Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business | Lee Kong Chian School of Business

2-2020

How smart is institutional trading?

Jingi HA
Singapore Management University, jingiha.2014@pbs.smu.edu.sg

Jianfeng HU
Singapore Management University, JIANFENGHU@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Finance and Financial Management Commons, and the Portfolio and Security Analysis Commons

Citation
HA, Jingi and Jianfeng HU. How smart is institutional trading?. (2020). 1-45.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/6591

This Working Paper is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

How Smart Is Institutional Trading?

JinGi Ha and Jianfeng Hu *

February 9, 2020

ABSTRACT

We estimate daily aggregate order flow of individual stocks from institutional investors as well as for hedge funds and other institutions separately. We find that the estimated institutional order imbalance positively predicts stock return on the next day and outperforms other institutional order flow estimates. The institutional order flow from hedge funds generates greater and more persistent price impact than the order flow from other institutions. We also find that hedge funds trade on well-known anomalies while the other institutions do not. Our findings suggest that the superior trading skills of institutional investors can be largely attributed to hedge funds.

Keywords: Institutional trading; Hedge funds; Trading behavior

*We would like to thank Ekkehart Boehmer, Si Cheng, Tarun Chordia, Jarred Harford, Dashan Huang, George Jiang, Jin-Mo Kim, Jeong-Bon Kim, Roger Loh, Alberto Manconi, Massimo Massa, Kazuhiko Ohashi, Gilbert Park, Elvira Sojli, Avanidhar Subrahmanyam, Johan Sulaeman, Yuehua Tang, Shu Tao, Charles Trzcinka, Dimitri Vayanos, Baolian Wang, John Wei, Jialin Yu, Joe Zhang, Xiaoyan Zhang, Hong Zhang, Bart Zhou and the seminar participants at Singapore Management University, the Fifth Asian Bureau Finance and Economic Research Conference, 2017 SMU Research Camp, 2017 Asian Finance Association Annual Conference, 2017 European Finance Association Annual Conference, 2017 Auckland Finance Meeting, and 2018 Financial Management Association Annual Conference for comments. All remaining errors are ours. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant and IREC, The Institute of Finance and Banking, and Seoul National University. Jianfeng Hu also acknowledges financial support from the Lee Kong Chian Fund for Excellence. JinGi Ha is at Auckland University of Technology (jingi.ha@aut.ac.nz); Jianfeng Hu is at Singapore Management University (jianfenghu@smu.edu.sg).

Electronic copy available at: https://ssrn.com/abstract=2907612

I. Introduction

Institutional investors are at the center of academic finance research as the professional asset management business continues to see unprecedented growth in the United States and worldwide. Many researchers study institutional investors' stock picking skills by analyzing their performance using publicly available low frequency data such as quarterly institutional holdings or monthly fund returns.[1] The low frequency data, however, are not able to capture the exact timing of institutional trades. Therefore, it remains challenging to examine the role of institutional trading in price discovery, the corresponding transactions cost, or how institutional investors may trade based on short-lived information. To circumvent limitations on institutional trading data, some studies estimate institutional order flow (IOF, hereafter) using high frequency data on the premise that institutions are more likely to place large orders, e.g. Lee (1992), Lee and Radhakrishna (2000), Battalio and Mendenhall (2005), Malmendier and Shanthikumar (2007), and Hvidkjaer (2008). However, as order splitting becomes common, arising from a concern for trading costs, the reliability of these size-based algorithms has been cast under serious doubt in recent sample periods, as pointed out by Cready, Kumas, and Subasi (2014). Alternatively, several recent studies employ proprietary data sets of institutional trading records for empirical analysis.[2] However, these proprietary data have limited coverage and can contain sample biases that impair external validity. In this article, we present a new method for estimating *aggregate* order flow at the daily level for institutional investors and for subgroups of institutions only.

[1]For example, Carhart (1997), Wermers (2000), Chen, Jegadeesh, and Wermers (2000), Barras, Scaillet, Wermers (2010), and Edelen, Ince, and Kadlec (2016) find weak evidence on institutional investors' stock picking skills using the low frequency data. On the contrary, Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Kosowski, *et al.* (2006), Avramov and Wermers (2006), Alexander, Cici, and Gibson (2007), Kacperczyk, Sialm, and Zheng (2008), and Elton, Gruber, and Blake (2011) show that institutions have skills in assembling portfolios using the same databases.

[2]For example, Puckett and Yan (2011), Anand, *et al.* (2012), and Henry and Koski (2017), Chakrabarty, Moulton, and Trzcinka (2017) use actual institutional trade data from Abel Noser, which covers about 10% of total institutional trading as estimated by Puckett and Yan (2011). Irvine, Lipson, and Puckett (2007) and Busse, Clifton, and Jegadeesh (2012) use data from Plexus Group. Griffin, Harris, and Topaloglu (2003), Boehmer and Kelley (2009), and Hendershott, Livdan, and Schurhoff (2015) use the Consolidated Equity Audit Trail Data from NYSE. Also, Foster, Gallagher, and Looi (2011) and Frijns, *et al.* (2018) use actual institutional trade data in Australia and Finland, respectively.

Electronic copy available at: https://ssrn.com/abstract=2907612

We estimate daily aggregate institutional order flow by extrapolating the quarterly relation between institutional ownership change in 13F and observable order flow variables constructed using TAQ and Ancerno data. Specifically, and similar to the IOF estimation method proposed by Campbell, Ramadorai, and Schwatz (CRS, 2009), we first evaluate the quarterly association of change in aggregate institutional holdings and aggregate order imbalance from several transaction size bins obtained from both TAQ and Ancerno data. Then, we fit the quarterly relation using the daily TAQ order flow and the Ancerno order flow to retrieve daily aggregate institutional order flow. The underlying assumption of our method is that institutional investors systematically differ from retail investors in the way of submitting orders of different sizes, similar to what is presented by CRS. This assumption is largely left unverifiable by CRS using TAQ data because there is not a trader-type flag in TAQ. Unlike CRS, however, we bring actual institutional trading records from Ancerno to enhance the linkage between institutional trading and trade size. The addition of Ancerno data therefore provides a nontrivial contribution to estimation accuracy.

As the full set of institutional order flow is unobservable, we evaluate the performance of our proposed order flow measure in a joint-hypothesis test using return prediction regressions. There is consensus in the literature that institutions are apt to be better informed than retail investors because institutions have direct communication channels with firms, better relationships with sell-side analysts, and better capabilities for processing financial information. Therefore, institutional trading can significantly contribute to the price formation process. In the same spirit of prior studies, we hypothesize that a less noisy IOF is more informative about the future stock price. In our sample of common stocks listed on NYSE, AMEX, and Nasdaq markets between January 1999 and March 2012, we find that our proposed IOF estimate, termed HH, significantly and positively predicts the returns on the following day with a coefficient of 0.122 and a t-statistic of 11.71 in the Fama-MacBeth (1973) regressions. A one-standard deviation change in HH is associated with a subsequent return of 1.928 basis points (bp). This return predictive ability outperforms the other IOF

Electronic copy available at: https://ssrn.com/abstract=2907612

measures we consider in the study, including the aggregate institutional order flow in Ancerno data, Campbell, Ramadorai, and Schwatzs' estimate, and the estimate based on a size cut-off rule following Lee and Radhakrishna (2000). We obtain the same conclusion in direct horse race tests of multiple IOFs in the same return prediction model as well as in investment analysis. These results suggest that combining TAQ and Ancerno data yields a more accurate IOF estimate in the cross-section.

Because we can identify institution types in 13F and Ancerno, it is possible to apply the same method to a subgroup of institutions such as: hedge funds, mutual funds, short-term institutional investors, and active institutions. These types of institutions can differ considerably on investment strategies, tools, and payoff structures. Therefore, looking at institutional investors as a whole may discard important heterogeneity across institution types. Taking advantage of estimation flexibility in our method, we choose to estimate and examine institutional order flow from hedge funds and non-hedge funds as an example to shed more light on the stock trading skills of institutional investors.

We investigate the systematic difference between aggregate hedge fund and non-hedge fund trading in three ways: First, we conjecture that hedge funds execute their orders at lower costs than non-hedge funds, hence generating smaller contemporaneous price impact because hedge fund managers can time market liquidity (Cao, *et al.*, 2013) and are likely to provide liquidity (Jame, 2017). However, this hypothesis is not well supported by our data. We find that hedge fund order flow is more balanced than non-hedge fund order flow but it is more likely due to less trading demand from the smaller hedge fund sector. Moreover, both types of institutions seem to have similar skills in managing trading costs as the estimated price impact from institutional order imbalance is significantly smaller than the overall market order imbalance. Between hedge funds and non-hedge funds, however, the difference in contemporaneous price impact per unit of trade is not statistically significant.

Second, there is a well-documented performance difference in the literature examined at monthly or quarterly horizons. The superior performance of hedge funds is persistent

Electronic copy available at: https://ssrn.com/abstract=2907612

(Jagannathan, Malakhov, and Novikov, 2010) and their confidential holdings exhibit positive performance for up to twelve months (Agarwal, *et al.*, 2013 and Argon, Hertzel, and Shi, 2013). To understand hedge funds' investment skills better, we test if their trades contain more informative about short-term future returns than trades from other institutions. Indeed, in Fama-MacBeth regressions, we find that the hedge fund order flow generates a coefficient of 1.603 with a t-statistic of 17.18 while the non-hedge fund order flow has a slightly negative and insignificant coefficient when predicting the next day's returns. Moreover, the pricing effect of hedge fund trades is permanent without subsequent reversals, while the non-hedge fund trades generate significantly negative price impact over longer horizons. This result is robust in several subsample tests based on firm size, bid-ask spread, time periods, and exchange markets. The investment strategies based on hedge fund order flow and non-hedge fund order flow generate average daily return differentials of 0.082 (t-statistic = 10.95) and 0.043 (t-statistic = 5.95), respectively.

Finally, we examine how the two types of institutions respond to well-known return anomalies. Akbas, *et al.* (2015), Kokkonen and Suominen (2015) show that fund flow into hedge funds mitigates mispricing. However, direct within-month arbitrage trading activities by hedge funds in response to predictable return anomalies are not documented in existing studies. We construct a mispricing index updated every day, similar to the one used by Stambaugh, Yu, Yuan (2012, 2015). In panel regressions of net IOFs on the lagged mispricing index, we find that hedge fund trades are positively associated with the mispricing index from the previous day, implying that they buy undervalued stocks and sell overvalued stocks. On the contrary, the coefficient estimate of lagged mispricing is negative and significant when we apply net non-hedge fund order flow as the dependent variable, suggesting that other institutions, on average, trade in a direction that exacerbates return anomalies.

We make two contributions to the finance literature. First, we introduce a new method of estimating institutional order flow for individual stocks at the daily level. Empirical analysis shows that this new method has more robust and stronger performance in terms of return

Electronic copy available at: https://ssrn.com/abstract=2907612

predictability than prior methods proposed by Lee and Radhakrishna (2000) and Campbell, Ramadorai, and Schwartz (2009). Our method can be applied in empirical studies that examine institutional trading behavior at the daily, or longer, frequency. The method is also flexible to be applied to subgroups of institutions. Second, by investigating the contribution of different types of funds to price discovery, we find that hedge funds have superior trading skills in timing short-term future returns rather than contemporaneous market liquidity compared to other institutional investors. We also find that hedge funds actively trade on well-known stock return anomalies on a daily basis, while the other institutions, on average, trade against those anomalies. These findings complement the studies using longer-horizon observations such as Akbas, *et al.* (2015) and Caglayan, Celiker, and Sonaer (2018) by providing direct evidence at a finer granularity.

The rest of the paper is organized as follows: Section II describes our estimation method of aggregate institutional order flow and sample selection. Section III tests the performance of our IOF estimate. Section IV investigates the systematic difference between hedge fund and non-hedge fund trading. Finally, Section V presents our conclusions..

II. Data and variable description

A. Sample selection

We merge four databases into our sample using eight-digit CUSIP and Monthly TAQ master file: (1) daily CRSP data for stock return, share volume turnover ratio, relative bid-ask spread, market capitalization, a primary exchange code, and share type code; (2) TAQ data to calculate total order imbalance in different trade size bins; (3) Thomson Reuters 13F Ownership data for quarterly holding of institutional investors; and (4) Ancerno institutional trade data to calculate stock-day Ancerno order imbalance. The sample includes $10,221,329$ stock-day observations from January 1999 to March 2012. Every observation in our sample meets the following criteria to be eligible: (1) it is a common stock; (2) it is primarily listed

Electronic copy available at: https://ssrn.com/abstract=2907612

on NYSE, AMEX, or Nasdaq; (3) it is available in Monthly TAQ master file, 13F and CRSP; (4) the relative close bid-ask spread is positive and less than one half; and (5) the stock price is greater than \$5. We winsorize all types of estimated institutional order flow measures, relative bid-ask spread, and share volume turnover ratio at one and 99 percent levels in the cross-section for each day.

B. Estimating institutional order flow

B.1. Cut-off rule from Lee and Radhakrishna (2000)

Many studies have identified individual and institutional trades using transaction size. For example, Lee (1992) considers transactions less than \$10,000 to be initiated by individual investors. More recently, Malmendier and Shanthikumar (2007) categorize individual and institutional investors by using \$20,000 and \$50,000 thresholds. The cut-off rule is intuitively appealing because institutional investors are more likely to be large investors who can buy or sell a large amount of stocks in a single transaction. Lee and Radhakrisha (LR, 2000) show that, according to TORQ data for 144 NYSE stocks from November 1990 to January 1991, the cut-off rule effectively separates institutional trades from individual trades. Also, Griffin, Harris, and Topaloglu (2003) report that institutional trades consist of 85.99% of block trades (more than 10,000 shares in a transaction) and 18.14% of small trades (less than 500 shares in a transaction) in a hundred of Nasdaq stocks from May 2000 to February 2001. However, the cut-off rule is subject to a misclassification concern rooted in order splits by institutional investors. Using the TORQ data, LR disprove it by showing that 94% of total orders are executed in a single transaction. When it comes to more recent periods, however, Campbell, Ramadorai, and Schwatz (2009) find that institutional trades significantly involve small- sized trades (less than \$2,000 in a transaction) for small firms, using quarterly institutional holding and TAQ data from 1993 to 2000. In addition, Cready, Kumas, and Subasi (2014) document that the majority of institutional trades are small-sized trades (less than 500 shares in a transaction) in Ancerno institutional data from 2003

Electronic copy available at: https://ssrn.com/abstract=2907612

to 2010. These empirical findings challenge the accuracy of the cut-off rule.

To construct a benchmark IOF measure following LR, we apply a $5,000 cut-off rule to identify institutional trades in TAQ data from January 1999 to March 2012.[3] First, we determine trade directions using the Lee and Ready (1991) algorithm.[4] We put a positive (negative) sign on a buyer-initiated (seller-initiated) transaction. Then, we categorize transactions greater than $5,000 as institutional trades. Lastly, we aggregate the number of signed shares of institutional trades at stock-day level and divide by the number of shares outstanding for cross-sectional normalization to produce an estimated institutional order imbalance, LR.

B.2. Institutional order flow from Campbell, Ramadorai, and Schwatz (2009)

Campbell, Ramadorai, and Schwatz (2009) develop a non-linear model to estimate institutional order flow at daily frequency, extrapolated from the quarterly relation of institutional ownership change in 13F and order imbalances in different trade sizes from TAQ. They highlight two advantages of the method in comparison to the cut-off rule. One is that it exploits information on net order flow in the full range of transaction sizes, rather than based on a single threshold, to account for the fact that institutional investors can split orders either to hide their trades or to reduce transaction costs. Another advantage is that their method allows size-categorized order imbalance to be positively or negatively associated with an institutional ownership change because institutions can strategically choose to take or provide liquidity, depending on the size of the trades. For example, an institution can submit several small-sized *limit* orders. In this case, order flow in the small-sized transaction would be *negatively* associated with the change of institutional ownership.

[3] LR show that the $5,000 cut-off can capture 82% of institutional trades, the highest percentage among their chosen cut-offs. Using alternative thresholds does not alter our conclusion about the effectiveness of the cut-off rule. These results are available upon request.

[4] Lee and Ready (1991) algorithm classifies a trade as buyer-initiated (seller-initiated) if its execution price is higher (lower) than a mid-point of National Best Bid and Offer (NBBO) prices. If the execution price is the same as the mid-point, the trade is classified as buyer-initiated (seller-initiated) if the execution price is higher (lower) than the previous trade price.

Electronic copy available at: https://ssrn.com/abstract=2907612

The estimation is based on the following equation:

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

where for a stock i in a quarter q, α is a set of four quarter dummies, Y is aggregate institutional ownership in 13F, U is aggregate unclassified trades scaled by shares outstanding for which the Lee and Ready (1991) algorithm cannot determine the direction, and F_Z is aggregate order imbalance scaled by shares outstanding in a trade-size bin Z. CRS assign trades into nineteen size bins which have lower limit points at \$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. Since large trades for a small stock are rare, the coefficients on order flow in large trade sizes are likely swung by the small number of unusual trades. To address the potential issue in estimation, CRS borrow a yield curve model for $\beta(Z, Y_{i,q-1})$ from Nelson and Siegel (NS, 1987):

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

where τ is a constant to be estimated. The NS model smooths out the coefficient variation across transaction size by putting less weight on order flow in larger trade-size bins. Also, CRS divide the sample into quintile portfolios based on NYSE breakpoints of market capitalization at the start of each quarter, and estimate the coefficients in Equation (1) for each quintile separately using non-linear least squares estimation that maximizes the adjusted R-squared over different values of τ.[5]

Next, taking the estimates from Equation (1), CRS calculate the expected change of

[5]There could be a convergence problem in the estimation of Equation (1). For example, we have observed the adjusted R-squared asymptotically approaching to its maximum as τ increases. For such cases when τ approaches infinity, we set the arbitrary maximum value of τ to 100,000.

Electronic copy available at: https://ssrn.com/abstract=2907612

institutional ownership, $\mathrm{E}[\Delta Y_{i,d}]$, as institutional order flow on a day.

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

where d indexes a day. The frequency conversion is possible under an exogeneity assumption that the error terms, $\epsilon_{i,d}$, are not correlated with all of its lead and lag terms within a quarter. Also, CRS set to zero the quarterly variables such as ρ and ϕ as well as a set of daily dummies, α_d, because intra-quarter events may invalidate the exogeneity assumption.

B.3. Institutional order flow from Ancerno institutional trading data

Ancerno (AN) institution data is transaction-level institutional trading data provided by Abel Noser Solutions, formerly Ancerno, Ltd., which offers consulting services for execution as well as transaction cost analyses to institutional asset owners, investment managers, and brokers. A material advantage to using AN is that it contains comprehensive records of an institution's trading history since that institution started to employ Abel Noser's consulting services. The data contain a client manager/broker/trader code, a trade date, CUSIP of a traded stock, a trade direction, an execution price, the number of execution shares, commissions, and fees for each trade[6]. However, AN has a sample bias because it covers a limited portion of institutional trades. Puckett and Yan (2011) estimate that the data fail to account for approximately 90% of total institutional trades. We also find that the time-series correlation of total 13F institutional ownership changes with the aggregate Ancerno institutional order flow is 0.249 at quarterly frequency for our sample period from January 1999 to March 2012. Those statistics cast doubt on whether AN trades sufficiently represent the entire set of institutional trades on its own.

As an alternative IOF measure, we calculate the imbalance in Ancerno trades. Specifically, we calculate the difference between buyer execution shares and seller execution shares

[6] For more detailed information, see Puckett and Yan (2011), Anand, *et al.* (2012), Busse, Clifton, and Jegadeesh (2012), Cready, Kumas, and Subasi (2014), Chakrabarty, Moulton, and Trzcinka (2017), Henry and Koski (2017), or Jame (2017).

Electronic copy available at: https://ssrn.com/abstract=2907612

in Ancerno for a stock on the same day scaled by number of shares outstanding, termed AN. Note that Ancerno does not keep stock-day records for zero volume from their clients. We set the AN volume on such stock-day observations to missing.[7]

B.4. Our proposed institutional order flow

We estimate institutional order flow in the same spirit as CRS. Exploiting the quarterly association of an institutional ownership change with TAQ order flow as well as AN institutional order flow, we fit their relation to retrieve the institutional ownership change on a daily basis. The utilization of AN institutional order flow has a non-trivial impact on our estimation. The actual institutional order flow allows us to reduce potential measurement errors in CRS estimation stemming from noisy TAQ order flow that groups together individual and institutional trades.

We use non-linear least square fit over different values of τ to evaluate the coefficients in the following model:

$$\begin{aligned}\Delta Y_{i,q} &= \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \beta^{UY} Y_{i,q-1} \times U_{i,q} \\ &+ \sum_{Z=1}^{19} \beta^F(Z, Y_{i,q-1}) F_{Z,i,q} + \sum_{Z=1}^{19} \beta^D(Z, Y_{i,q-1}) D_{Z,i,q} + \epsilon_{i,q}, \end{aligned} \tag{4}$$

where $D_{Z,i,q}$ is aggregate AN institutional order imbalance calculated as the difference of buyer execution shares and seller execution shares in a transaction size bin Z for a stock i at a quarter q and other variables are the same as in Equation (1). $\beta^F(Z, Y_{i,q-1})$ and $\beta^D(Z, Y_{i,q-1})$ have the same functional form as $\beta(Z, Y_{i,q-1})$ in Equation (2). We also use the same nineteen size bins as CRS, and estimate Equation (4) in each size quintile based on NYSE breakpoints separately.

Then, we recover the expected institutional ownership change, $\mathrm{E}[\Delta Y_{i,d}]$, on a daily basis,

[7] We also do the same test, setting the value of such zero-volume observations to zero. The test results are qualitatively the same as those reported in this paper.

Electronic copy available at: https://ssrn.com/abstract=2907612

by fitting the estimates of Equation (4) into the following equation:

$$\begin{aligned}\Delta Y_{i,d} &= \alpha_d + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,d} + \beta^{UY} Y_{i,q-1} \times U_{i,d} \\ &+ \sum_{Z=1}^{19} \beta^F (Z, Y_{i,q-1}) F_{Z,i,d} + \sum_{Z=1}^{19} \beta^D (Z, Y_{i,q-1}) D_{Z,i,d} + \epsilon_{i,d}. \end{aligned} \tag{5}$$

The estimated daily change of institutional ownership is termed HH.

B.5. Hedge fund and non-hedge fund order flow

We estimate two institutional order flows for hedge funds and non-hedge funds to explore the heterogenous trading behavior among institutions. Our estimation is, by construction, flexible enough to measure the expected institutional order flow for different types of institutional investors: One can evaluate institutional order flow for any type of institution by putting quarterly ownership of a target institution group in the estimation model of Equation (4). In this paper, we choose hedge funds and non-hedge funds as an example.

We replicate the estimation of Equation (4) to get coefficients regarding hedge funds and non-hedge funds, replacing the aggregate institutional ownership with aggregate hedge fund ownership and non-hedge fund ownership, respectively.[8] Then, we fit the estimates of Equation (4) into the retrieval procedure in Equation (5) to measure the ownership change of hedge funds and non-hedge funds at daily frequency as hedge fund order flow (HH^{HF}) and non-hedge fund order flow (HH^{NHF}), respectively.

C. Control variables

We calculate total order imbalance (TOI), share volume turnover ratio ($TURN$) and relative bid-ask spread ($SPRD$) for a stock on a day as control variables in return prediction regressions. The detailed definitions are the following:

[8] We thank Yuehua Tang for providing an identifier for hedge funds in Thomson Reuters 13F Ownership data used in Agarwal, *et al.* (2013).

Electronic copy available at: https://ssrn.com/abstract=2907612

- TOI: The number of buyer-initiated shares minus the number of seller-initiated shares in TAQ based on the Lee and Ready (1991) algorithm, scaled by the number of shares outstanding in CRSP.
- $TURN$: Daily trading volume over the number of shares outstanding.
- $SPRD$: The difference between the bid and ask prices scaled by the average of bid and ask prices.

We use the risk-adjusted mid quote stock return (RET) with respect to the Fama-French (2015) five factors as our measure of stock return.

D. Summary statistics

Panel A of Table I presents the time-series average of cross-sectional statistics for institutional order flow (IOF) measures and control variables. Our sample covers, on average, 3,073.2 stocks per trading day for 3,326 trading days from January 1999 to March 2012.

[Place Table I about here]

First, and for purposes of comparison, four versions of IOF measures are reported. Our proposed IOF (HH) has a mean of 0.025, which is five times its median of 0.005. HH has a standard deviation of 0.158 and is positively skewed with a minimum of -0.535 and a maximum of 0.671. AN is well-balanced as the mean and median are close to zero at 0.004 and 0.002, respectively. The average number of observations per day for AN is 2,043.1 due to limited coverage of AN data. CRS has a mean of 0.024 similar to HH, but a smaller standard deviation of 0.074. LR has a mean of 0.014, smaller than HH.

Statistics for hedge and non-hedge IOFs are reported next. The average numbers of observations per day for HH^{HF} and HH^{NHF} are 2,893.5 and 3,071.6, respectively. HH^{HF} has a mean of 0.006 and a standard deviation of 0.020, both smaller than those of HH^{NHF} (mean = 0.019; standard deviation = 0.121). The result is consistent with the fact that the

Electronic copy available at: https://ssrn.com/abstract=2907612

average ownership of hedge funds (9.30%) is about one-third of non-hedge funds (30.00%) in our sample period.

Lastly, descriptive statistics for control variables are presented. Total order imbalance (TOI) has very similar statistics to LR: a mean of 0.015 and a standard deviation of 0.179. The mean bid-ask spread ($SPRD$) is 0.009 and the mean turnover ratio ($TURN$) is 0.649. The mean and median of risk-adjusted mid-quote return are 0.000 and 0.027, respectively.

Panel B of Table I presents the time-series average of cross-sectional correlation between estimated IOFs and control variables. Although we use both TAQ and AN order imbalances to estimate HH under the same dependence structure, the loadings of HH on TAQ and AN data seem to differ considerably. The correlation between HH and AN is 0.875, almost seven times the correlation between HH and TOI at 0.127. This average daily correlation of 0.875 is also much higher than the quarterly correlation between aggregate AN institutional order flow and 13F ownership change at 0.249. The quarterly correlation between aggregate TAQ order imbalance and 13F ownership change is 0.099. HH's correlations with CRS and LR are 0.352 and 0.142, respectively. Therefore, we find significant impact of including the AN data in our estimation of institutional order flow. The resulting HH is sufficiently different from the existing IOF estimates. In addition, non-hedge fund IOF (HH^{NHF}) behaves similarly to HH with a high correlation of 0.957 between the two, because non-hedge funds make up the majority of the investment industry and also because 63.8 percent of their trades are correlated with hedge funds' trades.

We also find AN institutional order flow is weakly correlated with CRS, LR, and TOI. The correlation of CRS with LR is 0.306, similar to its correlation with TOI as LR and TOI are almost identical with a correlation coefficient of 0.946. All estimated institutional order flows except AN are positively associated with stock liquidity. For example, HH is negatively (positively) associated with relative bid-ask spread (share volume turnover ratio) with a correlation of -0.052 (0.148). AN is independent of stock liquidity. Additionally, IOF measures make a positive impact on the contemporaneous price. In particular, CRS

Electronic copy available at: https://ssrn.com/abstract=2907612

has the weakest influence on the current stock price with a correlation of 0.069.

Interestingly, correlation tests suggest that hedge funds are more sensitive to liquidity and their trades make weaker contemporaneous price impact than non-hedge funds. Specifically, the correlation of hedge funds' order flow is significantly more negative (positive) with relative bid-ask spread (share volume turnover ratio) than the correlation of non-hedge funds' order flow. The difference is -0.011 (0.128) with a t-statistic of -22.27 (106.77). The correlations of HH^{HF} and HH^{NHF} with contemporaneous return are 0.088 and 0.119, respectively. The difference in the correlations is -0.030 with a t-statistic of -45.83. We will dive deeper into the impact of hedge fund order flow on contemporaneous prices in Section IV.A.

III. Performance of institutional order flows

We compare our proposed institutional order flow HH with the other IOF estimates in this section. As the total institutional order flow is not observable on a daily basis, we evaluate the effectiveness of different IOF measures by testing a joint hypothesis that a more accurate IOF is more informative about future stock prices in the cross-section. Institutional trading can contain price information for the following reasons: First, many institutions, as large shareholders, have direct communication with firms. Therefore, they can ask firms to disclose more information for monitoring purposes (Boone and White, 2015). Second, they maintain good relationships with sell-side analysts - a relationship that allows institutions to receive additional tips on a particular stock from analysts (Irvine, Lipson, and Puckett, 2007). Third, they are capable of correctly interpreting financial information, such as earnings. Cohen, Gompers, and Vuolteenaho (2002) find evidence that institutions trade in the profitable direction following earnings news. Additionally, Piotroski and Roulstone (2004) and Campbell, Ramadorai, and Schwartz (2009) show that institutions can anticipate future earnings surprises. Market microstructure studies largely confirm the role of institutional trading in price discovery. For example, Hendershott, Livan, and Schurhoff (2015)

Electronic copy available at: https://ssrn.com/abstract=2907612

find that institutional trading enhances price efficiency by spreading information prior to a news announcement. Boehmer and Kelly (2009) also find that stocks with high institutional ownership have more efficient prices. Based on the premise that institutional trading conveys price information, we examine the predictive power of different measures of IOF in the following subsection. We also compare the performance of investment strategies based on different IOFs in the subsequent subsection.

A. Cross-sectional return prediction

We begin our return prediction analysis with Fama-MacBeth (1973) regressions in our sample consisting of all common stocks listed on NYSE, AMEX, and Nasdaq between January 1999 and March 2012. Table II reports the estimated coefficients of the following model:

$$\begin{aligned} R_{i,t} &= \alpha_t + \sum_{k=1}^{5} \beta_{t,k}^{I} IOF_{i,t-k} + \sum_{k=1}^{5} \beta_{t,k}^{T} TOI_{i,t-k} \\ &+ \gamma_t^{B} SPRD_{i,t-1} + \gamma_t^{T} TURN_{i,t-1} + \sum_{k=1}^{5} \gamma_{t,k}^{R} R_{i,t-k} + \epsilon_{i,t}, \end{aligned}$$

where for stock i on day t, R is risk-adjusted mid-quote stock return with respect to Fama-French (2015) five factors; IOF is an institutional order flow measure including HH, AN, CRS, and LR. We include lagged TOI in the model to control for the aggregate market order imbalance. Also, lagged relative bid-ask spread ($SPRD$) and lagged share volume turnover ratio ($TURN$) control for liquidity effects in the model, and lagged returns are included to control for return reversals. To account for serial correlations, we use Newey-West (1987) standard errors with eight lags to calculate the t-statistics.

[Place Table II about here]

The first column shows that the coefficient of HH is 0.122 at the first lag with a t-

Electronic copy available at: https://ssrn.com/abstract=2907612

statistic of 11.71. In terms of economic significance, a one-standard deviation increase of HH is associated with an increase of 1.928 basis point (bp) in the next day's stock return. However, this large price impact gradually reverses in subsequent days as the coefficients at longer lags are all negative and significant. After five days, the positive initial price impact completely dies out. In the second column, the number of observations decreases from $10,015,095$ to $5,033,554$ because the AN data do not fully cover all the stock-day observations in our sample. Similar to HH, AN also has a positive coefficient (0.097, t-statistic = 6.39) at the first lag. The economic significance of AN is weaker than HH as a one-standard deviation increase in AN is associated with an increase of 1.329 bp in stock return on the next day. The coefficients of AN also turn negative at longer lags, though only marginally significant, indicating that the price impact from AN lasts for more than a week. In the third column, CRS has significantly negative coefficients at all lags, indicating this IOF measure captures price pressure but not information content in the cross-section.[9] The last column shows that the first lagged term of LR has significantly negative price impact, but it flips to a positive sign in the second lag. The other lags are statistically indifferent from zero.

All the control variables other than total order imbalance have the expected signs of their coefficients. Total order flow has mixed signs in the first lagged term across columns with longer lags always having negative and significant coefficients. Spread and turnover ratio both have positive and significant coefficients, consistent with Amihud and Mendelson (1986) and Gervais, Kaniel, and Mingelgrin (2001), respectively. We also observe significant return reversals in all regressions.

The results in Table II suggest that HH is a powerful predictor for stock return. Focusing on the price impact on the next day only, HH outperforms the other IOF estimates in both statistical and economic significance. The strong predictive power of HH indicates that

[9]In unreported tests, we find that CRS positively predicts future returns on the following day in regressions without the control variables. We have also achieved similar results in CRS (2009) using the same vector auto-regression specification.

Electronic copy available at: https://ssrn.com/abstract=2907612

AN institutional trades significantly complements the TAQ data in extracting valuable price information. However, the pricing effect from HH turns negative at longer horizons while AN does not in a restricted sample. To compare the performance of HH with other IOF measures in fair tests, we conduct additional regression analysis by using HH and another IOF estimate in the same regression and report results in Table III. For brevity, we present the estimated coefficients of institutional order flow only, while the regressions always include the full set of control variables.

[**Place Table III about here**]

The first column presents the horse race result between HH and AN. In the same restricted sample, HH has a coefficient of 0.073 with a t-statistic of 3.24 at the first lag and the subsequent reversals also become smaller in magnitude compared to the result in Table II. AN, however, fails to retain coefficients significantly different from zero at any lag in the horse race. It seems the predictive ability of AN is subsumed by HH. In the second and third columns, the coefficients of HH remain positive and significant at the first lag while CRS and LR yield only negative and significant coefficients, similar to those in Table II.

In summary, the predictive power of our proposed IOF estimate, HH, outperforms the other IOF estimates in the cross-section. The estimated coefficient of HH is always positive and significant at the first lag, regardless of inclusion of competing institutional order flow. Note that the pricing effect of HH is not persistent and reverses in the four subsequent trading days. This IOF behavior implies institutions are likely to rely on short-lived information for their trades as a whole. AN institutional order flow is also positively associated with stock return on the following day and its price impact is persistent at least for five trading days. When it comes to the horse race with HH in Table III, however, AN loses its predictive ability for stock returns. CRS and LR only generate a negative price impact on future stock returns in the regression, controlling for stock illiquidity and return reversals. If

Electronic copy available at: https://ssrn.com/abstract=2907612

institutional investors are better informed than an average investor, their order flow should positively correlate with future stock returns. Consistent with this view, we find that HH generates persistent and robust return predictive ability for returns, while the other IOF measures do not.

B. Investment analysis

In this subsection, we perform univariate sorts to examine the profitability of order flow strategies. We sort all stocks into decile portfolios based on one of the IOF measures each day and calculate the average equal-weighted portfolio returns on the following day. Table IV reports the average decile portfolio returns as well as the return differentials and Fama-French (2015) alphas between the top and bottom order imbalance decile portfolios. All t-statistics in Table IV are calculated based on Newey-West (1987) standard errors with eight lags.

[Place Table IV about here]

We find that the investment strategy of buying stocks in the top HH decile and selling stocks in the bottom decile generates an average return of 5.2 bp on the following day with a t-statistic of 6.81. The alpha with respect to the Fama-French (2015) five factors (FF5) is at the same magnitude with an even larger t-statistic. The long-short strategy based on AN also produces a significantly positive average return of 5.9 bp on the next day with a t-statistic of 6.96.[10] Strategies based on CRS and LR, however, do not generate positive abnormal returns. In the case of CRS, the return differential is even significantly negative. The univariate portfolio sort results in this subsection are consistent with Fama- MacBeth regression results in the previous subsection. While the profitability analysis sheds light on the information content of order flows, such investment strategies may not be implementable in practice because the IOF model parameters are estimated using the full sample data.

[10] The performance difference between the HH-based and AN-based strategies is -0.007 with a t-statistic of -1.10, statistically indifferent from zero. Note that there is difference in sample sizes: $10,221,329$ for HH, CRS, and LR versus $6,790,481$ for AN.

Electronic copy available at: https://ssrn.com/abstract=2907612

IV. Trading behavior of hedge/non-hedge funds

One advantage of our IOF estimation method is its flexibility to estimate order flow for different types of institutions. As an example, we study potential heterogenous trading behavior between hedge funds and other institutional investors in this section. Specifically, we examine the following three aspects in stock trading:

1. Execution cost:

   We conjecture that hedge funds execute their orders at low costs. Aragon and Strahan (2012) find that a hedge fund investor can act as a liquidity provider per Lehman's bankruptcy in 2008. Cao, *et al.* (2013) also find evidence that hedge funds can time market liquidity. The direct comparison of execution costs between hedge funds and other institutions in a large cross-section is still absent in finance literature due to unobservable order flow. To undertake this task, we examine the contemporaneous price impact of hedge fund and non-hedge fund order flow estimates, constructed using our method.

2. Informed trading:

   We expect that institutional order flow from a hedge fund is more informative about future stock price. Prior studies using observations at low frequencies also support this view. For example, Jagannathan, Malakhov, and Novikov (2010) find persistency of superior performance in hedge funds after mitigating a self-reporting concern in hedge fund databases. Agarwal, *et al.* (2013) and Argon, Hertzel, and Shi (2013) report positive and significant abnormal returns from the 13F confidential holdings of hedge funds. More recently, Cao, *et al.* (2017) provide evidence that stocks bought by hedge funds subsequently improve in price efficiency. However, analysis at a daily frequency using hedge fund order flow in a large cross-section is scarce. To fill the gap in the literature, we investigate the return predictability of hedge fund order flow and non-hedge fund order flow in: 1) cross-sectional regressions replicating Table II; and 2) an

Electronic copy available at: https://ssrn.com/abstract=2907612

investment analysis replicating Table IV.

3. Arbitrage trading:

   We conjecture that hedge funds trade on mispriced stocks. Akbas, *et al.* (2015) and Kokkonen and Suominen (2015) find a positive relation between monthly flow to hedge funds and return spread between under-valued and over-valued stock portfolios, suggesting that hedge fund flow attenuates market mispricing. Taking advantage of the finer granularity of our estimated hedge fund IOF, we conduct a stock-level analysis on arbitrage trading by hedge funds and other institutions at a daily frequency.

A. Trading cost management

In this subsection, we compare the trading costs associated with hedge fund order flow (HH^{HF}) versus non-hedge fund order flow (HH^{NHF}) in the same sample of common stocks on NYSE, AMEX, and Nasdaq between January 1999 an March 2012. If institutions actively time market liquidity or provide liquidity in execution, their order imbalance can generate a relatively smaller contemporaneous price impact. We test the trading skills of hedge funds and non-hedge funds in Table V, which presents the estimated coefficients of the following model:

$$\begin{aligned} R_{i,t} &= \alpha_t + \beta_t^{HF} HH_{i,t}^{HF} + \beta_t^{NHF} HH_{i,t}^{NHF} \\ &+ \beta_t^{T} TOI_{i,t} + \gamma_t^{B} SPRD_{i,t-1} + \gamma_t^{T} TURN_{i,t-1} + \sum_{k=1}^{5} \gamma_{t,k}^{R} R_{i,t-k} + \epsilon_{i,t}, \end{aligned}$$

where HH^{HF} and HH^{NHF} are hedge fund order flow and non-hedge fund order flow, respectively, and the other variables are the same as defined in Table I.

In univariate regressions, we find that both hedge fund and non-hedge fund IOFs have large and significant contemporaneous price pressure. Specifically, the coefficient of HH^{HF} is 11.142 with a t-statistic of 50.44 in Column (1). HH^{NHF} has a coefficient of 2.636 with a t-statistics of 60.89 in Column (2). In Column (3), we include both HH^{HF} and HH^{NHF}

Electronic copy available at: https://ssrn.com/abstract=2907612

in the same regression model. The multivariate regression results show that the coefficient of HH^{HF} decreases to 2.043 with a much smaller t-statistic of 5.53, but the coefficient of HH^{NHF} remains at the same magnitude with a t-statistic of 38.23. The difference between the two coefficients is 0.32. However, this number is not statistically different from zero. The price pressure further decreases in Column (4) when we include TAQ order imbalance, illiquidity, and lagged returns in the regression, but the result is qualitatively the same to Column (3). We find the order imbalances of both hedge funds and non-hedge funds generate positive contemporaneous price impact in the market. Although the hedge fund price impact is smaller than that of non-hedge funds, the difference is not statistically significant. Interestingly, the coefficient on the total market order imbalance, TOI, is much larger (4.403, t-statistic = 90.02) than those of the two institutional order imbalances, suggesting that institutions in general can have better skills in managing contemporaneous price impact. The results in this subsection suggest weak evidence, if any, that hedge funds manage trading costs better than their non-hedge fund counterparts at the aggregate level. It is possible that the liquidity timing skills documented by Aragon and Strahan (2012) and Cao et al. (2013) are not be limited to hedge funds only.

B. Cross-sectional return prediction

In this subsection, we turn to return predictability of hedge fund and non-hedge fund trades. We replicate the regression analysis in Table II. For brevity, Table VI report the coefficient estimates of order flow only, while the regressions always include the full set of control variables.

[Place Table VI about here]

In Column (1), we examine the information content of hedge fund order flow first. The estimated coefficient is 1.508 at the first lag with a t-statistic of 18.49. Similar to the total institutional order flow, HH^{HF} also experiences significant reversal in its pricing effect on

Electronic copy available at: https://ssrn.com/abstract=2907612

subsequent days as the coefficient estimates become negative and significant on days $t-3$ to $t-5$. However, even after five days, the cumulative price impact is still positive and significant. In Column (2), we find that non-hedge fund order flow also has predictive power for stock returns on the subsequent day with a coefficient of 0.139 and a t-statistic of 10.25. However, the pricing effect of non-hedge fund trades is only transitory, similar to the HH price impact in Table II. This is not surprising given the high correlation between HH and HH^{NHF}. When we use both HH^{HF} and HH^{NHF} in the same regression in Column (3), we find that the discrepancies are enhanced. Specifically, HH^{HF} obtains similar predictive ability at the first lag without significant reversals later, while HH^{NHF} loses its initial positive price impact with almost the same negative pricing effects at longer horizons.[11] These results indicate that hedge fund trades have permanent price impact while non-hedge fund trades have transitory price pressure only, supporting our conjecture that hedge funds have an information advantage over other institutions.

Next, we investigate the predictive power of HH^{HF} and HH^{NHF} in subsamples of stocks as a robustness test. In Table VII, we begin by conditioning based on firm size in the first three columns. We sort all stocks into tertile portfolios based on daily market capitalization and report the estimated coefficients of institutional net order flows from our base regression model for each portfolio. The control variables are included in the regressions, but omitted in reporting for brevity. The results can be summarized as follows: First, focusing on the pricing effect on the subsequent day, HH^{HF} has positive and significant coefficient estimates in all three firm groups, while HH^{NHF} only has significant pricing effects for the middle-sized firm group. Second, HH^{HF}'s pricing effect on the subsequent day adversely depends on the firm size for both statistical and economic significance. Specifically, HH^{HF} has a coefficient of 4.260 (t-statistic = 16.60) for small stocks and 0.719 (t-statistic = 5.53) for large stocks. In terms of economic significance, a one-standard deviation increase in HH^{HF} leads to an increase of 7.25 bp in the next day's stock return for small stocks and 1.38 bp

[11] Increasing the number of lagged IOF generates similar results in this test.

Electronic copy available at: https://ssrn.com/abstract=2907612

for large stocks. Third, the pricing effect of HH^{HF} is permanent in all three groups, with a small reversal in the small firm group, no reversal in the median group, and a continuation of positive effects in the large firm group over subsequent days.

[Place Table VII about here]

In the next three columns, we examine subsamples of stocks based on liquidity measured by the relative bid-ask spread. We find the following results: First, HH^{HF} has positive and significant coefficient estimates at the first lag in all three spread groups, while HH^{NHF} only generates insignificant or negative price impact at the first lag in all stock groups. Second, hedge fund's pricing effects are stronger for stocks with wide spreads. For example, the coefficient estimate of HH^{HF} is 3.110 (t-statistic = 13.79) for stocks with wide spreads and 0.834 (t-statistic = 6.59) for stocks with narrow spreads. A one-standard deviation increase in HH^{HF} is associated with an increase in the subsequent return by 6.50 bp for stocks with wide spreads and by only 1.37 bp for those with narrow spreads. The difference between the spread groups is consistent with the notion that stocks with wide spreads are less transparent because of higher arbitrage costs.

We then break the sample into an early period from 1999 to 2004, and a late period from 2005 to 2012 to investigate the time-series variation in return predictability. We report estimated coefficients in the subsequent two columns for these subperiods. HH^{HF} positively and significantly predicts the next day's return in both subperiods. The coefficient of HH^{HF} at the first lag slightly reduces from 1.742 (t-statistic = 10.64) in the early period to 1.488 (t-statistic = 14.44) in the late period. The coefficients suggest that a one-standard deviation increase of HH^{HF} leads to a return increase of 3.54 bp in the early period and 2.91 bp in the late period. At the second lag, HH^{HF} shifts from a positive and significant pricing effect in the early period to a mild reversal in the late period. HH^{NHF} has insignificant coefficient estimates in both periods. The difference between two subperiods is consistent with the idea that liquidity improvement is accompanied by the rise of high frequency trading that, in

Electronic copy available at: https://ssrn.com/abstract=2907612

turn increases price efficiency over time[12].

Finally, we separate the sample according to the primary stock exchange in the last two columns. Because the Nasdaq market is traditionally a dealers' market and has a different market structure from NYSE and AMEX, we examine Nasdaq listed and non-Nasdaq listed stocks separately. The results show that HH^{HF} has positive and significant predictive ability in both Nasdaq and non-Nasdaq subsamples, while the coefficient of HH^{NHF} is negative and significant in the non-Nasdaq subsample and insignificant in the Nasdaq subsample. Specifically, HH^{HF} has a coefficient of 1.718 with a t-statistic of 9.24 for NYSE and AMEX stocks, and a coefficient of 1.956 with a t-statistic of 14.91 for Nasdaq stocks.

Given the strong return predictive ability from HH^{HF}, we analyze investment strategies based on lagged HH^{HF} and HH^{NHF}. Replicating Table IV, we sort all stocks into decile portfolios based on IOF measures for hedge funds and non-hedge funds every day and calculate the average equal-weighted portfolio returns on the next day. Table VIII reports the average decile portfolio returns as well as the return differentials and Fama-French (2015) alphas between the top and bottom order imbalance decile portfolios.

[**Place Table VIII about here**]

We find that both strategies based on hedge fund and non-hedge fund trades generate positive and significant abnormal returns on the following day. The HH^{HF} strategy produces a positive alpha of 8 bp with a t-statistic of 11.04. Similarly, the HH^{NHF} strategy also generates a positive and significant alpha of 4.2 bp (t-statistic = 6.00), much lower than the HH^{HF} strategy. These alphas should be interpreted with caution again because the parameter estimation utilizes the full sample analyzed.

The results in this subsection are summarized into three findings: First, hedge fund trades can predict stock returns on the subsequent day and make a permanent price impact. Second, the return predictability of hedge funds outperforms other institutions. Lastly, the superior power of return prediction is robust in diverse subsamples. Such findings also suggest that

[12] See Hendershott, Jones, and Menkveid (2011) and Broggard, Hendershott, and Riordan (2014).

Electronic copy available at: https://ssrn.com/abstract=2907612

hedge funds may capture fundamental information and exploit their information advantage in trading.

C. *Arbitrage trading*

In this subsection, we study how institutions exploit well-known arbitrage opportunities documented in the literature. We construct a mispricing index, similar to Stambaugh, Yu, Yuan (2012, 2015), that is updated daily. We identify mispricing based on the following nine return anomalies with detailed calculation methods in the Appendix.

1. Net stock issues as in Ritter (1991) and Loughram and Ritter (1995)
2. Composite equity issues as in Daniel and Titman (2006)
3. Total accruals as in Sloan (1996)
4. Net operating assets as in Hirshleifer, et al. (2004)
5. Momentum as in Jegadeesh and Titman (1993)
6. Gross profitability as in Novy-Marx (2013)
7. Asset growth as in Cooper, Gulen, and Schill (2008)
8. Return on assets as in Fama and French (2006) and Chen, Novy-Marx, and Zhang (2011)
9. Investment-to-asset ratio as in Titman, Wei, and Xie (2004)

Using daily CRSP-Compustat merged data from January 1999 to March 2012, we update accounting-based anomalies on the reporting date. To construct the mispricing index, we first rank all stocks into decile portfolios from over-valued stocks to under-valued stocks based on each anomaly factor each day. Then, we assign the portfolio rank to every stock in that portfolio as an index for that particular anomaly. Therefore, stocks with higher mispricing index values are more likely to be under-valued. We take the average of the ranks of the nine anomaly signals as the mispricing index for a stock on a particular day. We only consider stock-day observations for which at least five anomaly factors can be calculated. Consistent

Electronic copy available at: https://ssrn.com/abstract=2907612

with the anomaly literature, we examine only common stocks with prices above $5 on NYSE, AMEX, and Nasdaq markets and exclude all stocks in financial and utility industries. After merging the anomaly index into our main sample, we have $1,800,557$ stock-day observations.

We examine arbitrage trading of hedge funds and non-hedge funds using a panel regression. If institutional investors exploit arbitrage opportunities implied by these anomalies, their order flow should be positively associated with the mispricing index. Table IX report the least squares dummy variables (LSDV) estimates which remove firm fixed effects in the following model:

$$\begin{aligned} OF_{i,t} &= \beta MISP_{i,t-1} + \gamma^F \text{FirstWeek}_{i,t-1} + \gamma^L \text{LastWeek}_{i,t-1} \\ &+ \gamma^{FX} \text{FirstWeek}_{i,t-1} \times MISP_{i,t-1} + \gamma^{LX} \text{LastWeek}_{i,t-1} \times MISP_{i,t-1} + \epsilon_{i,t}, \end{aligned}$$

where for each firm i on day t, OF is the estimated institutional order flow as indicated on the top of each column, FirstWeek (LastWeek) is a dummy variable for days in the first (last) week in a month, and $MISP$ is the mispricing index we compose.

[Place Table IX about here]

In the univariate regression of HH^{HF} on lagged mispricing index, we find that the coefficient estimate of the mispricing index is 0.833 with a t-statistic of 4.34. When we include the week dummies and their interactions with the mispricing index, we find the association between HH^{HF} and $MISP$ remains positive and significant. Moreover, unconditionally, hedge funds seem more likely to purchase stocks in the first week of a month and sell stocks in the last week of a month given the significant coefficient estimates of 0.049 and -0.110 on the two week dummies, respectively. We also find the hedge fund purchase at the beginning of a month is not likely to be motivated by mispricing as the interaction of $MISP$ and the first-week dummy has a negative and significant coefficient estimate, fully offsetting the positive coefficient on $MISP$.

Turning to non-hedge funds, on the other hand, we find that $MISP$ has a coefficient of

Electronic copy available at: https://ssrn.com/abstract=2907612

-4.315 with a t-statistic of -3.74. This negative relation is also robust in the multivariate regression with week dummies. Interestingly, although the impact of $MISP$ is completely the opposite for HH^{HF} and HH^{NHF}, the seasonality of order flow is similar for both types of funds. HH^{NHF} also becomes significantly more positive in the first week and significantly more negative in the last week of a month.

The results in this table suggest that hedge fund managers trade on asset pricing anomalies on a daily basis. However, the non-hedge fund institutions trade in the opposite way, potentially exacerbating anomalies despite well-documented patterns of abnormal returns from such trading strategies.

V. Conclusion

In this article, we propose a new method to estimate net institutional order flow (IOF) that extrapolates the quarterly relation between institutional ownership change and quarterly order flow information from TAQ and Ancerno to a higher frequency–daily in our case. We test the performance of our estimated IOF estimate, termed HH, by comparing its return predictive ability to three other IOF measures for common stocks listed on NYSE, AMEX, and Nasdaq markets from January 1999 to March 2012. The alternative IOF methods include: the size cut-off rule as in Lee and Radhakrisha (LR, 2000); Campbell, Ramadorai, and Schwartz's (CRS, 2009) method, which is similar to ours, but uses only the TAQ order imbalance; and the aggregate imbalance in Ancerno data (AN). We find that HH outperforms the other IOF measures in predicting future stock returns in the cross-section. The superior performance of HH comes from the fact that we utilize the trade information made available in both TAQ and Ancerno. The CRS method relies exclusively on the heterogeneity of institutional trading patterns across trade sizes that can be discerned using TAQ data. Though TAQ records all transactions on public exchanges, the measurement error in trader identity is enormous. Ancerno, on the other hand, provides cleanly identified

Electronic copy available at: https://ssrn.com/abstract=2907612

institutional trade records, while its coverage is rather limited. Combining these two data sources, we construct a more informative measure of institutional order flow than presented previous studies.

Our estimation method can also be applied to net order flow estimation for particular types of institutions. As an example, we investigate trading skills of hedge fund and non-hedge fund investors, as such direct empirical comparison at a relatively high frequency in a large cross-section is still absent from the literature. Our results indicate that hedge funds have superior trading skills as opposed to non-hedge fund institutions. Specifically, hedge fund trades are more informative about future stock prices with a positive and permanent impact, while non-hedge fund trades exhibit only transitory price pressure. Moreover, we find that hedge fund trades are positively associated with a mispricing index constructed using well-known stock return anomalies on the previous day, indicating that hedge funds purchase under-valued stocks and sell over-valued stocks. In sharp contrast, the non-hedge fund order flow is negatively associated with the same mispricing index. The results suggest that hedge funds exploit arbitrage opportunities from stock return anomalies while non-hedge funds trade in the opposite direction to exacerbate the anomalies. Overall, our findings indicate that the superior trading skills of institutional investors are largely driven by hedge funds. We do find that all institutional investors generate smaller contemporaneous price impact than retail investors, suggesting that the skills of managing trading costs may be widespread in the asset management industry.

The proposed estimation method for institutional order flow can be applied in other empirical studies that require institutional order flow estimates at a daily frequency. We apply the method to hedge funds and non-hedge funds in this study, but the method can also be applied to other types of institutional trading, such as long-term and short-term investors, or active and passive investors. Our analysis is based on stock-level institutional order imbalance, and it would be interesting to combine it with fund-level analysis to gain a more complete picture of the wider-reaching effects of institutional trading. We leave these

Electronic copy available at: https://ssrn.com/abstract=2907612

questions to future studies.

Electronic copy available at: https://ssrn.com/abstract=2907612

Appendix. Return anomalies

1. Net stock issues (Ritter (1991) and Loughram and Ritter (1995)): We calculate net stock issues as the growth rate of split-adjusted shares outstanding in the past 252 trading days between days $t - 252$ and t.

2. Composite equity issues (Daniel and Titman (2006)): We obtain composite equity issues by subtracting log change of shares outstanding from a cumulative log share adjustment factor in the past five-year window. The adjustment factor is equal to the log number of shares if one reinvestment all cash distributions back into the stock.

3. Total accurals (Sloan (1996)): We measure total accruals as non-cash working capital less than the depreciation expense, scaled by total assets in the previous fiscal year.

4. Net operating assets (Hirshleifer, et al. (2004)): We measure net operating assets by subtracting all operating assets from all operating liability and scaling it by the total assets in the previous fiscal year.

5. Momentum (Jegadeesh and Titman (1993)): Momentum factor is calculated as the cumulative return between days $t - 252$ and $t - 42$.

6. Gross profitability (Novy-Marx (2013)): We use gross profitability scaled by total assets in the same fiscal year.

7. Asset growth (Cooper, Gulen, and Schill (2008)): Asset growth is defined as the growth rate of total assets from the previous fiscal year to the current fiscal year.

8. Return on assets (Fama and French (2006) and Chen, Novy-Marx, and Zhang (2011)): We calculate return on assets as quaterly earnings scaled by book equity value in the last quarter.

9. Investment-to-asset (Titman, Wei, and Xie (2004)): We measure investment-to-asset by adding gross property, plant, and equipment to change in inventories. We scale it by book equity value in the last fiscal year.

Electronic copy available at: https://ssrn.com/abstract=2907612

REFERENCES

[1] Agarwal, Vikas, W. E. I. Jiang, Yuehua Tang, and Baozhong Yang, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, The Journal of Finance 68, 739-783.

[2] Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, Journal of Financial Economics 118, 355-382.

[3] Alexander, G. J., G. Cici, and S. Gibson, 2007, Does motivation matter when assessing trade performance? An analysis of mutual funds, Review of Financial Studies 20, 125-150.

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

[5] Anand, A., P. Irvine, A. Puckett, and K. Venkataraman, 2012, Performance of institutional trading desks: An analysis of persistence in trading costs, Review of Financial Studies 25, 557-598.

[6] Aragon, George O., Michael Hertzel, and Zhen Shi, 2013, Why do hedge funds avoid disclosure? Evidence from confidential 13f filings, Journal of Financial and Quantitative Analysis 48, 1499-1518.

[7] Aragon, G. O., and P. E. Strahan, 2012, Hedge funds as liquidity providers: Evidence from the lehman bankruptcy, Journal of Financial Economics 103, 570-587.

[8] Avramov, D., and R. Wermers, 2006, Investing in mutual funds when returns are predictable, Journal of Financial Economics 81, 339-377.

[9] Barras, L., O. Scaillet, and R. Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, Journal of Finance 65, 179-216.

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

[11] Boehmer, Ekkehart, and Eric K. Kelley, 2009, Institutional investors and the informational efficiency of prices, Review of Financial Studies 22, 3563-3594.

[12] Boone, Audra L., and Joshua T. White, 2015, The effect of institutional ownership on firm transparency and information production, Journal of Financial Economics 117, 508-533.

[13] Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-frequency trading and price discovery, Review of Financial Studies 27, 2267-2306.

Electronic copy available at: https://ssrn.com/abstract=2907612

[14] Busse, Jeffrey A., T. Clifton Green, and Narasimhan Jegadeesh, 2012, Buy-side trades and sell-side recommendations: Interactions and information content, Journal of Financial Markets 15, 207-232.

[15] Caglayan, Mustafa Onur, Umut Celiker, and Gokhan Sonaer, 2018, Hedge fund vs. Non-hedge fund institutional demand and the book-to-market effect, Journal of Banking & Finance 92, 51-66.

[16] Campbell, John, Tarun Ramadorai, and Allie Schwatz, 2009, Caught on tape, Journal of Financial Economics 92, 26.

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

[18] Cao, Charles, Bing Liang, Andrew W. Lo, and Lubomir Petrasek, 2017, Hedge fund holdings and stock market efficiency, The Review of Asset Pricing Studies rax015-rax015.

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

[20] Chakrabarty, Bidisha, Pamela C. Moulton, and Charles Trzcinka, 2017, The performance of short-term institutional trades, Journal of Financial and Quantitative Analysis 52, 1403-1428.

[21] Chen, H. L., N. Jegadeesh, and R. Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers, Journal of Financial and Quantitative Analysis 35, 343-368.

[22] Chen, Long, Robert Novy-Marx, and Lu Zhang, 2011, An alternative three-factor model, Working paper.

[23] Cohen, R. B., P. A. Gompers, and T. Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions, Journal of Financial Economics 66, 409-462.

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

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

[26] Daniel, K., and S. Titman, 2006, Market reactions to tangible and intangible information, Journal of Finance 61, 1605-1643.

[27] Douglas Foster, F., David R. Gallagher, and Adrian Looi, 2011, Institutional trading and share returns, Journal of Banking & Finance 35, 3383-3399.

Electronic copy available at: https://ssrn.com/abstract=2907612

[28] Edelen, Roger M., Ozgur S. Ince, and Gregory B. Kadlec, 2016, Institutional investors and stock return anomalies, Journal of Financial Economics 119, 472-488.

[29] Elton, E. J., M. J. Gruber, and C. A. Blake, 2011, Holdings data, security returns, and the selection of superior mutual funds, Journal of Financial and Quantitative Analysis 46, 341-367.

[30] Fama, E. F., and K. R. French, 2006, Profitability, investment and average returns, Journal of Financial Economics 82, 491-518.

[31] Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, Journal of Financial Economics 116, 1-22.

[32] Fama, Eugene F., and James D. Macbeth, 1973, Risk, return, and equilibirum: Empirical tests, Journal of Political Economy 81, 607-636.

[33] Frijns, Bart, Thanh D. Huynh, Alireza Tourani-Rad, and P. Joakim Westerholm, 2018, Institutional trading and asset pricing, Journal of Banking & Finance 89, 59-77.

[34] Gervais, S., R. Kaniel, and D. H. Mingelgrin, 2001, The high-volume return premium, Journal of Finance 56, 877-919.

[35] Griffin, J. M., J. H. Harris, and S. Topaloglu, 2003, The dynamics of institutional and individual trading, Journal of Finance 58, 2285-2320.

[36] Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, The Journal of Finance 47, 1977-1984.

[37] Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity?, Journal of Finance 66, 1-33.

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

[39] Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988, The Journal of Finance 48, 93-130.

[40] Henry, Tyler R., and Jennifer L. Koski, 2017, Ex-dividend profitability and institutional trading skill, The Journal of Finance 72, 461-494.

[41] Hirshleifer, D., K. W. Hou, S. H. Teoh, and Y. L. Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, Journal of Accounting & Economics 38, 297-331.

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

[43] Irvine, P., M. Lipson, and A. Puckett, 2007, Tipping, Review of Financial Studies 20, 741-768.

Electronic copy available at: https://ssrn.com/abstract=2907612

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

[45] Jame, Russell, 2017, Liquidity provision and the cross-section of hedge fund returns, Management Science.

[46] Jegadeesh, N., and S. Titman, 1993, Returns to buying winners and seling losers - implications for stock-market efficiency, Journal of Finance 48, 65-91.

[47] Kacperczyk, M., C. Sialm, and L. Zheng, 2008, Unobserved actions of mutual funds, Review of Financial Studies 21, 2379-2416.

[48] Kokkonen, Joni, and Matti Suominen, 2015, Hedge funds and stock market efficiency, Management Science 61, 2890-2904.

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

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

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

[52] Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, Journal of Finance 46, 733-746.

[53] Loughran, T., and J. R. Ritter, 1995, The new issues puzzle, Journal of Finance 50, 23-51.

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

[55] Nelson, C. R., and A. F. Siegel, 1987, Parsimonious modeling of yield curves, Journal of Business 60, 473-489.

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

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

[58] Piotroski, Joseph D., and Darren T. Roulstone, 2004, The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices, The Accounting Review 79, 1119-1151.

Electronic copy available at: https://ssrn.com/abstract=2907612

[59] Puckett, Andy, and Xuemin Yan, 2011, The interim trading skills of institutional investors, Journal of Finance 66, 601-633.

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

[61] Sloan, R. G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, Accounting Review 71, 289-315.

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

[63] Stambaugh, Robert F., Jianfeng Yu, and Y. U. Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, Journal of Finance 70, 1903-1948.

[64] Titman, S., K. C. J. Wei, and F. X. Xie, 2004, Capital investments and stock returns, Journal of Financial and Quantitative Analysis 39, 677-700.

[65] Wermers, R., 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, Journal of Finance 55, 1655-1695.

Electronic copy available at: https://ssrn.com/abstract=2907612

Table I. Summary statistics

This table shows the time-series averages of cross-sectional statistics for the sample during January 1999 to March 2012. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data. For each institutional order flow estimate, we calculate the daily order imbalance as the aggregate buyer-initiated volume minus the aggregate seller-initiated volume scaled by shares outstanding. HH is our proposed institutional order flow described in Section II.B. AN is calculated using Ancerno data. CRS is the estimate proposed by Campbell, Ramadoral, and Schwatz (2009). LR is the estimate of Lee and Radhakrishna (2000) using $ 5,000 cut-off rule. HH^{HF} and HH^{NHF} are our proposed hedge fund and non-hedge fund order flow, respectively, described in Section II.B. TOI is daily total order flow in TAQ based on Lee and Ready (1993) algorithm. $SPRD$ is daily relative spreads measured as twice the distance between daily close offer and bid prices scaled by the quote midpoint. $TURN$ is daily turnover ratio defined as trading volume over the number of shares outstanding. RET is daily risk-adjusted mid-quote stock return with respect to Fama-French (2015) five factors. Panel A reports discriptive statistics. Panal B presents correlation coefficients between institutional order flows and other variables.

Panel A. Descriptive Statistics

	Number of Dates	Avg Num. of Stocks	Mean	Stdev	Min	Med	Max
HH	3326	3073.2	0.025	0.158	-0.535	0.005	0.671
AN	3326	2042.1	0.004	0.137	-0.574	0.002	0.547
CRS	3326	3073.2	0.024	0.074	-0.163	0.006	0.357
LR	3326	3073.2	0.014	0.151	-0.558	0.002	0.646
HH^{HF}	3326	2893.5	0.006	0.020	-0.053	0.002	0.096
HH^{NHF}	3326	3071.6	0.019	0.121	-0.411	0.004	0.521
TOI	3326	3073.2	0.015	0.179	-0.694	0.003	0.779
$SPRD$	3326	3073.2	0.009	0.013	0.000	0.005	0.126
$TURN$	3326	3073.2	0.649	0.788	0.003	0.414	4.877
RET	3326	3068.9	0.000	0.027	-0.122	-0.001	0.153

Panel B. Correlation

	HH	AN	CRS	LR	HH^{HF}	HH^{NHF}	TOI
HH	1.000						
AN	0.875	1.000					
CRS	0.352	0.003	1.000				
LR	0.142	0.055	0.306	1.000			
HH^{HF}	0.712	0.512	0.459	0.134	1.000		
HH^{NHF}	0.957	0.886	0.397	0.164	0.638	1.000	
TOI	0.127	0.061	0.303	0.946	0.144	0.173	1.000
$SPRD$	-0.052	0.002	-0.107	-0.054	-0.064	-0.053	-0.056
$TURN$	0.148	-0.005	0.339	0.149	0.282	0.154	0.153
RET	0.096	0.103	0.069	0.241	0.088	0.119	0.305

Electronic copy available at: https://ssrn.com/abstract=2907612

Table II. Daily return prediction in the cross-section

This table presents Fama-MacBeth (1973) regression results for the following equation.

$$R_{i,t} = \alpha_t + \sum_{k=1}^{5} \beta_{t,k}^{I} IOF_{i,t-k} + \sum_{k=1}^{5} \beta_{t,k}^{T} TOI_{i,t-k} + \gamma_t^{B} SPRD_{i,t-1} + \gamma_t^{T} TURN_{i,t-1} + \sum_{k=1}^{5} \gamma_{t,k}^{R} R_{i,t-k} + \epsilon_{i,t},$$

where for stock i on day t, R is risk-adjusted mid-quote stock return with respect to Fama-French (2015) five factors, IOF is an institutional order flow estimate indicated on the top of each column. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP and Thomson Reuters's 13F data from January 1999 to March 2012. All variables are the same as defined in Table I. 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.

	HH	AN	CRS	LR
Intercept	-0.013**	-0.038***	-0.008	-0.015***
	(-2.57)	(-6.62)	(-1.61)	(-2.88)
IOF_{t-1}	0.122***	0.097***	-0.155***	-1.349***
	(11.71)	(6.39)	(-6.36)	(-14.31)
IOF_{t-2}	-0.020**	-0.006	-0.091***	0.125***
	(-2.29)	(-0.42)	(-5.04)	(2.83)
IOF_{t-3}	-0.032***	-0.024*	-0.074***	0.027
	(-3.98)	(-1.71)	(-4.10)	(0.59)
IOF_{t-4}	-0.048***	-0.031**	-0.098***	-0.000
	(-5.19)	(-2.39)	(-5.30)	(-0.00)
IOF_{t-5}	-0.039***	-0.018	-0.105***	0.027
	(-4.57)	(-1.37)	(-5.67)	(0.62)
TOI_{t-1}	0.123***	-0.039***	0.152***	1.337***
	(11.91)	(-3.39)	(13.99)	(15.22)
TOI_{t-2}	-0.102***	-0.070***	-0.092***	-0.210***
	(-12.47)	(-7.39)	(-10.89)	(-5.11)
TOI_{t-3}	-0.049***	-0.024**	-0.040***	-0.074*
	(-5.80)	(-2.35)	(-4.71)	(-1.77)
TOI_{t-4}	-0.026***	-0.013	-0.017**	-0.033
	(-3.13)	(-1.27)	(-2.00)	(-0.79)
TOI_{t-5}	-0.026***	-0.003	-0.017**	-0.045
	(-2.94)	(-0.28)	(-1.97)	(-1.13)
$SPRD_{t-1}$	2.364***	7.080***	2.264***	2.409***
	(10.38)	(9.29)	(9.97)	(10.56)
$TURN_{t-1}$	0.069***	0.056***	0.083***	0.066***
	(10.11)	(9.20)	(11.46)	(10.38)
RET_{t-1}	-1.215***	-0.727***	-1.205***	-1.545***
	(-7.51)	(-4.35)	(-7.56)	(-10.11)
RET_{t-2}	-1.081***	-1.042***	-1.060***	-1.041***
	(-9.21)	(-7.02)	(-9.12)	(-8.83)
RET_{t-3}	-0.653***	-0.423***	-0.663***	-0.648***
	(-6.78)	(-3.36)	(-6.96)	(-6.72)
RET_{t-4}	-0.404***	-0.409***	-0.430***	-0.435***
	(-3.92)	(-3.15)	(-4.20)	(-4.23)
RET_{t-5}	-0.205**	-0.257**	-0.235**	-0.226**
	(-2.00)	(-2.01)	(-2.32)	(-2.21)
Adjusted R^2	0.025	0.039	0.026	0.027
Observation	10,015,095	5,033,554	10,015,095	10,015,095

Electronic copy available at: https://ssrn.com/abstract=2907612

Table III. Horse races in daily return predictive ability

This table presents Fama-MacBeth (1973) regression results with competing institutional order flow estimates. We replicate the regression analysis in Table II using HH and another institutional order flow in the same regression. For brevity, we only report the coefficient estimates of order flow, while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Nasdaq from January 1999 to March 2012. 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.

	AN	CRS	LR
Intercept	-0.035***	-0.009*	-0.013**
	(-5.93)	(-1.69)	(-2.58)
HH_{t-1}	0.073***	0.188***	0.156***
	(3.24)	(18.23)	(15.59)
HH_{t-2}	-0.018	-0.012	-0.023***
	(-1.02)	(-1.41)	(-2.60)
HH_{t-3}	-0.022	-0.027***	-0.034***
	(-1.22)	(-3.24)	(-4.26)
HH_{t-4}	-0.040**	-0.038***	-0.046***
	(-2.03)	(-3.95)	(-4.96)
HH_{t-5}	-0.026	-0.027***	-0.039***
	(-1.49)	(-3.03)	(-4.66)
IOF_{t-1}	0.008	-0.289***	-1.407***
	(0.27)	(-10.98)	(-14.81)
IOF_{t-2}	0.014	-0.087***	0.124***
	(0.58)	(-4.73)	(2.83)
IOF_{t-3}	0.006	-0.060***	0.037
	(0.22)	(-3.17)	(0.83)
IOF_{t-4}	0.018	-0.073***	0.011
	(0.68)	(-3.88)	(0.24)
IOF_{t-5}	0.018	-0.083***	0.042
	(0.75)	(-4.32)	(0.98)
Adjusted R^2	0.041	0.027	0.028
Observation	5,033,554	10,015,095	10,015,095

Electronic copy available at: https://ssrn.com/abstract=2907612

Table IV. Investment strategies based on institutional order flow

This table reports the performance of investment strategies based on estimated institutional order flows (IOFs) as indicated on the top of each column. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. We sort all stocks into decile portfolios based on IOFs every day. The average portfolio returns on the next day are reported. Also reported are the return differentials and alphas with respect to the Fama-French (2015) five factors between the top and bottom decile portfolios. All variables are the same as defined in Table I. 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.

	HH	AN	CRS	LR
Low (%)	0.103	0.106	0.167	0.191
2	0.077	0.026	0.147	0.126
3	0.115	0.006	0.151	0.105
4	0.110	0.019	0.084	0.096
5	0.026	0.047	0.009	0.072
6	0.062	0.072	0.052	0.038
7	0.102	0.099	0.079	0.032
8	0.113	0.114	0.095	0.062
9	0.125	0.143	0.098	0.098
High	0.155	0.165	0.105	0.168
HML (%)	0.052***	0.059***	-0.062***	-0.023
	(6.81)	(6.96)	(-6.86)	(-1.59)
FF5 Alpha (%)	0.049***	0.056***	-0.063***	-0.007
	(6.92)	(6.99)	(-8.13)	(-0.67)

Electronic copy available at: https://ssrn.com/abstract=2907612

Table V. Contemporaneous price impact of institutional trading

This table presents Fama-MacBeth (1973) regression results of risk-adjusted mid-quote stock return on contemporaneous hedge fund order flow (HH^{HF}), non-hedge fund order flow (HH^{NHF}) and various control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. All variables are the same as defined in Table I. 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)	(3)	(4)
Intercept	-0.034***	-0.016***	-0.027***	-0.038***
	(-6.89)	(-3.40)	(-5.44)	(-6.53)
HH_t^{HF}	11.142***		2.043***	1.008***
	(50.44)		(5.53)	(2.91)
HH_t^{NHF}		2.636***	2.363***	1.418***
		(60.89)	(38.23)	(29.41)
TOI_t				4.403***
				(90.02)
$SPRD_{t-1}$				5.252***
				(17.99)
$TURN_{t-1}$				-0.149***
				(-14.55)
RET_{t-1}				-2.953***
				(-19.74)
RET_{t-2}				-1.894***
				(-16.52)
RET_{t-3}				-1.230***
				(-13.16)
RET_{t-4}				-0.825***
				(-8.31)
RET_{t-5}				-0.543***
				(-5.49)
Adjusted R^2	0.011	0.016	0.018	0.127
Observation	9,610,012	10,202,019	9,417,101	9,417,101

Electronic copy available at: https://ssrn.com/abstract=2907612

Table VI. Private information in hedge fund and non-hedge fund

This table replicates Fama-MacBeth (1973) regression in Table II using our proposed hedge fund (HH^{HF}) and non-hedge fund (HH^{NHF}) order flow estimates. For brevity, we only report the coefficient estimates of order flow variables, while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. All variables are the same as defined in Table I. 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)	(3)
Intercept	-0.016***	-0.013**	-0.016***
	(-3.00)	(-2.54)	(-3.06)
HH_{t-1}^{HF}	1.508***		1.603***
	(18.49)		(17.18)
HH_{t-2}^{HF}	-0.070		0.065
	(-1.03)		(0.79)
HH_{t-3}^{HF}	-0.175***		-0.026
	(-2.74)		(-0.32)
HH_{t-4}^{HF}	-0.322***		-0.096
	(-4.65)		(-1.24)
HH_{t-5}^{HF}	-0.233***		-0.020
	(-3.38)		(-0.24)
HH_{t-1}^{NHF}		0.137***	-0.015
		(10.25)	(-0.98)
HH_{t-2}^{NHF}		-0.024**	-0.028**
		(-2.07)	(-2.03)
HH_{t-3}^{NHF}		-0.035***	-0.032**
		(-3.34)	(-2.37)
HH_{t-4}^{NHF}		-0.063***	-0.055***
		(-5.36)	(-4.28)
HH_{t-5}^{NHF}		-0.055***	-0.049***
		(-5.04)	(-3.71)
Adjusted R^2	0.027	0.025	0.027
Observation	9,414,134	10,009,538	9,408,616

Electronic copy available at: https://ssrn.com/abstract=2907612

Table VII. Return prediction in subsamples

This table examines the predictive ability of estimated institutional order flows in subsamples based on firm characteristics. We replicate the regression analysis in Table VI. For brevity, we only report the coefficient estimates of our proposed hedge fund order flow (HH^{HF}) and non-hedge fund order flow (HH^{NHF}), while the regressions always include the full set of control variables. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. In first (second) three columns, we separate our sample into tertile subsamples based on market capitalization (relative bid-ask spread). In subsequent two columns, we divide the sample into into an early (1999-2004) and a late (2005-2012) subperiods. In the last two columns, we compare stocks listed on NYSE and AMEX versus Nasdaq. All variables are the same as defined in Table I. 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 the 1, 5, and 10 percent lavel, respectively.

	Size			Spread			Time		ExMarket	
	Small	Middle	Large	Wide	Medium	Narrow	Early	Late	NYMX	NASDAQ
Intercept	-0.033***	-0.037***	-0.019***	-0.024***	-0.054***	-0.038***	-0.032***	-0.003	-0.044***	-0.003
	(-4.30)	(-5.11)	(-3.08)	(-3.11)	(-8.64)	(-6.44)	(-3.58)	(-0.48)	(-6.29)	(-0.41)
HH_{t-1}^{HF}	4.260***	1.066***	0.719***	3.110***	1.773***	0.834***	1.742***	1.488***	1.718***	1.956***
	(16.60)	(6.67)	(5.53)	(13.79)	(12.30)	(6.59)	(10.64)	(14.44)	(9.24)	(14.91)
HH_{t-2}^{HF}	-0.560**	0.013	0.447***	-0.178	-0.008	0.286**	0.340**	-0.161*	0.197	-0.029
	(-2.36)	(0.09)	(3.51)	(-0.84)	(-0.07)	(2.45)	(2.40)	(-1.78)	(1.39)	(-0.25)
HH_{t-3}^{HF}	-0.269	-0.122	0.320**	-0.011	0.100	0.055	0.112	-0.139	0.138	-0.089
	(-1.22)	(-0.93)	(2.53)	(-0.05)	(0.81)	(0.48)	(0.87)	(-1.36)	(0.88)	(-0.78)
HH_{t-4}^{HF}	0.053	-0.033	0.128	0.037	-0.028	-0.246**	-0.142	-0.058	0.021	0.018
	(0.22)	(-0.25)	(1.02)	(0.19)	(-0.22)	(-2.30)	(-1.11)	(-0.62)	(0.11)	(0.16)
HH_{t-5}^{HF}	-0.308	-0.133	0.388***	-0.247	0.110	0.154	0.045	-0.073	0.205	-0.057
	(-1.30)	(-1.01)	(3.33)	(-1.30)	(0.90)	(1.29)	(0.32)	(-0.74)	(1.26)	(-0.50)
HH_{t-1}^{NHF}	-0.198***	0.068**	-0.023	0.005	-0.050**	-0.020	-0.045	0.010	-0.062***	0.007
	(-3.52)	(2.38)	(-1.32)	(0.11)	(-2.11)	(-1.04)	(-1.62)	(0.61)	(-3.08)	(0.31)
HH_{t-2}^{NHF}	-0.023	-0.034	-0.045**	-0.015	-0.011	-0.045**	-0.007	-0.046***	-0.021	-0.025
	(-0.43)	(-1.31)	(-2.52)	(-0.37)	(-0.50)	(-2.42)	(-0.28)	(-3.13)	(-1.08)	(-1.15)
HH_{t-3}^{NHF}	-0.063	-0.035	-0.039**	-0.110***	-0.049**	-0.018	-0.032	-0.032**	-0.028	-0.036*
	(-1.21)	(-1.39)	(-2.43)	(-2.76)	(-2.13)	(-1.00)	(-1.40)	(-2.02)	(-1.57)	(-1.78)
HH_{t-4}^{NHF}	-0.127**	-0.130***	-0.030*	-0.081**	-0.073***	-0.022	-0.054**	-0.056***	-0.047**	-0.071***
	(-2.42)	(-5.42)	(-1.78)	(-2.02)	(-3.26)	(-1.33)	(-2.33)	(-4.13)	(-2.17)	(-3.46)
HH_{t-5}^{NHF}	-0.019	-0.085***	-0.036**	-0.078**	-0.044**	-0.051***	-0.067***	-0.035**	-0.015	-0.081***
	(-0.37)	(-3.61)	(-2.20)	(-2.06)	(-2.11)	(-2.75)	(-2.81)	(-2.46)	(-0.86)	(-3.97)
Adjusted R^2	0.031	0.035	0.062	0.032	0.036	0.049	0.024	0.030	0.043	0.028
Observation	2,810,254	3,289,716	3,308,646	3,277,090	3,225,654	2,905,872	4,481,597	4,927,019	4,226,645	5,181,971

Electronic copy available at: https://ssrn.com/abstract=2907612

Table VIII. Investment analysis of hedge/non-hedge fund order flow

This table documents the performance of investment strategies based on our proposed hedge fund order flow (HH^{HF}) and non-hedge fund order flow (HH^{NHF}) described in Section II.B. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. We sort all the stocks into decile portfolios based on IOFs every day. The average portfolio returns on the next day are reported. Also reported are the return differentials and alphas with respect to the Fama-French (2015) five factors between the top and bottom decile portfolios. All variables are the same as defined in Table I. 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.

	HH^{HF}	HH^{NHF}
Low (%)	0.111	0.107
2	0.083	0.084
3	0.098	0.129
4	0.045	0.114
5	0.028	0.016
6	0.061	0.053
7	0.077	0.092
8	0.099	0.113
9	0.128	0.130
High	0.192	0.150
HML (%)	0.082***	0.043***
	(10.95)	(5.95)
FF5 Alpha (%)	0.080***	0.042***
	(11.04)	(6.00)

Electronic copy available at: https://ssrn.com/abstract=2907612

Table IX. Institutional trading on asset pricing anomalies

This table presents panel regression results with firm-fixed effect using the least squares dummy variable (LSDV) method for the following equation,

$$\begin{aligned} OF_{i,t} &= \beta MISP_{i,t-1} + \gamma^F \text{FirstWeek}_{i,t-1} + \gamma^L \text{LastWeek}_{i,t-1} \\ &+ \gamma^{FX} \text{FirstWeek}_{i,t-1} \times MISP_{i,t-1} + \gamma^{LX} \text{LastWeek}_{i,t-1} \times MISP_{i,t-1} + \epsilon_{i,t}, \end{aligned}$$

where for each firm i on day t, OF is an institutional order flow estimate as indicated on the top of each column, FirstWeek (LastWeek) is a dummy variable for the first (last) week in a month, and $MISP$ is a mispricing index composed from nine categorical variables for well-known stock return anomalies including net stock issues of Loughran and Ritter (1995); composite equity issues of Daniel and Titman (2006); momentum of Jegadeesh and Titman (1993); gross profitability of Novy-Marx (2013); return on assets of Chen, Novy-Marx, and Zhang (2010); total accurals of Sloan (1996); net operating assets of Hirshleifer, Hou, Teoh, and Zhang (2004); asset growth of Cooper, Gulen, and Schill (2008); investment-to-assets of Titman, Wei, and Xie (2004). HH^{HF} and HH^{NHF} are our proposed hedge fund and non-hedge fund order flow, respectively, described in Section II.B. The sample includes all common stocks listed on NYSE, AMEX, and Nasdaq with prices above five dollars that have information in TAQ, CRSP, and Thomson Reuters's 13F data from January 1999 to March 2012. All coefficient estimates are multiplied by 100. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

	HH^{HF}		HH^{NHF}	
$MISP$	0.833***	0.988***	-4.315***	-3.434***
	(4.34)	(4.53)	(-3.74)	(-2.62)
First-Week		0.049**		0.359***
		(2.18)		(2.66)
Last-Week		-0.110***		-0.298**
		(-5.57)		(-2.52)
First-Week $\times MISP$		-1.213***		-5.104**
		(-3.09)		(-2.16)
Last-Week $\times MISP$		0.123		-0.442
		(0.36)		(-0.21)
Firm Fixed Effect	Yes	Yes	Yes	Yes
Adjusted R^2	0.051	0.051	0.026	0.026
Observation	1,730,798	1,730,798	1,790,153	1,790,153

Electronic copy available at: https://ssrn.com/abstract=2907612