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

JinGi Ha, Jianfeng Hu, and Yuehua Tang *

October 14, 2025

ABSTRACT

We develop a new measure of daily aggregate hedge fund trades in individual U.S. stocks from 1993 to 2022. Using this measure, we find that hedge fund pre-announcement trading significantly predicts abnormal returns around corporate events and public news arrivals. We do not find similar results using non-hedge fund or aggregate institutional trading. We uncover that hedge funds increasingly utilize alternative data, such as satellite images, to generate alpha. Furthermore, hedge fund trading positively predicts future stock returns in the cross-section, both in-sample and out-of-sample. Finally, hedge fund trading enhances price efficiency, reducing stock variance ratios and post-earnings announcement drift.

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

*We would like to thank George Aragon, Laurent Barras, Maxime Bonelli, Darwin Choi, William Cong, Thomas Ernst, Dashan Huang, Russell Jame, Weikai Li, Byoung Uk Kang, Cristian Tiu, Russ Wermers, Hong Zhang, and the seminar participants at 2025 EFA (Paris), 2025 Lapland Investment Fund Summit, 2025 World Symposium on Investment Research, 2025 SMU Research Summer Camp, 2025 Asia-Pacific Association of Derivatives, Aarhus University, Deakin University, Monash University, Singapore Management University, and Zhejiang University for comments and suggestions. We are grateful to Russell Jame for sharing the list of hedge fund firms in the Abel Noser data. 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. Jianfeng Hu is with Singapore Management University (jianfenghu@smu.edu.sg); JinGi Ha is with Soongsil University (jgha@ssu.ac.kr); and Yuehua Tang is with 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 and 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 hedge 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 and the resulting impact on capital markets.

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 impact, offering new insights into the mechanisms behind hedge fund trading strategies and

[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.

their influence on market efficiency at a higher frequency.

Early studies on institutional trading estimate institutional order flow using high-frequency NYSE Trade and Quote (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 (2009), hereafter referred to as CRS, 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 corporate earnings.

In this paper, we build on the CRS method to estimate daily hedge fund (HF) and non-hedge fund (NHF) trades across a large sample of stocks from 1993 to 2022 by combining TAQ data with HF and NHF holding changes in 13F filings. Using proprietary institutional transaction-level data from Abel Noser that covers a subset of both HFs and NHFs, we find 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 variation in trade size allows us to disentangle HF and NHF trades by calibrating TAQ order flows across different trade-size bins to quarterly changes in holdings for HFs and NHFs separately. Our extension of the CRS model to HFs and NHFs thus enhances the precision of institutional trade estimation.[5] Our measures enable 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, previously unavailable due to data limitations, provides a clearer understanding of hedge fund trading dynamics relative to those of other institutional investors.

Using this new measure, we first investigate the potential informational sources 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).

[5] For converting quarterly to daily frequency, the CRS method requires that daily aggregate institutional trading error terms be uncorrelated with their leads and lags within a quarter. Using proprietary transaction-level data from Abel Noser, we confirm that this condition holds for both HFs and NHFs in the data. The averages of cross-sectional autocorrelations of the "residual" daily trading (defined as the difference between actual and estimated daily trades) for HFs and NHFs are close to zero.

funds may possess. We start by examining how HF and NHF trades respond to salient corporate events, including earnings announcements, analyst rating updates, and price jumps. Hedge fund trades before the event positively correlate with cumulative abnormal returns (CARs) around all types of corporate events. The economic magnitude is sizable. For instance, a one-standard-deviation increase in HF net purchases in the week immediately preceding an event corresponds to an increase of 50.3 basis points in three-day CARs around earnings announcements and 87.5 basis points around price jumps. 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 edge through trading ahead of such events. NHF trades, however, tend to be negatively associated with the abnormal returns around these events, highlighting their relative disadvantage in processing information around corporate disclosures.[6] Importantly, we do not find similar results around corporate events when using the total institutional order flow measure of Campbell, Ramadorai, and Schwartz (2009), which highlights the importance of separately estimating order flows for HFs and NHFs.

Our daily HF trading measure differs substantially from the net arbitrage trading (NAT) measure of Chen, Da, and Huang (2019), which is a quarterly measure defined as the difference between quarterly abnormal HF holdings and abnormal short interest. All our main results are obtained with NAT included as a control variable. Moreover, our HF trading measure is also distinct from the daily institutional trading measure based on the daily change in the security lending supply proposed by Barardehi, Da, Dixon, and Wang (2025) (BDDW). The correlation between our measure and theirs is low (below 0.03), and our results continue to hold when controlling for BDDW's institutional trading measure.

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) and Appel and

[6] Our evidence that NHF investors are uninformative in aggregate is consistent with the extensive literature documenting that the majority of mutual funds have negative (or close to zero) alphas (e.g., Jensen (1968), Gruber (1996), Carhart (1997), Kosowski, Timmermann, Wermers, and White (2006), Fama and French (2010), and Lewellen (2011)).

Fos (2024)). 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)).

One might be concerned that hedge funds may employ different execution strategies in trading on event vs. non-event days. We conduct two tests and do not find this to be the case. First, we do not observe a significant shift in the size distribution of hedge fund trades in the pre-event period, event period, and non-event days in the Abel Noser sample. Second, to formally test whether hedge funds use different trade sizes on event vs. non-event days, we regress returns on past size-bin level order flows and their interactions with an event dummy and find that the interaction terms are insignificant. Thus, we do not find a structural shift in hedge fund execution methods on event or news days in our data.

Importantly, we further 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, HF 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 HFs to integrate innovative sources such as big data into their investment processes, a capability less prevalent among NHFs, 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 the pervasiveness of hedge funds' informed trading. We do so by studying the return predictability of HF versus NHF trades using Fama-MacBeth regressions.

We find that daily aggregate HF trades significantly predict future stock returns over the next five trading days, with positive coefficients across various time lags, indicating a permanent price impact. The economic impact is sizable, with a one–standard-deviation rise in hedge fund net purchases corresponding to an increase of approximately 4 basis points (bps) in next-day stock returns. 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.

Reinforcing our interpretation of return predictability as evidence of informed trading, we find that hedge fund order flow predicts returns not only at the daily level but also over the monthly horizon. The predictability is stronger in stocks with higher trading frictions or weaker information environments. Importantly, our measure retains predictive power in a strict out-of-sample setting. Specifically, for each year we estimate the CRS model with the preceding ten years of data, use it to construct an out-of-sample HF measure, and document strong predictive power of this measure for subsequent stock returns. Finally, to address the limitation that 13F filings do not capture HF short-selling activity, we construct an alternative measure of HF trading by applying the CRS framework to the difference between quarterly 13F hedge fund holdings and quarterly short interest. This short-sale adjusted HF trading measure yields similar results to our baseline daily HF measure. Collectively, these findings further suggest that hedge funds engage in informed trading and that our estimation method effectively captures their informational edge.

In the final set of analyses, we examine 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 trading both in the cross-section and post-earnings announcements. Thus, hedge funds play a pivotal role in adjusting stock prices to reflect new information, leading to a quicker and more accurate incorporation of this information. In all these tests, we also include NHF 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.

Our study makes several contributions to the literature. First, to the best of our knowledge, we are the first to develop a daily aggregate HF 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 13F holdings data, 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 Abel Noser; however, these cover a highly selective sample of fewer than 100 hedge fund firms (e.g., Jame (2018) and Çötelioğlu, Franzoni, and Plazzi (2021)). In addition, Collin-Dufresne and Fos (2015) examine daily trades from 13D filings of specialized activist investors (mostly hedge funds) and their relation to market liquidity. Our paper advances the literature by developing high-frequency measures of hedge fund trading across a large sample of stocks, which shed new light on their trading behavior. In particular, our focus on daily HF trading differs from and complements prior studies at the quarterly frequency that combine 13F holdings and short interest data (e.g., Jiao, Massa, and Zhang (2016) and Chen, Da, and Huang (2019)), and the study on daily institutional trading based on security lending supply by Barardehi, Da, Dixon, and Wang (2025).[7]

[7]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 small group of hedge funds from Abel Noser.

Second, we provide the first systematic evidence that hedge funds engage in informed trading *ahead of* salient corporate events and firm-specific news. While it is a commonly held belief that hedge funds trade on superior information ahead of earnings announcements or other news events, there is surprisingly little direct empirical evidence to support this view. Most existing evidence remains suggestive using low-frequency data (see footnote 1). Our study addresses this gap by showing that hedge fund order flow reliably predicts returns in the days leading up to a comprehensive set of salient events for a large cross-section of stocks. 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 thus 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 (e.g., Goldstein, Spatt, and Ye (2025)). While recent studies have examined the impact of alternative data on stock prices, corporate managers, and mutual fund managers, there remains little direct evidence on which group of investors benefits most from the availability of alternative data.[8] To our knowledge, this is the first study to document that hedge funds have adopted alternative data in recent years to sustain their informational advantage.

Lastly, we shed new light on hedge funds' impact on the informational efficiency of stock prices at a higher frequency. In particular, we offer novel daily-level evidence of hedge funds' pivotal role in enhancing price efficiency, both broadly and around key events such as earnings announcements. In this regard, our study complements earlier evidence from lower-frequency analyses (e.g., Boehmer and Kelley (2009) and Cao, Liang, Lo, and Petrasek (2018)).

Importantly, their focus is on anomaly mispricing. Our paper differs by focusing on informed trading via fundamental information and by studying the full set of hedge fund firms covered in the 13F data.

[8]For recent studies on alternative data, see, e.g., Zhu (2019), Katona, Painter, Patatoukas, and Zeng (2024), Dessaint, Foucault, and Fresard (2024), and Bonelli and Foucault (2024).

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, 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. We further restrict the sample to include only stocks with a market capitalization above the 5% breakpoint of NYSE-listed stocks each quarter to avoid micro-cap stocks. 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 restrict TAQ observations to regular transactions between 9:30:00 to 15:59:59 EST.[9] We then employ the algorithm of Lee and Ready (1991) with no quote lag to identify trade direction. Specifically, a trade is classified as buyer-initiated (seller-initiated) if its trade price is above (below) the midpoint of the contemporaneous bid-ask spread. For trades at the midpoint price, we apply the tick test: if the trade price is higher (lower) than the previous trade price, it is classified as buyer-initiated (seller-initiated).[10] Order imbalance is calculated as the difference between buyer-initiated and seller-initiated trading volume in a trade-size bin on a given day, scaled by shares outstanding.

We identify hedge fund firms as 13F-filing institutions whose primary business is spon-

[9] Following the literature, 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').

[10] Chakrabarty, Pascual, and Shkilko (2015) show that the Lee and Ready (1991) algorithm performs well in classifying trades using NASDAQ's TotalView-ITCH data.

soring 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 (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 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:

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

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. 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. We use the same nineteen size bins introduced previously. 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), we follow CRS and apply a yield curve function from Nelson and Siegel (1987) to model the structure of β across trade-size bins:

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

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. Also 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 only report these estimates in Table A1 of the Internet Appendix. The estimated coefficients are highly significant 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 by setting the lagged level of quarterly institutional ownership to its in-sample mean. Figure A2 of the Internet Appendix plots the implied trade-size coefficients standardized across trade-size bins within each market capitalization group.

More importantly, to verify that our estimates reflect different execution styles of HF and

NHF, we compute the paired differences between the Nelson and Siegel (1987) sensitivities ($\beta(Z, v)$) of HF and NHF across trade size bins and stock sizes. Figure 2 plots the t-test results for the differences, $\beta^{HF}(Z, v) - \beta^{NHF}(Z, v)$, assuming constant Z, v, and τ.This plot reveals distinct patterns in execution methods across hedge funds and non-hedge funds over the three subperiods.

During the 1993-2002 period, for large-cap stocks, hedge funds exhibit a greater tendency to take liquidity in small-sized trades (bins 1-6) but provide liquidity in medium- and large-sized trades (bins 9-19) compared to non-hedge funds. For medium-cap stocks, we observe a similar pattern for small-sized trades that hedge funds take liquidity, but the difference becomes statistically insignificant after bin 5. For small-cap stocks, hedge funds demonstrate a higher propensity to provide liquidity in large trades (bins 14-19) relative to non-hedge funds, but show no significant differences in executing trades in other size bins. During the 2003-2012 period, hedge and non-hedge funds display similar behavior in small-sized trades across all stock sizes. However, for large-cap stocks, hedge funds show a greater likelihood of providing liquidity in medium- and large-sized trades (bins 11-19). During the 2013-2022 period, while hedge funds and non-hedge funds exhibit similar trade patterns for large-cap stocks across all trade size bins, significant divergences exist for small- and medium-cap stocks. In small-cap stocks, hedge funds tend to take liquidity in medium- and large-sized trades (bins 13-19). For medium-cap stocks, hedge funds are more inclined to provide liquidity in medium-sized trades (bins 10-14) and take liquidity for the largest trades in bin 19. Overall, the two types of institutional investors exhibit distinct execution methods regarding trade size, and these differences vary substantially over time and across market-cap quintiles.

[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,

$E[\Delta Y_{i,d}]$, taking the estimates in Equation (1):

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

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. 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.

We note that the frequency conversion from quarterly to daily horizon is valid under an exogeneity assumption that the error terms, $\epsilon_{i,d}$, are not correlated with all of its leads and lags within a quarter. To validate this assumption, we employ daily HF and non-HF trades from the Abel Noser (AN) trade data. Specifically, we estimate quarterly coefficients from a regression model based on Equation (1), using quarterly aggregated HF or NHF trades in AN data as the dependent variable.[11] We then retrieve daily HF and NHF order flow by applying the estimated coefficients to daily AN trade size bins using a CRS-style procedure, and define the "residual" daily trading as the difference between actual and retrieved daily trades. We find that the time-series averages of cross-sectional Spearman autocorrelations of the residuals are close to zero: −1.25% for HF and 2.24% for NHF. These results support the validity of applying the CRS method to estimate daily HF and NHF trades.

D. *Summary statistics*

Table I provides 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

[11] In this specification, $Y_{i,q-1}$ is set to zero as we do not know the ownership level of the AN HF/NHF traders, and $F_{Z,i,q}$ represents trade size bin aggregates from the AN data.

described in Section II.C. TOF is calculated for all valid TAQ transactions 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 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 A2 of the Internet Appendix. 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.

[Place Table I about here]

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.

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).[12] 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. Table II reports the time-series

[12] We document detailed calculation methods for these firm characteristics in the Appendix at the end of the paper.

averages of cross-sectional coefficients for the following model:

$$\begin{aligned}\text{OF}_{i,m} &= \alpha_m + \beta_m^B \text{MktBeta}_{i,m} + \beta_m^C \text{MktCap}_{i,m} + \beta_m^{MB} \text{MB}_{i,m} + \beta_m^R \text{REV}_{i,m} + \beta_m^M \text{MOM}_{i,m} \\ &\quad + \beta_m^S \text{SPRD}_{i,m} + \beta_m^T \text{TURN}_{i,m} + \beta_m^I \text{IdioRisk}_{i,m} + \beta_m^{M2} \text{MISP excl. MOM}_{i,m} + \epsilon_{i,m}. \quad (4)\end{aligned}$$

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 monthly 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:

$$\begin{aligned} \text{CAR}_{i,t-1,t+1} \;&=\; \alpha_i + \alpha_y + \sum_{k=2}^{6} \beta_k^{HF} \text{HF}_{i,t-k} + \sum_{k=2}^{6} \beta_k^{NHF} \text{NHF}_{i,t-k} + \sum_{k=2}^{6} \gamma_k^{T} \text{TOF}_{i,t-k} \\ &+ \sum_{k=2}^{6} \gamma_k^{R} \text{Return}_{i,t-k} + \gamma^{N} \text{NAT}_{i,q-1} + \gamma^{S} \text{SPRD}_{i,t-2} + \gamma^{S} \text{AMI}_{i,t-2} + \epsilon_{i,t}, \quad (5) \end{aligned}$$

where for each event i on day t in a quarter q of a year y, CAR is the cumulative abnormal return from days $t-1$ to $t+1$, NAT is the quarterly net arbitrage trading measure proposed by Chen, Da, and Huang (2019) (i.e., the difference between quarterly abnormal hedge fund holdings and abnormal short interest), SPRD is relative bid-ask spread, AMI is Amihud (2002) illiqudity measure, and the other variables are the same as defined in Table I. 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 8.093

(t-stat = 5.16) for earnings announcements, 1.873 (t-stat = 2.73) for analyst rating updates, and 11.956 (t-stat = 5.83) for permanent price jumps. The economic magnitude is sizable. For instance, a one–standard deviation increase in hedge fund net purchases over days $d-6$ to $d-2$ corresponds to a total increase of 50.3 basis points in cumulative abnormal returns around earnings announcements and 87.5 basis points around 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.

The negative coefficients for NHFs partly reflect the high correlation between HF and NHF order flows. In unreported tests where NHF is included as the sole regressor, we find that NHF positively predicts event returns at the first lag; however, reversals at longer lags offset the initial impact, and the total predictive power over the five-day pre-event window is statistically insignificant. Our finding on the informativeness of NHFs is consistent with the consensus in the extensive literature that mutual funds in aggregate do not generate positive alpha (e.g., Jensen (1968), Gruber (1996), Carhart (1997), and Fama and French (2010)).

We then analyze the three event types across two subperiods, separated by the enactment of Regulation Fair Disclosure (RegFD) in 2000: 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. For instance, a one–standard deviation increase in hedge fund net purchases over days $d-6$ to $d-2$ corresponds to a total increase of 133.9 basis points around permanent price jumps in 1993-1999. This effect decreases to 85.5 basis points in the 2000-2022 period. 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]

One potential concern is that hedge funds may employ special execution techniques to exploit short-term information around event days that differ from their regular trading patterns identified by the linkage between 13F and TAQ data. If so, this may challenge the validity of our hedge fund trade flow estimated using the full sample of trades in this event study. Specifically, the relation between hedge fund trading and order flows of the 19 trade size bins in the TAQ data may depend on the types of days (e.g., event vs. non-event days). We conduct two tests and verify that this is not the case. First, we plot the trade distribution across size bins for pre-corporate event period (event day $t-6$ to $t-2$), event period ($t-1$ to $t+1$) and non-event days using the Abel Noser sample in Figure A3 of the Internet Appendix. The pre-event and event periods correspond to the measurement windows of independent and dependent variables in our main analysis. We do not observe a significant shift in the distribution of hedge fund trades across the three types of days. Second, as informed traders, if hedge funds use different trade sizes on event days from normal (non-event) days, we should expect to see significant differences in the predictive ability of order flows from several size bins between event days and normal days. In Table A3 of the Internet Appendix, we formally test this hypothesis by regressing returns on past size-bin order flows and their interactions with an event dummy. We cannot reject the null hypothesis that these interactions have zero coefficients. Therefore, we do not find a structural shift in hedge fund execution methods on event days in our data.

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.[13] 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 RavenPack database that cover more than 5,000 firms to ensure representation of smaller firms (Kolasinski and Yang (2018)).[14] 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.

[13] 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.

[14] 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.

[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. Cross-sectional variation conditioning on firm characteristics

In this subsection, we study the cross-sectional variation in the reaction of hedge fund (HF) and non-hedge fund (NHF) trades to corporate events and RavenPack news across subsamples of stocks. Specifically, we first condition on liquidity measures, including firm size, bid-ask spread, and the Amihud illiquidity measure. To do so, we sort all stocks into two groups based on the daily cross-sectional median of each liquidity measure from the full sample. Panels A to C of Table V report the estimated coefficients of HF and NHF from the regression model in Table III for each subsample. For brevity, we use the average HF and NHF from $d-6$ to $d-2$ instead of five individual lagged terms in this analysis and examine the combined samples for corporate events and news arrivals without disentangling sub-events.

[Place Table V about here]

The results reveal several interesting patterns. First, in predicting stock price reactions, HF shows positive and significant coefficients across all liquidity-sorted subsamples, whereas NHF exhibits negative and significant coefficients in all subsamples. Second, the predictability power of HF trading is significant across all event types, including corporate events and both fundamental and non-fundamental RavenPack news. Third, the predictability of HF trading varies inversely with liquidity measures. Specifically, the five-day average of HF trades in the corporate event sample has coefficients of 38.52 for small stocks and 17.06 for large stocks; 28.28 for stocks with wide spreads and 15.52 for stocks with narrow spreads; 46.42 for illiquid stocks with high Amihud measures and 15.83 for liquid stocks with low

Amihud measures. A stronger effect in illiquid stocks suggests that hedge funds are more likely to have an edge in illiquid stocks, which have higher information asymmetry and higher arbitrage costs.

In Panels D and E of Table V, we repeat the analysis over subsamples of stocks sorted by the information environment, as measured by the number of sell-side analysts from I/B/E/S and institutional ownership from Thomson Reuters' 13F data. As with liquidity proxies, we separate all stocks into two groups based on the daily cross-sectional median of each information proxy from the full sample. The results show that HF exhibits positive predictive power in all subsamples, while the coefficient for NHF is negative in all subsamples. For corporate events and RavenPack fundamental news, HF trading's predictability is notably stronger for stocks with low analyst coverage and low institutional ownership. These findings support the notion that more opaque stocks are associated with higher arbitrage risks and information asymmetry, areas where hedge funds have a distinct advantage.

Taken together, the results indicate that hedge funds' ability to trade ahead of short-lived information is robust across various stock subsamples, with a particularly strong advantage in stocks characterized by higher arbitrage costs and less transparency.

D. 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 days. Satellite data provides valuable insights into forthcoming quarterly reports (e.g., 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.[15] 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 VI presents the estimated coefficients from the following model, using a difference-in-differences approach with firm and year fixed effects:

$$
\begin{aligned}
\text{Return}_{i,d} &= \alpha_i + \alpha_y + \beta_1 \text{TREAT} \times \text{POST} \times \text{HF}_{d-1} + \beta_2 \text{TREAT} \times \text{POST} \times \text{NHF}_{d-1} \\
&+ \beta_3 \text{TREAT} \times \text{POST} \times \text{TOF}_{d-1} + \beta_4 \text{TREAT} \times \text{HF}_{d-1} + \beta_5 \text{TREAT} \times \text{NHF}_{d-1} \\
&+ \beta_6 \text{TREAT} \times \text{TOF}_{d-1} + \beta_7 \text{POST} \times \text{HF}_{d-1} + \beta_8 \text{POST} \times \text{NHF}_{d-1} + \beta_9 \text{POST} \times \text{IOF}_{d-1} \\
&+ \beta_{10} \text{TREAT} \times \text{POST} + \beta_{11} \text{TREAT} + \beta_{12} \text{POST} + \beta_{13} \text{HF}_{d-1} + \beta_{14} \text{NHF}_{d-1} + \beta_{15} \text{TOF}_{d-1} \\
&+ \gamma \text{Controls}_{d-1} + \epsilon_{i,d}, \qquad (6)
\end{aligned}
$$

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. 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 the net arbitrage trading measure, lagged stock returns and illiquidity measures. Standard errors are clustered by firm and year.

In Table VI, 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

[15]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).

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 VI 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 in a consistent manner 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 VII reports the estimated coefficients of the following model:

$$\begin{aligned} \text{Return}_{i,d} \quad = \quad & \alpha_d + \sum_{k=1}^{5} \beta_{d,k}^{HF} \text{HF}_{i,d-k} + \sum_{k=1}^{5} \beta_{d,k}^{NHF} \text{NHF}_{i,d-k} + \sum_{k=1}^{5} \gamma_{d,k}^{T} \text{TOF}_{i,d-k} \\ & + \sum_{k=1}^{5} \gamma_{d,k}^{R} \text{Return}_{i,d-k} + \gamma_d^N \text{NAT}_{i,q-1} + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \epsilon_{i,d}. \quad (7) \end{aligned}$$

where for stock i on day d in a quarter q, and all variables are the same as defined before. To account for serial correlations, we use Newey−West (1987) standard errors with ten lags to calculate the t-statistics. For brevity, Table VII 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 17.16. Economically, this translates into a notable effect: a one standard-deviation increase in HF net purchases corresponds to a 3.9 bp increase in the next day's stock return. This positive effect persists across multiple lags, with coefficients of 0.046 (t-stat = 0.73), 0.143 (t-stat = 2.27), 0.109 (t-stat = 1.71), and 0.154 (t-stat = 2.42) 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 = −12.25), 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 2.887 (t-stat = 15.27), declining to 1.294 (t-stat = 11.87) in the 2003–2012 period, and further tapering to 0.444 (t-stat = 4.86) in the 2013–2022 period. Similarly, NHF exhibits reversals across all subperiods, with coefficients of −0.572 (t-stat = −12.57) in 1993–2002, −0.201 (t-stat = −7.64) in 2003–2012, and −0.078 (t-stat = −2.14) in 2013–2022. This attenuation in both statistical and economic significance over time may reflect the evolution of a more transparent and efficient market.

[Place Table VII about here]

In Table B1 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 support 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. To reinforce our interpretation of the return predictability as a result of informed trading, we further investigate the cross-sectional variation of the predictability based on firm characteristics in Table B2 of the Internet Appendix. The predictive power of hedge fund order flow remains robust across subsamples sorted by liquidity (firm size, bid-ask spread, and Amihud illiquidity) and information environment (analyst coverage and institutional ownership). Notably, the predictive effect is stronger in illiquid stocks and in stocks with lower analyst coverage or institutional ownership. These patterns suggest that hedge funds' informational advantage is most pronounced in markets where traditional information channels are weaker and trading costs are higher.

B. *Out-of-sample test*

To further evaluate the robustness of our hedge fund order flow estimates, we conduct an out-of-sample (OOS) test designed to assess predictive performance in a setting that more closely mirrors real-time investment decisions. We implement a ten-year rolling window approach to construct daily hedge fund (OOS HF) and non-hedge fund (OOS NHF) order flows. Specifically, for any given year, we estimate model coefficients using the preceding ten years of data, and fit the real-time TAQ order flow data on the lagged parameters to obtain the OOS HF and NHF. This procedure is repeated annually for the full OOS window from 2003 to 2022. Despite relying exclusively on historical information, OOS HF and OOS NHF exhibit high correlations with their in-sample counterparts at around 65%, indicating consistency in signal extraction while preserving out-of-sample integrity.

Using these OOS estimates, we replicate the regression analyses in Table VII, and the results are summarized in Table VIII. For the whole OOS sample from 2003 to 2022, the coefficient for OOS HF at the first lag is positively significant, with a value of 0.798 (t-stat = 9.47). Economically, this indicates that a one-standard-deviation increase in hedge fund order flow is associated with an increase in subsequent day's stock return of approximately 2.2 basis points. The positive predictive influence persists over the following days, evidenced by coefficients of 0.118 (t-stat = 1.52), 0.178 (t-stat = 2.29), 0.151 (t-stat = 1.93), and 0.062

(t-stat = 0.88) from lags two to five. In contrast, OOS NHF displays a significantly negative coefficient of −0.106 at the first lag (t-stat = −4.58), with continued negative impacts from subsequent lags, consistent with the transitory price pressure observed in earlier sections.

Subperiod results further underscore the stability of our approach. The first-lag coefficients for OOS HF are 0.839 (t-stat = 5.96) in 2003–2012 and 0.757 (t-stat = 8.38) in 2013–2022, reflecting sustained predictive power across distinct market conditions. OOS NHF coefficients remain significantly negative in both periods, with values of −0.113 (t-stat = −4.14) and −0.100 (t-stat = −2.73), respectively. These findings confirm that our methodology produces a robust measure of hedge fund trading activity that retains its informativeness even in an out-of-sample context. The results reinforce the value of our estimates in forecasting returns and highlight the persistent informational content embedded in hedge fund order flow.

[**Place Table VIII about here**]

C. Comparison with other measures of institutional trading

To further benchmark our daily HF trading measure against existing proxies for daily institutional trading, we conduct a series of comparison analyses with two widely cited alternatives: the daily institutional order flow estimate of Campbell, Ramadorai, and Schwartz (2009), and the lendable-shares-based measure introduced by Barardehi, Da, Dixon, and Wang (2025) (BDDW). While these measures are commonly used to capture institutional trading activities, neither is designed to distinguish between hedge funds and other institutions. In contrast, our method explicitly isolates hedge fund order flow and seeks to identify informed trading with greater precision. Additionally, to address the limitation that 13F holdings do not capture short-selling activities by hedge funds, we construct a short-sale adjusted measure (HF−SI) and compare its performance with the original hedge fund (HF) trade measure.

As reported in Section C of the Internet Appendix, the results consistently demonstrate the superior performance of our measure. In cross-sectional return predictability regressions, hedge fund order flow significantly outperforms both CRS and BDDW measures, including in direct horse race specifications where all measures are included simultaneously. The coefficient on our hedge fund measure remains large and statistically significant, whereas those on CRS and BDDW are generally insignificant. Event study analyses further reinforce this pattern: before earnings announcements, analyst rating changes, and permanent price jumps, hedge fund order flow is consistently and significantly related to cumulative abnormal returns, while the alternative measures are not able to capture such predictive signals. Similarly, in the context of firm-specific news from RavenPack, hedge fund order flow shows strong and consistent return predictability across both fundamental and non-fundamental news types, while the alternative measures perform inconsistently. These findings underscore the informativeness and robustness of our measure in capturing the institutional trades from hedge funds that reflect superior information.

To address the limitation that 13F filings do not capture hedge fund short-selling activity, we construct an alternative measure of hedge fund trading by subtracting quarterly short interest (SI) in Compustat from hedge fund holdings in 13F filings. The implicit assumption of this analysis is that all short interest originates from hedge funds and hedge funds do not supply lendable shares. As detailed in Section D of the Internet Appendix, we estimate the HF−SI measure at the daily frequency using the same CRS framework as our baseline. Across a range of tests, including cross-sectional return predictability, corporate event studies, and firm-specific news reactions, this short-sale adjusted HF trading measure yields results that closely track those of our baseline HF measure, albeit with slightly attenuated magnitudes in some specifications. These findings suggest that while 13F data do not explicitly account for short-selling of hedge funds, this omission does not materially bias the informativeness of our daily hedge fund order flow estimates.

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

$$\mathrm{y}_{i,d} = \alpha_i + \alpha_q + \beta^{HF}\mathrm{HF}_{i,d} + \beta^{NHF}\mathrm{NHF}_{i,d} + \gamma \mathrm{ControlVariables}_{i,d} + \epsilon_{i,d}, \tag{8}$$

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. $\mathrm{CAR}_{d+1,d+61}$ 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 ($\mathrm{CAR}_{d-21,d}$). 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 IX presents the results for PEAD. Hedge fund trades on the day of the earnings announcement (HF_d) exhibit a negative and significant coefficient of −7.980, with a t-statistic of −2.16. 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 3.183, with a t-statistic of 2.03, 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: −1.042 (t-stat = −10.16) for hedge fund trades and −0.654 (t-stat = −13.36) 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 ($|HF|_d$) is negatively and significantly associated with ΔVR, with a coefficient of −2.626 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 ($|NHF|_d$) 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 IX 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:

$$\Delta\text{VR}_{i,d} = \alpha_d + \beta_d^{HF}|\text{HF}|_{i,d-1} + \beta_d^{NHF}|\text{NHF}|_{i,d-1} + \gamma\text{ControlVariables}_{i,d-3} + \epsilon_{i,d}, \quad (9)$$

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

Table X 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 Tables I and A2 and Figure A2, supporting the interpretation that non-hedge fund trading behavior has become increasingly similar to that of hedge funds over time.

[Place Table X 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 being 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 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] Appel, Ian and Fos, Vyacheslav, 2024, Short campaigns by hedge funds, The Review of Financial Studies **37**, pp. 1460–1493.

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

[9] 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.

[10] Barardehi, Yashar, Da, Zhi, Dixon, Peter, and Wang, Junbo L., 2025, You Can Only Lend What You Own: Inferring Daily Institutional Trading from Security Lending Supply, Available at SSRN: https://ssrn.com/abstract=5025126, pp. 1–30.

[11] 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.

[12] 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.

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

[14] 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.

[15] Bonelli, M. and Foucault, T., 2024, Does Big Data Devalue Traditional Expertise? Evidence from Active Funds, HEC Paris Research Paper FIN-2023-1492, pp. 1–30.

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

[17] Campbell, John, Ramadorai, Tarun, and Schwartz, Allie, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, Journal of Financial Economics **92**, pp. 66–91.

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

[19] 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.

[20] Cao, Charles, Liang, Bing, Lo, Andrew W., and Petrasek, Lubomir, 2018, Hedge fund holdings and stock market efficiency, The Review of Asset Pricing Studies **8**, pp. 77–116.

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

[22] Chakrabarty, Bidisha, Pascual, Roberto, and Shkilko, Andriy, 2015, Evaluating Trade Classification Algorithms: Bulk Volume Classification versus the Tick Rule and the Lee-Ready Algorithm, Journal of Financial Markets **25**, pp. 52–79.

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

[24] 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.

[25] 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.

[26] Collin-Dufresne, Pierre and Fos, Vyacheslav, 2015, Do prices reveal the presence of informed trading?, The Journal of Finance **70**, pp. 1555–1582.

[27] 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.

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

[29] 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.

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

[31] 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.

[32] Fama, Eugene F and French, Kenneth R, 2010, Luck versus skill in the cross-section of mutual fund returns, The journal of finance **65**, pp. 1915–1947.

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

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

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

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

[37] 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.

[38] Goldstein, Itay, Spatt, Chester S, and Ye, Mao, 2025, The next chapter of big data in finance, The Review of Financial Studies **38**, pp. 605–622.

[39] 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.

[40] Gruber, Martin J, 1996, Another puzzle: The growth in actively managed mutual funds, The journal of finance **51**, pp. 783–810.

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

[42] 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.

[43] 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.

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

[45] 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.

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

[47] 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.

[48] Jensen, Michael C, 1968, The performance of mutual funds in the period 1945-1964, The Journal of finance **23**, pp. 389–416.

[49] 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.

[50] 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.

[51] 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.

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

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

[54] 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.

[55] Kosowski, R., Timmermann, A., Wermers, R., and White, H., 2006, Can mutual fund ”stars” really pick stocks? New evidence from a bootstrap analysis, Journal of Finance **61**, pp. 2551–2595.

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

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

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

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

[60] Lewellen, Jonathan, 2011, Institutional investors and the limits of arbitrage, Journal of Financial Economics **102**, pp. 62–80.

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

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

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

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

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

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

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

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

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

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

[75] 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.

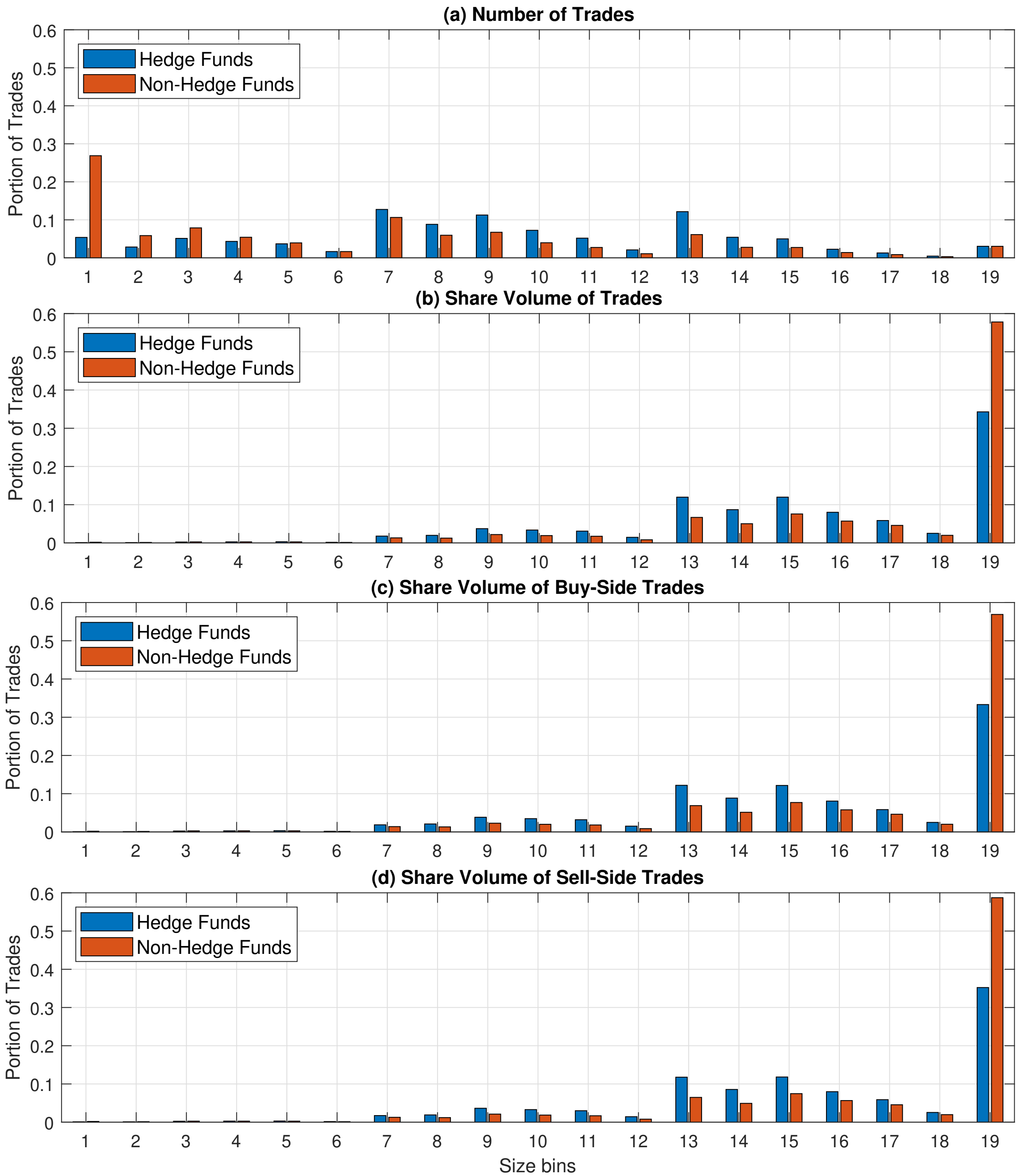


Figure 2. This figure shows t-statistics for the difference between the Nelson and Siegel (1987) sensitivities ($\beta(Z, v)$ in Equations (1) and (2)) of hedge funds and non-hedge funds, across trade size bins and stock sizes, based on the results in Table A1 of the Internet Appendix. The trade size 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' 13F data.

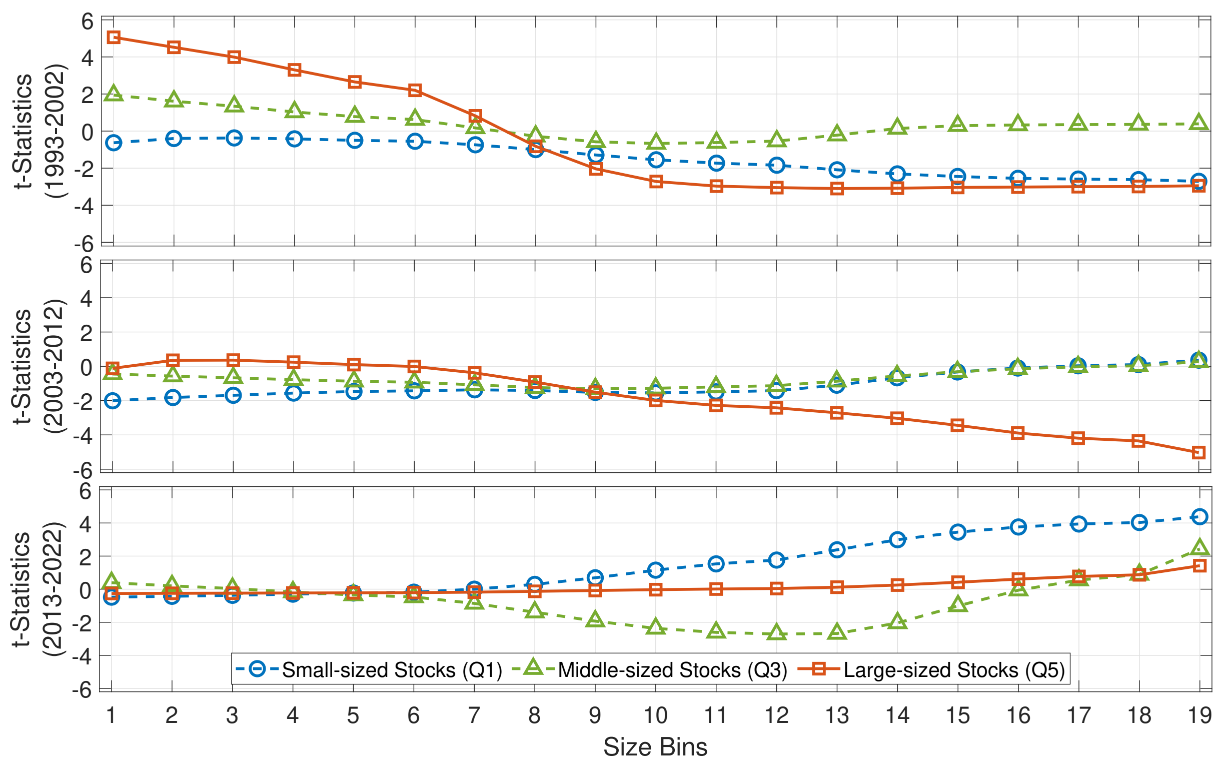


Table I. Summary Statistics

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' 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' 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,

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

where for each corporate event i announced on day t in a year y, CAR is cumulative abnormal return from day $t-1$ to $t+1$, NAT is net arbitrage trading measure proposed by Chen, Da, and Huang (2019), SPRD is relative bid-ask spread, AMI is Amihud (2002) illiqudity measure, and the other variables are the same as defined in Table I. 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' 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}	8.093***	11.352***	7.159***	1.873***	5.264***	1.363*	11.956***	28.440***	10.042***
	(5.16)	(2.71)	(4.28)	(2.73)	(2.63)	(1.94)	(5.83)	(4.41)	(4.70)
HF_{d-3}	2.791*	3.188*	2.843	0.881	4.889**	0.374	6.319***	17.872***	4.836***
	(1.94)	(1.66)	(1.63)	(1.30)	(2.21)	(0.53)	(4.97)	(4.89)	(4.09)
HF_{d-4}	5.592***	6.959*	5.590***	2.412***	2.841***	2.326***	7.812***	12.710***	7.420***
	(5.46)	(1.73)	(4.96)	(4.13)	(2.89)	(3.52)	(4.63)	(4.71)	(3.70)
HF_{d-5}	1.840	-1.309	2.454	2.986***	5.422***	2.649***	3.806**	14.113***	2.788
	(1.00)	(-0.44)	(1.12)	(3.62)	(3.14)	(3.16)	(2.28)	(4.91)	(1.54)
HF_{d-6}	1.792	-1.553	2.030	1.089	2.745	1.083	5.109***	15.408**	3.760***
	(1.40)	(-0.70)	(1.34)	(1.56)	(0.88)	(1.52)	(3.57)	(2.28)	(2.98)
NHF_{d-2}	-1.637***	-2.510***	-1.265***	-0.610**	-1.143**	-0.473	-3.081***	-4.595***	-2.825***
	(-4.36)	(-4.63)	(-2.93)	(-2.27)	(-2.53)	(-1.48)	(-4.70)	(-5.00)	(-3.45)
NHF_{d-3}	-1.080*	-0.937***	-1.133	-0.452*	-1.504***	-0.214	-1.914***	-4.293***	-1.399***
	(-1.83)	(-2.66)	(-1.45)	(-1.85)	(-3.49)	(-0.78)	(-3.89)	(-3.09)	(-3.00)
NHF_{d-4}	-1.405***	-1.416**	-1.474**	-0.928***	-1.586***	-0.749***	-2.054***	-2.842***	-1.938***
	(-3.01)	(-2.33)	(-2.45)	(-4.41)	(-2.97)	(-3.49)	(-4.21)	(-4.21)	(-3.15)
NHF_{d-5}	-0.829	-0.470	-0.937	-0.912***	-1.499	-0.778***	-1.643***	-3.155***	-1.379**
	(-1.30)	(-0.49)	(-1.17)	(-2.95)	(-1.40)	(-3.14)	(-3.03)	(-3.76)	(-2.14)
NHF_{d-6}	-0.391	-0.585	-0.239	-0.634***	-0.390	-0.753***	-1.978***	-3.543***	-1.649***
	(-0.72)	(-0.78)	(-0.35)	(-3.49)	(-0.96)	(-3.50)	(-4.89)	(-5.33)	(-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	0.012	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' 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}	5.445***	1.943	8.050***	2.431	-1.136	4.896***	-5.125	3.934**	1.875	7.562***	-0.983
	(3.56)	(1.51)	(2.98)	(1.32)	(-0.62)	(3.08)	(-1.23)	(2.34)	(0.67)	(3.31)	(-0.34)
HF_{d-3}	3.950***	3.132*	3.280	2.009	0.710	0.101	1.265	2.522**	1.268	2.039	-2.288
	(3.87)	(1.86)	(1.09)	(0.80)	(0.29)	(0.04)	(0.35)	(2.19)	(0.55)	(1.23)	(-1.00)
HF_{d-4}	3.596***	2.158	6.653***	9.970***	3.444**	-0.752	0.823	1.713	4.542	5.089***	3.545
	(3.19)	(1.30)	(3.74)	(3.36)	(2.13)	(-0.22)	(0.21)	(0.87)	(1.36)	(3.27)	(1.10)
HF_{d-5}	1.243	5.192***	10.885***	1.128	0.966	3.318**	12.455**	1.418	-1.457	1.930	4.203*
	(1.02)	(4.07)	(3.12)	(0.31)	(0.61)	(2.31)	(2.46)	(0.65)	(-0.77)	(1.02)	(1.72)
HF_{d-6}	-1.275	4.816***	0.851	-3.281	1.690*	2.552	1.566	1.068	3.498*	-0.376	-0.348
	(-0.76)	(3.13)	(0.34)	(-0.91)	(1.79)	(1.23)	(0.42)	(0.93)	(1.90)	(-0.20)	(-0.12)
NHF_{d-2}	-1.597***	-0.816***	-1.491	-0.322	0.747	-0.769	0.018	-0.949	0.393	-1.336**	0.673
	(-2.89)	(-3.15)	(-1.31)	(-0.32)	(1.32)	(-1.60)	(0.03)	(-1.19)	(0.49)	(-2.20)	(1.06)
NHF_{d-3}	-0.593	-0.277	-0.900	-2.027*	-0.469	-0.136	-0.460	-0.462	0.048	-0.693	0.436
	(-1.60)	(-0.58)	(-1.15)	(-1.71)	(-0.59)	(-0.21)	(-0.36)	(-0.65)	(0.09)	(-1.20)	(0.69)
NHF_{d-4}	-1.121**	-0.578	-0.226	-1.934**	-0.109	-0.148	0.767	-0.879*	0.155	-1.111*	-0.424
	(-2.00)	(-1.24)	(-0.35)	(-2.37)	(-0.15)	(-0.16)	(1.08)	(-1.83)	(0.19)	(-1.93)	(-0.56)
NHF_{d-5}	-0.220	-0.795**	-2.203**	0.474	0.046	0.242	-2.801**	-0.420	-0.186	-1.066*	-1.478*
	(-0.53)	(-2.08)	(-2.35)	(0.38)	(0.06)	(0.53)	(-2.22)	(-0.57)	(-0.39)	(-1.86)	(-1.90)
NHF_{d-6}	0.081	-1.734***	-0.306	1.895*	-0.934**	-1.247*	-2.345**	-0.020	-1.041	-0.359	-0.009
	(0.19)	(-3.95)	(-0.28)	(1.73)	(-2.26)	(-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	0.017	0.009	0.007	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. Cross-sectional Variation in Predicting Event Returns

This table replicates the regression analysis in Table III using event study subsamples based on liquidity and information proxies. The liquidity proxies are market capitalization, relative bid-ask spread, and Amihud illiquidity. The information proxy is analyst coverage and institutional ownership. For brevity, we only report the coefficient estimates of hedge fund order flow (HF) and non-hedge fund order flow (NHF) averaged over the past five trading days from $d-6$ to $d-2$, 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' 13F data from 1993 to 2022. All variables are the same as defined in Table I. Panels A, B, C are for liquidity proxies and Panels D and E are 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.

	Corporate Events		RavenPack Fundamental News		RavenPack Non-fundamental News	
	Low	High	Low	High	Low	High
Panel A. Market Capitalization						
$\text{Avg}_{d-6,d-2}$(HF)	38.516***	17.056***	13.334***	7.605***	19.991***	9.460***
	(4.77)	(6.99)	(3.91)	(4.96)	(3.06)	(4.11)
$\text{Avg}_{d-6,d-2}$(NHF)	-18.042***	-4.562***	-8.315***	-1.545***	-11.530***	-2.312***
	(-5.17)	(-4.82)	(-4.49)	(-2.79)	(-3.62)	(-2.65)
Adjusted R^2	0.005	0.007	0.005	0.005	0.009	0.005
Observation	516,820	896,640	392,871	787,645	319,719	754,507
Panel B. Bid-Ask Spread						
$\text{Avg}_{d-6,d-2}$(HF)	15.522***	28.275***	6.656***	7.561**	7.942***	15.588***
	(6.63)	(4.71)	(4.63)	(2.49)	(4.70)	(3.21)
$\text{Avg}_{d-6,d-2}$(NHF)	-5.075***	-9.698***	-1.744***	-4.636***	-2.104***	-7.310***
	(-4.89)	(-5.35)	(-3.05)	(-4.06)	(-2.82)	(-3.07)
Adjusted R^2	0.007	0.005	0.004	0.005	0.005	0.007
Observation	846,443	567,017	757,512	423,004	700,953	373,273
Panel C. Amihud Illiquidity						
$\text{Avg}_{d-6,d-2}$(HF)	15.826***	46.416***	6.897***	13.852***	8.400***	29.793***
	(7.24)	(4.25)	(4.90)	(3.49)	(4.48)	(3.35)
$\text{Avg}_{d-6,d-2}$(NHF)	-4.582***	-14.050***	-1.702***	-7.552***	-1.992***	-13.087***
	(-5.65)	(-3.10)	(-3.53)	(-3.38)	(-3.21)	(-2.97)
Adjusted R^2	0.007	0.006	0.003	0.008	0.005	0.013
Observation	904,870	508,586	792,050	388,455	758,429	315,792
Panel D. Analyst Coverage						
$\text{Avg}_{d-6,d-2}$(HF)	19.747***	15.064***	8.401***	6.532***	6.088*	9.059***
	(3.29)	(8.16)	(3.40)	(5.09)	(1.70)	(4.65)
$\text{Avg}_{d-6,d-2}$(NHF)	-11.338***	-3.855***	-5.053***	-1.378***	-5.411***	-1.974***
	(-5.02)	(-7.01)	(-4.12)	(-4.11)	(-2.65)	(-3.50)
Adjusted R^2	0.007	0.005	0.007	0.004	0.009	0.004
Observation	481,622	762,973	397,556	682,086	322,425	646,299
Panel E. Institutional Ownership						
$\text{Avg}_{d-6,d-2}$(HF)	22.487***	17.538***	6.020**	6.556***	3.843	9.901***
	(3.83)	(7.03)	(2.44)	(4.31)	(1.01)	(4.82)
$\text{Avg}_{d-6,d-2}$(NHF)	-9.877***	-4.684***	-3.775***	-1.734***	-4.518*	-2.187***
	(-4.81)	(-7.04)	(-2.83)	(-4.05)	(-1.80)	(-4.49)
Adjusted R^2	0.005	0.006	0.006	0.004	0.009	0.005
Observation	547,287	807,224	441,122	672,338	401,455	621,061

Table VI. Alternative Data and Institutional Trading

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

$$\begin{aligned}\text{Return}_{i,d} \quad = \quad & \alpha_i + \alpha_y + \beta_1\text{TREAT} \times \text{POST} \times \text{HF}_{d-1} + \beta_2\text{TREAT} \times \text{POST} \times \text{NHF}_{d-1} \\ & +\beta_3\text{TREAT} \times \text{POST} \times \text{TOF}_{d-1} + \beta_4\text{TREAT} \times \text{HF}_{d-1} + \beta_5\text{TREAT} \times \text{NHF}_{d-1} \\ & +\beta_6\text{TREAT} \times \text{TOF}_{d-1} + \beta_7\text{POST} \times \text{HF}_{d-1} + \beta_8\text{POST} \times \text{NHF}_{d-1} + \beta_9\text{POST} \times \text{IOF}_{d-1} \\ & +\beta_{10}\text{TREAT} \times \text{POST} + \beta_{11}\text{TREAT} + \beta_{12}\text{POST} + \beta_{13}\text{HF}_{d-1} + \beta_{14}\text{NHF}_{d-1} + \beta_{15}\text{IOF}_{d-1} \\ & +\beta_{16}\text{MktCap}_{d-1} + \beta_{17}\text{Return}_{d-1} + \beta_{18}\text{AMI}_{d-1} + \beta_{19}\text{SPRD}_{d-1} + \beta_{20}\text{NAT}_{q-1} + \epsilon_{i,d},\end{aligned}$$

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 III. 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}	0.342***	0.327***
	(7.62)	(9.20)
TREAT × POST × NHF_{d-1}	-0.542	-0.581
	(-1.48)	(-1.53)
TREAT × POST × TOF_{d-1}	-0.151	-0.189
	(-0.69)	(-0.85)
TREAT × HF_{d-1}	-0.022	-0.065
	(-0.04)	(-0.13)
TREAT × NHF_{d-1}	0.248	0.295
	(1.08)	(1.19)
TREAT × TOF_{d-1}	0.019	0.029
	(0.18)	(0.28)
POST × HF_{d-1}	-0.573	-0.600
	(-0.76)	(-0.78)
POST × NHF_{d-1}	0.412	0.464
	(1.19)	(1.32)
POST × TOF_{d-1}	0.114	0.137
	(0.76)	(0.85)
TREAT × POST	-0.004	-0.003
	(-0.90)	(-0.67)
TREAT	0.003	0.002
	(1.44)	(1.10)
POST	-0.003**	-0.002
	(-2.00)	(-1.29)
HF_{d-1}	0.458	0.513
	(0.68)	(0.77)
NHF_{d-1}	-0.182	-0.227
	(-0.68)	(-0.83)
IOF_{d-1}	0.013	0.006
	(0.21)	(0.09)
MktCap_{d-1}		-0.033***
		(-3.61)
Return_{d-1}		-1.603***
		(-2.97)
AMI_{d-1}		19.064***
		(3.32)
SPRD_{d-1}		0.695
		(0.03)
NAT_{q-1}		0.300**
		(2.38)
Fixed Effects	Yes	Yes
Adjusted R^2	0.004	0.007
Observation	186,956	186,879

Table VII. Return Predictability

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

$$\begin{aligned}\text{Return}_{i,d} \quad = \quad & \alpha_d + \sum_{k=1}^{5} \beta_{d,k}^{HF} \text{HF}_{i,d-k} + \sum_{k=1}^{5} \beta_{d,k}^{NHF} \text{NHF}_{i,d-k} + \sum_{k=1}^{5} \gamma_{d,k}^{T} \text{TOF}_{i,d-k} \\ & + \sum_{k=1}^{5} \gamma_{d,k}^{R} \text{Return}_{i,d-k} + \gamma_d^N \text{NAT}_{i,q-1} + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \epsilon_{i,d}. \quad (10)\end{aligned}$$

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' 13F data from 1993 to 2022. All variables are the same as defined in Tables I and III. 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***	2.887***	1.294***	0.444***
	(17.16)	(15.27)	(11.87)	(4.86)
HF_{d-2}	0.046	0.080	0.196**	-0.139*
	(0.73)	(0.59)	(2.01)	(-1.69)
HF_{d-3}	0.143**	0.373***	0.081	-0.023
	(2.27)	(2.65)	(0.86)	(-0.28)
HF_{d-4}	0.109*	0.292**	0.130	-0.096
	(1.71)	(2.15)	(1.27)	(-1.19)
HF_{d-5}	0.154**	0.127	0.228**	0.108
	(2.42)	(0.90)	(2.46)	(1.28)
NHF_{d-1}	-0.283***	-0.572***	-0.201***	-0.078**
	(-12.25)	(-12.57)	(-7.64)	(-2.14)
NHF_{d-2}	-0.043**	-0.048	-0.125***	0.044
	(-2.16)	(-1.32)	(-4.51)	(1.21)
NHF_{d-3}	-0.041**	-0.037	-0.078***	-0.007
	(-2.08)	(-1.03)	(-2.80)	(-0.19)
NHF_{d-4}	-0.046**	-0.083**	-0.067***	0.010
	(-2.47)	(-2.38)	(-2.64)	(0.29)
NHF_{d-5}	-0.088***	-0.078**	-0.097***	-0.090**
	(-4.62)	(-2.16)	(-3.90)	(-2.50)
Adjusted R^2	0.031	0.022	0.029	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 VIII. Out-of-Sample Return Predictability

This table reproduces the regression analysis from Table VII using out-of-sample hedge fund order flow (OOS HF) and non-hedge fund order flow (OOS NHF). The coefficients of the CRS model for HF and NHF are estimated using a ten-year rolling window, updated annually from 2003 to 2022. For brevity, we only report the coefficient estimates of OOS HF and OOS 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' 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.

	2003-2022	2003-2012	2013-2022
OOS HF_{d-1}	0.798***	0.839***	0.757***
	(9.47)	(5.96)	(8.38)
OOS HF_{d-2}	0.118	-0.044	0.280***
	(1.52)	(-0.33)	(3.75)
OOS HF_{d-3}	0.178**	0.275**	0.080
	(2.29)	(2.04)	(1.06)
OOS HF_{d-4}	0.151*	0.176	0.126*
	(1.93)	(1.28)	(1.70)
OOS HF_{d-5}	0.062	0.020	0.105
	(0.88)	(0.16)	(1.51)
OOS NHF_{d-1}	-0.106***	-0.113***	-0.100***
	(-4.58)	(-4.14)	(-2.73)
OOS NHF_{d-2}	-0.084***	-0.047*	-0.121***
	(-4.02)	(-1.85)	(-3.68)
OOS NHF_{d-3}	-0.078***	-0.077***	-0.079**
	(-3.55)	(-2.89)	(-2.28)
OOS NHF_{d-4}	-0.075***	-0.074***	-0.077**
	(-3.84)	(-3.07)	(-2.48)
OOS NHF_{d-5}	-0.097***	-0.077***	-0.117***
	(-4.93)	(-3.18)	(-3.74)
Adjusted R^2	0.036	0.028	0.043
Number of Stocks	2,608.7	2,600.6	2,616.9
Number of Days	5,035.0	2,517.0	2,518.0

Table IX. Institutional Trading and Price Efficiency around Earnings Announcements

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

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

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 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. $\text{CAR}_{d+1,d+61}$ 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 ($\text{CAR}_{d-21,d}$). 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 the lowest decile) and 9 (for the highest decile). 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. $|\text{OF}|_{d-21,d-1}$ 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 $-q$ ($\text{MktCap}_{d-p,d-q}$), relataive bid-ask spread averaged over days $-p$ to $-q$ ($\text{SPRD}_{d-p,d-q}$), Amihud illiquidity averaged over days $-p$ to $-q$ ($\text{AMI}_{d-p,d-q}$), cumulative abnormal return over days $-p$ to $-q$ ($\text{CReturn}_{d-p,d-q}$), and standard deviation of abnormal returns over days $-p$ to $-q$ ($\text{SReturn}_{d-p,d-q}$). 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 = $\text{CAR}_{d+1,d+61}$		(2) y = Jump Ratio Rank		(3) y = ΔVR	
HF_d	-7.980**	$\|\text{HF}\|_{d-21,d-1}$	-1.042***	$\|\text{HF}\|_d$	-2.626**
	(-2.16)		(-10.16)		(-2.41)
NHF_d	3.183**	$\|\text{NHF}\|_{d-21,d-1}$	-0.654***	$\|\text{NHF}\|_d$	-0.565
	(2.03)		(-13.36)		(-1.17)
TOF_d	-0.490	$\|\text{TOF}\|_{d-21,d-1}$	1.305***	$\|\text{TOF}\|_d$	-0.422**
	(-0.74)		(34.31)		(-2.52)
				$\text{VR}_{d-42,d-22}$	-7.157***
					(-8.79)
NAT_{q-1}	8.048***	NAT_{q-1}	-0.533***	NAT_{q-1}	0.238
	(3.49)		(-2.64)		(0.34)
InstOwn_{d-1}	2.209***	InstOwn_{d-22}	2.752***	InstOwn_{d-22}	0.795***
	(3.98)		(27.50)		(4.15)
NumAnalyst_{d-1}	-0.472**	NumAnalyst_{d-22}	1.017***	NumAnalyst_{d-22}	0.241***
	(-2.16)		(36.24)		(3.48)
$\text{MktCap}_{d-21,d-1}$	-0.110	$\text{MktCap}_{d-42,d-22}$	-0.093***	$\text{MktCap}_{d-42,d-22}$	-0.091*
	(-0.52)		(-4.76)		(-1.68)
$\text{SPRD}_{d-21,d-1}$	-11.726	$\text{SPRD}_{d-42,d-22}$	38.830***	$\text{SPRD}_{d-42,d-22}$	6.982
	(-0.42)		(21.46)		(1.03)
$\text{AMI}_{d-21,d-1}$	0.581	$\text{AMI}_{d-42,d-22}$	0.922***	$\text{AMI}_{d-42,d-22}$	0.630**
	(0.61)		(7.86)		(2.01)
$\text{CReturn}_{d-21,d-1}$	-2.972*	$\text{CReturn}_{d-42,d-22}$	0.185**	$\text{CReturn}_{d-42,d-22}$	0.216
	(-1.74)		(2.44)		(1.26)
$\text{SReturn}_{d-21,d-1}$	30.661***	$\text{SReturn}_{d-42,d-22}$	5.418***	$\text{SReturn}_{d-42,d-22}$	9.729***
	(2.68)		(3.97)		(3.27)
Fixed Effects	Yes	Fixed Effects	Yes	Fixed Effects	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 X. Institutional Trading and Variance Ratio

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

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

where for stock i on day d, ΔVR is a two-day change in variance ratio: $\text{VR}_{i,d} - \text{VR}_{i,d-2}$. 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 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' 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)
$\|\text{HF}\|_{d-1}$	-3.301***	-5.773***	-1.539***	-2.598***
	(-5.71)	(-3.60)	(-3.26)	(-6.08)
$\|\text{NHF}\|_{d-1}$	-0.108	0.658**	-0.300	-0.680***
	(-0.75)	(1.98)	(-1.55)	(-3.50)
$\|\text{TOF}\|_{d-1}$	-0.568***	0.038	-0.258***	-1.482***
	(-8.46)	(0.26)	(-2.96)	(-20.14)
VR_{d-3}	-2.549***	-1.984***	-2.359***	-3.301***
	(-42.55)	(-16.02)	(-28.58)	(-43.05)
$\|\text{NAT}\|_{q-1}$	0.880***	1.904***	0.100	0.638***
	(3.80)	(2.81)	(0.63)	(4.47)
InstOwn_{d-3}	-0.022	0.039	-0.007	-0.096***
	(-0.84)	(0.58)	(-0.26)	(-2.85)
NumAnalyst_{d-3}	-0.170***	-0.061***	-0.097***	-0.352***
	(-13.64)	(-2.68)	(-5.86)	(-16.48)
MktCap_{d-3}	-0.740***	-0.653	-0.526*	-1.042***
	(-3.01)	(-0.94)	(-1.76)	(-6.94)
SPRD_{d-3}	31.401***	-3.073	34.105***	63.025***
	(9.43)	(-1.41)	(5.01)	(10.37)
AMI_{d-3}	0.320	-5.555**	2.622***	3.871***
	(0.35)	(-2.22)	(3.77)	(5.43)
Return_{d-3}	1.718***	-0.772	1.786***	4.128***
	(5.59)	(-1.14)	(4.23)	(10.89)
Intercept	0.993***	0.460***	0.686***	1.829***
	(20.98)	(5.73)	(10.91)	(24.64)
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 at the monthly frequency and explores cross-sectional variation of the predictability based on firm characteristics. Section C compares our proposed hedge fund order flow estimates with daily institutional order flow measures proposed by Campbell, Ramadorai, and Schwartz (2009) and Barardehi, Da, Dixon, and Wang (2025). Section D presents additional robustness tests addressing the limitation that 13F filings do not capture hedge fund short-selling activity, by estimating and evaluating a short-sale adjusted hedge fund trading measure.

A. Estimate daily hedge fund trades

A.1. Difference in hedge fund and non-hedge fund trading

Replicating Figure 1 from the 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:

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

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,

[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').

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:

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

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 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.

The estimated coefficients are highly significant 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. We then plot the implied trade-size coefficients in Figure A2. Following CRS, when computing the implied sensitivities, we set the lagged level of quarterly institutional ownership to its in-sample mean and standardize the net flow coefficients by subtracting their mean and dividing by their standard deviation.

[Place Table A1 about here]

Figure A2 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.

[Place Figure A2 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:

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

where for stock i in a day d, HF and NHF represent daily aggregated order flow from hedge funds and non-hedge funds, respectively, as estimated in Section A. TOF is the daily 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. NAT is the quarterly net arbitrage trading measure proposed by Chen, Da, and Huang (2019), 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 3.130 (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 7.8 basis points (bp) increase in contemporaneous return, whereas NHF's price impact reaches 20.1 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 11.8 bp increase in contemporaneous return, compared to NHF's 24.9 bp. From 2003-2012, HF's impact turns negative, while NHF's impact rises to 27.9

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]

A.4. Hedge fund trading on event days vs. non-event days

In this subsection, we address concerns regarding the validity of hedge fund and non-hedge fund trade flow estimation during event periods. Some may argue that hedge funds could use different execution techniques during event periods, potentially differing from those employed in non-event periods. If this is the case, it could invalidate the estimated coefficients of the CRS model derived from our full sample. To mitigate this concern, we conduct two key tests in this subsection.

First, we examine whether hedge fund trades in the Abel Noser sample behave differently on event days versus non-event days, as illustrated in Figure A3. We replicate Figure 1 from our main paper, categorizing trading days into news and no-news days. A news day is defined as a trading day with corporate events or RavenPack news. Corporate events are specified as in Table III of our main paper, and RavenPack news includes both fundamental and non-fundamental news articles as described in Table IV of our main paper. The analysis reveals no significant shift in the distribution of hedge fund trades between news and no-news days, suggesting that hedge fund trading patterns largely remain consistent across different market conditions.

[Place Figure A3 about here]

Second, we investigate whether the predictive power of total order flow in the 19 size bins differs on event days compared to non-event days, as shown in Table A3. If hedge funds adjust their trade sizes on event days compared to normal days, we would expect to see significant differences in the predictive ability of total order flows across size bins between event and non-event days. To formally test this, we create a dummy variable, NewsDay, which is set to one for news days (those with corporate events or RavenPack news) and zero otherwise. We then estimate the following equation, which includes interactions between the NewsDay dummy and total order flow in the size bins, using Fama-MacBeth (1973) regressions across the full cross-section of stocks in our sample:

$$\begin{aligned}\text{Return}_{i,d} \quad = \quad & \alpha_d + \sum_{Z=1}^{19} \beta_Z^I \text{Avg}_{d-5,d-1}(F_{Z,i,d}) \times \text{NewsDay}_{i,d} + \sum_{Z=1}^{19} \beta_Z^A \text{Avg}_{d-5,d-1}(F_{Z,i,q}) + \beta^N \text{NewsDay}_{i,d} \\ & + \sum_{k=1}^{5} \gamma_{d,k}^R \text{Return}_{i,d-k} + \gamma_q^N \text{NAT}_{i,q-1} + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \epsilon_{i,d},\end{aligned}$$

where for stock i in a day d in a quarter q, $\text{Avg}_{d-5,d-1}(F_Z)$ is the average of F_Z over the past five trading days. For brevity, Table A3 only reports the coefficient estimates for the interaction terms between the 19 total order flows and the news event dummy, while the regressions always include the full set of control variables. We find that 17 out of 19 interaction terms are statistically indifferent from zero and the other two interaction terms are only weakly significant: $\text{Avg}_{d-5,d-1}(F_1)$ has an estimated coefficient of -11.728 (t-stat: -2.18), and $\text{Avg}_{d-5,d-1}(F_{12})$ has an estimated coefficient of 17.004 (t-stat: 1.84). The results suggest that the relationship between trade sizes and returns does not significantly change during event periods.

[**Place Table A3 about here**]

Taken together, these findings suggest that there is no structural shift in hedge fund execution methods on event days in our data. This robustness check reinforces the validity of our hedge fund trade estimation approach, ensuring its consistency across various market conditions, including both event and non-event periods.

B. Hedge funds as informed traders

B.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 B1 reports the estimated coefficients of the following model:

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

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 B1, 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 B1 about here]

B.2. 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 B2, 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 B2 Panel A reports the estimated coefficients of HF and NHF from the regression model in Table VII of our main paper 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 3.069 (t-stat = 17.67) for small stocks and 1.177 (t-stat = 12.52) for large stocks; 1.117 (t-stat = 12.10) for stocks with narrow spreads and 2.326 (t-stat = 15.81) for stocks with wide spreads; 0.762 (t-stat = 8.79) for liquid stocks with low Amihud measures and 4.489 (t-stat = 19.77) 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.

[**Place Table B2 about here**]

In Panel B of Table B2, we examine subsamples of stocks based on the information environment, measured by the number of sell-side analysts and institutional 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 VII of our main paper. 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. For instance, HF has a first-lag coefficient of 2.387 (t-stat = 16.58) for stocks with low analyst coverage and 0.761 (t-stat = 8.82) for stocks with high analyst coverage. 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.

C. Comparison with daily institutional trading measures

Existing literature provides several daily institutional order flow (IOF) measures, notably by Campbell, Ramadorai, and Schwartz (2009) (CRS) and Barardehi, Da, Dixon, and Wang (2025) (BDDW). It is widely recognized that institutional investors possess superior informational advantages, including direct communication channels with firms, stronger relationships with sell-side analysts, and advanced capabilities for financial information processing. Consequently, these alternative IOF measures are expected to capture informed trading activities, similar to our proposed hedge fund order flow estimates. In the main analysis, we explicitly differentiate hedge fund trades from non-hedge fund trades by incorporating both as control variables.[2] This section further complements our main analysis by explicitly comparing our hedge fund estimates with the alternative institutional trading measures proposed by CRS and BDDW.

C.1. Estimating institutional order flow

Campbell, Ramadorai, and Schwartz (2009) construct a daily measure of institutional trading (CRS) by combining data from the Trade and Quotes (TAQ) database and quarterly 13F filings. Their methodology involves estimating daily institutional trading activities from the coefficients derived from a regression of quarterly institutional ownership changes against cumulative trades across different size bins. They apply a Nelson-Siegel functional form to estimate how trade size affects the likelihood that a given trade is institutional. In our analysis, we replicate their approach by using the coefficients presented in Table 4 of their original study to construct a daily institutional order flow measure.

Barardehi, Da, Dixon, and Wang (2025) propose an alternative daily institutional trad-

[2] We also include the net arbitrage trading (NAT) measure of Chen, Da, and Huang (2019) in the main paper to highlight the distinctiveness of our hedge fund order flow estimates compared to general arbitrage activities.

ing measure (BDDW) based on changes in lendable shares. Their approach leverages the fact that institutions can only lend shares they possess, making changes in lendable shares a proxy for institutional ownership adjustments. Specifically, we calculate the daily institutional order flow as the daily change in lendable shares (dLend) divided by the ratio of lendable shares to institutional ownership (Lratio) from the previous quarter-end. Following Barardehi, Da, Dixon, and Wang (2025), to correct for settlement timing differences between equity and security lending markets, we shift the dLend/Lratio measure backward by three days before September 5, 2017, two days before May 28, 2024, and one day thereafter.

C.2. Summary statistics

Panel A of Table C1 presents the time-series averages of cross-sectional statistics for our sample, covering an average of $4,028$ stocks per trading day over $2,581$ trading days from 2007 to 2022.

For comparative purposes, statistics for four versions of institutional order flow (IOF) measures are reported. First, statistics for hedge fund (HF) and non-hedge fund (NHF) order flow estimates are shown. HF exhibits a mean of 0.011 and a standard deviation of 0.032, both notably smaller than those of NHF, which has a mean of 0.022 and a standard deviation of 0.074. The CRS measure is well-balanced, with a mean and median near zero, at -0.002 and -0.001, respectively. BDDW also has a mean close to zero (0.000) but a substantially smaller standard deviation (0.020) compared to CRS (standard deviation = 0.174). Additionally, the TAQ-based order imbalance (TOF) has a mean of 0.042.

Panel B of Table C1 presents the time-series averages of cross-sectional correlations among the measures. Although HF, NHF, and CRS are all derived using TAQ order imbalance data (TOF), their loadings on TOF differ significantly. Specifically, the correlations of HF and NHF with TOF are positive at 0.176 and 0.279, respectively. In contrast, CRS is negatively correlated with TOF (-0.172). Furthermore, correlations of HF and NHF with BDDW are both low at 0.021. These statistics indicate that our proposed institutional order flow measures substantially differ from the alternative IOF estimates.

We also report summary statistics for a short-sale adjusted hedge fund order flow measure, HF−SI, which is estimated similarly to HF and NHF using the CRS model but with the quarterly difference between hedge fund holdings and short interest (SI) as the dependent variable. We assume that all SI activities originate from hedge funds. The resulting daily HF−SI measure has an average of 0.009 and a standard deviation of 0.035, and the correlation between HF and HF−SI is 0.732. The strong correlation and similar distributional properties suggest that adjusting for short interest does not substantially alter the estimated hedge fund trading activity. Further discussion on this adjustment and its implications is provided in Section D, where we address the limitation that 13F holdings do not capture hedge fund shorting activity and conduct additional tests to assess the robustness of our findings.

[Place Table C1 about here]

C.3. Cross-sectional return prediction

We compare the return predictability of daily institutional trading measures using a Fama-MacBeth (1973) regression:

$$\begin{aligned}\text{Return}_{i,d} \quad = \quad & \alpha_d + \beta_d^H \text{Avg}_{i,d-5,d-1}(\text{HF}) + \beta_d^N \text{Avg}_{i,d-5,d-1}(\text{NHF}) + \beta_d^C \text{Avg}_{i,d-5,d-1}(\text{CRS}) \\ & + \beta_d^B \text{Avg}_{i,d-5,d-1}(\text{BDDW}) + \gamma_d^T \text{Avg}_{i,d-5,d-1}(\text{TOF}) + \gamma_d^N \text{NAT}_{i,q-1} \\ & + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \sum_{k=1}^{5} \gamma_{d,k}^R \text{Return}_{i,d-k} + \epsilon_{i,d}.\end{aligned}$$

where for stock i on day d in a quarter q, $\text{Avg}_{d-5,d-1}$ is the average value over the previous five trading days from $d-5$ to $d-1$, NAT is the quarterly net arbitrage trading measure of Chen, Da, and Huang (2019), SPRD is relative bid-ask spread, AMI is Amihud (2002) illiquidity measure, and Return is a risk-adjusted mid-quote return with respect to the Fama−French

(1993) factors and Carhart (1997) momentum factor. To account for serial correlations, we use Newey−West (1987) standard errors with eight lags to calculate the t-statistics.

To fairly assess the predictive performance of HF against CRS and BDDW, we conduct regression analyses that include HF, NHF, and alternative IOF measures simultaneously, reporting the results in the first column. The coefficient of $\text{Avg}_{d-5,d-1}(\text{HF})$ is significantly positive at 0.959 with a t-statistic of 5.59. In contrast, the coefficients for $\text{Avg}_{d-5,d-1}(\text{CRS})$ and $\text{Avg}_{d-5,d-1}(\text{BDDW})$ are not statistically different from zero in the horse race. Further, separate regression analyses reported in the second, third and fourth columns confirm the robustness of HF's predictive ability, maintaining a positive and significant coefficient. Meanwhile, the CRS and BDDW measures consistently fail to achieve statistical significance, mirroring the results from the horse race.

[Place Table C2 about here]

C.4. Corporate event studies

In this subsection, we compare the reaction of hedge fund (HF) and non-hedge fund (NHF) trades with alternative daily institutional trading measures in response to salient corporate events. Consistent with Table III of our main paper, the event sample includes quarterly earnings announcements (EA), analyst rating updates (ARU) representing unscheduled corporate announcements, and permanent price jumps (EPM) unrelated to earnings, reflecting other material information. Due to data availability constraints for the BDDW measure, we limit our analysis to the period from 2007 to 2022.

Table C3 presents the estimated coefficients from the following model, using ordinary least squares regressions with firm and year fixed effects:

$$\begin{aligned}\text{CAR}_{i,t-1,t+1} &= \alpha_i + \alpha_y + \beta^I \text{Avg}_{i,t-6,t-2}(\text{IOF}) + \beta^T \text{Avg}_{i,t-6,t-2}(\text{TOF}) \\ &\quad + \sum_{k=2}^{6} \gamma_k^R \text{Return}_{i,t-k} + \gamma^N \text{NAT}_{i,q-1} + \gamma^S \text{SPRD}_{i,t-2} + \gamma^S \text{AMI}_{i,t-2} + \epsilon_{i,t},\end{aligned}$$

where for each event i on day t in a quarter q of a year y, CAR is the cumulative abnormal return from days $t-1$ to $t+1$, NAT is the quarterly net arbitrage trading measure proposed by Chen, Da, and Huang (2019) (i.e., the difference between quarterly abnormal hedge fund holdings and abnormal short interest), and all other explanatory variables are the same as defined in Tables A2 and C1 with event subscription i instead of firm subscription. We cluster the standard errors around firm and year in calculating the t-statistics. For brevity, we only report the coefficient estimates of daily institutional order flow estimates (IOF), averaged over the previous five trading days from $d-6$ to $d-2$, including hedge fund order flow (HF), non-hedge fund order flow (NHF), institutional trading measures from Campbell, Ramadorai, and Schwartz (2009) (CRS) and Barardehi, Da, Dixon, and Wang (2025) (BDDW) with the full set of control variables in the regressions.

Table C3 summarizes the regression outcomes across event types. The averaged HF order flow consistently exhibits positive and statistically significant coefficients across all event categories. Specifically, $\text{Avg}_{d-6,d-2}$(HF) coefficients are 18.201 (t-stat = 4.23) for earnings announcements, 7.854 (t-stat = 3.19) for analyst rating updates, and 25.531 (t-stat = 5.13) for permanent price jumps. Conversely, $\text{Avg}_{d-6,d-2}$(NHF) consistently demonstrates negative and significant associations with cumulative abnormal returns around these corporate events. The $\text{Avg}_{d-6,d-2}$(CRS) measure exhibits positive and significant coefficients for analyst rating updates and permanent price jumps, although with lower statistical significance compared to HF; for earnings announcements, its coefficient is negative and insignificant. The $\text{Avg}_{d-6,d-2}$(BDDW) measure does not exhibit significant predictive power for any event type, suggesting limited informativeness regarding corporate event returns. These findings remain robust in the last column of horse race, where all institutional trading measures are included simultaneously.

[Place Table C3 about here]

C.5. Institutional trading around news

In this subsection, we compare the reaction of institutional trades to a comprehensive set of firm-specific news stories. Consistent with Table IV in our main paper, the news data are sourced from RavenPack, with the analysis period restricted to 2007–2022 due to the availability constraints of the BDDW measure. To ensure broad coverage and representation of smaller firms, we include all news "groups" within the business "topic" category from RavenPack, specifically those covering more than 5,000 firms (Kolasinski and Yang (2018)).

We replicate the methodology of Table C3, substituting corporate events with news events, to assess the predictive power of various institutional trading measures. The results from these RavenPack news event studies are presented in Tables C4 and C5. Table C4 specifically addresses news related to firm fundamentals, including earnings, analysts, mergers and acquisitions, assets, credit, dividends, equity actions, labor issues, products and services, revenues, and partnerships. Hedge fund trades, averaged over the preceding five trading days (days $d-6$ to $d-2$), consistently show positive and statistically significant predictive power for stock returns across most fundamental news categories, with the exceptions being news related to credit, equity actions, and partnerships. In contrast, non-hedge fund trades averaged over the same period typically exhibit a negative price impact on news-event days across all fundamental categories. Additionally, the estimated coefficients for $\text{Avg}_{d-6,d-2}(\text{CRS})$ and $\text{Avg}_{d-6,d-2}(\text{BDDW})$ are generally insignificant across most news groups, with limited exceptions such as dividends for CRS and labor issues or revenues for BDDW.

Table C5 analyzes non-fundamental news groups, including insider information, investor relations, marketing, order imbalance, price targets, and stock prices. The findings remain qualitatively consistent with fundamental news categories: hedge fund trades yield positive coefficients, whereas other institutional trades generally display negative or statistically insignificant coefficients across the majority of non-fundamental news types.

[Place Table C4 about here]

[Place Table C5 about here]

D. Hedge fund trades adjusted for short interests

The hedge fund trading measure constructed in the main analysis relies on quarterly changes in hedge fund (HF) holdings from 13F filings. However, 13F data do not capture hedge fund shorting activities, potentially introducing measurement errors. To address this limitation, we construct an adjusted hedge fund order flow measure, HF−SI, which subtracts short interest (SI) from HF holdings prior to estimating the daily order flow. Specifically, we assume that all SI activities originate from hedge funds and re-estimate the CRS model using the quarterly HF−SI as the dependent variable. Although this assumption may overstate hedge fund shorting, it allows us to examine whether short interest alters the predictive power of our hedge fund order flow measure.

To construct the short-interest adjustment, we obtain short interest data from the Compustat Short Interest file, which reports monthly short interest for stocks listed on NYSE, AMEX, and NASDAQ. For each stock, we calculate quarterly short interest (SI) as the ratio of shares sold short at the end of the quarter to the total number of shares outstanding. If a stock is not covered in the short interest file, its SI is set to zero. To mitigate data errors or extreme outliers, we remove quarterly SI observations greater than 1. Hedge fund ownership is identified using the methodology of Agarwal, Jiang, Tang, and Yang (2013) and is measured as the ratio of hedge fund-held shares (from Thomson Reuters 13F Ownership data) to total shares outstanding at the end of each quarter. Consistent with the procedure in Section A, we define the dependent variable Y in Equation (1) as the quarterly difference between hedge fund ownership and short interest (HF−SI). We then retrieve the expected daily change in HF−SI order flow by applying the estimated coefficients from Equation (1) to Equation (3) of the main paper.

D.1. Cross-sectional return prediction

We examine the return predictability of HF−SI measures using a Fama-MacBeth (1973) regression:

$$\begin{aligned}\text{Return}_{i,d} &= \alpha_d + \sum_{k=1}^{5} \beta_{d,k}^{HF} \text{HF} - \text{SI}_{i,d-k} + \sum_{k=1}^{5} \beta_{d,k}^{NHF} \text{NHF}_{i,d-k} + \sum_{k=1}^{5} \gamma_{d,k}^{T} \text{TOF}_{i,d-k} \\ &\quad + \sum_{k=1}^{5} \gamma_{d,k}^{R} \text{Return}_{i,d-k} + \gamma_d^N \text{NAT}_{i,q-1} + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \epsilon_{i,d}. \quad (3)\end{aligned}$$

where for stock i on day d in a quarter q, NAT is the quarterly net arbitrage trading measure of Chen, Da, and Huang (2019), SPRD is relative bid-ask spread, AMI is Amihud (2002) illiquidity measure, and Return is a risk-adjusted mid-quote return with respect to the Fama−French (1993) factors and Carhart (1997) momentum factor. To account for serial correlations, we use Newey−West (1987) standard errors with eight lags to calculate the t-statistics.

Table D1 reports the Fama–MacBeth (1973) estimates for HF−SI. Focusing on the first lag, the HF−SI measure exhibits statistically significant predictive power for future returns across all columns, albeit slightly weaker than the HF measure. For example, in the full sample period (1993–2022), the coefficient on the first lag of HF−SI is 1.188 (t-stat = 15.16), compared to 1.540 (t-stat = 17.16) for HF in Table VII of the main paper. In terms of economic significance, the first lag of HF−SI is associated with return predictability of 2.37, 4.36, 2.24, and 0.53 basis points in the periods 1993–2022, 1993–2002, 2003–2012, and 2013–2022, respectively, whereas HF is associated with 3.70, 4.61, 3.07, and 1.61 basis points over the same periods. These results suggest that subtracting short interest does not materially enhance the informativeness of hedge fund trading. Overall, the findings indicate that short-sale adjustments have only a marginal effect on the return predictability of hedge fund order flow.

[Place Table D1 about here]

D.2. Corporate event studies

This subsection examines the response of the short-sale adjusted hedge fund trades (HF−SI) to corporate events. Replicating Table III of the main paper, Table D2 reports the estimated coefficients for the HF−SI measure. Similar to the baseline results, the HF−SI measure significantly predicts CARs, although the magnitude of the coefficients is slightly smaller than that of the HF measure. For example, in the case of earnings announcements over the 1993–2022 period, the coefficient for HF−SI is 6.319 (t-stat = 4.89), while that for HF is 8.093 (t-stat = 5.16). These findings further confirm the robustness of hedge fund trading informativeness, even when accounting for potential short-selling activity.

[Place Table D2 about here]

D.3. HF trading around news

This subsection examines the response of the short-sale adjusted hedge fund trades (HF−SI) to a comprehensive collection of firm-specific news stories from RavenPack between 2000 and 2022.

Tables D3 and D4 replicate the event-study analyses in Table IV of the main paper, evaluating whether HF−SI is associated with future returns around firm-specific news events. Table D3 presents the predictive power for cumulative abnormal returns (CAR) across various RavenPack news groups based on firm fundamentals, while Table D4 focuses on non-fundamental news categories. The HF−SI measure continues to exhibit predictive power across several news categories, particularly those related to firm fundamentals. For example, for earnings-related news, the coefficient for HF−SI_{d-2} is 3.864 (t-stat = 2.87), compared to 5.445 (t-stat = 3.56) for HF_{d-2} in Table IV of the main paper. These findings reinforce the conclusion that omitting short interest does not substantially diminish the informativeness of hedge fund order flow measures.

[**Place Table D3 about here**]

[**Place Table D4 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] Barardehi, Yashar, Da, Zhi, Dixon, Peter, and Wang, Junbo L., 2025, You Can Only Lend What You Own: Inferring Daily Institutional Trading from Security Lending Supply, Available at SSRN: https://ssrn.com/abstract=5025126, pp. 1–30.

[4] Campbell, John, Ramadorai, Tarun, and Schwartz, Allie, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, Journal of Financial Economics **92**, pp. 66–91.

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

[6] 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.

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

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

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

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

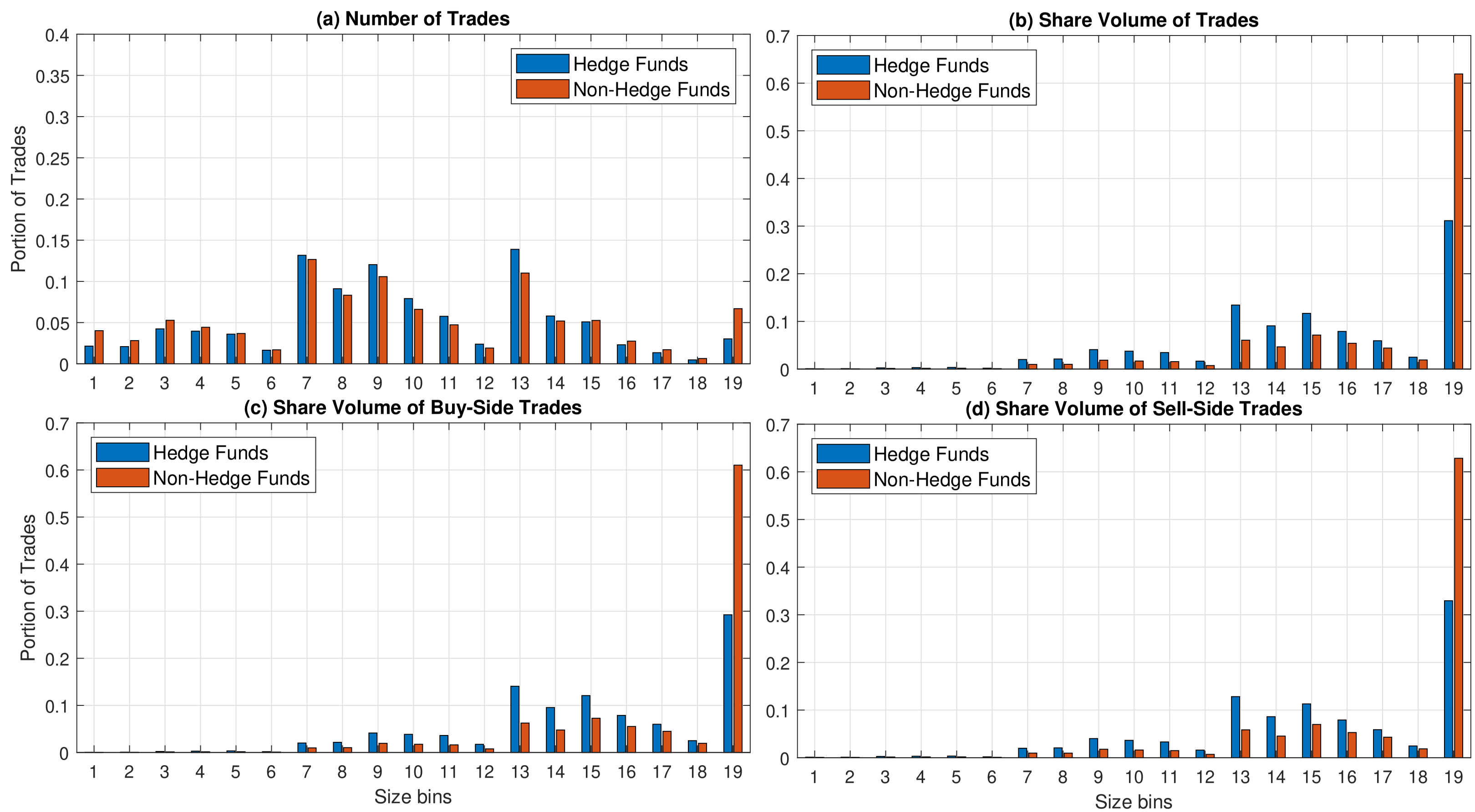
[11] Nelson, C. R. and Siegel, A. F., 1987, Parsimonious modeling of yield curves, Journal of Business **60**, pp. 473–489.

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

[13] 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.

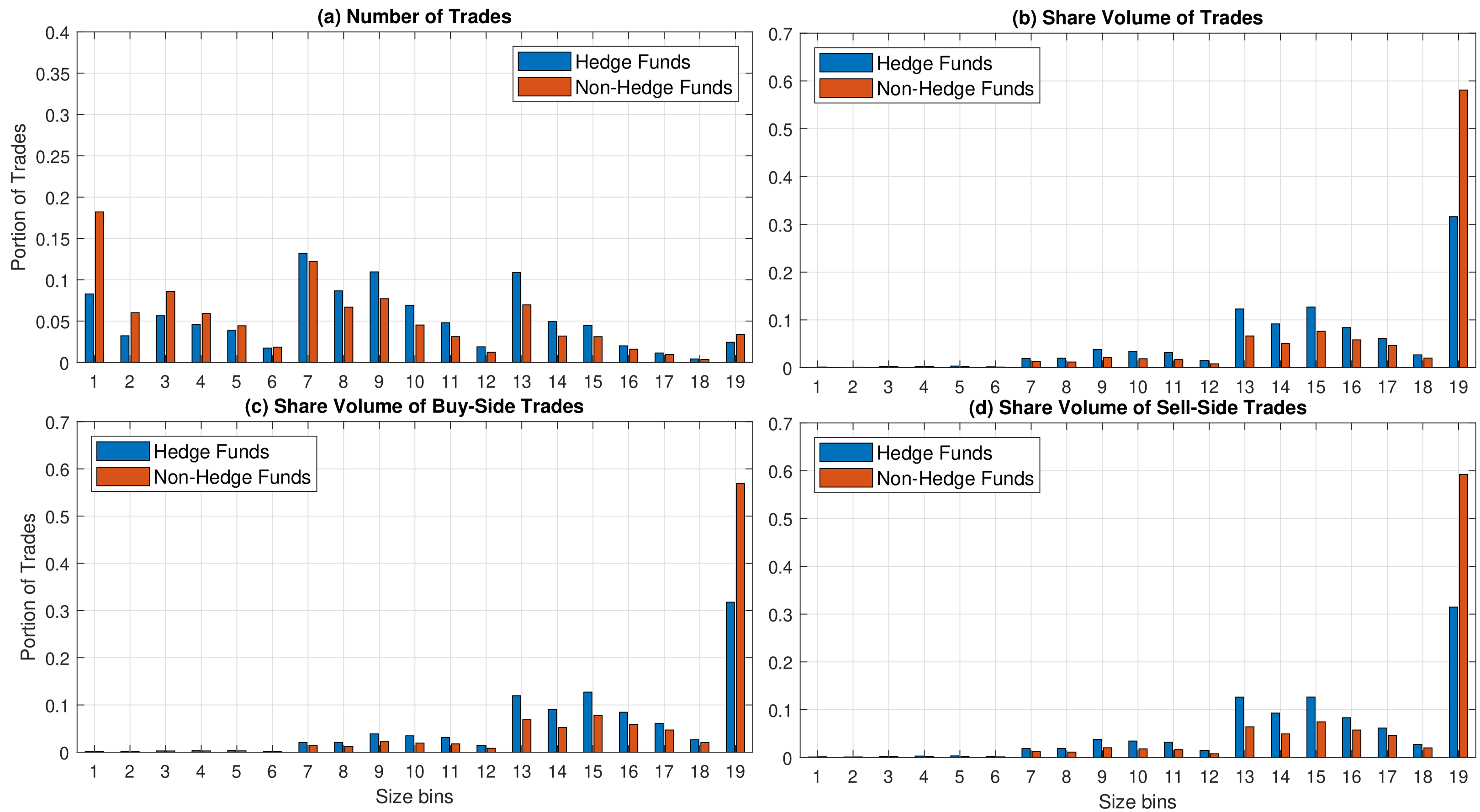
Panel A. 1999-2003



(Continued)

Figure A1 – *Continued*

Panel B. 2004-2007



(Continued)

Figure A1 – *Continued*

Panel C. 2008-2012

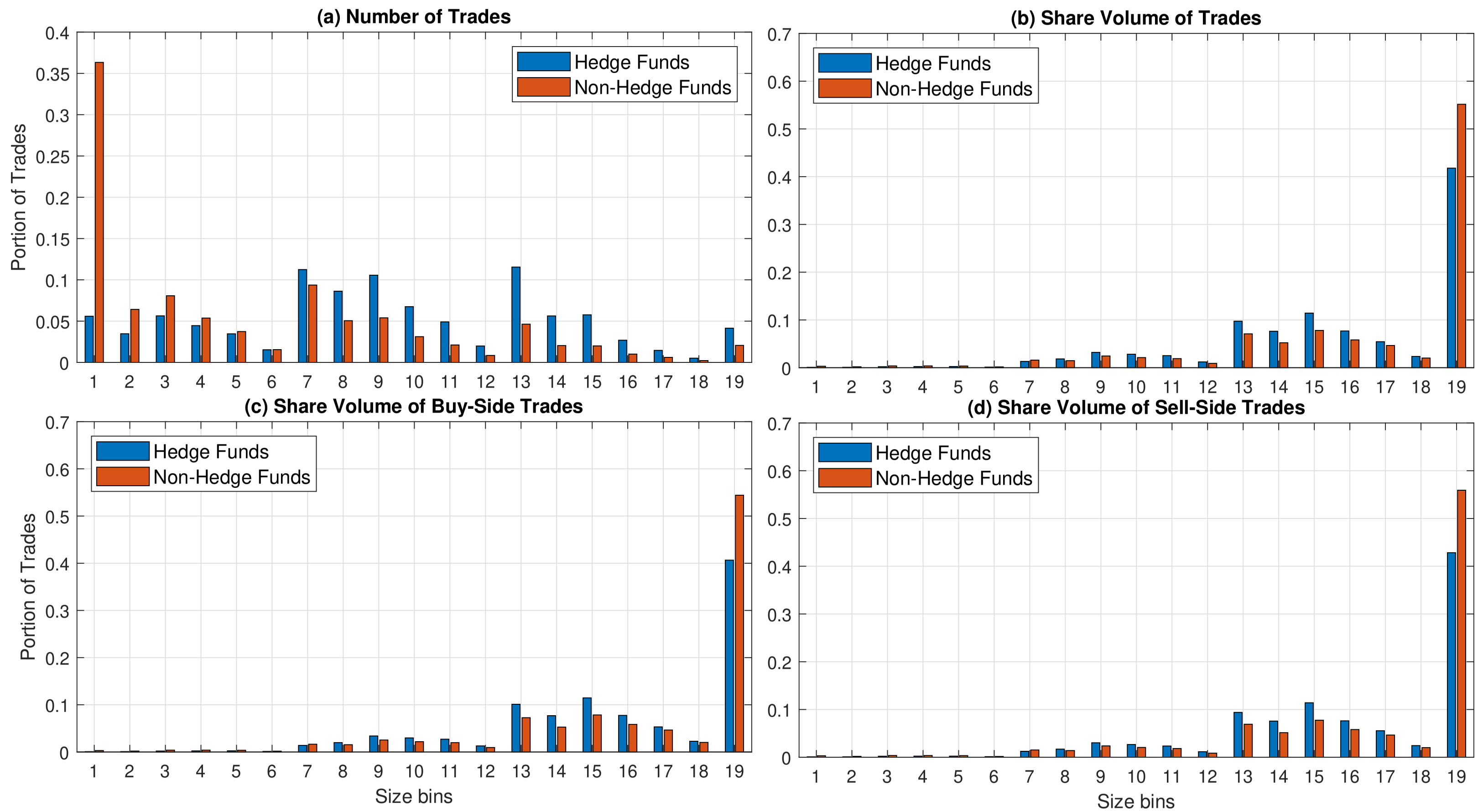


Figure A2. This figure shows the variation in coefficients reflecting the sensitivities of the publicly observable Lee and Ready (1991) order imbalance to hedge fund or non-hedge fund ownership across trade size bins and stock sizes, based on the results in Table A1. The coefficients are standardized 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.

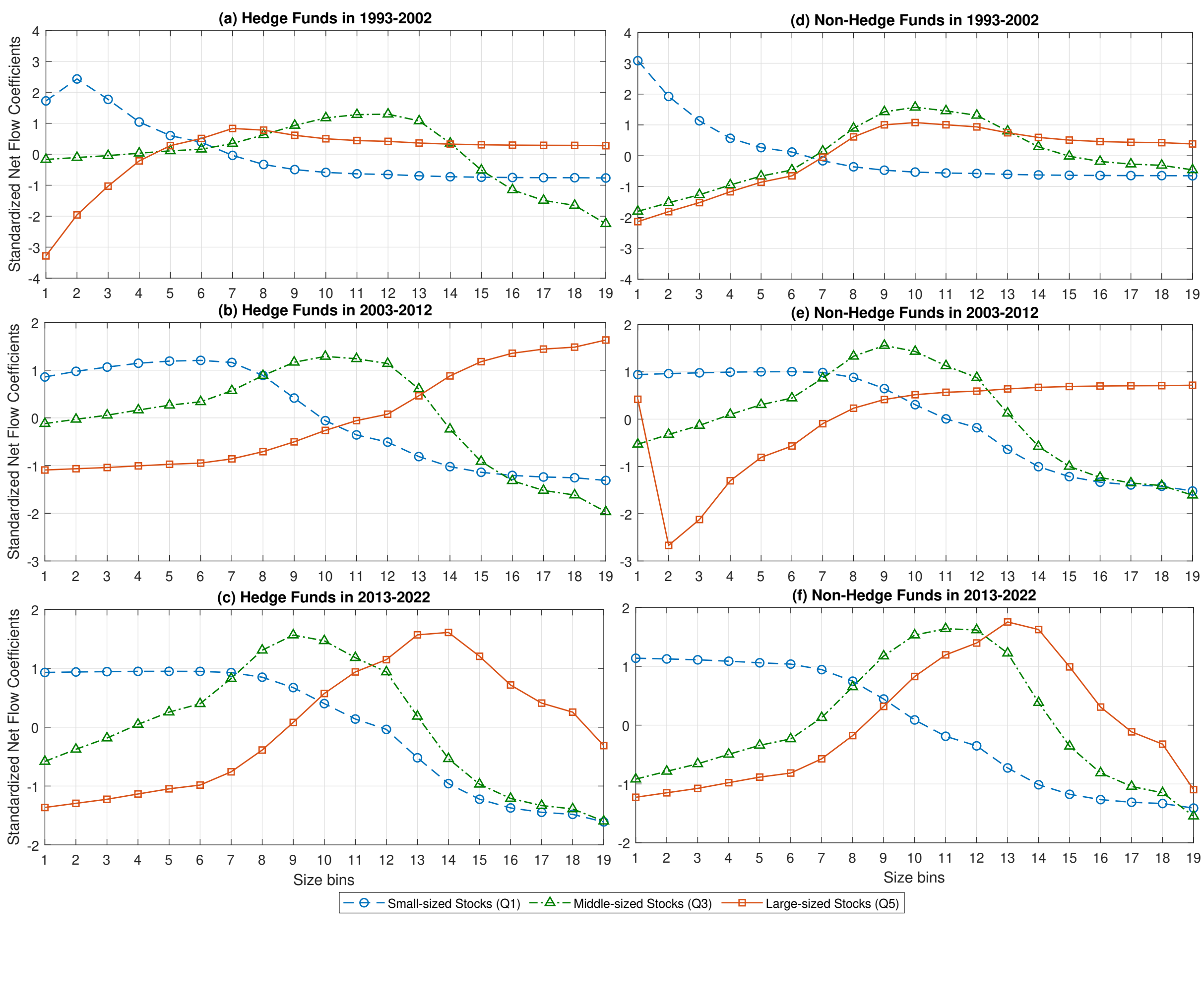


Figure A3. This figure shows the distribution of trading activity in the Abel Noser institutional trade database across trade size bins for news, pre-news, and no-news days separately. News days are defined as trading days from one day before to one day after corporate events or RavenPack news. Pre-news days are trading days from six days before to two days before the events or news, and no-news days are trading days without such events. 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.

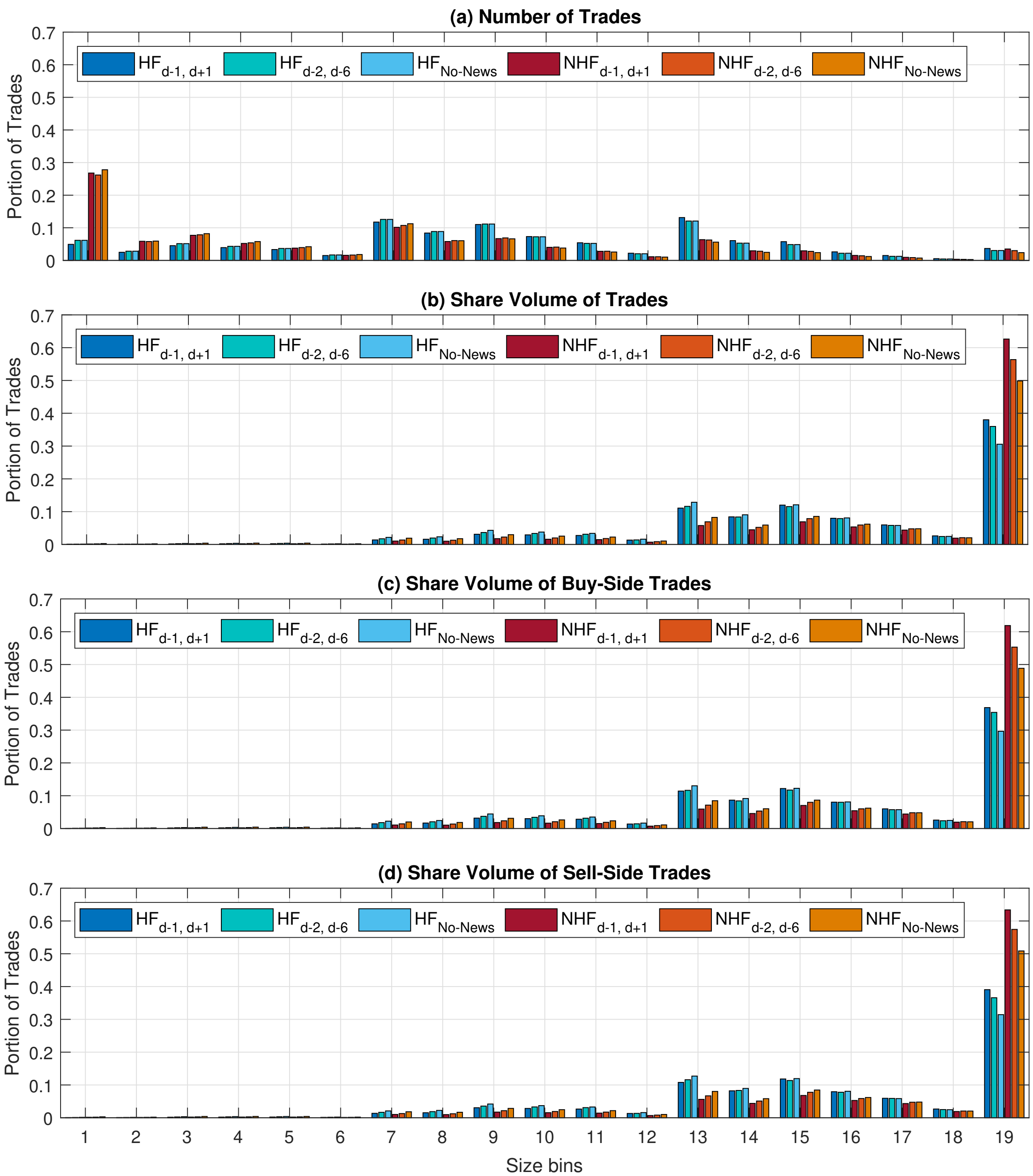


Table A1. Estimated coefficients of Campbell, Ramadorai, and Schwartz (2009) for hedge funds and non-hedge funds

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). The trade size bins are defined as Figure A2. $\beta(Z, Y_{i,q-1}) = b_{01} + b_{02} Y_{i,q-1} + (b_{11} + b_{12} Y_{i,q-1} + b_{21} + b_{22} Y_{i,q-1})[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22} Y_{i,q-1}) e^{-Z/\tau}$. 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. Corresponding t-statistics based on Newey−West (1987) standard errors are reported in parentheses. Superscripts ^{a}, ^{b}, and ^{c} indicate statistical significance at 1, 5, and 10 percent level, respectively.

Panel A. Estimates in 1993-2002

	Hedge Funds					Non-hedge Funds					Difference				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.052^a	-0.048^a	-0.049^a	-0.037^a	-0.041^a	-0.022^a	-0.011^a	-0.012^a	-0.025^a	-0.029^a	-0.030	-0.037	-0.037	-0.012	-0.012
	(-28.34)	(-12.12)	(-8.40)	(-5.22)	(-5.95)	(-21.16)	(-5.15)	(-3.88)	(-6.60)	(-6.53)	(-0.93)	(-0.80)	(-0.66)	(-0.19)	(-0.18)
ρ	0.067^a	0.055^a	0.049^a	0.062^a	0.070^a	-0.031^a	0.021^a	0.014^b	-0.011	-0.156^a	0.098^c	0.034	0.035	0.073	0.226^a
	(10.94)	(5.40)	(3.87)	(4.66)	(5.88)	(-11.97)	(4.22)	(2.29)	(-1.52)	(-21.24)	(1.91)	(0.48)	(0.44)	(0.85)	(2.61)
β^U	0.139^a	0.445^a	0.365^a	0.320^a	0.256^a	0.449^a	1.931^a	2.510^a	1.565^a	1.538^a	-0.310^a	-1.486^a	-2.145^a	-1.245^a	-1.281^a
	(14.09)	(17.91)	(11.11)	(9.73)	(9.82)	(17.81)	(22.85)	(20.69)	(10.58)	(8.43)	(-11.44)	(-16.86)	(-17.06)	(-8.22)	(-6.95)
β^{UY}	-0.060^a	-0.071^a	-0.047^a	-0.048^a	-0.032^a	-0.039^a	-0.053^a	-0.055^a	-0.031^a	-0.026^a	-0.021^a	-0.018^a	0.008	-0.017^a	-0.006
	(-27.30)	(-17.64)	(-9.60)	(-9.80)	(-6.99)	(-32.67)	(-23.88)	(-20.77)	(-10.66)	(-7.51)	(-8.37)	(-3.87)	(1.37)	(-2.97)	(-1.01)
b_{01}	-0.056^a	-0.184^a	-0.134^a	-0.135^a	0.013	-0.017	-0.017	-0.174^b	-0.016	0.360^a	-0.039^a	-0.167^a	0.039	-0.119	-0.347^a
	(-10.82)	(-7.75)	(-4.58)	(-4.05)	(0.93)	(-1.27)	(-0.48)	(-2.11)	(-0.18)	(3.05)	(-2.67)	(-3.90)	(0.45)	(-1.25)	(-2.92)
b_{02}	0.020^a	0.020^a	0.019^a	0.006	0.011^a	0.012^a	0.010^a	0.009^a	0.010^a	0.008^a	0.007^a	0.009^a	0.010^b	-0.004	0.003
	(21.50)	(5.81)	(4.65)	(1.31)	(4.34)	(24.55)	(11.74)	(5.25)	(5.60)	(3.33)	(7.01)	(2.70)	(2.21)	(-0.72)	(0.94)
b_{11}	-0.130	0.116^a	0.172^a	0.045	-1.626^a	0.263	13.598^a	-0.383^b	-3.016^a	-5.939^a	-0.393	-13.482^a	0.555^a	3.060^a	4.312^a
	(-1.18)	(4.61)	(5.15)	(1.24)	(-6.30)	(0.78)	(8.25)	(-2.11)	(-7.08)	(-10.72)	(-1.10)	(-8.17)	(3.00)	(7.16)	(7.05)
b_{12}	0.457^a	-0.039^a	-0.061^a	-0.033^a	0.531^a	0.483^a	-0.010	-0.031^a	0.010	0.078^a	-0.026	-0.029	-0.030^a	-0.043^a	0.452^a
	(16.46)	(-10.44)	(-12.45)	(-6.03)	(11.08)	(24.04)	(-0.25)	(-7.49)	(1.16)	(7.37)	(-0.76)	(-0.69)	(-4.60)	(-4.20)	(9.22)
b_{21}	0.703^a	0.727^a	0.640^a	1.050^a	2.633^a	0.550	-18.689^a	2.458^a	9.759^a	12.914^a	0.152	19.416^a	-1.818^b	-8.709^a	-10.281^a
	(3.95)	(6.88)	(4.00)	(5.47)	(6.07)	(1.05)	(-8.32)	(3.36)	(8.09)	(8.67)	(0.27)	(8.63)	(-2.43)	(-7.13)	(-6.62)
b_{22}	-0.884^a	0.030^c	0.056^b	0.022	-0.979^a	-0.808^a	-0.139^b	0.019	-0.136^a	-0.262^a	-0.076	0.169^a	0.037	0.158^a	-0.717^a
	(-21.00)	(1.89)	(2.42)	(0.78)	(-12.30)	(-27.89)	(-2.45)	(1.16)	(-5.58)	(-9.03)	(-1.48)	(2.87)	(1.31)	(4.23)	(-8.47)
τ	900	87,700	74,100	90,900	4,700	800	1,200	26,500	16,700	18,000					
Adj R^2	0.038	0.054	0.047	0.043	0.067	0.062	0.087	0.072	0.089	0.132					
# of Obs	143,964	37,284	23,916	18,846	15,656	143,964	37,284	23,916	18,846	15,656					

(Continued)

Table A1 – *Continued*

Panel B. Estimates in 2003-2012

	Hedge Funds					Non-hedge Funds					Difference				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.049[a]	-0.060[a]	-0.053[a]	-0.051[a]	-0.040[a]	-0.011[a]	-0.017[a]	-0.019[a]	-0.018[a]	-0.034[a]	-0.038	-0.043	-0.034	-0.033	-0.006
	(-26.96)	(-11.69)	(-7.88)	(-7.11)	(-6.16)	(-11.06)	(-6.24)	(-4.92)	(-3.74)	(-6.67)	(-1.20)	(-0.82)	(-0.54)	(-0.47)	(-0.08)
ρ	0.024[a]	-0.027[a]	0.006	0.013	-0.042[a]	0.013[a]	-0.002	-0.013[a]	-0.037[a]	-0.093[a]	0.011	-0.025	0.019	0.050	0.051
	(4.34)	(-2.66)	(0.52)	(1.04)	(-3.95)	(4.18)	(-0.34)	(-1.68)	(-4.28)	(-10.57)	(0.20)	(-0.32)	(0.21)	(0.53)	(0.54)
β^U	0.853[a]	0.918[a]	0.945[a]	0.767[a]	0.474[a]	0.874[a]	2.346[a]	2.233[a]	2.310[a]	0.763[a]	-0.020	-1.428[a]	-1.287[a]	-1.543[a]	-0.289
	(30.48)	(16.49)	(14.26)	(11.54)	(8.51)	(19.87)	(17.56)	(12.22)	(10.21)	(2.83)	(-0.39)	(-9.87)	(-6.62)	(-6.55)	(-1.05)
β^{UY}	-0.042[a]	-0.036[a]	-0.035[a]	-0.026[a]	-0.023[a]	-0.035[a]	-0.045[a]	-0.039[a]	-0.041[a]	-0.011[b]	-0.007[a]	0.009[a]	0.005	0.016[a]	-0.011[a]
	(-21.21)	(-9.68)	(-7.83)	(-5.52)	(-4.53)	(-33.94)	(-19.33)	(-13.02)	(-11.18)	(-2.54)	(-3.07)	(1.96)	(0.86)	(2.66)	(-1.70)
b_{01}	-0.255[a]	-0.150[a]	-0.156[a]	-0.089[a]	0.115	-0.287[a]	-0.426[a]	-0.254	-1.544[a]	1.010[a]	0.031	0.276[a]	0.098	1.455[a]	-0.895[a]
	(-8.95)	(-2.75)	(-1.86)	(-1.80)	(1.46)	(-5.63)	(-2.86)	(-1.19)	(-4.37)	(7.06)	(0.54)	(1.74)	(0.43)	(4.08)	(-5.47)
b_{02}	0.004[b]	-0.002	0.004	0.005	-0.018[b]	0.010[a]	0.010[a]	0.007[b]	0.027[a]	-0.012[a]	-0.006[a]	-0.013[a]	-0.003	-0.022[a]	-0.006
	(2.19)	(-0.63)	(0.78)	(1.26)	(-2.50)	(8.40)	(4.10)	(2.00)	(4.62)	(-4.88)	(-2.59)	(-2.85)	(-0.41)	(-3.26)	(-0.81)
b_{11}	0.284[a]	0.058	0.130	0.498[a]	-0.221[a]	0.580[a]	0.315	0.427	1.382[a]	11.904[a]	-0.295[a]	-0.257	-0.297	-0.884[a]	-12.125[a]
	(8.47)	(0.66)	(1.33)	(1.81)	(-2.65)	(11.03)	(1.60)	(1.48)	(3.69)	(1.69)	(-4.73)	(-1.20)	(-0.98)	(-1.90)	(-1.72)
b_{12}	-0.016[a]	-0.014[b]	-0.012[a]	-0.014	0.013[a]	-0.014[a]	-0.010[a]	-0.016[a]	-0.028[a]	-0.138	-0.002	-0.004	0.005	0.015	0.151
	(-6.31)	(-2.44)	(-1.76)	(-0.76)	(1.76)	(-10.26)	(-2.79)	(-3.35)	(-4.64)	(-1.15)	(-0.82)	(-0.63)	(0.56)	(0.77)	(1.26)
b_{21}	0.621[a]	1.077[a]	0.539	0.025	0.054	0.774[a]	4.842[a]	3.799[a]	8.169[a]	-24.782[b]	-0.153	-3.765[a]	-3.261[b]	-8.144[a]	24.836[b]
	(4.38)	(3.09)	(1.19)	(0.04)	(0.14)	(3.45)	(5.51)	(3.07)	(4.66)	(-2.44)	(-0.58)	(-3.98)	(-2.47)	(-4.32)	(2.45)
b_{22}	-0.002	0.009	-0.011	-0.062	0.042	0.015[b]	-0.054[a]	-0.022	-0.088[a]	0.267	-0.016	0.062[b]	0.011	0.026	-0.226
	(-0.16)	(0.37)	(-0.35)	(-1.28)	(1.11)	(2.39)	(-3.53)	(-1.05)	(-3.04)	(1.57)	(-1.27)	(2.23)	(0.29)	(0.47)	(-1.29)
τ	11,800	14,100	48,500	4,100	100,000	24,000	22,700	27,300	100,000	800					
Adj R^2	0.036	0.052	0.047	0.047	0.057	0.048	0.076	0.074	0.064	0.069					
# of Obs	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					Difference				
	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large	Small	Q2	Q3	Q4	Large
ϕ	-0.052[a]	-0.048[a]	-0.049[a]	-0.037[a]	-0.041[a]	-0.022[a]	-0.011[a]	-0.012[a]	-0.025[a]	-0.029[a]	-0.030	-0.037	-0.037	-0.012	-0.012
	(-28.34)	(-12.12)	(-8.40)	(-5.22)	(-5.95)	(-21.16)	(-5.15)	(-3.88)	(-6.60)	(-6.53)	(-0.93)	(-0.80)	(-0.66)	(-0.19)	(-0.18)
ρ	0.067[a]	0.055[a]	0.049[a]	0.062[a]	0.070[a]	-0.031[a]	0.021[a]	0.014[b]	-0.011	-0.156[a]	0.098[c]	0.034	0.035	0.073	0.226[a]
	(10.94)	(5.40)	(3.87)	(4.66)	(5.88)	(-11.97)	(4.22)	(2.29)	(-1.52)	(-21.24)	(1.91)	(0.48)	(0.44)	(0.85)	(2.61)
β^U	0.139[a]	0.445[a]	0.365[a]	0.320[a]	0.256[a]	0.449[a]	1.931[a]	2.510[a]	1.565[a]	1.538[a]	-0.310[a]	-1.486[a]	-2.145[a]	-1.245[a]	-1.281[a]
	(14.09)	(17.91)	(11.11)	(9.73)	(9.82)	(17.81)	(22.85)	(20.69)	(10.58)	(8.43)	(-11.44)	(-16.86)	(-17.06)	(-8.22)	(-6.95)
β^{UY}	-0.060[a]	-0.071[a]	-0.047[a]	-0.048[a]	-0.032[a]	-0.039[a]	-0.053[a]	-0.055[a]	-0.031[a]	-0.026[a]	-0.021[a]	-0.018[a]	0.008	-0.017[a]	-0.006
	(-27.30)	(-17.64)	(-9.60)	(-9.80)	(-6.99)	(-32.67)	(-23.88)	(-20.77)	(-10.66)	(-7.51)	(-8.37)	(-3.87)	(1.37)	(-2.97)	(-1.01)
b_{01}	-0.056[a]	-0.184[a]	-0.134[a]	-0.135[a]	0.013	-0.017	-0.017	-0.174[b]	-0.016	0.360[a]	-0.039[a]	-0.167[a]	0.039	-0.119	-0.347[a]
	(-10.82)	(-7.75)	(-4.58)	(-4.05)	(0.93)	(-1.27)	(-0.48)	(-2.11)	(-0.18)	(3.05)	(-2.67)	(-3.90)	(0.45)	(-1.25)	(-2.92)
b_{02}	0.020[a]	0.020[a]	0.019[a]	0.006	0.011[a]	0.012[a]	0.010[a]	0.009[a]	0.010[a]	0.008[a]	0.007[a]	0.009[a]	0.010[b]	-0.004	0.003
	(21.50)	(5.81)	(4.65)	(1.31)	(4.34)	(24.55)	(11.74)	(5.25)	(5.60)	(3.33)	(7.01)	(2.70)	(2.21)	(-0.72)	(0.94)
b_{11}	-0.130	0.116[a]	0.172[a]	0.045	-1.626[a]	0.263	13.598[a]	-0.383[b]	-3.016[a]	-5.939[a]	-0.393	-13.482[a]	0.555[a]	3.060[a]	4.312[a]
	(-1.18)	(4.61)	(5.15)	(1.24)	(-6.30)	(0.78)	(8.25)	(-2.11)	(-7.08)	(-10.72)	(-1.10)	(-8.17)	(3.00)	(7.16)	(7.05)
b_{12}	0.457[a]	-0.039[a]	-0.061[a]	-0.033[a]	0.531[a]	0.483[a]	-0.010	-0.031[a]	0.010	0.078[a]	-0.026	-0.029	-0.030[a]	-0.043[a]	0.452[a]
	(16.46)	(-10.44)	(-12.45)	(-6.03)	(11.08)	(24.04)	(-0.25)	(-7.49)	(1.16)	(7.37)	(-0.76)	(-0.69)	(-4.60)	(-4.20)	(9.22)
b_{21}	0.703[a]	0.727[a]	0.640[a]	1.050[a]	2.633[a]	0.550	-18.689[a]	2.458[a]	9.759[a]	12.914[a]	0.152	19.416[a]	-1.818[b]	-8.709[a]	-10.281[a]
	(3.95)	(6.88)	(4.00)	(5.47)	(6.07)	(1.05)	(-8.32)	(3.36)	(8.09)	(8.67)	(0.27)	(8.63)	(-2.43)	(-7.13)	(-6.62)
b_{22}	-0.884[a]	0.030[c]	0.056[b]	0.022	-0.979[a]	-0.808[a]	-0.139[b]	0.019	-0.136[a]	-0.262[a]	-0.076	0.169[a]	0.037	0.158[a]	-0.717[a]
	(-21.00)	(1.89)	(2.42)	(0.78)	(-12.30)	(-27.89)	(-2.45)	(1.16)	(-5.58)	(-9.03)	(-1.48)	(2.87)	(1.31)	(4.23)	(-8.47)
τ	900	87,700	74,100	90,900	4,700	800	1,200	26,500	16,700	18,000					
Adj R^2	0.038	0.054	0.047	0.043	0.067	0.062	0.087	0.072	0.089	0.132					
# of Obs	143,964	37,284	23,916	18,846	15,656	143,964	37,284	23,916	18,846	15,656					

Table A2. Contemporaneous Price Impact

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

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

where for stock i in a day d in a quarter q, TOF is total order flow, NAT is net arbitrage trading measure proposed by Chen, Da, and Huang (2019), SPRD is daily relative spread, AMI is Amihud (2002) illiquidity, and all the other variables are the same as defined in Table I of our main paper. 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 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	3.130***	7.885***	-1.595***	3.115***
	(10.46)	(12.20)	(-6.14)	(15.71)
NHF_d	2.795***	3.919***	3.526***	0.944***
	(33.02)	(42.98)	(27.07)	(9.10)
Adjusted R^2	0.043	0.033	0.048	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 of Total Order Flow in 19 Trade Size Bins

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

$$\begin{aligned}\text{Return}_{i,d} &= \alpha_d + \sum_{Z=1}^{19} \beta_Z^I \text{Avg}_{d-5,d-1}(F_{Z,i,d}) \times \text{NewsDay}_{i,d} + \sum_{Z=1}^{19} \beta_Z^A \text{Avg}_{d-5,d-1}(F_{Z,i,q}) + \beta^N \text{NewsDay}_{i,d} \\ &+ \sum_{k=1}^{5} \gamma_{d,k}^R \text{Return}_{i,d-k} + \gamma_q^N \text{NAT}_{i,q-1} + \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \epsilon_{i,d},\end{aligned}$$

where for stock i in a day d in a quarter q, $\text{Avg}_{d-5,d-1}(F_Z)$ is the average of F_Z over the past five trading days and NewsDay is a dummy variable equal to one for a news day with corporate events or RavenPack news, and zero otherwise. For brevity, we only report the coefficient estimates of interaction terms for 19 total order flows and a news event dummy, 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 III of our main paper. 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.

	Corporate Event Sample	RavenPack News Sample
$\text{Avg}_{d-5,d-1}(\text{F}_{01}) \times \text{NewsDay}_d$	-11.728**	-48.769
	(-2.18)	(-0.96)
$\text{Avg}_{d-5,d-1}(\text{F}_{02}) \times \text{NewsDay}_d$	-3.274	-6.852
	(-0.19)	(-0.42)
$\text{Avg}_{d-5,d-1}(\text{F}_{03}) \times \text{NewsDay}_d$	64.537	4.603
	(0.96)	(0.28)
$\text{Avg}_{d-5,d-1}(\text{F}_{04}) \times \text{NewsDay}_d$	74.173	31.906
	(1.33)	(1.00)
$\text{Avg}_{d-5,d-1}(\text{F}_{05}) \times \text{NewsDay}_d$	-78.117	-7.612*
	(-0.94)	(-1.82)
$\text{Avg}_{d-5,d-1}(\text{F}_{06}) \times \text{NewsDay}_d$	-97.663	-7.107
	(-1.18)	(-1.06)
$\text{Avg}_{d-5,d-1}(\text{F}_{07}) \times \text{NewsDay}_d$	7.673	2.679
	(0.48)	(1.08)
$\text{Avg}_{d-5,d-1}(\text{F}_{08}) \times \text{NewsDay}_d$	9.329	-0.136
	(1.35)	(-0.21)
$\text{Avg}_{d-5,d-1}(\text{F}_{09}) \times \text{NewsDay}_d$	3.601	-0.685
	(0.50)	(-1.13)
$\text{Avg}_{d-5,d-1}(\text{F}_{10}) \times \text{NewsDay}_d$	14.710	-1.032
	(0.79)	(-1.40)
$\text{Avg}_{d-5,d-1}(\text{F}_{11}) \times \text{NewsDay}_d$	-16.899	0.665
	(-1.31)	(0.79)
$\text{Avg}_{d-5,d-1}(\text{F}_{12}) \times \text{NewsDay}_d$	17.004*	1.159
	(1.84)	(0.74)
$\text{Avg}_{d-5,d-1}(\text{F}_{13}) \times \text{NewsDay}_d$	4.791	-0.427
	(0.75)	(-0.93)
$\text{Avg}_{d-5,d-1}(\text{F}_{14}) \times \text{NewsDay}_d$	0.039	0.385
	(0.00)	(0.69)
$\text{Avg}_{d-5,d-1}(\text{F}_{15}) \times \text{NewsDay}_d$	-3.976	0.047
	(-0.69)	(0.10)
$\text{Avg}_{d-5,d-1}(\text{F}_{16}) \times \text{NewsDay}_d$	-6.965	-0.049
	(-0.65)	(-0.08)
$\text{Avg}_{d-5,d-1}(\text{F}_{17}) \times \text{NewsDay}_d$	15.764	-0.614
	(0.75)	(-0.87)
$\text{Avg}_{d-5,d-1}(\text{F}_{18}) \times \text{NewsDay}_d$	170.853	-3.233
	(0.82)	(-0.83)
$\text{Avg}_{d-5,d-1}(\text{F}_{19}) \times \text{NewsDay}_d$	-0.688	-0.018
	(-1.05)	(-0.12)
Adjusted R^2	0.077	0.044
Number of Stocks	2,952.4	2,952.4
Number of Days	7,544.0	7,544.0

Table B1. Return Predictability at Monthly Frequency

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

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

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 Table II of the main paper. 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

Table B2. Return Predictability Conditioning on Firm Characteristics

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 VII of our main paper. 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 of our main paper. 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}	3.069***	1.177***	1.117***	2.326***	0.762***	4.489***
	(17.67)	(12.52)	(12.10)	(15.81)	(8.79)	(19.77)
HF_{d-2}	0.037	0.236***	0.102	-0.049	-0.030	0.468***
	(0.30)	(3.45)	(1.37)	(-0.43)	(-0.45)	(3.20)
HF_{d-3}	0.314**	0.137*	0.173**	0.048	-0.011	0.783***
	(2.32)	(1.94)	(2.46)	(0.44)	(-0.17)	(5.44)
HF_{d-4}	0.080	0.204***	0.090	0.085	0.043	0.405***
	(0.67)	(2.71)	(1.30)	(0.76)	(0.61)	(2.95)
HF_{d-5}	0.240**	0.221***	0.110	0.291***	-0.002	0.597***
	(1.98)	(3.01)	(1.52)	(2.73)	(-0.03)	(4.16)
NHF_{d-1}	-0.490***	-0.179***	-0.192***	-0.417***	-0.134***	-0.367***
	(-8.70)	(-7.43)	(-8.01)	(-9.96)	(-5.75)	(-5.79)
NHF_{d-2}	-0.199***	-0.042**	-0.031	-0.077**	0.020	-0.087
	(-4.17)	(-2.03)	(-1.43)	(-2.07)	(0.97)	(-1.62)
NHF_{d-3}	-0.282***	-0.019	-0.027	-0.074*	0.027	-0.166***
	(-5.81)	(-0.95)	(-1.33)	(-1.94)	(1.40)	(-3.12)
NHF_{d-4}	-0.205***	-0.038*	-0.004	-0.148***	0.002	-0.042
	(-4.63)	(-1.95)	(-0.21)	(-4.02)	(0.09)	(-0.85)
NHF_{d-5}	-0.358***	-0.031	-0.047**	-0.193***	0.005	-0.190***
	(-8.01)	(-1.53)	(-2.27)	(-5.34)	(0.24)	(-3.77)
Adjusted R^2	0.034	0.047	0.042	0.035	0.047	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 B2 – *Continued*

Panel B. Information environment	Analyst Coverage		Ownership	
	Low	High	Low	High
HF_{d-1}	2.387***	0.761***	2.476***	1.068***
	(16.58)	(8.82)	(15.14)	(13.12)
HF_{d-2}	0.050	0.053	-0.128	0.178***
	(0.44)	(0.69)	(-1.05)	(2.67)
HF_{d-3}	0.112	0.137*	0.301**	0.097
	(1.02)	(1.75)	(2.54)	(1.42)
HF_{d-4}	0.348***	0.015	0.052	0.141**
	(3.28)	(0.20)	(0.45)	(2.07)
HF_{d-5}	0.263**	0.109	0.139	0.174**
	(2.48)	(1.39)	(1.21)	(2.51)
NHF_{d-1}	-0.363***	-0.109***	-0.345***	-0.186***
	(-8.92)	(-4.52)	(-7.91)	(-8.03)
NHF_{d-2}	-0.056	0.016	-0.067*	-0.048**
	(-1.40)	(0.71)	(-1.74)	(-2.28)
NHF_{d-3}	-0.070*	-0.006	-0.137***	-0.000
	(-1.87)	(-0.29)	(-3.36)	(-0.01)
NHF_{d-4}	-0.081**	-0.005	-0.068*	-0.033*
	(-2.28)	(-0.24)	(-1.92)	(-1.67)
NHF_{d-5}	-0.151***	-0.018	-0.163***	-0.050***
	(-4.10)	(-0.80)	(-4.30)	(-2.59)
Adjusted R^2	0.033	0.046	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 C1. Summary Statistics

This table shows the time-series averages of cross-sectional statistics for the sample from 2007 to 2022. HF and NHF represent daily hedge fund and non-hedge fund order flows, respectively, estimated following Campbell, Ramadorai, and Schwartz (2009) by taking the estimated coefficients in Table A1. HF−SI is a short-sale adjusted hedge fund trading measure, estimated in the same manner as HF and NHF, using the quarterly difference between hedge fund holdings and short interest (SI) as the dependent variable in the model of Campbell, Ramadorai, and Schwartz (2009). CRS is a daily institutional trading measure proposed by Campbell, Ramadorai, and Schwartz (2009), using the estimated coefficients in Table 4 of their paper. BDDW is a daily institutional trading proxy proposed by Barardehi, Da, Dixon, and Wang (2025). TOF is daily total order flow in TAQ based on the Lee and Ready (1991) algorithm. Return is the daily risk-adjusted mid-quote stock return with respect to the Fama−French−Carhart four factors. The sample includes all common stocks listed on the NYSE, AMEX, and Nasdaq with market capitalization above the 5% NYSE breakpoints that have information in TAQ, CRSP, and Thomson Reuters's 13F data. Panel A reports descriptive statistics, and Panel B reports correlation coefficients among key variables.

Panel A. Descriptive statistics

	Average	Standard Deviation	Minimum	Median	Maximum
HF	0.011	0.032	-0.090	0.004	0.171
HF−SI	0.009	0.035	-0.113	0.002	0.164
NHF	0.022	0.074	-0.205	0.007	0.363
CRS	-0.002	0.174	-0.600	-0.001	0.572
BDDW	0.000	0.020	-0.150	0.000	0.174
TOF	0.042	0.154	-0.596	0.022	0.810
Return	0.000	0.026	-0.236	0.000	0.445

Panel B. Correlation

	HF	HF−SI	NHF	CRS	BDDW	TOF	Return
HF	1.000						
HF−SI	0.732	1.000					
NHF	0.754	0.664	1.000				
CRS	-0.080	-0.042	-0.118	1.000			
BDDW	0.021	0.009	0.021	-0.016	1.000		
TOF	0.176	0.174	0.279	-0.172	0.038	1.000	
Return	0.064	0.046	0.074	-0.045	0.054	0.124	1.000

Table C2. Return Predictability

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

$$\begin{aligned}\text{Return}_{i,d} &= \alpha_d + \beta_d^H \text{Avg}_{i,d-5,d-1}(\text{HF}) + \beta_d^N \text{Avg}_{i,d-5,d-1}(\text{NHF}) + \beta_d^C \text{Avg}_{i,d-5,d-1}(\text{CRS}) \\ &+ \beta_d^B \text{Avg}_{i,d-5,d-1}(\text{BDDW}) + \gamma_d^T \text{Avg}_{i,d-5,d-1}(\text{TOF}) + \gamma_d^N \text{NAT}_{i,q-1} \\ &+ \gamma_d^B \text{SPRD}_{i,d-1} + \gamma_d^A \text{AMI}_{i,d-1} + \sum_{k=1}^{5} \gamma_{d,k}^R \text{Return}_{i,d-k} + \epsilon_{i,d}.\end{aligned}$$

We report the coefficient estimates of trading measures, averaged over the past five trading days from $d-5$ to $d-1$, including hedge fund order flow (HF), non-hedge fund order flow (NHF), and institutional trading measures from Campbell, Ramadorai, and Schwartz (2009) (CRS) and Barardehi, Da, Dixon, and Wang (2025) (BDDW). All variables are the same as defined in Tables A2 and C1. 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 2007 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.

	Horse Race	HF	CRS	BDDW
$\text{Avg}_{d-5,d-1}(\text{HF})$	0.959***	0.919***		
	(5.59)	(5.58)		
$\text{Avg}_{d-5,d-1}(\text{NHF})$	-0.295***	-0.303***		
	(-4.96)	(-5.10)		
$\text{Avg}_{d-5,d-1}(\text{CRS})$	-0.008		-0.014	
	(-0.66)		(-1.29)	
$\text{Avg}_{d-5,d-1}(\text{BDDW})$	0.485			0.243
	(1.19)			(0.63)
$\text{Avg}_{d-5,d-1}(\text{TOF})$	0.064***	0.064***	0.055**	0.061***
	(2.82)	(2.82)	(2.19)	(2.62)
NAT_{q-1}	0.066**	0.078**	0.086***	0.064*
	(1.99)	(2.42)	(2.64)	(1.95)
SPRD_{d-1}	8.097***	10.051***	8.391***	8.385***
	(10.21)	(10.80)	(10.22)	(10.11)
AMI_{d-1}	0.137***	0.014**	0.077**	0.032***
	(3.07)	(2.39)	(2.23)	(2.80)
Return_{d-1}	-0.632***	-0.261	-0.306*	-0.635***
	(-3.86)	(-1.55)	(-1.80)	(-3.89)
Return_{d-2}	-0.231*	-0.168	-0.134	-0.240*
	(-1.84)	(-1.36)	(-1.09)	(-1.86)
Return_{d-3}	-0.342***	-0.303**	-0.301**	-0.393***
	(-2.99)	(-2.45)	(-2.50)	(-3.33)
Return_{d-4}	-0.382***	-0.323***	-0.306**	-0.381***
	(-3.30)	(-2.70)	(-2.52)	(-3.27)
Return_{d-5}	-0.266**	-0.233*	-0.235**	-0.365***
	(-2.38)	(-1.74)	(-2.02)	(-2.65)
Intercept	-0.014***	-0.014***	-0.008**	-0.009***
	(-4.34)	(-4.41)	(-2.55)	(-3.21)
Adjusted R^2	0.033	0.035	0.032	0.033
Number of Stocks	2,030.9	2,577.8	2,360.9	2,102.1
Number of Days	4,027.0	4,028.0	4,028.0	4,027.0

Table C3. Predicting CAR around Corporate Events

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

$$\text{CAR}_{i,t-1,t+1} = \alpha_i + \alpha_y + \beta^I \text{Avg}_{i,t-6,t-2}(\text{IOF}) + \beta^T \text{Avg}_{i,t-6,t-2}(\text{TOF}) + \sum_{k=2}^{6} \gamma_k^R \text{Return}_{i,t-k} + \gamma^N \text{NAT}_{i,q-1} + \gamma^S \text{SPRD}_{i,t-2} + \gamma^S \text{AMI}_{i,t-2} + \epsilon_{i,t},$$

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 A2 and C1 with event subscription i instead of firm subscription. For brevity, we only report the coefficient estimates of daily institutional order flow estimates (IOF), averaged over the previous five trading days from $d-6$ to $d-2$, including hedge fund order flow (HF), non-hedge fund order flow (NHF), institutional trading measures from Campbell, Ramadorai, and Schwartz (2009) (CRS) and Barardehi, Da, Dixon, and Wang (2025) (BDDW) 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' 13F data from 2007 to 2022. The corporate events we study are quarterly earnings announcements (EA) from I/B/E/S, analysts rating updates (ARU) from I/B/E/S, and extreme price movement (EPM) 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. ^{a}, ^{b}, and ^{c} indicate statistical significance at the 1, 5, and 10 percent level, respectively.

	HF			CRS			BDDW			Horse Race		
	EA	ARU	EPM	EA	ARU	EPM	EA	ARU	EPM	EA	ARU	EPM
$\text{Avg}_{d-6,d-2}$(HF)	18.201^a	7.854^a	25.531^a							19.350^a	4.099	26.803^a
	(4.23)	(3.19)	(5.13)							(4.74)	(1.37)	(5.14)
$\text{Avg}_{d-6,d-2}$(NHF)	-3.272^b	-3.097^b	-7.586^a							-3.418^b	-1.139	-6.516^a
	(-2.25)	(-2.47)	(-4.15)							(-2.24)	(-1.01)	(-3.86)
$\text{Avg}_{d-6,d-2}$(CRS)				-0.022	0.354^b	0.509^a				0.08	0.25	0.562^a
				(-0.08)	(2.02)	(3.12)				(0.21)	(1.10)	(3.60)
$\text{Avg}_{d-6,d-2}$(BDDW)							4.73	5.329	0.057	4.733	8.078	0.665
							(1.04)	(0.91)	(0.01)	(1.12)	(1.31)	(0.15)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.002	0.005	0.005	0.002	0.005	0.004	0.002	0.005	0.004	0.002	0.005	0.005
Observations	150,606	244,398	254,583	138,006	222,160	232,595	123,718	194,495	203,948	119,676	188,908	197,175

Table C4. Predicting CAR around RavenPack Fundamental News

This table reproduces Table C3 to examine the predictive ability of daily institutional 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 2007-2022. 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 daily institutional order flow estimates, averaged over the past five trading days from $d-6$ to $d-2$, with the full set of control variables in the regressions. 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' 13F data. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on firm and year clustered standard errors are reported in parentheses. ^{a}, ^{b}, and ^{c} indicate statistical significance at 1, 5, and 10 percent level, respectively.

	Earnings	Analysts	M&A	Assets	Credit	Dividends	Equity Actions	Labor Issues	Products Services	Revenues	Partnership
	Panel A. HF										
$\text{Avg}_{d-6,d-2}$(HF)	12.237^a	17.692^a	26.009^a	15.919^b	7.154	8.481^b	13.676	9.541^b	8.607^b	20.340^a	3.572
	(3.09)	(3.34)	(4.89)	(2.08)	(0.98)	(2.39)	(0.99)	(2.11)	(2.26)	(5.58)	(0.87)
$\text{Avg}_{d-6,d-2}$(NHF)	-2.810^b	-3.801	-2.163	-5.543	-1.847	-1.625	-5.002^b	-3.077	-0.721	-5.236^a	-0.716
	(-2.36)	(-1.59)	(-1.00)	(-1.58)	(-0.59)	(-1.33)	(-2.05)	(-1.62)	(-0.42)	(-3.68)	(-0.44)
Adjusted R^2	0.003	0.005	0.019	0.013	0.007	0.003	0.045	0.005	0.004	0.002	0.010
Observations	181,610	92,978	43,549	15,901	56,891	57,528	58,987	75,405	68,898	137,521	24,178
	Panel B. CRS										
$\text{Avg}_{d-6,d-2}$(CRS)	-0.169	0.464	-0.244	0.929	0.327	1.060^a	0.584	0.277	0.323	0.381	-0.044
	(-0.77)	(0.78)	(-0.38)	(1.61)	(0.72)	(4.88)	(1.22)	(0.75)	(1.15)	(1.30)	(-0.12)
Adjusted R^2	0.003	0.004	0.014	0.008	0.004	0.003	0.046	0.004	0.005	0.002	0.011
Observations	167,277	85,959	39,655	14,761	52,535	54,532	53,819	69,541	64,054	125,703	22,316
	Panel C. BDDW										
$\text{Avg}_{d-6,d-2}$(BDDW)	4.412	-11.354	-52.070	6.348	16.358	-4.420	22.928	6.852^a	6.203	20.490^a	-6.817
	(0.71)	(-1.30)	(-1.39)	(0.57)	(0.46)	(-0.67)	(1.24)	(3.20)	(0.64)	(2.69)	(-1.29)
Adjusted R^2	0.003	0.003	0.020	0.014	0.024	0.003	0.059	0.003	0.006	0.002	0.014
Observations	149,023	76,339	35,254	132,052	46,991	49,913	46,951	61,378	55,595	112,862	20,284
	Panel D. Horse Race										
$\text{Avg}_{d-6,d-2}$(HF)	12.101^a	18.270^a	23.332^a	17.904^b	7.040	9.070^b	16.167	7.844^b	7.897^c	21.676^a	4.741
	(4.08)	(3.13)	(3.76)	(2.33)	(1.16)	(2.01)	(1.55)	(2.50)	(1.73)	(7.04)	(1.01)
$\text{Avg}_{d-6,d-2}$(NHF)	-2.149^c	-3.346	-1.718	-5.762^c	-0.604	-1.894	-0.693	-1.998	-0.238	-5.733^a	-1.175
	(-1.73)	(-1.44)	(-0.60)	(-1.69)	(-0.22)	(-1.22)	(-0.27)	(-1.40)	(-0.12)	(-4.16)	(-0.53)
$\text{Avg}_{d-6,d-2}$(CRS)	-0.155	0.770	0.100	0.453	0.338	1.209^a	0.189	-0.077	0.430	0.458	0.006
	(-0.72)	(1.09)	(0.13)	(0.73)	(0.94)	(4.77)	(0.28)	(-0.26)	(1.42)	(1.57)	(0.02)
$\text{Avg}_{d-6,d-2}$(BDDW)	4.882	-11.725	-45.336	5.403	21.834	-3.509	31.978	6.752^a	5.475	22.007^a	-5.868
	(0.81)	(-1.56)	(-1.20)	(0.52)	(0.57)	(-0.54)	(1.44)	(2.93)	(0.58)	(3.04)	(-1.14)
Adjusted R^2	0.003	0.004	0.018	0.007	0.004	0.004	0.059	0.003	0.006	0.003	0.016
Observations	144,618	74,560	33,300	12,798	45,341	48,692	45,069	59,632	54,001	109,398	19,510

Table C5. Predicting CAR around RavenPack Non-Fundamental News

This table reproduces Table C3 to examine the predictive ability of daily institutional 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 2007-2022. News groups of exploration, indexes, industrial-accidents, regulatory, stock-picks are excluded. For brevity, we only report the coefficient estimates of daily institutional order flow estimates, averaged over the past five trading days from $d-6$ to $d-2$, with the full set of control variables in the regressions. 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' 13F data. All coefficient estimates are multiplied by 100. Corresponding t-statistics based on firm and year clustered standard errors are reported in parentheses. ^{a}, ^{b}, and ^{c} indicate statistical significance at 1, 5, and 10 percent level, respectively.

	Insider	Investor Relation	Marketing	Order Imbalance	Price Target	Stock Price
	Panel A. HF					
$\text{Avg}_{d-6,d-2}$(HF)	5.059^a	6.239^b	6.000^c	1.794	8.013^a	4.803
	(2.92)	(2.07)	(1.93)	(0.35)	(3.02)	(1.19)
$\text{Avg}_{d-6,d-2}$(NHF)	-1.268^a	-1.474	-2.411^b	-0.007	-2.823	-3.852
	(-3.05)	(-1.64)	(-2.19)	(-0.01)	(-1.15)	(-1.08)
Adjusted R^2	0.004	0.002	0.002	0.003	0.008	0.037
Observations	464,766	211,943	79,508	39,276	67,338	90,329
	Panel B. CRS					
$\text{Avg}_{d-6,d-2}$(CRS)	0.147	0.136	0.158	0.543	-0.074	1.470
	(1.44)	(0.95)	(0.77)	(1.36)	(-0.28)	(1.58)
Adjusted R^2	0.004	0.001	0.002	0.003	0.009	0.049
Observations	428,120	195,805	72,869	36,420	60,578	87,014
	Panel C. BDDW					
$\text{Avg}_{d-6,d-2}$(BDDW)	-2.878	1.371	-6.940^c	9.377^c	18.478^a	60.843
	(-1.27)	(0.39)	(-1.71)	(1.73)	(2.64)	(0.99)
Adjusted R^2	0.004	0.002	0.002	0.002	0.010	0.044
Observations	386,850	175,918	64,459	30,590	57,803	83,986
	Panel D. Horse Race					
$\text{Avg}_{d-6,d-2}$(HF)	4.987^a	7.934^a	5.875^c	6.376^b	7.727^b	-2.443
	(2.97)	(2.72)	(1.71)	(2.53)	(2.16)	(-0.57)
$\text{Avg}_{d-6,d-2}$(NHF)	-1.208^a	-2.406^a	-2.803^b	-1.459	-1.152	0.505
	(-2.70)	(-3.35)	(-2.01)	(-1.49)	(-0.46)	(0.14)
$\text{Avg}_{d-6,d-2}$(CRS)	0.171^b	0.194	0.200	0.907	-0.527	1.204^c
	(2.26)	(1.04)	(1.05)	(1.51)	(-1.28)	(1.90)
$\text{Avg}_{d-6,d-2}$(BDDW)	-2.923	2.356	-7.629^c	4.517	20.067^a	73.490
	(-1.32)	(0.65)	(-1.75)	(0.96)	(2.65)	(1.22)
Adjusted R^2	0.004	0.002	0.002	0.002	0.011	0.055
Observations	375,558	170,874	62,761	29,669	56,037	82,215

Table D1. Return Predictability

This table replicates the return predictability analysis from Table VII of the main paper to evaluate the return predictability of hedge fund order flow net of short interest (HF−SI). The regression specification and control variables are consistent with Table VII of the main pape. For brevity, we only report the coefficient estimates of HF−SI and NHF, while the regressions always include the full set of control variables. 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
$HF-SI_{d-1}$	1.188***	2.573***	0.905***	0.092**
	(15.16)	(15.51)	(11.00)	(2.32)
$HF-SI_{d-2}$	0.026	0.027	0.141*	-0.090
	(0.50)	(0.23)	(1.82)	(-1.45)
$HF-SI_{d-3}$	0.143***	0.305**	0.062	0.062
	(2.74)	(2.50)	(0.84)	(0.96)
$HF-SI_{d-4}$	0.086*	0.267**	0.142*	-0.150***
	(1.70)	(2.25)	(1.91)	(-2.59)
$HF-SI_{d-5}$	0.157***	0.133	0.308***	0.029
	(2.98)	(1.09)	(4.04)	(0.50)
NHF_{d-1}	-0.173***	-0.482***	-0.075***	0.036
	(-7.61)	(-11.11)	(-3.08)	(0.98)
NHF_{d-2}	-0.031*	-0.027	-0.090***	0.022
	(-1.72)	(-0.80)	(-3.72)	(0.63)
NHF_{d-3}	-0.035*	-0.011	-0.046**	-0.048
	(-1.91)	(-0.31)	(-1.99)	(-1.33)
NHF_{d-4}	-0.029*	-0.066**	-0.047**	0.025
	(-1.70)	(-1.99)	(-2.20)	(0.76)
NHF_{d-5}	-0.071***	-0.070**	-0.089***	-0.053*
	(-4.16)	(-2.06)	(-4.12)	(-1.70)
Adjusted R^2	0.031	0.022	0.029	0.041
Number of Stocks	2,824	3,610	2,466	2,400
Number of Days	7,544	2,509	2,517	2,518

Table D2. Predicting CAR around Corporate Events

This table replicates the analysis in Table III of the main paper to examine the predictability of hedge fund order flow net of short interest (HF−SI) for CAR around corporate events. The regression specification and control variables are consistent with Table III of the main paper. For brevity, we only report the coefficient estimates of HF−SI and NHF with the full set of control variables in the regressions. 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
$HF-SI_{d-2}$	6.319***	9.637***	5.659***	0.702	4.179***	0.212	7.591***	24.683***	5.872***
	(4.89)	(3.06)	(4.04)	(1.26)	(2.89)	(0.41)	(4.79)	(4.40)	(3.79)
$HF-SI_{d-3}$	2.494***	1.795	2.693**	1.053	4.160**	0.660	4.879***	15.100***	3.776***
	(2.63)	(0.94)	(2.53)	(1.57)	(2.11)	(0.94)	(4.23)	(4.58)	(3.36)
$HF-SI_{d-4}$	2.029**	6.476*	1.495*	1.101***	1.048	1.052**	3.426***	10.800***	2.589**
	(2.33)	(1.79)	(1.67)	(2.74)	(1.42)	(2.27)	(3.38)	(5.53)	(2.42)
$HF-SI_{d-5}$	0.275	-1.684	0.620	1.150*	4.274***	0.769	1.182	12.609***	-0.050
	(0.21)	(-0.56)	(0.41)	(1.81)	(3.00)	(1.22)	(0.84)	(4.99)	(-0.03)
$HF-SI_{d-6}$	1.395	-2.087	1.631	-0.304	1.784	-0.450	3.528***	13.461**	2.399***
	(1.17)	(-1.28)	(1.22)	(-0.59)	(0.67)	(-0.90)	(3.34)	(2.45)	(2.96)
NHF_{d-2}	-1.276***	-2.155***	-0.936**	-0.281	-0.915***	-0.108	-1.952***	-3.680***	-1.707**
	(-3.37)	(-5.84)	(-2.00)	(-1.39)	(-2.59)	(-0.48)	(-3.54)	(-4.85)	(-2.49)
NHF_{d-3}	-0.973**	-0.663**	-1.062*	-0.367*	-1.330***	-0.135	-1.496***	-3.538***	-1.100**
	(-2.33)	(-2.02)	(-1.94)	(-1.74)	(-3.70)	(-0.59)	(-3.23)	(-2.59)	(-2.43)
NHF_{d-4}	-0.534	-1.365**	-0.367	-0.571***	-1.267***	-0.367*	-0.876***	-2.324***	-0.584
	(-1.11)	(-2.02)	(-0.62)	(-3.03)	(-2.76)	(-1.91)	(-2.64)	(-3.92)	(-1.54)
NHF_{d-5}	-0.455	-0.394	-0.480	-0.519*	-1.200	-0.373*	-0.763	-2.789***	-0.351
	(-0.81)	(-0.41)	(-0.70)	(-1.86)	(-1.21)	(-1.69)	(-1.45)	(-3.48)	(-0.57)
NHF_{d-6}	-0.244	-0.585	-0.049	-0.239*	-0.213	-0.283*	-1.464***	-3.024***	-1.170***
	(-0.55)	(-0.76)	(-0.09)	(-1.76)	(-0.58)	(-1.75)	(-4.71)	(-6.96)	(-3.30)
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	0.003	0.004	0.004	0.007	0.012	0.006
Observation	296,314	90,802	205,512	500,811	140,815	359,996	511,958	153,375	358,583

Table D3. Predicting CAR around RavenPack Fundamental News

This table reproduces Table III of the main paper to examine the predictive ability of hedge fund order flow net of short interest (HF−SI) for CAR around corporate news events that RavenPack designates as indicated on the top of each column. The regression specification and control variables are consistent with Table III of the main paper. For brevity, we only report the coefficient estimates of HF−SI and NHF, while the regressions always include the full set of control variables. 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.

	Earnings	Analysts	M&A	Assets	Credit	Dividend	Equity Actions	Labor Issues	Product Services	Revenue	Partnership
HF−SI$_{d-2}$	3.864***	1.717	6.028**	2.152	-2.744	2.252*	-4.207	2.868**	0.891	4.511***	-2.058
	(2.87)	(1.52)	(2.08)	(1.25)	(-1.08)	(1.77)	(-0.87)	(1.98)	(0.52)	(3.66)	(-0.91)
HF−SI$_{d-3}$	1.622	2.384*	1.282	0.106	0.120	-1.562	2.770	3.415**	-0.602	1.089	2.403
	(1.08)	(1.81)	(0.46)	(0.05)	(0.13)	(-1.57)	(1.44)	(2.57)	(-0.45)	(1.12)	(1.45)
HF−SI$_{d-4}$	0.710	0.996	2.186	2.831*	2.196**	-0.309	1.730	1.364	4.035**	1.994**	0.149
	(0.64)	(0.89)	(1.16)	(1.72)	(2.36)	(-0.19)	(0.57)	(0.87)	(2.06)	(1.99)	(0.08)
HF−SI$_{d-5}$	0.113	1.406	5.444**	1.130	-0.376	0.362	5.017	-1.175	2.355*	1.370	1.109
	(0.11)	(1.34)	(2.10)	(0.62)	(-0.29)	(0.24)	(0.93)	(-1.44)	(1.79)	(1.07)	(0.61)
HF−SI$_{d-6}$	0.605	3.552***	-0.019	-1.806	0.897	0.586	6.862**	-0.233	-0.184	2.274	-1.677
	(0.68)	(2.95)	(-0.01)	(-0.65)	(0.65)	(0.45)	(2.06)	(-0.20)	(-0.11)	(1.57)	(-0.76)
NHF$_{d-2}$	-1.166**	-0.605*	-0.806	-0.210	1.018**	0.023	0.367	-0.598	0.660	-0.615	0.957
	(-2.32)	(-1.95)	(-0.78)	(-0.19)	(2.02)	(0.05)	(0.52)	(-0.82)	(0.83)	(-1.25)	(1.32)
NHF$_{d-3}$	-0.147	0.030	-0.367	-1.354	-0.477	0.291	-0.744	-0.755	0.525	-0.433	-0.600
	(-0.45)	(0.05)	(-0.56)	(-1.33)	(-1.00)	(0.50)	(-0.80)	(-1.41)	(1.15)	(-0.98)	(-0.87)
NHF$_{d-4}$	-0.251	-0.265	1.105*	-0.240	0.250	-0.206	0.199	-0.512	0.326	-0.315	0.328
	(-0.52)	(-0.63)	(1.73)	(-0.30)	(0.55)	(-0.37)	(0.24)	(-1.08)	(0.55)	(-0.75)	(0.43)
NHF$_{d-5}$	0.171	0.012	-0.694	0.545	0.370	0.957**	-0.853	0.191	-0.924**	-0.869	-0.875
	(0.47)	(0.04)	(-0.91)	(0.47)	(0.56)	(2.12)	(-0.61)	(0.38)	(-2.00)	(-1.62)	(-1.07)
NHF$_{d-6}$	-0.243	-1.333***	0.205	1.609	-0.668	-0.725	-3.196***	0.387	0.025	-0.832*	0.367
	(-0.75)	(-3.25)	(0.19)	(1.63)	(-1.50)	(-1.62)	(-2.79)	(0.88)	(0.04)	(-1.70)	(0.46)
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	0.003	0.003	0.013	0.009	0.006	0.003	0.018	0.004	0.004	0.002	0.012
Observation	258,175	163,753	68,563	23,935	73,519	77,395	86,633	111,976	109,182	180,730	32,175

Table D4. Predicting CAR around RavenPack Non-Fundamental News

This table reproduces Table III of the main paper to examine the predictive ability of hedge fund order flow net of short interest (HF−SI) for CAR around corporate news events that RavenPack designates as indicated on the top of each column. The regression specification and control variables are consistent with Table III of the main paper. For brevity, we only report the coefficient estimates of HF−SI and NHF, while the regressions always include the full set of control variables. 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.

	Insider	Investor Relation	Marketing	Order Imbalance	Price Target	Stock Price
$HF{-}SI_{d-2}$	0.085	1.206*	1.291	1.215	-1.701	1.179
	(0.17)	(1.68)	(1.28)	(0.85)	(-1.61)	(0.56)
$HF{-}SI_{d-3}$	0.647	0.417	1.095	-0.407	1.578	1.514
	(1.53)	(0.65)	(1.01)	(-0.18)	(0.91)	(0.81)
$HF{-}SI_{d-4}$	0.483	0.665	1.161	-1.223	0.336	2.313**
	(1.51)	(0.98)	(1.39)	(-0.85)	(0.18)	(2.03)
$HF{-}SI_{d-5}$	-0.148	0.823	1.593*	1.061	0.014	0.867
	(-0.50)	(1.49)	(1.86)	(0.60)	(0.01)	(0.62)
$HF{-}SI_{d-6}$	0.125	0.846	0.599	-0.288	0.934	-0.901
	(0.29)	(1.48)	(0.73)	(-0.20)	(0.57)	(-0.47)
NHF_{d-2}	0.327**	-0.122	-0.353	-0.526	-0.001	-0.983
	(2.24)	(-0.34)	(-0.73)	(-1.26)	(-0.00)	(-0.97)
NHF_{d-3}	-0.308**	-0.051	-0.555	-0.331	0.272	-0.792
	(-2.02)	(-0.20)	(-1.39)	(-0.81)	(0.28)	(-0.86)
NHF_{d-4}	-0.008	-0.405	-1.005***	1.335***	0.336	0.272
	(-0.05)	(-1.48)	(-3.25)	(3.49)	(0.54)	(0.41)
NHF_{d-5}	-0.109	-0.237	0.054	-0.049	-0.305	-0.683*
	(-1.10)	(-1.08)	(0.13)	(-0.16)	(-0.49)	(-1.79)
NHF_{d-6}	-0.088	-0.288	-0.372	0.082	0.136	-1.591
	(-0.45)	(-1.06)	(-1.30)	(0.22)	(0.19)	(-1.59)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.004	0.002	0.002	0.002	0.008	0.018
Observation	506,436	235,378	92,265	65,511	62,152	97,899