.

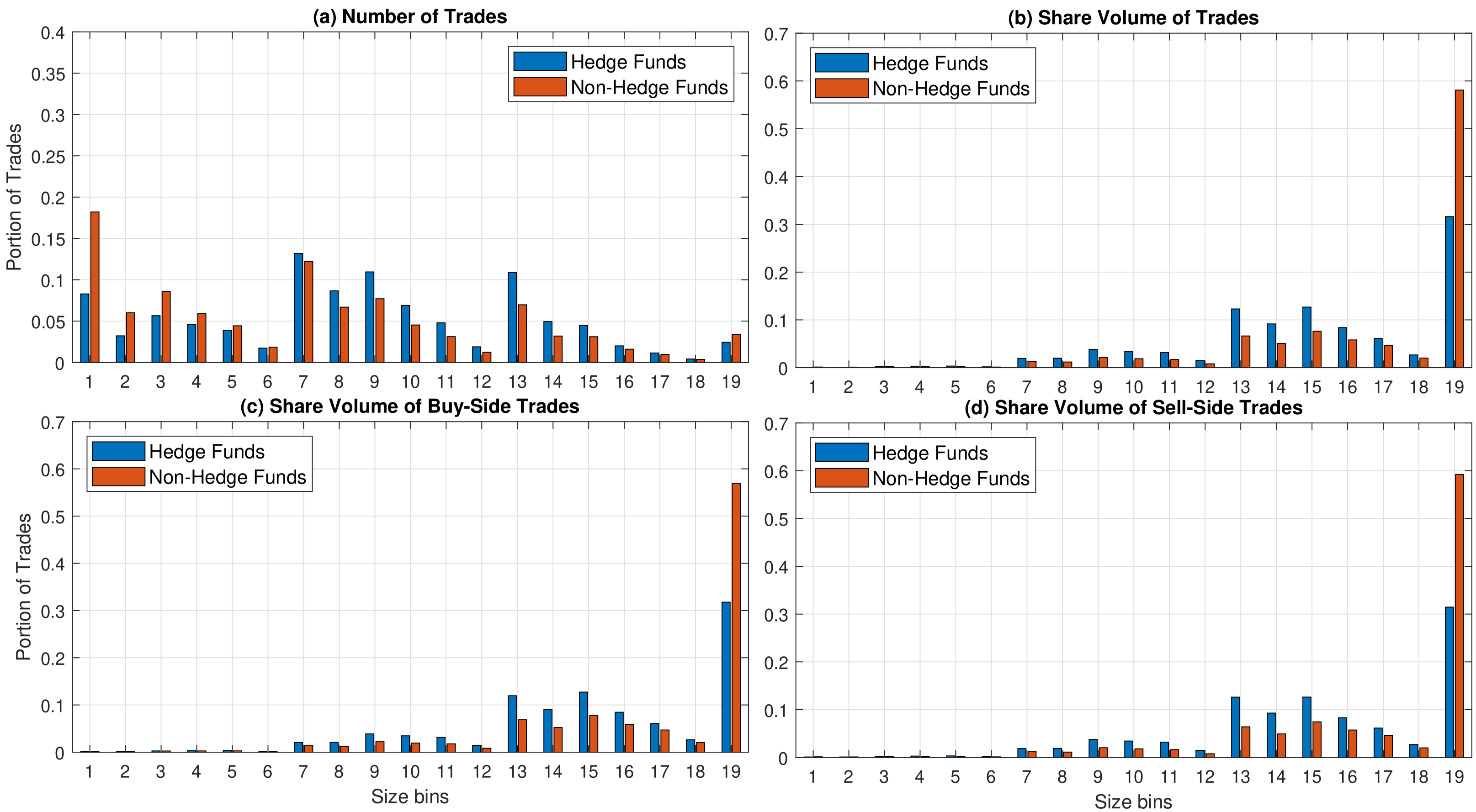
Panel A. 1999-2003



(Continued)

Figure A1 – *Continued*

Panel B. 2004-2007



(Continued)

Figure A1 – *Continued*

Panel C. 2008-2012

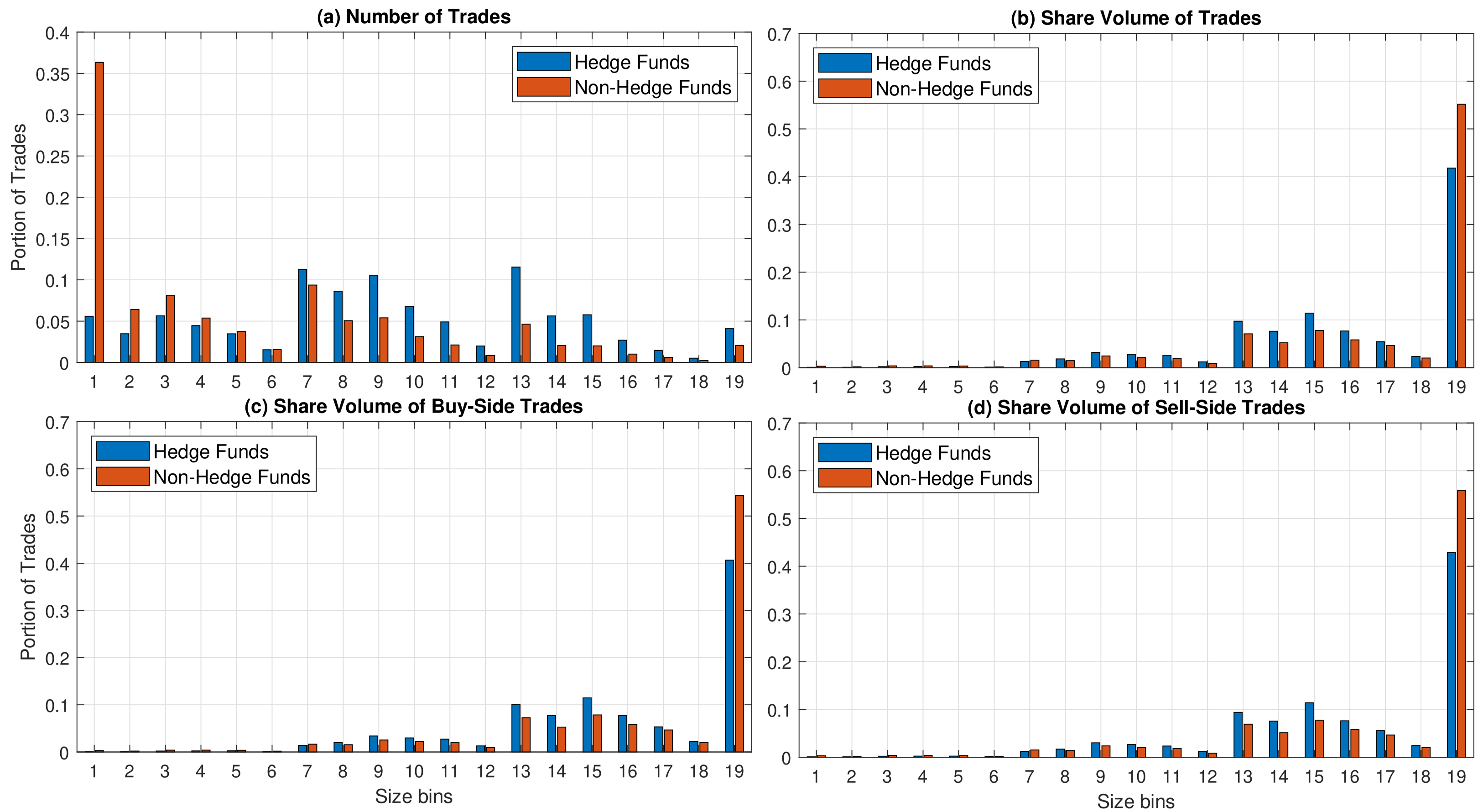


Table A1. Estimated coefficients of Campbell, Ramadorai, and Schwartz (2009) for hedge funds and non-hedge funds

This table presents non-linear least squares estimates of Campbell, Ramadorai, and Schwartz (CRS, 2009):

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \beta^{UY} Y_{i,q-1} \times U_{i,q} + \sum_{Z=1}^{19} \beta(Z, Y_{i,q-1}) F_{Z,i,q} + \epsilon_{i,q},$$

where for a stock i in a quarter q, α is a set of four quarter dummies, Y is aggregate hedge fund or non-hedge fund ownership in 13F, F_Z is aggregate Lee and Ready (1991) order imbalance scaled by shares outstanding in a trade-size bin Z, and U is aggregate unclassified trades scaled by shares outstanding for which the Lee and Ready (1991) algorithm cannot determine the direction. Hedge funds are identified in Thomson Reuters's 13F data, following Agarwal, Jiang, Tang, and Yang 2013. $\beta(Z, Y_{i,q-1}) = b_{01} + b_{02}Y_{i,q-1} + (b_{11} + b_{12}Y_{i,q-1} + b_{21} + b_{22}Y_{i,q-1})[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22}Y_{i,q-1})e^{-Z/\tau}$. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq that have information in TAQ, CRSP, and Thomson Reuters's 13F data. Estimated coefficients are reported separately for five size quintiles using NYSE breakpoints at the end of each quarter. Panels A, B, and C report estimates in 1993-2002, 2003-2012, and 2013-2022, respectively.

Panel A. Estimates in 1993-2002

	Hedge Funds					Non-hedge Funds				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.052***	-0.048***	-0.049***	-0.037***	-0.041***	-0.034***	-0.046***	-0.058***	-0.046***	-0.016***
	(-28.34)	(-12.12)	(-8.40)	(-5.22)	(-5.95)	(-18.84)	(-12.15)	(-12.49)	(-9.32)	(-3.96)
ρ	0.067***	0.055***	0.049***	0.062***	0.070***	0.111***	0.015	0.062***	0.064***	0.015
	(10.94)	(5.40)	(3.87)	(4.66)	(5.88)	(13.36)	(1.20)	(4.63)	(4.81)	(1.25)
β^U	0.139***	0.445***	0.365***	0.320***	0.256***	0.082***	0.243***	0.196***	0.190***	0.185***
	(14.09)	(17.91)	(11.11)	(9.73)	(9.82)	(14.57)	(16.67)	(11.05)	(10.77)	(11.60)
β^{UY}	-0.060***	-0.071***	-0.047***	-0.048***	-0.032***	-0.006***	-0.009***	-0.007***	-0.008***	-0.011***
	(-27.30)	(-17.64)	(-9.60)	(-9.80)	(-6.99)	(-18.93)	(-14.30)	(-8.85)	(-9.63)	(-11.23)
b_{01}	-0.056***	-0.184***	-0.134***	-0.135***	0.013	-0.494***	-0.780***	-0.487***	-0.529***	-0.385***
	(-10.82)	(-7.75)	(-4.58)	(-4.05)	(0.93)	(-9.00)	(-7.85)	(-4.04)	(-3.92)	(-3.42)
b_{02}	0.020***	0.020***	0.019***	0.006	0.011***	0.002	0.026***	0.008	0.006	0.006
	(21.50)	(5.81)	(4.65)	(1.31)	(4.34)	(0.99)	(5.97)	(1.43)	(0.94)	(0.95)
b_{11}	-0.130	0.116***	0.172***	0.045	-1.626***	0.647***	0.643***	0.222	0.329**	-0.691***
	(-1.18)	(4.61)	(5.15)	(1.24)	(-6.30)	(12.03)	(5.42)	(1.42)	(2.14)	(-5.19)
b_{12}	0.457***	-0.039***	-0.061***	-0.033***	0.531***	-0.007***	-0.031***	-0.008	-0.004	0.060***
	(16.46)	(-10.44)	(-12.45)	(-6.03)	(11.08)	(-2.78)	(-5.93)	(-1.12)	(-0.47)	(7.60)
b_{21}	0.703***	0.727***	0.640***	1.050***	2.633***	0.757***	3.631***	2.134***	4.856***	7.213***
	(3.95)	(6.88)	(4.00)	(5.47)	(6.07)	(3.77)	(6.77)	(2.91)	(5.77)	(9.76)
b_{22}	-0.884***	0.030*	0.056**	0.022	-0.979***	0.064***	-0.039	0.026	-0.194***	-0.436***
	(-21.00)	(1.89)	(2.42)	(0.78)	(-12.30)	(5.28)	(-1.54)	(0.78)	(-4.67)	(-9.29)
τ	900	87,700	74,100	90,900	4,700	31,400	33,800	28,000	71,000	100,000
Adjusted R^2	0.038	0.054	0.047	0.043	0.067	0.036	0.056	0.057	0.050	0.043
Observation	143,964	37,284	23,916	18,846	15,656	65,892	23,769	16,101	14,784	14,248

(Continued)

Table A1 – *Continued*

Panel B. Estimates in 2003-2012	Hedge Funds					Non-hedge Funds				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.049***	-0.060***	-0.053***	-0.051***	-0.040***	-0.011***	-0.017***	-0.019***	-0.018***	-0.034***
	(-26.96)	(-11.69)	(-7.88)	(-7.11)	(-6.16)	(-11.06)	(-6.24)	(-4.92)	(-3.74)	(-6.67)
ρ	0.024***	-0.027***	0.006	0.013	-0.042***	0.013***	-0.002	-0.013*	-0.037***	-0.093***
	(4.34)	(-2.66)	(0.52)	(1.04)	(-3.95)	(4.18)	(-0.34)	(-1.68)	(-4.28)	(-10.57)
β^U	0.853***	0.918***	0.945***	0.767***	0.474***	0.874***	2.346***	2.233***	2.310***	0.763***
	(30.48)	(16.49)	(14.26)	(11.54)	(8.51)	(19.87)	(17.56)	(12.22)	(10.21)	(2.83)
β^{UY}	-0.042***	-0.036***	-0.035***	-0.026***	-0.023***	-0.035***	-0.045***	-0.039***	-0.041***	-0.011**
	(-21.21)	(-9.68)	(-7.83)	(-5.52)	(-4.53)	(-33.94)	(-19.33)	(-13.02)	(-11.18)	(-2.54)
b_{01}	-0.255***	-0.150***	-0.156*	-0.089*	0.115	-0.287***	-0.426***	-0.254	-1.544***	1.010***
	(-8.95)	(-2.75)	(-1.86)	(-1.80)	(1.46)	(-5.63)	(-2.86)	(-1.19)	(-4.37)	(7.06)
b_{02}	0.004**	-0.002	0.004	0.005	-0.018**	0.010***	0.010***	0.007**	0.027***	-0.012***
	(2.19)	(-0.63)	(0.78)	(1.26)	(-2.50)	(8.40)	(4.10)	(2.00)	(4.62)	(-4.88)
b_{11}	0.284***	0.058	0.130	0.498*	-0.221***	0.580***	0.315	0.427	1.382***	11.904*
	(8.47)	(0.66)	(1.33)	(1.81)	(-2.65)	(11.03)	(1.60)	(1.48)	(3.69)	(1.69)
b_{12}	-0.016***	-0.014**	-0.012*	-0.014	0.013*	-0.014***	-0.010***	-0.016***	-0.028***	-0.138
	(-6.31)	(-2.44)	(-1.76)	(-0.76)	(1.76)	(-10.26)	(-2.79)	(-3.35)	(-4.64)	(-1.15)
b_{21}	0.621***	1.077***	0.539	0.025	0.054	0.774***	4.842***	3.799***	8.169***	-24.782**
	(4.38)	(3.09)	(1.19)	(0.04)	(0.14)	(3.45)	(5.51)	(3.07)	(4.66)	(-2.44)
b_{22}	-0.002	0.009	-0.011	-0.062	0.042	0.015**	-0.054***	-0.022	-0.088***	0.267
	(-0.16)	(0.37)	(-0.35)	(-1.28)	(1.11)	(2.39)	(-3.53)	(-1.05)	(-3.04)	(1.57)
τ	11,800	14,100	48,500	4,100	100,000	24,000	22,700	27,300	100,000	800
Adjusted R^2	0.036	0.052	0.047	0.047	0.057	0.048	0.076	0.074	0.064	0.069
Observation	97,289	24,289	16,665	14,533	13,497	97,289	24,289	16,665	14,533	13,497

(Continued)

Table A1 – *Continued*

Panel C. Estimates in 2013-2022

	Hedge Funds					Non-hedge Funds				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.034***	-0.046***	-0.058***	-0.046***	-0.016***	-0.015***	-0.033***	-0.040***	-0.022***	-0.024***
	(-18.84)	(-12.15)	(-12.49)	(-9.32)	(-3.96)	(-12.64)	(-13.76)	(-12.96)	(-7.01)	(-8.45)
ρ	0.111***	0.015	0.062***	0.064***	0.015	0.021***	0.021***	0.023***	0.012	-0.013
	(13.36)	(1.20)	(4.63)	(4.81)	(1.25)	(5.32)	(3.43)	(2.92)	(1.40)	(-1.51)
β^U	0.082***	0.243***	0.196***	0.190***	0.185***	0.107***	0.448***	0.384***	0.518***	0.220***
	(14.57)	(16.67)	(11.05)	(10.77)	(11.60)	(15.48)	(18.52)	(11.07)	(13.26)	(5.44)
β^{UY}	-0.006***	-0.009***	-0.007***	-0.008***	-0.011***	-0.006***	-0.008***	-0.007***	-0.009***	-0.004***
	(-18.93)	(-14.30)	(-8.85)	(-9.63)	(-11.23)	(-26.82)	(-17.91)	(-11.48)	(-13.18)	(-5.98)
b_{01}	-0.494***	-0.780***	-0.487***	-0.529***	-0.385***	-0.897***	-1.224***	-1.443***	-1.773***	-1.007***
	(-9.00)	(-7.85)	(-4.04)	(-3.92)	(-3.42)	(-13.34)	(-7.19)	(-5.85)	(-6.24)	(-3.32)
b_{02}	0.002	0.026***	0.008	0.006	0.006	0.020***	0.024***	0.029***	0.032***	0.020***
	(0.99)	(5.97)	(1.43)	(0.94)	(0.95)	(10.99)	(7.38)	(6.83)	(6.95)	(4.00)
b_{11}	0.647***	0.643***	0.222	0.329**	-0.691***	1.140***	0.691***	0.859***	1.459***	0.127
	(12.03)	(5.42)	(1.42)	(2.14)	(-5.19)	(17.28)	(3.53)	(3.00)	(4.58)	(0.35)
b_{12}	-0.007***	-0.031***	-0.008	-0.004	0.060***	-0.019***	-0.013***	-0.019***	-0.030***	-0.009
	(-2.78)	(-5.93)	(-1.12)	(-0.47)	(7.60)	(-9.83)	(-3.63)	(-3.81)	(-5.80)	(-1.46)
b_{21}	0.757***	3.631***	2.134***	4.856***	7.213***	1.016***	10.060***	11.999***	10.174***	7.364***
	(3.77)	(6.77)	(2.91)	(5.77)	(9.76)	(4.20)	(11.18)	(8.37)	(5.79)	(3.87)
b_{22}	0.064***	-0.039	0.026	-0.194***	-0.436***	0.033***	-0.092***	-0.124***	-0.043	-0.050
	(5.28)	(-1.54)	(0.78)	(-4.67)	(-9.29)	(4.03)	(-5.02)	(-4.73)	(-1.43)	(-1.51)
τ	31,400	33,800	28,000	71,000	100,000	22,000	43,000	50,700	70,600	100,000
Adjusted R^2	0.036	0.056	0.057	0.050	0.043	0.061	0.081	0.076	0.081	0.044
Observation	65,892	23,769	16,101	14,784	14,248	65,892	23,769	16,101	14,784	14,248

Table A2. Contemporaneous Price Impact

This table presents time-series averages of coefficient estimates from cross-sectional regressions of the following equation,

$$\begin{aligned}\text{Return}_{i,d} &= \alpha_d + \beta_d^{HF}\text{HF}_{i,d} + \beta_d^{NHF}\text{NHF}_{i,d} \\ &\quad + \sum_{k=1}^{5}\gamma_{d,k}^{T}\text{TOF}_{i,d-k} + \sum_{k=1}^{5}\gamma_{d,k}^{R}\text{Return}_{i,d-k} + \gamma_d^{B}\text{SPRD}_{i,d-1} + \gamma_d^{A}\text{AMI}_{i,d-1} + \epsilon_{i,d},\end{aligned}$$

where for stock i in a day d, $SPRD$ is daily relative spread, AMI is Amihud (2002) illiquidity, and all the other variables are the same as defined in Table I. For brevity, we only report the coefficient estimates of hedge fund order flow (HF), non-hedge fund order flow (NHF), and total order flow (TOF), while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively. Panels A, B, C and D report estimates in 1993-2022, 1993-2002, 2003-2012, and 2013-2022, respectively.

	1993-2022	1993-2002	2003-2012	2013-2022
HF_d	3.131***	7.889***	-1.599***	3.116***
	(10.46)	(12.21)	(-6.14)	(15.70)
NHF_d	2.795***	3.917***	3.528***	0.945***
	(33.02)	(42.96)	(27.08)	(9.10)
Adjusted R^2	0.043	0.033	0.047	0.048
Number of Stocks	2,973.1	3,704.3	2,600.6	2,616.9
Number of Days	7,544.0	2,509.0	2,517.0	2,518.0

Table A3. Return Predictability at Monthly Frequency

This table presents time-series averages of coefficient estimates from cross-sectional regressions of the following equation,

$$\begin{aligned}\text{Return}_{i,m} &= \alpha_m + \beta_m^{HF}\text{HF}_{i,m-1} + \beta_m^{NHF}\text{NHF}_{i,m-1} + \gamma_m^T\text{TOF}_{i,m-1} + \gamma_m^R\text{Return}_{i,m-1} \\ &\quad + \gamma_m^B\text{SPRD}_{i,m-1} + \gamma_m^A\text{AMI}_{i,m-1} + \gamma_m^M\text{MISP}_{i,m-1} + \gamma_m^I\text{IdioRisk}_{i,m-1} + \epsilon_{i,m}.\end{aligned}$$

where for stock i in a month m, HF and NHF represent monthly aggregated order flow from hedge funds and non-hedge funds, respectively, as estimated in sec:estimate. TOF is the monthly aggregated total order flow in TAQ, calculated using the Lee and Ready (1991) algorithm. SPRD is the relative spread at the end of the month, and AMI is the Amihud (2002) illiquidity measure over the month. IdioRisk refers to idiosyncratic risk from a Fama–French three-factor model over the past three years. MISP is a mispricing index proposed by Stambaugh, Yu, and Yuan (2012), while MISP excl. MOM is a mispricing index that excludes the momentum factor. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq market capitalization above the 10% NYSE breakpoints that have information in TAQ, CRSP, Compustat, and Thomson Reuters's 13F data. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at 1, 5, and 10 percent level, respectively.

	(1)	(2)
HF_{m-1}	0.525***	0.504**
	(2.71)	(2.58)
NHF_{m-1}	-0.251***	-0.246***
	(-4.16)	(-4.06)
TOF_{m-1}	0.019	0.019
	(0.60)	(0.62)
Return_{m-1}	-0.032***	-0.032***
	(-6.48)	(-6.48)
SPRD_{m-1}	0.458***	0.448***
	(5.99)	(5.90)
AMI_{m-1}	1.100***	1.101***
	(2.95)	(2.95)
MISP_{m-1}	0.019***	
	(5.22)	
MISP excl. MOM_{m-1}		0.021***
		(6.32)
IdioRisk_{m-1}	0.009	0.011
	(0.96)	(1.16)
Intercept	-0.014***	-0.014***
	(-6.41)	(-7.01)
Adjusted R^2	0.026	0.026
Number of Stocks	1,958.2	1,957.9
Number of Months	359.0	359.0