WARWICK

ECONOMICS

Institutional Liquidity Demand and the Internalization of Retail Order Flow : The Tail Does Not Wag the Dog

Yashar H. Barardehi, Dan Bernhardt, Zhi Da & Mitch Warachka

January 2022 No: 1394

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Institutional Liquidity Demand and the Internalization of Retail Order Flow: The Tail Does Not Wag the Dog*

Yashar H. Barardehi Dan Bernhardt
Zhi Da Mitch Warachka

January 17, 2022

Abstract

The decision of wholesalers to internalize retail order flow primarily reflects institutional liquidity demand. We first use the Tick Size Pilot to highlight this decision's influence on the retail trade imbalances denoted $Mroib$ by Boehmer et al. (2021). We then show that wholesalers internalize more retail order flow when institutional demand is higher, leading $Mroib$ to be inversely related to institutional order flow. Intraday returns move in the same direction as institutional price pressure but the opposite direction of $Mroib$. Moreover, $|Mroib|$ is highest when institutional trading costs are highest. Distant future returns display strong $\cup$-shaped patterns conditional on $Mroib$, consistent with a permanent liquidity premium driving the positive relation between these returns and the magnitude of $|Mroib|$.

Keywords: Retail Trade, Institutional Trade, Payment for Order Flow, Liquidity, Microstructure

*Barardehi (barardehi@chapman.edu) is at the Argyros School of Business & Economics, Chapman University and the U.S. Securities and Exchange Commission. Bernhardt (danber@illinois.edu) is at Department of Economics at the University of Illinois and the University of Warwick. Da (zda@nd.edu) is at the Mendoza College of Business, University of Notre Dame. Warachka (warachka@chapman.edu) is at the Argyros School of Business & Economics, Chapman University. The Securities and Exchange Commission disclaims responsibility for any private publication or statement of any SEC employee or Commissioner. This article expresses the authors' views and does not necessarily reflect those of the Commission, the Commissioners, or other members of the staff. Any errors are our own.

1 Introduction

The question of whether retail investors are informed has provoked an active debate and motivated competing strands of the asset pricing literature.[1] This debate continues, in part, due to the difficulty in observing retail trading activity.[2] Boehmer, Jones, Zhang, and Zhang (2021, henceforth BJZZ) address this unobervability issue by recognizing that in U.S. equity markets trade executions at sub-penny prices other than the quote mid-point *necessarily* originate from retail investors. The authors use this insight to develop a normalized "marketable retail order flow imbalance" measure denoted $Mroib$ from publicly-available data sources. BJZZ show that this imbalance predicts the cross-section of stock returns for several weeks, and attribute nearly half of its return predictability to informed trading by retail investors.

Although $Mroib$ is constructed from retail trade executions, our paper provides comprehensive evidence that institutional liquidity demand is the key determinant of $Mroib$. We show that the sub-penny-executed transactions underlying this imbalance reflect the incentives of wholesalers to "internalize" retail order flow—institutional investor demand for liquidity is the "dog" whose "tail" is the retail trade imbalance captured by BJZZ's algorithm. In practice, the vast majority of retail orders submitted to retail brokerages are routed to wholesalers who decide which orders to execute internally and which ones to reroute to trading venues. We show that these internalization decisions underlie most variation in $Mroib$ imbalances, and that greater imbalances reflect internalization choices in response to pressing liquidity demands of institutions facing high trading costs.[3] Figure 1 illustrates that (1) net institutional trade imbalance is inversely related to the imbalance in internal-

[1] See Barber and Odean (2000, 2008), Kumar and Lee (2006), Kaniel et al. (2008), Barber et al. (2008), Foucault et al. (2011), Kaniel et al. (2012), Kelley and Tetlock (2013), and Barrot et al. (2016), among others.

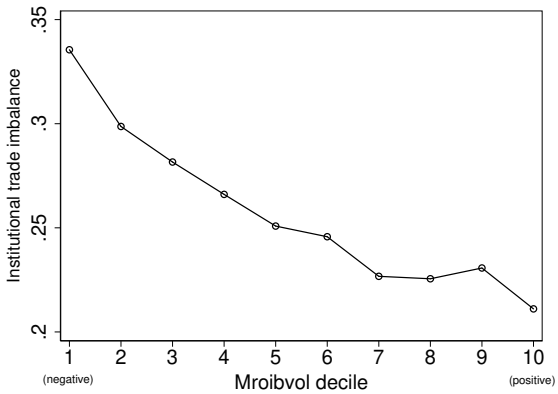
[2] This difficulty has led some researchers to rely on proprietary, non-representative datasets. Boehmer et al. (2021) contain a summary of these datasets.

[3] Hu (2009) reports that execution costs measured using ANcerno data are larger for buy trades that sell trades in down markets. That execution costs are higher when $Mroib$ is extremely negative than when it is extremely positive is consistent with the fact that liquidity is lower in down markets (Chordia et al. 2002).

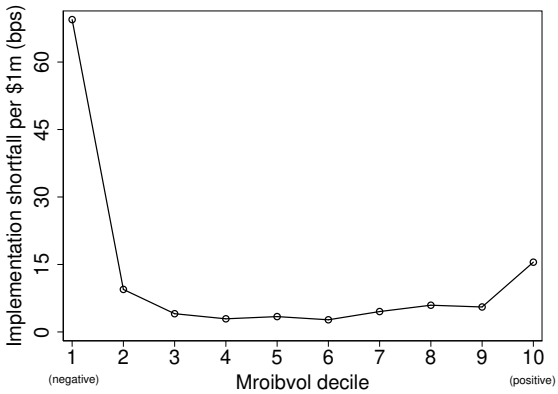
ized retail trade volume, and (2) institutional trading costs are highest when imbalances in internalized retail trade are most extreme. Our companion paper (Barardehi et al. 2022) shows that the absolute value of $Mroib$ offers a stock-level liquidity measure that captures institutional trading costs and is priced in the cross-section of stock returns, even in the recent post-RegNMS era where traditional liquidity measures are not priced.

Figure 1: ***Mroib* vs. Institutional Trade Imbalance and Implementation Shortfall.** This figure plots mutual-fund net trade imbalance and institutional-trade implementation shortfalls against the net imbalance in retail order flow internalization. Weekly cross-sections are sorted into deciles of net imbalance in the volume of internalized retail trades ($Mroibvol$), with delice 1 and 10 reflecting the most negative and positive quantities of $Mroibvol$, respectively. Average institutional trade imbalance and implementation shortfall measures are calculated by $Mroibvol$ deciles each week. The time-series means of these averages are plotted against $Mroibvol$ deciles. The sample includes NMS-listed stocks that can be matched with ANcerno data from 2010–2014.

Panel A: Institutional trade imbalance



Panel B: Institutional implementation shortfall



To understand the underpinnings of $Mroib$, one needs to understand the retail order execution process in U.S. equity markets. First, a retail investor chooses an order type (market, marketable limit, non-marketable limit) as illustrated in Figure 3. Market and marketable limit orders seek immediate execution at or above prevailing best-quoted prices, while non-marketable limit orders require execution at prices that are better than prevailing best-quoted prices. Next, the retail broker decides whether to route the order to an exchange or an off-exchange venue (wholesaler or dark pool). When an order is routed to a wholesaler, the wholesaler decides whether to internalize the order and make the requisite payment for

order flow (PFOF) to the retail broker or to reroute the order as illustrated in Figure 4. The "best execution" duty of brokers requires internalized order execution at prices no worse than the best quoted prices. This normally leads to minimal, under 0.1¢, price improvements for retail order execution relative to these quoted prices (BJZZ).[4]

A wholesaler's motive to internalize (purchase) retail order flow and pay PFOF stems from their off-exchange market-making activities. In such market-making, a wholesaler trades as a principal against order flow from opposing sides, buying low and selling high. Off-exchange interactions of wholesalers with retail investors occur largely via internalization. In contrast, wholesaler off-exchange interactions with institutional investors primarily occur on Alternative Trading Systems (ATSs) such as dark pools, where trading is anonymous, or on Single Dealer Platforms (SDPs) owned by the wholesaler. In contrast to an ATS, a wholesaler trading on its SDP acts as a principal against a select set of institutional investors who access the wholesaler's SDP at a premium, thereby revealing their willingness to pay for liquidity.

Most importantly for *Mroib*, wholesalers do not internalize all retail orders. Instead, wholesalers optimize internalization depending on the type and amount of retail order flow and the liquidity demands of institutions. Contrary to interpretations in BJZZ, in addition to marketable orders, wholesalers internalize a *non-trivial* share of non-marketable limit orders. In fact, SEC Rule 606 filing disclosures reveal that about 25% of non-marketable limit orders are internalized.[5] Moreover, even though non-marketable limit orders are less profitable to internalize (as they are priced closer to the National Best Bid-Offer midpoint), wholesalers offer them PFOF that is, on average, over double that offered to marketable orders. This higher

[4]For example, FINRA Regulatory Notice 21-23 reminds member entities that "...firms that provide payment for order flow for the opportunity to internalize customer orders cannot allow such payments to interfere with their best execution obligations ...." U.S. Securities & Exchange Commission (2021) describes "best execution" as that "at the most favorable terms reasonably available under the circumstances, generally, the best reasonably available price." Observing the regulatory oversight, retail brokers justify receiving PFOF on internalized retail order flow by offering price improvements, relative to the best quoted price, to retail investors.

[5]According to FINRA Regulatory Notice 01-30, since November 17, 2000 "...all broker/dealers that route customer orders in equity and option securities ...[must] ... make publicly available quarterly reports about the routing of customer orders." This requirement is known as Rule 606 of Regulation NMS.

PFOF reflects competition from exchanges that provide higher liquidity rebates to non-marketable (liquidity-providing) orders than marketable (liquidity-consuming) orders. We show that the willingness of wholesalers to internalize non-marketable limit orders and offer them high PFOF signals high institutional liquidity demand that makes such costly internalization profitable. These non-marketable limit orders are the marginal order type internalized by wholesalers, internalized typically only when liquidity demands of institutions are high.

We provide a simple framework that describes the profit-maximizing internalization choices of a wholesaler who interacts with retail investors via internalization and with institutional investors via their SDP. The wholesaler faces non-trivial inventory costs from holding positions that deviate from their preferred holdings. We show that in the absence of institutional liquidity demand, the wholesaler internalizes roughly equal amounts of retail buy and sell orders, resulting in a balanced (near zero) level of $Mroib$. In contrast, high institutional demand leads the wholesaler to internalize more retail orders on the opposing side of institutional orders, especially more costly non-marketable limit orders, resulting in greater imbalances in $Mroib$. Thus, we show how variation in institutional liquidity demand drives the variations in $Mroib$ imbalances, especially those associated with greater PFOF.

We use the Tick Size Pilot to demonstrate formally that wholesaler decisions to internalize limit orders determine $Mroib$. We first examine the effects of an exogenous increase in the quoted spread that preserves the minimum penny tick size. An exogenously widened spread increases the off-exchange liquidity provision profits of wholesalers, and we show that this increases internalization. Conversely, a joint increase in the minimum quoted spread *and* the tick size (a) sharply reduces internalization and (b) greatly increases $|Mroib|$. These effects reflect that the larger tick discourages the submission of market orders by retail investors by raising their risk of execution at far less favorable prices. The reduction in retail market orders reduces internalization and results in non-marketable limit orders comprising a larger share of internalized order flow. The accompanying increase in $|Mroib|$ reflects

the relatively greater influence of non-marketable limit orders, whose costly internalization becomes justified when there is unusually high institutional liquidity demand.

After replicating BJZZ's results, we extend their analysis along multiple dimensions. First, using data from ANcerno and FINRA, we document that a more negative $Mroib$, which results from the internalization of more retail sell orders than retail buy orders, is associated with more institutional buy volume than sell volume. Second, we show that more negative $Mroib$ is also associated with the covering of short positions through buying. Third, we show that institutional order flow imbalances and $Mroib$ are both more extreme when institutional trading costs are higher. For example, when implementation shortfalls are higher, quoted spreads are wider, depth is lower, and off-exchange trading at the mid-point is abnormally lower. The association between more extreme $Mroib$ and high institutional trading costs is consistent with internalization by wholesalers being a more costly source of liquidity that is accessed largely when institutional liquidity demand is high. Fourth, consistent with this high demand, but not informed retail trading, contemporaneous *intraday* prices move in the same direction as institutional trade imbalances and therefore in the *opposite* direction of $Mroib$.[6]

We next provide cross-sectional regressions of stock returns on $Mroib$. We find that higher $Mroib$ is associated with higher near-term future weekly returns (e.g., subsequent 12 weeks) but *lower* distant future weekly returns (e.g., weeks 39 through 60), which is inconsistent with informed retail trading,. The near-term return predictability of $Mroib$ is reconciled by price reversals following price pressure from persistent institutional trading, especially institutional buying (Hendershott and Seasholes 2007, Akepanidtaworn et al. 2021). Thus, negative current $Mroib$ (institutional buying, retail selling) tends to be associated with lower future returns for several weeks. Decomposing contemporaneous daily returns into intraday and overnight returns sheds more light on the liquidity-driven nature of these dynamics, as we document intraday institutional buy price pressure, followed by overnight reversals.

[6]To clarify, BJZZ's algorithm constructs $Mroib$ exclusively from transactions executed during regular trading hours. Thus, intraday returns are the relevant metric for examining $Mroib$'s price impact.

Finally, we uncover a characteristic liquidity premium component that drives $Mroib$'s relationship with long-term future weekly returns. While contemporaneous and first week returns monotonically increase with $Mroib$, the relation between $Mroib$ and distant future weekly returns is ∪-shaped. By week 12, the returns associated with the highest and lowest $Mroib$ deciles are over double those of the median $Mroib$. This ∪-shaped pattern extends beyond a year and indicates that larger internalized retail trade imbalances, arising from *both* buy and sell orders, reflect higher institutional trading costs that are compensated for by higher expected returns.[7] This result motivates Barardehi et al. (2022)'s use of $|Mroib|$ as a stock-specific measure of *institutional* liquidity costs. While traditional liquidity measures are no longer priced in recent years, this institutional liquidity measure yields large monthly liquidity premia even after implementing the most conservative filters commonly used in empirical asset pricing.

2 Literature Review

Our paper contributes to the literature on order flow internalization in equity markets. Easley et al. (1996) and Bessembinder and Kaufman (1997) report that internalized orders are less informed. Ernst et al. (2021) find evidence of delayed release of information regarding internalized order flow to the market, providing incentives for institutions to pursue such order flow to conceal their intended position size. By establishing how retail order internalization varies with institutional liquidity demand, our empirical contribution highlights the dependence of internalization on the incentives of wholesaler and details how retail investors provide liquidity to institutional investors through internalization.

Consequently, we also contribute to the literature on liquidity provision by retail investors (e.g., Kaniel et al. 2008; Kaniel et al. 2012). Consistent with Kelley and Tetlock (2013), we

[7] The persistence of institutional buy orders prevents the ∪-shaped pattern from appearing earlier. At longer horizons the liquidity premium delivers the ∪-shaped pattern as price pressure from institutional trading recedes.

find that the use of retail non-marketable retail to meet institutional liquidity demand underlies return predictability of retail order flow. Barrot et al. (2016) uses proprietary data to identify liquidity provided by retail investors that does not receive compensation because (i) retail investors trade before the price pressures from institutional trading are fully realized (ii) retail investors don't unwind their positions before price pressures revert. We extend these insights using comprehensive data that covers all NMS common shares. We also identify the U.S. equity market mechanisms that link a subset of retail trades with institutional trades and provide direct evidence that institutional liquidity demand determines the internalization of retail order flow by wholesalers.

Theoretical models have identified conditions under which order flow internalization harms market quality, liquidity, or welfare (e.g., Battalio and Holden 1995; Bernhardt et al. 2001; Parlour and Rajan 2003; Parlour and Seppi 2003). However, empirical studies motivated by these predictions find modest support (e.g., Battalio et al. 1997; Battalio et al. 2003; Peterson and Sirri 2003; Battalio 2012). We contribute to this debate by showing that non-marketable limit orders are the marginal order type in the internalization process. Were these non-marketable orders to reach the limit order book, they would improve liquidity and (if round lot orders) tighten bid-ask spreads on exchanges. However, PFOF arrangements facilitate the execution of non-marketable limit orders at marginally-improved prices by wholesalers, preventing these orders from reaching exchanges when the demand for liquidity is high.

While a simple interpretation of *Mroib*'s return predictability suggests that retail investors are informed, such an interpretation is at odds with institutional details and our extended empirical findings,[8] which are all reconciled by properties of institutional liquidity demand. Indeed, we find that large negative and large positive internalized retail order flow imbalances both predict higher distant future weekly returns than intermediate values

[8]See also Barber et al. (2021), who show that when internalized trade volume is abnormally high (top decile), extremely positive *Mroib* (top quintile) is followed by negative abnormal returns, suggesting that informed retail trade does not underlie this outcome.

of $Mroib$, indicating that cross-sectional variation in distant future returns is driven by a liquidity premium captured by the absolute value of $Mroib$.

Finally, we contribute to the literature identifying and understanding differences between intraday and overnight returns. Cliff et al. (2008) and Berkman et al. (2012) document that overnight returns are positive and intraday returns are negative, on average. Hendershott et al. (2020) show that CAPM holds overnight but not intraday, and attribute intraday deviations from CAPM to noise trading. Bogousslavsky (2021) finds that arbitragers tend to close positions near the end of a trading day. Intraday return variation induced by closing arbitrage positions allows a mispricing factor to explain intraday returns but not overnight returns. Lou et al. (2019) report that intraday and overnight returns exhibit strong persistence vis à vis past intraday and overnight returns, respectively, but strong reversals relative to overnight and intraday returns. They posit that clientele effects underlie these patterns. However, we establish that persistence in institutional order flow leads to the accumulation of price pressure during consecutive trading days that is partially reversed each night. This partial overnight reversal in conjunction with daytime persistence explains the distinct autocorrelations of daytime versus overnight returns. This reflects that liquidity premia are primarily embedded into intraday returns, driving deviations of intraday returns from CAPM.

3 Institutional Details and Hypothesis Development

The execution of a retail order in U.S. equity markets follows one of several paths depending on the decisions of three entities: a retail investor, a retail broker (broker-dealer), and an off-exchange market maker (wholesaler).[9] The retail investor chooses an order type and may, but rarely does, indicate a preferred trading venue. Instead, the retail broker decides whether to route the retail order to (i) a registered exchange, (ii) an off-exchange trading venue (ATS), or (iii) a wholesaler. In practice, wholesalers handle a large fraction of re-

[9] A wholesaler is an example of an over-the-counter (OTC) market maker.

tail orders. Thus, wholesaler decisions to internalize retail order flow are crucial and these decisions are influenced by several factors.

Order type choice: A retail order is "directed" if a retail investor specifies a particular trading venue(s). Directed orders comprise a tiny fraction of the orders received by brokers. For example, about 0.01% of the orders received by TD Ameritrade in the first quarter of 2020 were directed. By default, the retail broker determines where to route "non-directed" orders.

Retail investors choose an order type from alternatives that include market, marketable limit, and non-marketable limit order types.[10] Figure 3 illustrates the relevant order types in our analysis.[11] As the average (equally-weighted) quoted spread in our sample is 6¢, suppose the current best bid and ask prices are \$9.97 and \$10.03, respectively, when a retail trader submits an order. A market order demands immediate execution at the best available price. Ignoring price improvement, a market buy order would be executed at the best possible ask price at the time of execution, which is \$10.03 if the best ask price does not change in the interim. A marketable limit order also seeks immediate execution at the best price, but specifies a price equal to the current best quote. Thus, should the best price move against the investor, the limit order may not be executed. A marketable buy limit order at \$10.03 will either be executed at the best price (\$10.03 or lower) or enter the limit order book at \$10.03 if the best ask price increased above \$10.03 in the interim. A non-marketable buy limit order specifies a price below \$10.03, while an *attractive* non-marketable buy limit order specifies a price below \$10.03 but above the quote mid-point of \$10.00.[12]

Order routing by a broker-dealer: A retail broker's decisions regarding order routing reflects its monetary incentives. Routing to exchanges is rewarded by make-take rebates, while

[10] Our categorization of order types is consistent with Rule 606 filings. The "other" order type category includes orders that are similar to non-marketable limit orders such as stop or stop-loss orders.

[11] Our examples describe buy orders but the same description applies to sell orders.

[12] Rapid updates in the order book preclude permanent distinctions between marketable and non-marketable limit orders. A change in the best quoted price can cause a non-marketable order to become marketable or vice versa. We refer to order types based on their status at the time a wholesaler receives them.

routing to wholesalers is rewarded by payment for order flow (PFOF). PFOF represents a negotiated cash payment from a wholesaler to a retail broker that allows the wholesaler to trade, as a principal, against retail order flow from the retail broker.[13] In practice, retail brokers commonly route *all* of their order flow to wholesalers according to common negotiated terms. As retail brokers must follow "best execution" regulatory requirements, the wholesaler offers price improvement (PI) to the retail investor submitting the order to ensure its execution price is never worse than the best quoted price displayed on exchanges at the time of the transaction.[14] This sub-penny (more generally, sub-tick) price improvement (PI) is typically no more than 0.1¢ (or 0.1 tick) per share according to BJZZ.[15]

It is important to distinguish quoted prices from execution prices. For limit orders, a retail investor *quotes* a price for the order, while market orders seek execution at the best available prices. Rule 612 of RegNMS generally requires all orders, including limit orders, to be quoted at penny increments. However, retail trade execution is governed by the "best execution" duties of retail brokers. Thus, the execution of any retail limit or market order may involve sub-penny price improvements, and hence sub-penny execution prices.[16] Some exchanges also offer Retail Liquidity Programs (RLPs) that allow for sub-penny execution prices when retail orders can be matched inside the bid-ask spread, but such transactions are rare.[17]

Internalization by a wholesaler: The process by which wholesalers trade against retail order flow is referred to as internalization. In May 2012, internalized retail order flow comprised roughly 8% of all trading volume on NMS stocks (Tuttle 2014). Wholesalers are usually registered brokers, but they are not subject to the rules of registered exchanges. Most notably,

[13]Retail brokers may route orders to off-exchange trading venues (ATSs) but the economic incentive to do so is unclear for retail brokers who are not affiliated with the ATS receiving the routed order flow.

[14]See FINRA Regulatory Notice 21-23 for details on best execution. On exchanges, the SEC's Order Protection Rule guarantees execution at the national best quoted price.

[15]To clarify, the best prices quoted on exchanges are often better than the best displayed orders (i.e., NBB or NBO) because odd-lot orders between these best prices are not displayed. See U.S. Securities & Exchange Commission (2021), Staff Report for details on Equity and Options Market Structure Conditions.

[16]See the response to question 13 in Rule 612 FAQs.

[17]See NYSE Retail Liquidity Program's Fact Sheet.

wholesalers can execute trades at sub-penny prices despite the 1¢ minimum tick size. This flexibility allows wholesalers to coordinate with retail brokers and execute retail orders at sub-penny prices after offering price improvements that fulfill their "best execution" duties.

While broker-wholesaler negotiations may specify per share PFOF and PI levels, wholesalers do not internalize all retail order flow routed to them via retail brokers. Table 1 reports the distribution of order types across all non-directed and all internalized orders, along with average PFOF for each order type. Market orders and marketable limit orders account for a disproportionately large share of non-directed orders receiving PFOF, indicating that wholesalers prefer to internalize market and marketable limit retail orders relative to non-marketable limit orders. The U.S. Securities & Exchange Commission (2010) indicates that essentially all retail market and marketable limit orders are routed to wholesalers who profit from executing these orders by acting as an intermediary and earning twice the half spread. Simple calculations reveal that the share of non-marketable limit orders receiving PFOF is only one fourth that of market orders. Assuming that all retail orders receiving PFOF also receive price improvement,[18] it follows that one quarter of non-marketable limit orders receive price improvement. The relatively low frequency with which non-marketable limit orders receive price improvement suggests that those receiving price improvement are attractive limit orders that specify a price on the same side of the quote midpoint as the best quote (see Figure 3). Thus, these orders are correctly classified by BJZZ's algorithm. More important, despite receiving price improvement far less often, non-marketable limit orders that are internalized receive over double the PFOF per share as market and marketable limit orders. That non-marketable limit orders are more costly to internalize and also less profitable due to the inside-quote pricing suggests that they are more likely to be internalized when high institutional liquidity demand offers high compensating profits.

Most wholesalers own Single Dealer Platforms (SDPs), also known as ping pools, where

[18]This assumption is suggested by the fact that per share payment for market and marketable limit orders was exactly 0.120000¢ for TD Ameritrade in the first quarter of 2020.

a select set of institutions and institutional brokers trade against that wholesaler.[19] An institution may "ping" a wholesaler, which signals an unusually high demand for liquidity, thereby encouraging the wholesaler to internalize more retail order flow in order to profit from providing liquidity to the institution. By 2017, over 2.5% of all trading in NMS stocks occurred in ping pools,[20] or roughly 30% of all internalized order retail flow. Thus, when institutional liquidity demand is high, institutions are prepared to pay more to wholesalers, allowing wholesalers to pay sufficient PFOF to compete with exchange rebates while offering marginal price improvements to retail investors.[21] Figure 4 illustrates the market-structure aspects relevant for retail order flow internalization/execution and PFOF.

Market and marketable limit orders are less expensive for a wholesaler to internalize than attractive non-marketable limit orders. Referring to the example, a market buy at $10.03 is more profitable to fill than an attractive non-marketable buy limit order at $10.01 or $10.02. Even so, a wholesaler may profit from executing an internalized marketable limit buy order or an attractive non-marketable limit buy order at a price at or below the $10.00 midpoint against a counter party who is a retail investor submitting sell orders or an institutional investor pinging the wholesaler to indicate a strong selling interest on its SDP.

Overall, internalizing attractive non-marketable limit orders is profitable when the demand for liquidity by institutions is sufficiently high. Thus, the marginal internalized order type is typically an attractive non-marketable limit order. Reflecting their marginal nature, these orders are internalized at roughly one quarter the frequency of market orders.

The above institutional details suggest the following hypotheses:

[19] Trading that does not occur on exchanges or ATSs has attracted the attention of regulators. For example, FINRA Regulatory Notice 18-28 describes the nature of SDP trading, a major component of non-ATS trading, and highlights the agency's transparency concerns that led to Regulatory Notice 19-29, which expanded the transparency of OTC trading volume in December 2019.

[20] Trader VIP Clubs, 'Ping Pools' Take Dark Trades To New Level, *Bloomberg*, Jan 16, 2018.

[21] Brokers can route orders to a wholesaler or route orders to exchanges and receive a rebate of up to 0.3¢ per share. That the average PFOF for non-marketable limit orders slightly exceeds 0.3¢ is consistent with competition from liquidity-making rebates offered by exchanges.

Hypothesis 1: *Informed trading is not the primary determinant of imbalances in the number or volume of retail orders receiving price improvement. Instead, off-exchange sub-penny trade executions reflect the economic incentives of wholesalers to internalize retail order flow.*

Hypothesis 2: *The internalization of non-marketable limit orders and the corresponding price improvements for these orders are determined by institutional liquidity demand. This demand reflects institutional trading costs, which determine the economic incentives of wholesalers to internalize retail order flow.*

Hypothesis 3: *The ability of $Mroib$ to predict future returns is a liquidity-driven phenomenon. In the near-term, it reflects the reversals from the short-term price pressure associated with institutional liquidity consumption, but a characteristic (long-run) liquidity premium component related to institutional trading costs underlies why more extreme $Mroib$ predicts higher distant future returns.*

4 Data

We follow BJZZ in constructing our main sample. Our sample spans January, 2010 through December, 2014, covering common shares listed on the NYSE, AMEX, and NASDAQ.[22] We use daily open and close price information from CRSP's Daily Stock File to calculate three measures of daily returns: the standard close-to-close (CC) return, the open-to-close intraday (ID) return, and the close-to-open overnight (ON) return. Our construction accounts for overnight adjustments and, to minimize variations due to bid-ask bounce, is based on the midpoint of the best quoted prices at close. We then aggregate daily log-return observations into overlapping 5-day rolling windows to construct daily cross-sections of 5-day (weekly) returns, as in BJZZ. We include a stock-day observation in our sample if it had a closing price of at least $1 at the end of the previous calendar month.

[22] We do not include 2015, which is in BJZZ's sample because our ANcerno institutional trade data ends in 2014. Unreported results verify that all findings that do not require ANcerno data are robust to adding 2015.

We use BJZZ's algorithm to construct measures of internalized retail order flow. Using TAQ data, we focus on round-lot orders executed off-exchange (transactions with exchange flag "D") that feature sub-penny execution prices.[23] A transaction is classified as a retail buy order if the sub-penny (sub-tick) increment exceeds 0.6¢ and is classified as a retail sell order if the sub-penny increment is less than 0.4¢.[24] We construct daily, normalized measures of imbalance in internalized retail order flow trade frequency and trade volume. $Mroibtrd = (Mrbtrd - Mrstrd)/(Mrbtrd + Mrstrd)$ divides the difference between the number of internalized retail buy and internalized retail sell orders by their sum, while $Mroibvol = (Mrbvol - Mrsvol)/(Mrbvol + Mrsvol)$ is the normalized difference in internalized trade volume. Table 2 reports the summary statistics for these measures, which closely match the BJZZ summary statistics. We then aggregate these daily observations of normalized internalized retail order flow imbalances into overlapping 5-day rolling windows to construct daily cross-sections of 5-day (weekly) internalized retail order flow imbalances.

ANcerno data provides institutional trade sizes, buy versus sell indicators, execution prices, and stock identifiers for the 2010–2014 period.[25] We aggregate institutional buy and sell trade separately at the stock-day level to construct institutional buy and sell volumes. Using these volumes, we construct the institutional analogue of $Mroibvol$ denoted $Inroibvol$. We also construct implementation shortfall measures. For each institutional buy trade, the implementation shortfall equals the execution price minus that day's open price divided by the open price and scaled by the trade's dollar value in millions. Similarly, for each in-

[23] As in BJZZ, our findings are robust to including odd-lots.

[24] This algorithm has minimal, if any, mis-classifications for transactions that correspond to non-binding penny quoted spreads. For example, sample 606 filings from the fourth quarter of 2020 for E*TRADE, TDAmeritrade and Charles Schwab reveal that PFOF for market and marketable limit orders ranged between 0.09 to 0.19 cents per share, while the average PFOF for non-marketable limit orders did not exceed 0.34 cents per share. Furthermore, the price improvement offered to retail traders is typically very small (0.01 cents per share as documented by BJZZ) compared to PFOF.

[25] In our sample, ANcerno trading volume accounts for about 7% of all trading volume reported by CRSP, on average. Days with no institutional trading reported by ANcerno are treated as missing observations. The low share of overall volume does not invalidate the representativeness of ANcerno data (Puckett and Yan 2011; Anand et al. 2012; Jame 2018). However, to reduce noise in some analyses we focus on stocks with a greater-than-average share of trading volume on ANcerno relative to their overall trading volume.

stitutional sell trade, the implementation shortfall equals that day's open price minus the execution price divided by the open price and scaled by the trade's dollar value in millions. The mean implementation shortfall at the stock-day level is given by the value-weighted average across buy and sell implementation shortfalls. To maintain consistency with BJZZ's analysis, we aggregate institutional trading outcomes over 5-day rolling windows to construct daily cross-sections of 5-day (weekly) institutional trading outcomes.

We construct stock characteristics using information from CRSP's Daily and Monthly Stock Files and Compustat. Using CRSP data, we construct each stock's return volatility using daily observations from the preceding month (VOLAT). A stock's book-to-market (BM) ratio equals its most recent book value of equity divided by its market capitalization from the previous month.[26] Past return measures include the previous month's return (RET_{-1}) and the compound return over the preceding 5 months ($RET_{(-6,-2)}$). The previous month's turnover (TO) equals the ratio of the previous month's share volume to shares outstanding, and Size equals the firm's market capitalization at the end of the previous month.

We obtain the identifying information for control and treatment stocks in the U.S. Tick Size Pilot program (TSP) from FINRA's website, focusing on Test Groups 1 and 2 of the program.[27] For each stock, we construct daily observations over the 10 trading days prior to implementation of TSP on 10/03/2016 as well as the 10 trading days after full implementation on 10/17/2016.[28] From Daily TAQ's Trades, Quotes, and NBBO files, we obtain trade and quote information to match off-exchange transactions executed at sub-penny prices with the national best bid and ask prices at the time of transaction based on millisecond timestamps. Then, for each stock-day, we construct the following outcome variables: (1) the absolute value

[26]Book value is defined as Compustat's shareholder equity value (`seq`) plus deferred taxes (`txdb`).

[27]Exogenous changes in the minimum quoted spread and price increments for these two test groups are most relevant for our analysis. The exogenous changes that differentiate Test Group 3 from Test Group 2 are not relevant for our analysis. Unreported results reveal qualitatively similar findings for the two groups.

[28]Implementation consists of three phase-ins with different subsets of control stocks experiencing tick size changes on 10/03/2016, 10/10/2016, and 10/17/2016. For more details about the Tick Size Pilot program see https://www.sec.gov/rules/sro/nms/2015/34-74892.pdf.

of $Mroibtrd$; (2) the absolute value of $Mroibvol$; (3) size-weighted average relative percentage price improvement, which divides the relative price improvement for a sub-penny-executed transaction (i.e., the difference between the best quoted price and the transaction price) by the mid-point of best bid and ask; (4) total dollar-denominated price improvement, which is the sum of dollar relative price improvements across all sub-penny-executed transactions; (5) the total share volume of trades receiving price improvement; and (6) the size-weighted average sub-tick (sub-penny) fraction of trades receiving price improvement.

5 Replication and Extension of BJZZ

We first replicate the findings of BJZZ. We then extend their analysis to provide evidence that liquidity provision by retail investors to institutional investors underlies both near-term and distant future weekly returns conditional on $Mroib$.[29] Table 2 provides summary statistics that closely match those reported in Table I of BJZZ, confirming that our construction of $Mtoibtrd$ and $Mroibvol$ parallels BJZZ's methodology. We estimate the predictability of weekly returns conditional on $Mroibvol$ by estimating:

$$R_{j,w+i} = c_w^0 + c_w^1 Mroibvol_{j,w-1} + c_w^{2\top} \text{controls}_{j,w-1} + u_{j,w+i}, \tag{1}$$

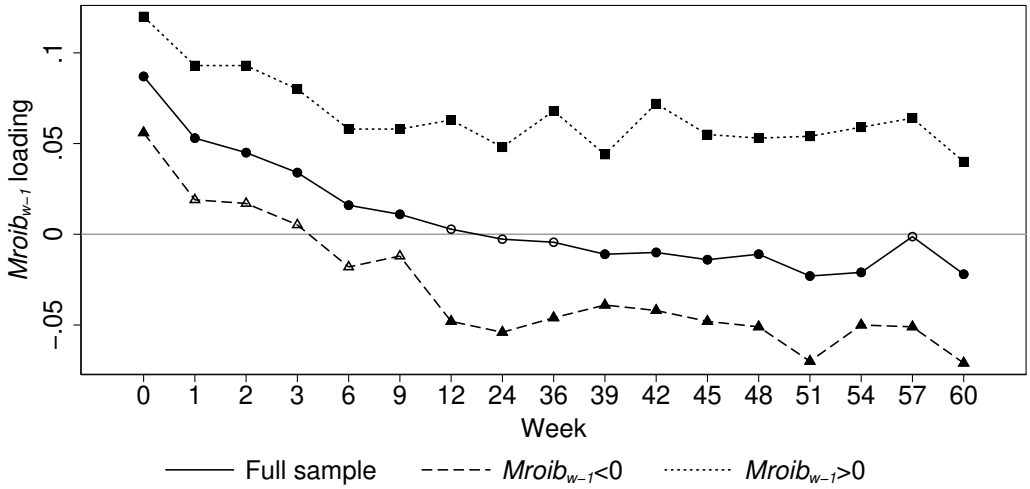
where $R_{j,w+i} \in \{CCR_{j,w+i}, IDR_{j,w+i}, ONR_{j,w+i}\}$ denotes weekly (rolling 5-day) close-to-close, intraday, and overnight returns, respectively, of stock j in week $w+i$. $Mroibvol_{j,w-1}$ denotes the imbalance in the trading volume of internalized retail order flow receiving price improvement in the previous week. We estimate equation (1) both unconditionally and conditional on the *sign* of $Mroibvol_{j,w-1}$ to examine its return predictability separately when this order flow imbalance is negative and positive. Control variables include the previous week's return (R_{w-1}) in percentage points, the previous month's return (RET_{-1}), the return over the five months prior to the last month ($RET_{(-7,-2)}$), return volatility (VOLAT), as

[29] Our sample period spans 2010–2014, while BJZZ's spans 2010–2015.

well as the natural logs of turnover (ln(TO)), market capitalization (ln(Size)), and book-to-market ratio (ln(BM)). Following BJZZ, we estimate equation (1) using Fama-Macbeth regressions, featuring Newey-West corrected standard errors with 6 lags.

Table 3 presents the estimation results for $i = 0$. The second column of this table corresponds to the second column of Table III in BJZZ. Our point estimate ($\hat{c}_w^1$) of 0.087% is nearly identical to BJZZ's estimate of 0.09%. The coefficients for the control variables are also similar to those reported by BJZZ. We next extend BJZZ by (i) estimating weekly return predictability of $Mroibvol_{j,w-1}$ for up to 60 weeks ahead, extending far beyond the 12 weeks that BJZZ consider, (ii) characterizing this return predictability conditioning on the sign of $Mroibvol_{j,w-1}$, and (iii) decomposing returns into intraday and overnight components.

Figure 2: **Internalized Order Flow and the Cross-sections of Future Weeks' Returns.** This figure shows the associations between internalized retail order flow and future week $w + i$ returns (in %), with $i \in \{0, 1, 2, 3, 6, 9, 12, 24, 36, 39, 42, 45, 48, 51, 54, 57, 60\}$. Returns reflect the quoted mid-points at close. Week $w + i$'s returns cross-section is decomposed by the sign of $Mroibvol_{w-1}$. According to equation (1), week $w + i$ returns in each sample are regressed on $Mroibvol_{w-1}$, whose loadings are plotted against time. Estimates are based Fama-Macbeth regressions, featuring Newey-West corrected standard errors with 6 lags. Sample includes NMS common shares from Jan 2010 – Dec 2014, excluding observations with previous month-end's closing price below $1. Statistically significant and insignificant loadings at the 10% type one error are identified by filled and hollow symbols, respectively.



Striking evidence obtains. As Figure 2 illustrates, the coefficients on $Mroibvol_{j,w-1}$ be-

come *uniformly negative* after 39 weeks, inconsistent with informed trading underlying the return predictability.[30] Moreover, although a negative $Mroibvol_{j,w-1}$ yields a positive coefficient for the current week's close-to-close return ($i = 0$), this coefficient declines and becomes *negative* even by week $w + 6$, again inconsistent with retail sell orders being informed, since "retail sell order flow" realizes weekly losses due to persistent price appreciation after 6 weeks. In contrast, a positive $Mroibvol_{j,w-1}$ always yields a positive coefficient for weekly returns across all horizons (see Table 4 for tabulated results).

Decomposing returns into intraday and overnight returns, uncovers further heterogeneous asymmetries in the coefficients according to the sign of $Mroibvol_{j,w-1}$. For overnight returns, $\hat{c}_w^1$ is positive following negative $Mroibvol_{j,w-1}$ (retail selling, institutional buying), but negative and insignificant following positive $Mroibvol_{j,w-1}$ (retail buying, institutional selling). Barclay and Hendershott (2003) and Jiang et al. (2012), among others, show that overnight price movements are information-driven. However, the insignificant negative relation between net retail buying imbalances and next week's overnight returns indicates that retail buys are also not informed.[31] Moreover, informed retail trading cannot explain why $\hat{c}_w^1$ switches signs for intraday returns when $Mroibvol_{j,w-1}$ switches signs.[32]

6 The Economics of Internalization

This section highlights the key economic features underlying our results. We first provide a framework that shows how the economic choices of wholesalers drives the association between institutional liquidity demand and $Mroib$. We show that in the absence of institutional liquidity demand, the internalization incentives of the wholesaler result in minimal $Mroib$ imbalances, whereas high institutional liquidity demand leads to highly imbalanced $Mroib$. We

[30] Our index starts at $i = 0$, while BJZZ's index starts at $k = 1$.

[31] Further, retail short selling is limited, suggesting that informed trading does not underlie the association between net retail selling imbalances and next week's overnight returns.

[32] Table 4 shows that the asymmetry in the predictability of close-to-close returns also holds for intraday and overnight returns, which is further at odds with retail investors being informed.

also argue that internalized retail orders receiving smaller price improvements are likely associated with greater PFOF and are more strongly related to heightened institutional liquidity demand faced by the wholesaler. We then exploit exogenous shocks driven by the U.S. Tick Size Pilot to draw causal inference about the economic choices of wholesalers as predicted by our simple framework. Finally, we establish that (1) $Mroib$ is negatively related to institutional order flow from both long-only investors and short sellers; and (2) more imbalanced $Mroib$ is associated with higher institutional trading costs and abnormally low stock liquidity.

6.1 Wholesaler Incentives, *Mroib*, and Institutional Liquidity

Our setting illustrates the economic incentives underlying a wholesaler's decisions regarding which retail order types to internalize, and the buy/sell composition of internalized orders.

Suppose that the public information value of a share is V, and there is a four tick spread, so the bid is $\$(V - 2t)$ and the ask is $\$(V + 2t)$. The distribution of retail orders routed by the broker-dealer to a wholesaler is given by

- n^s_{-2} marketable sell orders at $\$(V - 2t)$
- n^s_{-1} limit sell orders at $\$(V - t)$
- n^s_0 limit sell orders and n^b_0 limit buy orders at $\$V$
- n^b_{+1} limit buy orders at $\$(V + t)$
- n^b_{+2} marketable buy orders at $\$(V + 2t)$

The wholesaler has the option of internalizing a retail order in return for giving the broker-dealer payment for order flow, where a wholesaler's "best execution" duties mandate that retail traders also receive price improvement. Alternatively, a wholesaler can reroute a retail order directly to an exchange, in which case all rebates (or fees) go to the retail broker, where the rebate for liquidity-making limit orders exceeds that for liquidity-taking market orders.

The broker-dealer has market power leading her to obtain $PFOF_j > l_{\pm j}$ in return for routing a type $\pm j$ order $j \in \{\pm 2, \pm 1, 0\}$ to the wholesaler, where $l_{\pm j}$ is the expected liquidity rebate (accounting for execution risk) from rerouting a type $\pm j$ retail order to exchanges. To give perspective, in 2020, TDAmeritrade received PFOF of \$0.0012 for each liquidity-taking marketable order, and an average of about \$0.0034 for liquidity-making limit orders, which exceeds the maximum liquidity rebate of \$0.003 that an exchange can offer (an inverted exchange pays far less for liquidity-taking orders—see Battalio, Corwin and Jennings (2016) for details on make/take rebate/fee schedules). We use $PI_j > 0$ to denote the price improvement given to a type $\pm j$ order. We assume that $PFOF_j + PI_j < 0.005$; this ensures that the wholesaler obtains positive profits from simultaneously filling say a limit sell order quoted at $\$V$ and a limit buy order quoted at $\$(V + t)$, while leaving its inventory unchanged.

Finally, we assume that it is costly for the wholesaler to hold inventory that deviates by q from its preferred inventory level of 0. We assume that these costs rise convexly in q, i.e., $c(q) - c(q-1)$ is strictly increasing in q, where $c(1) - c(0) < l_{\pm j}$ indicates that tiny deviations from optimal inventory levels are not that costly.

We first highlight the economic forces for balanced levels of internalized retail trade in the absence of institutional trade. When a wholesaler is not "pinged" by an institution in this setting, it is strictly profitable for the wholesaler to internalize marketable sell orders and limit sell orders at $\$(V-t)$ simultaneously with marketable buy orders and limit buy orders at $\$(V+t)$, yielding the wholesaler a strictly positive profit of at least $2t - 2PFOF_1 - 2PI_1 > 0$. Thus, at least $min\{n^s_{-2} + n^s_{-1}, n^b_2 + n^b_1\}$ will be filled on both sides by the wholesaler's internalization. Assume without loss of generality that $n^s_{-2} + n^s_{-1} \geq n^b_2 + n^b_1$.

We make the following assumptions to minimize the number of cases that we have to enumerate, recognizing that all cases have qualitatively similar implications: (a) $n^s_{-2} \leq n^b_2 + n^b_1$, and (b) $n^s_{-2} + n^s_{-1} \leq n^b_2 + n^b_1 + n^b_0$. Then, after simultaneously filling these orders, the distribution of the remaining retail orders to be handled by the wholesaler takes the form

- 0 marketable sell orders at $\$(V-2t)$
- $n^s_{-2}+n^s_{-1}-(n^b_2+n^b_1)$ limit sell orders at $\$(V-t)$
- n^s_0 limit sell orders and n^b_0 limit buy orders at $\$V$
- 0 limit buy orders at $\$(V+t)$
- 0 marketable buy orders at $\$(V+2t)$

Next observe that it is optimal for the wholesaler to internalize any remaining limit sell orders at $\$(V-t)$ by holding inventory, stopping at the inventory imbalance of q^* where

$$\begin{aligned} t-(c(q^*)-c(q^*-1)) &\geq t-PFOF_1-PFOF_0-PI_1-PI_0 \\ &> t-(c(q^*+1)-c(q^*)). \end{aligned}$$

That is, the wholesaler stops when the marginal profit from internalizing by holding increased unbalanced inventory would be less than the marginal profit from simultaneously filling a non-marketable limit sell order at $\$(V-t)$ and a non-marketable limit buy order at $\$V$.

Note that the wholesaler would hold more unbalanced inventory when she pays greater price improvement for non-marketable limit orders (i.e., when she pays higher PI_1 and PI_0). Reducing PI_1 and PI_0 to the minimum positive increment possible while satisfying a wholesaler's "best execution" duties by raising the price improvement PI_2 for marketable orders to leave compensation to retail traders unchanged would raise wholesaler profits by reducing inefficient holding of excess unbalanced inventory (i.e., by reducing q^*). This suggests that wholesalers should offer less PI for non-marketable limit orders than marketable limit orders.

When $n^s_{-2}+n^s_{-1}-(n^b_2+n^b_1)>q^*$, the wholesaler fills all limit buy orders at $\$V$ along with the remaining limit sell orders at $\$(V-t)$. The dealer then submits all remaining limit orders[33] at $\$V$ to exchanges. This "residual" retail order flow is empirically undetectable

[33] That is, the n^s_0 limit sell orders, and the $n^b_0-q^*-(n^s_{-2}+n^s_{-1}-(n^b_2+n^b_1))$ remaining limit buy orders.

by BJZZ's algorithm, while the algorithm identifies internalized orders receiving sub-penny price improvements as the wholesaler fills them. Thus, absent institutional demand, for $n^s_{-2} + n^s_{-1} \leq n^b_2 + n^b_1 + q^*$, internalization order imbalances equal

$$|Mroibvol| = \frac{|(n^s_2 + n^s_1) - (n^b_{-2} + n^b_{-1})|}{n^b_2 + n^b_1 + n^s_{-2} + n^s_{-1}},$$

reaching a maximum at $n^s_{-2} + n^s_{-1} = n^b_2 + n^b_1 + q^*$ of

$$|Mroibvol| = \frac{q^*}{2(n^b_2 + n^b_1) + q^*}.$$

For $n^s_{-2} + n^s_{-1} > n^b_2 + n^b_1 + q^*$, $|Mroibvol|$ again decreases because the numerator remains q^*, while the denominator rises due to the "crossing" of limit buy orders at $\$V$.

These details yield the following key points. Absent institutional demand, we predict that (a) internalization of retail orders will be roughly balanced, (b) the wholesaler will pay more PFOF plus PI for non-marketable limit orders than for marketable orders due to the higher liquidity rebates offered by exchanges to liquidity-making orders than liquidity-taking orders, but (c) an optimal design of PI will feature less PI for non-marketable limit orders than for marketable orders, and (d) independently of these payments the profits from internalizing orders that are further from the public information value of the asset are higher.

Now suppose there is significant institutional demand. Internalized order flow is an expensive source of liquidity for institutions. To see why, first note the straightforward direct effect—an institution interested in buying shares must compensate a wholesaler for the profits that the wholesaler would otherwise obtain by internalizing retail sell orders. More subtly, the institution must also compensate a wholesaler for the foregone possibility of using the internalized retail sell orders to fill retail buy orders without distorting the wholesaler's inventory—the retail sell order flow that a wholesaler uses to fill an institutional buy order in a SDP cannot be used to simultaneously fill retail buy orders from which the wholesaler would also profit on via riskless principal trading. Finally, a wholesaler may have significant

bargaining power in her negotiations with the institution. This logic implies that an institution interested in consuming liquidity on a SDP must offer a wholesaler a price $p_b > \$(V+2t)$ to buy and a price $p_s < \$(V-2t)$ to sell.

Net institutional demand, when non-zero, is often large relative to retail order flow, reflecting the much larger positions that institutions take, and the fact that there is little point for an institution to ping a wholesaler for a small position. To begin, suppose that the net institutional demand is very high relative to retail order flow. Concretely, suppose that the institution is buying with a demand that exceeds $n_2^s + n_1^s + n_0^s + q_b^*$ where q_b^* solves

$$\begin{aligned} p_b - V - (c(q_b^*) - c(q_b^* - 1)) &\geq 0 \\ &> p_b - V - (c(q_b^* + 1) - c(q_b^*)). \end{aligned}$$

Then a wholesaler will internalize all retail orders on the other side that it receives $(n_{-2}^s + n_{-1}^s + n_0^s)$, and continue to accommodate the institutional demand only up to the point $(n_2^s + n_1^s + n_0^s + q_b^*)$ where the marginal profit from internalization equals the marginal increase in inventory costs. All retail orders submitted on the same side as the institutional demand are rerouted to other trading venues so $|Mroibvol|$ (and $|Mroibtrd|$) take their maximum values of one, and the sign of $Mroib$ is the opposite of that on net institutional demand.

From this point, as one reduces net institutional demand, one eventually reaches the level $(n_2^s + n_1^s + n_0^s + q_b^*)$ below which a wholesaler now accommodates all institutional demand. To do this, the wholesaler uses all retail orders on the other side plus distorting its inventory to the minimum extent needed, while still rerouting all retail orders on the same side as the institutional demand to exchanges. Thus, on this range, $|Mroib|$ remains maximal, not varying with net institutional demand, as the marginal order is accommodated out of wholesaler inventory.

With further reductions, one reaches a level of net institutional demand at which the marginal inventory cost just falls below the profit from filling a marketable retail order. This

level is reached earlier for more liquid stocks that have lower inventory imbalance costs. At this point, a wholesaler starts to internalize marketable retail orders on the same side as net institutional demand (rising one-for-one with decreases in net institutional demand until all such marketable orders are filled), causing $|Mroib|$ to fall with reductions in net institutional demand.

As net institutional demand falls further so does *Mroib*, as first more attractive limit orders on the same side as net institutional demand are internalized, and then limit orders at $\$V$ on the opposite side begin to be rerouted to other trading venues instead of being internalized. It follows that small *Mroib* imbalances are an indication of the balanced or near balanced net institutional demand, while sufficiently large *Mroib* imbalances indicate unbalanced net institutional liquidity demand of the opposite sign of *Mroib*.

6.2 Minimum Tick Sizes and Internalization

In this section, we exploit the design of the U.S. Tick Size Pilot to conclude that variation in $Mroibtrd$ and $Mroibvol$ reflects the internalization decisions of wholesalers. We first establish that the profitability of off-exchange liquidity provision drives wholesalers to internalize retail orders. We then show that non-marketable orders are the marginal order type considered for internalization, and that such orders tend to receive the least price improvement.

The SEC implemented the U.S. Tick Size Pilot program (TSP) On October 3, 2016. This program offered an experimental design for studying the causal impact of the minimum tick size on trading outcomes. The program included 2,400 securities. To ensure that stocks were randomly assigned to control and treatment groups, stocks were sorted into 27 categories based on share price, market-capitalization, and trading volume terciles. Across these categories, stocks were randomly assigned to three treatment groups that each contained 400 stocks. Treatment stocks in Test Group 1 were subject to a a minimum quoted spread of 5¢ but could trade at price increments of 1¢—the *quote rule* (Rindi and Werner 2018). Treat-

ment stocks in Test Groups 2 and 3 were subject to a minimum quoted spread of 5¢ and had to trade at price increments of 5¢—the *trade rule* (Rindi and Werner 2018). Test Group 3 stocks were also subject to a Trade-At Prohibition provision that is less relevant for our study.

A key exception to the minimum tick size applied to retail trades. Although retail trades had to be quoted using the minimum tick size, they could be executed at sub-penny prices off-exchange. However, for Test Groups 2 and 3, the program required a minimum price improvement of \$0.005 should the broker-dealer/wholesaler decide to offer price improvement. BJZZ's algorithm is designed to detect sub-penny execution prices in a 1¢ tick size regime, but it can be scaled to detect sub-tick execution prices in any tick size regime. To accomplish this, for Test Groups 2 and 3, after the activation of the Trade Rule, we re-scale the command in BJZZ's algorithm that classifies trades according to small versus large sub-penny increments by a factor of 5. Thus, borrowing BJZZ's notation, we replace "$Z_{jt} = 100 * \text{mod}(P_{jt}, 0.01)$" by "$Z^5_{jt} = 20 * \text{mod}(P_{jt}, 0.05)$", where Z^5_{jt} reflects the *sub-tick* execution price (P_{jt}) increment when the tick size is \$0.05. With this scaling, $Z^5_{it} \in [0, 1]$ and transactions can be again classified into retail buy and retail sell trades as in Section 4.

The TPS provides an ideal setting to study the economics of retail order flow internalization by wholesalers since the experiment impacts (i) the order type choices of investors and (ii) the profitability of off-exchange liquidity provision (Rindi and Werner 2018). These impacts allow us to conclude that variation in $Mroibtrd$ and $Mroibvol$ is determined by wholesaler decisions to internalize specific retail orders and offer price improvement. We use the following Difference-in-Difference (DiD) methodology to examine the causal impact of a tick size change:

$$X_{j,d} = b^g_0 + b^g_1(\text{Post}_d) + b^g_2(\text{Treat}^g_j) + b^g_3(\text{Post}_j) \times (\text{Treat}^g_d) + u_{j,d}. \tag{2}$$

Here $d \in [-11, -1]$ indexes the 11 trading days ending on 10/02/2016, and $d \in [0, 10]$ indexes the 11 trading days beginning on 10/17/2016.[34] $X_{j,d}$ is stock j's outcome variable

[34]Our event window excludes the 10 trading days spanning 10/03/2016 through 10/16/2016 to account

on trading day d; Post_d is an indicator variable that equals 0 if $d < 0$ and 1 if $d \geq 0$. Treatment_j^g is an indicator variable that equals 0 if stock j is in the control group and 1 if stock j is in the treatment group for Test Group $g \in \{1, 2\}$. The coefficient b_3^g captures the treatment effects associated with Test Group g. To ensure that estimated treatment effects are unaffected by outliers, we use both OLS and quantile (median) regressions to estimate equation (6.2). Following the standard practice in the literature (e.g., Rindi and Werner 2019, Griffith and Roseman 2019, and Albuquerque et al. 2020), we condition estimates on quoted spread levels prior to the introduction of TSP.

Table 5 presents estimation results for Test Group 1, and Figure 5 provides complementary visual evidence. The quote test tends to marginally increase relative price improvement (as a percentage), especially for treatment stocks with tighter pre-TSP spreads.[35] However, the quote rule raises the average and median volume of sub-penny-executed trades by 9% and 63% relative to the corresponding intercept, respectively. As a result, similar increases are discernible in the total dollar-amount of daily price improvement.[36] This suggests that wholesalers internalize retail orders more aggressively in response to the quote rule. The effects are stronger for stocks with tighter pre-TSP quoted spreads—stocks that are more likely to have binding quote test restrictions. Lastly, the quote rule tends to reduce sub-penny increments for trades involving stocks with tight pre-TSP spreads. This indicates that investors switched to liquidity-providing limit orders that are more expensive to internalize, limiting the sub-penny price improvement a wholesaler is willing to pay. The opposite effect applies to stocks with wide pre-TSP spreads, suggesting that because the internalization of

for the staggered phase-in of tick size changes for treated stocks. There were three phase-ins of treated stocks in Test Groups 1 and 2 stocks: 5 stocks from each group on 10/03/2016, 92 stocks from each group on 10/10/2016, and the remaining 303 stocks on 10/17/2016.

[35]Note that the effects on "Relative %-PI" for stocks with non-binding pre-TSP spreads are negative. This likely reflects the fact that quoting within bid-ask spreads that exceed 5¢ is now restricted by the minimum 5¢ spread. As a result, some non-marketable limit orders must be quoted at prices closer to the best quoted price, mechanically reducing the distance between their execution price and the best quoted price on the same side of the quotes midpoint.

[36]Rindi and Werner (2018) find no discernible effect on consolidated volumes of treated stocks in TSP, indicating that our findings are likely orthogonal to any stock-level volume effect.

non-marketable limit orders became more profitable with a 5¢ minimum spread, wholesalers were more willing to provide to internalize these limit orders.

Consider a low spread stock for which the 5¢ minimum spread reflects an exogenously-widened quoted spread. For example, suppose marketable limit buy and sell orders were quoted at best prices of \$10.02 and \$9.99, respectively, before the spread was widened to \$10.03 and \$9.98. This widening of the spread increases depth at the best price, facilitating larger transactions (Rindi and Werner 2019). However, the aggregate amount of order flow that a wholesaler would otherwise have internalized is unaffected,[37] replacing the set of attractive non-marketable limit orders with marketable limit orders.[38] More importantly, widening the quoted spread increased the profitability of off-exchange liquidity provision at the midpoint, increasing the willingness of wholesalers to internalize order flow.

Table 5 reports that the intensity of sub-penny-executed retail trades, as measured by the total volume of price-improved trades, the total dollar price improvement, or size-weighted relative price improvement, all increase due to the minimum 5¢-spread. In contrast, the absolute values of $Mroibvol$ and $Mroibtrd$ decrease, moving in the *opposite* direction of retail order flow internalization intensity. This opposite reaction indicates that $Mroibvol$ and $Mroibtrd$ respond to the economic incentives of wholesalers regarding retail order flow internalization rather than retail trading per se.

Table 6 presents estimation results for Test Group 2 that introduced a 5¢ tick. Figure 6 provides complementary visual evidence. In contrast to the quote-rule treatment, this trade-rule treatment caused the absolute values of $Mroibtrd$ and $Mroibvol$ to increase dramatically, even though the treatment *sharply reduced* the volume of sub-penny-executed (internalized) trades. For stocks with tight spreads, median internalized trade volume fell by 47% relative

[37] Werner et al. (2019) find that the wider spread incentivized the submission of limit orders, resulting in a longer queue at the bid and ask, while volume was unchanged.

[38] For example, consider two stocks, one with a mandated 5 cent spread and the other with whose 5 cent spread was non-mandated (pre-existing). There can be attractive non-marketable limit orders with the latter but not the former.

to the corresponding intercept, while trade volume is unchanged for stocks with wide spreads.

The key feature of the trade rule is that it quintupled the trading increment. This impacted the composition of retail orders as market orders risked execution at prices 5¢ further from current best prices (i.e., by further than 1¢). This led retail traders to rely more on marketable limit orders in lieu of market orders. By the time a wholesaler begins handling orders flagged as marketable limit, some have become non-marketable due to changes in the order book in the interim. The overall effect is to increase the share of non-marketable limit orders, which reduces internalization.[39] Consistent with the predictions in Section 6, the increased submission of limit orders is associated with an economically significant 20-39% decrease in sub-penny price improvement. Importantly, the increases we find in the absolute values of $Mroibtrd$ and $Mroibvol$ allow us to attribute the increased variation in $Mroib$ to the increased internalization of non-marketable limit orders. We posit that these effects manifest themselves in the increased sensitivity of $Mroib$ to institutional liquidity demand, as non-marketable limit orders are the marginal retail orders used to provide liquidity to institutions through internalization. Section7.4 provides support for this prediction when $Mroib$ is constructed from retail orders that receive less price improvement, i.e., when it is constructed only from the marginal internalized orders.

These findings based on the TSP reinforce the conclusion that variations in $Mroibtrd$ and $Mroibvol$ reflect wholesaler incentives to internalize retail order flow, rather than informed trading. Indeed, in light of our previous findings, an informed retail trading interpretation would imply that wholesalers would not be profit maximizing—internalizing more toxic (informed retail) orders while also paying more PFOF + PI is at odds with notions of profit-maximization. In contrast, the willingness to pay more for internalizing these marginal orders is consistent with them being needed to provide liquidity when institutional demand is high.

[39] Our estimates likely understate the actual effect because wholesaler incentives to internalize order flow increase with a wider 5¢ spread.

6.3 Interactions Between Institutional and Retail Order Flow

Our next analysis links *Mroibvol* imbalances with the demand for liquidity by institutions on the opposite side.[40] We first examine the variation in *Mroibvol* against contemporaneous intraday and overnight returns. We find intraday prices move in the opposite direction of retail imbalances—contrary to premises of being driven by aggressive informed retail trade. Instead, intraday prices move in the same direction as institutional order flow imbalances. We then link *Mroibvol* imbalances to institutional order flow, institutional trading costs, and stock liquidity. We find that the most extreme retail and institutional order flow imbalances are associated with the highest levels of institutional trading costs, the highest spreads, and the least depth. We also find that institutional order flow from mutual fund or short seller trades is negatively correlated with retail order flow. These findings, together with the market microstructure details and our earlier empirical findings in Section 6.2, reinforce conclusions that *Mroibvol* imbalances are a symptom of wholesalers using retail order flow to provide liquidity to institutions.

Table 7 summarizes the relationships between *Mroibvol* and various contemporaneous outcomes across 10 *Mroibvol* portfolios. While close-to-close returns monotonically increase from −2bps in the bottom *Mroibvol* portfolio to 30bps in the top *Mroibvol* portfolio, this pattern is not due to price pressure from retail order flow. Decomposing daily returns into their intraday and overnight components reveals that intraday returns decrease monotonically from 10bps in the bottom *Mroibvol* portfolio to −14bps in the top *Mroibvol* portfolio.[41] As most internalized (price-improved) trades are market and marketable-limit orders, the *negative* association between *Mroibvol* and intraday returns is inconsistent with retail price pressure. This negative association is at odds with notions of informed retail trading as they would require a negative price impact from "informed" orders submitted by retail investors.

[40] Our analysis focuses on *Mroibvol* but, as in BJZZ, similar empirical results obtain for *Mroibtrd*.

[41] Recall that BJZZ's algorithm constructs retail order flow imbalances using off-exchange transactions executed during regular trading hours.

In sharp contrast to intraday returns, overnight returns are positively related to $Mroibvol$. Indeed, the signs of intraday and overnight returns differ for eight of the ten $Mroibvol$ deciles.[42] This opposing return pattern can be understood by examining institutional trading. Recall that institutional and retail imbalances are negatively correlated, as wholesalers internalize this portion of retail order flow to meet institutional demand. Average institutional order flow falls from 33.6% in the bottom $Mroibvol$ decile to 21.1% in the top $Mroibvol$ decile. Thus, when institutional order flow imbalances skew toward more buying, internalized retail order flow imbalances skew toward more selling. Moreover, short selling activity also occurs on the opposite side of internalized retail order flow as increased short interest (increased short selling) is associated with a larger positive internalized retail order flow imbalance. Importantly, directional (as opposed to liquidity provider) short sellers, whose aggregate positions are reflected in short interest data, are known to be informed investors (Desai et al. 2002; Engelberg et al. 2013; Boehmer and Wu 2013). Thus, the negative association between such short selling activity and $Mroibvol$ represents further evidence against the informativeness of retail orders priced at sub-pennies. Instead, these findings suggest that the intraday price movements reflect institutional price pressure that is followed by overnight reversals.

Our next analysis identifies the economic roots of internalized trade imbalances, showing that they reflect variation in the extent of institutional liquidity demand relative to extant liquidity on other venues. The arrival of uninformed institutional liquidity shocks provides wholesalers (OTC market makers) with opportunities to profitably intermediate between retail and institutional investors. Such liquidity provision opportunities are more lucrative when institutional trading costs are higher (spreads are wide, depth near best prices is low). Wholesalers can detect these liquidity demand shocks indirectly by observing unfilled block orders on their affiliated ATSs (dark pools), or directly through heightened participation in their SDPs (ping pools). Their endogenous response is to internalize more retail orders

[42] We also find that intraday returns are generally negative while overnight returns are generally positive, consistent with the asset pricing literature (Cliff et al. (2008), Berkman et al. (2012), and Lou et al. (2019)).

from the opposite side of the market, including even non-marketable limit orders (see Section 6.2). This imbalanced internalization lets wholesalers fill these profitable institutional orders off-exchange. This economic logic finds strong support in the data.

Table 7 documents that institutional implementation shortfalls are highest, spreads are widest, and depth is lowest when $Mroibvol$ is highest in absolute value, i.e., at the extreme deciles of $Mroibvol$. Concretely, implementation shortfall per \$1m worth of institutional order size is 69bps and 15bps when $Mroibvol$ is at its lowest and highest deciles, respectively, while balanced $Mroibvol$ is associated with roughly 3bps of such costs. Similarly, average dollar and relative quoted spreads in the lowest and highest $Mroibvol$ deciles are essentially double those when $Mroibvol$ is relatively balanced.

To hammer this $\cup$-shaped relationship home, we construct a stock-specific measure of abnormal realized off-exchange institutional liquidity. For each stock-day, we divide the number of large off-exchange mid-point executions[43] by the same stock's average of this quantity over the sample period. Higher values of this measure indicate greater liquidity. We find that extreme $Mroibvol$ is associated with the least abnormal off-exchange midpoint execution, indicating that internalization of retail order flow is more prevalent when off-exchange liquidity is abnormally scarce.

Table 7 also reveals that intraday and overnight returns in the extreme $Mroibvol$ deciles reflect more than the unwinding of price pressure. This observation is clearest in $Mroibvol$'s bottom decile where price pressure from institutional buying is 0.098%, but the contemporaneous overnight reversal of -0.116% exceeds this price pressure. To study this phenomena more accurately, a 5-day overnight return is constructed that omits the first close-to-open return and adds the overnight return on the sixth day. This adjustment properly aligns the timing of intraday price pressure and overnight reversals". We see that this adjustment *exacerbates* the disconnect between the intraday "price pressure" and the subsequent (next-

[43]TAQ data transactions with trade venue flag 'D' that are at least 1,000 shares and worth at least \$50k.

day) overnight "reversals" that average -0.134% when $Mroibvol$ is in decile 1. In fact, comparing intraday and "next-day" overnight returns when $Mroibvol$ is in decile 1 versus decile 5 reveals differences of $0.098 - (-0.063) = 0.161\%$ and $-0.0138 - 0.257 = -0.379\%$, respectively. The analogous differences when $Mroibvol$ is in decile 10 versus decile 5 are $-0.138 - (-0.063) = -0.075\%$ and $0.456 - 0.257 = 0.199\%$, respectively. Thus, weekly overnight returns revert by far more than is needed to offset intraday returns, especially when $Mroibvol$ is extremely negative. We next reconcile this pattern by establishing that institutional buy order flow is more persistent than institutional sell order flow. As a result, institutional buy order flow predicts returns and, in turn, is predicted by internalized retail order flow (with an inverse relation) over longer horizons.

7 Why Does *Mroib* Predict Returns?

7.1 Dynamics of Institutional and Retail Order Flows

This section shows that overnight reversals exceed intraday price pressure (during the same week) because overnight reversals also reflect the unwinding of institutional price pressure accumulated in prior weeks. This effect is more salient when more retail sell orders have been internalized, presumably to provide liquidity for institutional buy orders. A recent literature finds that long-only fund managers accumulate long positions slowly, but sell quickly, largely to fund purchases. This asymmetry is consistent with institutional buying, but not selling, being motivated by a fund manager's best ideas (Akepanidtaworn et al. 2021). This would lead long positions to be accumulated more gradually to conceal their presence, prolonging the unwinding of price pressure. Hendershott and Seasholes (1994) also document that the short positions of market makers, which reflect institutional buying, are associated with subsequent price reversals that last up to 11 trading days. In contrast, price reversals that follow the accumulation of long positions by market markers, which reflect institutional selling, only last for 7 trading days.

We estimate

$$\begin{aligned} X_{j,w} &= a^0 + \sum_{i=1}^{6} a_i^1 Inoibvol_{j,w-i} + \sum_{i=1}^{6} a_i^2 \left[I(Inoibvol_{j,w-i} < 0) \right] \\ &+ \sum_{i=1}^{6} a_i^3 \left[I(Inoibvol_{j,w-i} < 0) \times Inoibvol_{j,w-i} \right] + \epsilon_{j,w}, \end{aligned} \tag{3}$$

where $X \in \{Inoibvol, Mroibvol\}$; and $I(\cdot)$ is an indicator function that equals 1 if $Inoibvol < 0$ and equals 0 otherwise. The models are estimated using Fama-MacBeth regressions, with standard errors corrected using the Newey-West methodology with 6 lags. On average across stocks, ANcerno covers less than 7% of the total daily trading volume reported by CRSP.[44] This leads us to reduce the noise attributable to a lack of coverage by using the subset of stocks for which the share of ANcerno-reported volume relative to CRSP is above-average.

Columns (1)–(4) in Table 8 present the AR(k) estimates for $Inoibvol$, showing that past positive and negative institutional order flows, especially those with longer lags, predict current institutional order flows differently. In particular, the most recent week's positive and negative $Inoibvol$ predict current week's $Inoib$ quite similarly, with point estimates of 0.33 and 0.35 for positive and negative $Inoibvol_{w-1}$, respectively. However, these coefficients sharply diverge for $k > 1$, where the loadings of negative $Inoibvol_{w-i}$ become 30-70% smaller than those for their positive $Inoibvol_{w-i}$ counterparts. This finding is consistent with the more gradual accumulation of institutional buy positions found in the literature. This persistent institutional buying drives the accumulation of positive price pressure whose unwinding extends beyond the subsequent close-to-open to subsequent days, while institutional selling is less persistent.

Columns (5)–(8) in Table 8 highlight how past institutional order flow predicts future internalized retail order flow, reinforcing our earlier conclusion that wholesalers intermediate trades between institutional and retail investors. Consistent with the stronger auto-

[44] Hu et al. (2018) report similar coverage over a longer sample period. Nevertheless, modest coverage does not invalidate the representativeness of ANcerno data (Puckett and Yan 2011, Anand et al. 2012, and Jame 2018).

correlation for institutional buying, and retail sell orders being internalized to provide liquidity for institutional buy orders, $Inoibvol_{w-i}$ loads with negative and significant coefficients.[45] Mirroring the weaker auto-correlation in institutional order flow when $Inoibvol_{w-i} < 0$, the loadings for $Inoibvol_{w-i}$ become positive for $k > 2$. These dynamics indicate that the most negative $Mroibvol_w$ observations, i.e., those in decile 1 of Table 7, are disproportionately more likely to arise following persistent institutional buying pressure whose unwinding makes the current week's overnight returns more negative.

These statistical findings contain insights about the pecking order that institutions consider in the pursuit of liquidity. The negative correlation between past positive institutional order flow and current internalized retail order flow is consistent with institutions resorting to SDPs, and hence wholesaler PFOF, only after exhausting less expensive sources of liquidity.

7.2 Institutional Trading and Short-Term Return Predictability

We next establish that $Mroib$'s short-term return predictability is a liquidity-driven phenomenon. Due to the persistence of institutional order flow, especially institutional buying, overnight price reversals associated with extreme $Mroibvol$ magnitudes extend into future week(s). To the extent that institutional order flow is persistent, subsequent abnormal overnight price reversals remain nontrivial, creating distinguishable differences between close-to-close returns that follow extremely negative and extremely positive internalized retail order flow imbalances.

To highlight the persistence of institutional order flow, we estimate

$$\begin{aligned} Inoibvol_{j,w} &= c^0 + \sum_{i=1}^{6} c_i^1 Mroibvol_{j,w-i} + \sum_{i=1}^{6} c_i^2 \left[I(Inoibvol_{j,w-i} < 0) \right] \\ &+ \sum_{i=1}^{6} c_i^3 \left[I(Inoibvol_{j,w-i} < 0) \times Mroibvol_{j,w-i} \right] + \epsilon_{j,w}, \end{aligned} \tag{4}$$

with variable definitions and estimation approaches identical to those in equation (3). As

[45] The only exception to statistical significance appears in column (8) for $Inoibvol_{w-5}$.

Table 9 demonstrates, the first and second lags of internalized retail order flow load with significantly negative coefficients when these lagged internalized order flows correspond to positive institutional flow. That is, when institutional order flow is positive, greater internalization of retail sell orders relative to buy orders is associated with abnormally high institutional buy pressure up to two weeks forward. This effect, as discussed above, drives subsequent abnormally negative overnight returns that skews subsequent weeks' close-to-close returns downward. As such, $Mroibvol$ appears to predict future close-to-close returns, even though it merely captures price reversals following institutional buy pressure.[46]

7.3 Long-Term Return Predictability and Liquidity Premia

This section revisits the return predictability of $Mroibvol$ to offer a unifying explanation for the patterns documented in Section 5. We analyze $Mroibvol$'s long-term return predictability, especially cross-sectional return differences after conditioning on the sign of $Mroibvol_{w-1}$. We show that long-term return predictability reflects liquidity premia required by institutional investors to hold less liquid assets (Amihud and Mendelson 1986).

Table 10 documents the relationships between close-to-close, intraday, and overnight weekly returns conditional on $Mroibvol_{w-1}$. Consistent with $Mroibvol_{w-1}$'s short-term return predictability, close-to-close returns for week w monotonically increase from the bottom decile of $Mroibvol_{w-1}$ to the highest decile. Most of the return variation is concentrated in the extreme deciles (deciles 1, 9, and 10). Furthermore, consistent with $Mroibvol_{w-1}$'s declining impact on close-to-close returns in Table 4, the return difference between the bottom and top deciles of $Mroibvol_{w-1}$ rapidly decline in subsequent weeks, nearly disappearing by week $w+12$. Instead, a striking $\cup$-shaped pattern in close-to-close returns across the $Mroibvol_{w-1}$ portfolios begins to emerge by week $w+3$, strengthening sharply in subsequent weeks. For example, average week $w+12$'s close-to-close returns in deciles 1 and 10 of $Mroibvol_{w-1}$

[46] Lou et al. (2019) document that overnight and intraday returns display persistence relative to overnight and intraday returns, respectively, but reversals relative to intraday and overnight returns.

(0.15% and 0.18%, respectively) are more than double that in decile 6 (0.07%). Similar patterns extend to all future weeks. This ∪-shaped pattern implies that future returns are inversely related to negative $Mroibvol_{w-1}$ and positively related to positive $Mroibvol_{w-1}$. These distinct relationships reinforce the earlier negative and positive coefficients from regressing weekly returns on negative and positive $Mroibvol_{w-1}$, respectively (see Table 4).

To relate the ∪-shaped pattern to liquidity premia, focus on lower $Mroibvol_{w-1}$ deciles. In Section 7.2, we found evidence that the short-term negative overnight returns associated with extremely negative $Mroibvol_{w-1}$ likely reflect extended price reversals of previously-accumulated long institutional positions. These price reversals are temporary and reflect preceding price pressure from institutional trading. In contrast, a liquidity premium implies *long-term* return differences according to the level of liquidity. The strong association between liquidity measures, institutional trading costs, and retail order flow internalization indicates that stocks with more extreme $Mroibvol_{w-1}$ are less liquid. Hence, these stocks command higher *permanent* expected return premia (higher cross-sectional returns) as compensation for illiquidity. Week $w + i$ returns for the bottom decile of $Mroibvol_{w-1}$ demonstrate the net effect of temporary price reversals and characteristic liquidity premia. Initially, price pressure dominates but a liquidity premium eventually dominates cross-sectional returns, which become negatively correlated with negative $Mroibvol_{w-1}$. Conversely, when $Mroibvol_{w-1}$ is positive, disentangling short-term and long-term effects in close-to-close returns is more difficult since their impacts on returns have the same sign.

Decomposing close-to-close returns into intraday and overnight components enables us to identify when liquidity premia are realized during the day and contribute to the recent asset pricing literature that documents important time-of-day return disparities that are important to asset pricing anomalies. For example, Hendershott et al. (2020) report that CAPM predictions hold overnight but not during the day; and Lou et al. (2019) and Bogousslavsky (2021) find that most return anomalies accrue during the trading day rather than

overnight. Consistent with these findings, our decomposition of close-to-close returns reveals that the ∪-shaped pattern in future close-to-close returns across $Mroibvol_{w-1}$ portfolios are largely attributable to intraday returns. In fact, overnight returns follow a ∩-shaped pattern across the 10 $Mroibvol_{w-1}$ portfolios. Attributing the ∪-shaped pattern in intraday returns to liquidity premia allows us to identify an economic mechanism that explains why return anomalies differ between intraday and overnight returns. In particular, liquidity premia are realized during the trading day, but not overnight.

Our next analysis verifies that the long-term return predictability associated with negative and positive internalized retail order flow imbalances are not attributable to persistence in such internalization or contrarian trading. Following BJZZ, we decompose weekly internalized retail order flow imbalances and estimate the cross-sectional specification

$$Mroibvol_{j,w-1} = \lambda_w^0 + \lambda_w^1 Mroibvol_{j,w-2} + \lambda_w^2 R_{w-2} + \eta_{j,w-1} \tag{5}$$

to construct

$$\begin{aligned} \text{Persistence}_{j,w-1} &= \hat{\lambda}_w^1 Mroibvol_{j,w-2}, & (6)\\ \text{Contrarian}_{j,w-1} &= \hat{\lambda}_w^1 R_{j,w-2}, & (7)\\ \text{Other}_{j,w-1} &= \lambda_w^0 + \hat{\eta}_{j,w-1}. & (8) \end{aligned}$$

By construction, $Mroibvol_{j,w-1} = \text{Persistence}_{j,w-1} + \text{Contrarian}_{j,w-1} + \text{Other}_{j,w-1}$. Substituting these components for $Mroibvol_{j,w-1}$ in equation 1, we estimate

$$\begin{aligned} R_{j,w+i} &= d_w^0 + d_w^{1p}\left(\text{Persistence}_{j,w-1}\right) + d_w^{1c}\left(\text{Contrarian}_{j,w-1}\right) + d_w^{1o}\left(\text{Other}_{j,w-1}\right) \\ &+ {d_w^2}^{\top}\text{controls}_{j,w-1} + u_{j,w+i}, \end{aligned} \tag{9}$$

with $i \in \{0, 1, 2, 3, 6, 9, 12, 24, 36, 39, 42, 45, 48, 51, 54, 57, 60\}$. We fit Fama-MacBeth regressions with Newey-West-corrected standard errors using 6 lags.

Table 11 reports estimated predictive powers of these three components of internalized retail order flow for future weekly returns. First note that our "Persistence" and "Other" co-

efficients for week w's close-to-close returns of 0.32% and 0.08%, respectively, are quite close to their counterparts of 0.27% and 0.08% in BJZZ. Comparing the coefficients for the residual term, Other_{w-1}, in Table 11 to those of $Mroibvol_{w-1}$ in Table 4 reveals that the significance of an $Mroibvol_{w-1}$ coefficient almost always corresponds to the significance of Other_{w-1}'s counterpart coefficient. For example, with week $w+i$'s close-to-close return as the dependent variable, when $Mroibvol_{w-1}$ has a negative coefficient for $k > 39$, $Mroibvol_{w-1} < 0$ also has a negative coefficient. A similar finding applies when $Mroibvol_{w-1} > 0$ has a positive coefficient. Contrary to BJZZ's interpretation that significant Other_{w-1} coefficients indicate informed trading, our analysis indicates that the predictive power associated with this residual component is reconciled with liquidity-driven price dynamics.

7.4 Sub-penny Price Improvement Size and Return Predictability

We conclude by linking long-term return predictability to the magnitudes of sub-penny price improvements that wholesalers offer when internalizing retail orders. BJZZ find that the short-term predictive power of $Mroib$ is greater when the measure is constructed using retail transactions receiving *smaller* sub-penny price improvements.[47] BJZZ interpret this finding as evidence that wholesalers can identify informed retail orders, and offer them less price improvement, effectively price discriminating against informed retail order flow. Our analysis, in contrast, attributes smaller sub-penny price improvements to the profit-maximizing incentives of wholesalers. Specifically, when internalizing non-marketable limit orders, which are more costly and less profitable to execute (Section 6.1), wholesalers offer less price improvement. This ensures that their trade executions remain profitable when they intermediate between non-marketable limit retail orders and institutional orders on the opposite side (Section 6.2). The institutional price pressure on the opposite side of $Mroib$ in conjunction with low liquidity implies larger subsequent price reversals, strengthening the

[47] Unreported results replicate these findings.

positive correlation between $Mroib$ and short-term future returns, as found by BJZZ.

To provide evidence for our liquidity-based interpretations, we examine the relation between institutional trading costs and longer-term future returns using versions of $Mroibvol$ constructed from transactions with low versus high sub-penny price improvements.[48] Panel A in Figure 7 shows that institutional implementation shortfalls display a stronger ∪-shaped pattern in $Mroibvol$ when constructed using retail transactions with low sub-penny price increments compared to transactions with high-sub-penny transactions. Similar qualitative findings extend to other measures of liquidity analyzed in Table 7. Given the economic incentives of wholesalers, these findings indicate that most variation in $Mroib$ is attributable to the internalization of *non-marketable* orders, which, in turn, are more strongly correlated with institutional trading costs. Panel B in Figure 7 reinforces this interpretation by showing that the ∪-shaped pattern of week $w+12$ returns conditional on $Mroibvol_{w-1}$, documented in Table 10, is *entirely* attributable to the $Mroibvol_{w-1}$ imbalances constructed from internalized low sub-penny price improvements. As such, consistent with less liquid stocks commanding higher expected returns, extreme imbalances in internalized low-sub-penny trades are a manifestation of lower liquidity and higher trading costs, but not more informed (toxic) orders.

8 Conclusion

Our paper uncovers the economic mechanisms underlying the return predictability of imbalances in internalized retail order flow denoted $Mroib$ by Boehmer, Jones, Zhang, and Zhang (2021). Wholesalers acting as off-exchange market makers in U.S. equity markets internalize retail order flow and provide payment for order flow (PFOF) to execute retail orders routed to them by retail brokers. Institutional details and Rule 606 disclosures indicate that wholesalers are substantially more attracted to internalizing marketable orders, which are both

[48] To construct these two versions of $Mroibol$, we follow BJZZ. Small sub-penny increments are those falling in (0,0.2] or [0.8,1), and large sub-penny increments are those falling in (0.2,0.4] and [0.6,0.8).

less costly and more profitable to internalize than non-marketable orders. Nevertheless, a significant fraction of non-marketable orders are internalized.

Wholesaler decisions to internalize retail order flow are determined by the potential gains from providing liquidity off-exchange; and our paper provides extensive evidence that most variation in $Mroib$ reflects these strategic choices. We establish that $Mroib$ is inversely related to institutional order flow, and the absolute value of this imbalance is largest when institutional liquidity is most costly—i.e., when implementation shortfalls are highest, spreads are widest, depth is lowest, and mid-point off-exchange liquidity is abnormally scarce. Moreover, intraday returns reflect institutional price pressure and are inversely related to $Mroib$.

We show that the short-term predictive power of $Mroib$ for future returns is attributable to price reversals that follow institutional liquidity consumption. When institutional liquidity demand interacts with $Mroib$ through wholesalers, contemporaneous returns move in the opposite direction of $Mroib$, reflecting institutional price pressure. As this price pressure unwinds in the form of subsequent price reversals, current $Mroib$ appears to be positively related to future returns. However, future returns over longer horizons display a strong $\cup$-shaped pattern conditional on $Mroib$. Observing the economic and statistical relations between $Mroib$ and institutional trading costs, we attribute this $\cup$-shaped pattern to a characteristic liquidity premium component. This liquidity premium manifests itself in higher expected returns for stocks with higher average absolute $Mroib$ quantities. This motivates our companion paper (Barardehi et al. 2020), which uses $|Mroib|$ as a measure of daily stock liquidity. This institutional liquidity measure, unlike all standard market-microstructure-based, high-frequency liquidity measures, is priced during the 2010–2019 period.

Our analysis informs policymakers about the implications of existing retail order internalization practices for liquidity provision by retail investors. First, while sub-penny execution prices realized in this process may reflect best execution quality for retail orders, they have a potential cost. We show that non-marketable limit orders are the marginal order types in

the internalization process. When routed to exchanges, such orders would tighten bid-ask spreads. However, in less liquid markets where institutional trading costs are high, retail brokers, wholesalers, and institutional investors are motivated to internalize these orders, rather than add them to the limit order book. Second, internalization facilitates liquidity provision by retail investors to institutional investors. However, retail investors are only minimally compensated for their liquidity provision through price improvements and, recently, zero-commission trade execution (see Jain et al. 2020[49]). Instead, most of the compensation for the provision of liquidity by retail traders to institutions accrues to wholesalers via the profits from near-riskless principal trading and to broker-dealers via PFOF.

[49]They also show that the zero commissions encourage retail investors to post (non-displayed) odd lot orders inside the NBBO, reducing effective spreads, but the price improvement associated with internalization is reduced despite the improved execution prospects on exchanges.

References

Akepanidtaworn, K., R. Di Mascio, A. Imas, and L. Schmidt, 2020. Selling fast and buying slow: Heuristics and trading performance of institutional investors. *Journal of Finance*, forthcoming.

Albuquerque, R., S. Song, and C. Yao, 2020. The price effects of liquidity shocks: A study of the SEC's tick size experiment. *Journal of Financial Economics* 138: 700-734.

Amihud, Y. and H. Mendelson, 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17: 223-249.

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.

Barardehi, Y.H., D. Bernhardt, Z. Da, and M. Warachka, 2022. Institutional liquidity premia and the cross-section of stock returns post RegNMS. Working paper.

Barber, B. M., S. Lin, and T. Odean, 2021. Resolving a paradox: Retail trades positively predict returns but are not profitable. Working paper.

Barber, B. M., and T. Odean, 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55: 773-806.

Barber, B. M., and T. Odean, 2008. The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21: 785-818.

Barber, B. M., T. Odean, and N. Zhu, 2009. Do retail trades move markets? *Review of Financial Studies* 22: 151-186.

Barclay, M. J. and T. Hendershott, 2003. Price discovery and trading after hours. *Review of Financial Studies* 16: 1041-1073.

Barclay, M. J. and T. Hendershott, 2008. A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance* 15: 839-849.

Barrot, J., R. Kaniel, and D. A. Sraer, 2016. Are retail traders compensated for providing liquidity? *Journal of Financial Economics* 120: 146-168.

Battalio, R., 2012. Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?. *Journal of Finance* 52: 341-352.

Battalio, R. and C. W. Holden, 1995. A simple model of payment for order flow, internalization, and total trading cost. *Journal of Financial Markets* 4: 33-71.

Battalio, R., S. A. Corwin, and R. Jennings, 2016. Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality. *Journal of Finance* 121: 2193-2238.

Battalio, R., J. Greene and R. Jennings, 1997. Do Competing Specialists and Preferencing Dealers Affect Market Quality? *Review of Financial Studies* 10: 969-993.

Battalio, R., J. Greene, B. Hatch and R. Jennings, 2002. Does the Limit Order Routing Decision Matter? *Review of Financial Studies* 15: 159-194.

Berkman, H., P. D. Koch, L. Tuttle and Y. J. Zhang, 2012. Paying attention: Overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis* 47: 715-741.

Bernhardt, D. E. Hughson and G. Naganthan, 2001. Cream-skimming and Payment for Order Flow. Working paper.

Bessemnbinder, H. and H. M. Haufman, 1997. A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks. *Journal of Financial Economics* 46: 293-319.

Boehmer, E., C. M. Jones, X. Zhang and X. Zhang. Tracking retail investor activity. *Journal of Finance* 76: 2242-2305.

Boehmer, E. and J. Wu, 2013. Short selling and the price discovery process. *Review of Financial Studies* 26: 287-322.

Bogousslavsky, V., 2021. The cross-section of intraday and overnight returns. *Journal of Financial Economics* 141(1): 172-194.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson, 2019. Information, trading, and volatility: Evidence from firm-specific news. *Review of Financial Studies* 32: 992-1033.

Cao, C., E. Ghysels and F. Hatheway, 2000. Price discovery without trading: Evidence from the Nasdaq pre-opening. *Journal of Finance* 55: 1339-1365.

Cliff, M., M. J. Cooper, and H. Gulen, 2008. Return difference between trading and non-trading hours: Like night and day. Working paper.

Chordia, T., R. Roll, and A. Subrahmanyan, 2002. Market Liquidity and Trading Activity. *Journal of Finance* 56: 501-530.

Desai, H., K. Ramesh, S. Thiagarajan, and B. Balachandran, 2002. An investigation of the informational role of short interest in the Nasdaq market. *Journal of Finance* 57: 2263-2287.

Easley, D., N. M. Kiefer, and M. O'Hara, 1996. Cream-skimming or profit-sharing? The curious role of purchased order flow. *Journal of Finance* 51: 811-833.

Engelberg, J., A. Reed and M. Ringgenberg, 2012. How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics* 105: 260-278.

Ernst, T., J. Skobin and C. Spatt, 2021. The value of off-exchange data. Working paper.

Foucault, T., D. Sraer, and D. Thesmar, 2011. Individual investors and volatility. *Journal of Finance* 66: 1369-1406.

Griffith, T. G., and B. S. Roseman, 2019. Making cents of tick sizes: An investigation of the 2016 US SEC Tick Size Pilot. *Journal of Banking & Finance* 101: 104-121.

Hendershott T., D. Livdan, and D. Rösch. 2020. Asset pricing: A tale of night and day. *Journal of Financial Economics* 138: 635-662.

Hendershott, T. and M. S. Seasholes, 2007. Market maker inventories and stock prices.

American Economic Review 97: 210-214.

Hu, G., 2009. Measures of implicit trading costs and buy–sell asymmetry. Journal of Financial Markets 12, 418-437.

Hu, G., K. Jo, Y. Wang, and J. Xie, 2018. Institutional trading and Abel Noser data. *Journal of Corporate Finance* 52, 143-167.

Jain, P. K., S. Mishra, S. O'Donoghue, and L. Zhao, 2020. Trading Volume Shares and Market Quality: Pre- and Post-Zero Commissions. Working paper.

Jame, R., 2018. Liquidity provision and the cross section of hedge fund returns. *Management Science* 64, 2973-3468.

Jiang, C., T. Likitapiwat, and T. McInish, 2012. Information content of earnings announcements: Evidence from after-hours trading. *Journal of Financial and Quantitative Analysis* 47: 1303-1330.

Kaniel, R., S. Liu, G. Saar, and S. Titman, 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67: 639-680.

Kaniel, R., G. Saar, and S. Titman, 2008. Individual investor sentiment and stock returns. *Journal of Finance* 63: 273-310.

Kelley, E. K., and P. C. Tetlock, 2013. How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68: 1229-1265.

Kim, S., and D. Murphy, 2013. The impact of high-frequency trading on stock market liquidity measures. Working paper.

Kumar, A., and C. Lee, 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61: 2451-2486.

Lou, D., C. Polk, and S. Skouras, 2019. A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics* 134: 192-213.

Parlour, C. and U. Rajan, 2003. Payment for order flow. *Journal of Financial Economics*

68: 379-411.

Parlour, C. and D. Seppi, 2003. Liquidity-Based Competition for Order Flow. *Review of Financial Studies* 16: 301-343.

Peterson, M. A. and E. Sirri, 2003. Order Preferencing and Market Quality on U.S. Equity Exchanges. *Review of Financial Studies* 16: 385-415.

Puckett, A., and X. Yan, 2011. The interim trading skills of institutional investors. *Journal of Finance* 66: 601-633.

Rindi, B. and I.M. Werner 2018. U.S. Tick Size Pilot. Working paper.

Stoll, H. and R. Whaley, 1990. Stock market structure and volatility. *Review of Financial Studies* 3: 37-71.

U.S. Securities & Exchange Commission, 2010, Concept Release on Equity Market Structure.

U.S. Securities & Exchange Commission, 2021. Staff Report on Equity and Options Market Structure Conditions in Early 2021.

Werner, I.M., Y. Wen, B. Rindi, and S. Buti, 2019. Tick Size, Trading Strategies and Market Quality. Working paper.

Tables and Figures

Table 1: **Retail Orders Receiving Payment for Order Flow.** This table reports (1) distributions of retail order types among all non-directed orders received by retail brokers; (2) distributions of retail order types among non-directed orders that are internalized and receive PFOF; and (3) PFOF amount per 100 shares for different retail order types. All quantities are extracted from Charles Schwab, TD Ameritrade, and E*TRADE's 606 filing disclosures for the final quarter of 2020. When applicable, quantities reflect dollar-weighted averages across the top-5 wholesalers handling retail orders for the respective broker.

	Charles Schwab			TD Ameritrade			E*TRADE		
	Non-directed orders (%)	Orders receiving PFOF (%)	PFOF (cents per 100 shares)	Non-directed orders (%)	Orders receiving PFOF (%)	PFOF (cents per 100 shares)	Non-directed orders (%)	Orders receiving PFOF (%)	PFOF (cents per 100 shares)
Market	52.9	57.2	9.0	18.8	44.7	12.0	49.3	53.7	19.9
Marketable limit	4.8	14.1	9.0	9.2	24.2	12.0	5.8	12.9	18.8
Non-marketable limit	33.8	21.1	29.6	31.9	21.2	33.5	35.0	18.0	29.3
Other order types	8.5	7.6	10.0	40.2	9.9	9.4	9.9	15.5	15.8
Total	100	100	–	100	100	–	100	100	–

Table 2: **Summary Statistics for Internalized Retail Order Flow.** This table reports summary statistics for daily measures of internalized order flows for our sample of NYSE-, AMEX-, and NASDAQ-listed common shares during the 2010–2014 period. $Mrbvol$ and $Mrsvol$ denote trading volumes for internalized trades classified as retail buy and retail sell, respectively. $Mrbtrd$ and $Mrstrd$ denote the number of internalized trades classified as retail buy and retail sell, respectively. $Mroibvol$ and $Mroibtrd$ then denote normalized imbalances in internalized retail order flow based on trading volume and trade frequency, respectively.

	N	Mean	St. dev.	Median	Q1	Q3
$Mrbvol$	4,627,339	46,345	288,628	5,850	1,395	23,157
$Mrsvol$	4,627,339	46,249	270,718	6,333	1,559	24,346
$Mrbtrd$	4,627,339	108	389	23	6	79
$Mrstrd$	4,627,339	106	349	24	6	81
$Mroibvol$	4,627,339	−0.035	0.453	−0.025	−0.286	0.209
$Mrioibtrd$	4,627,339	−0.030	0.430	−0.008	−0.263	0.200
$Mroibvol > 0$	2,154,810	0.330	0.295	0.233	0.101	0.471
$Mroibvol < 0$	2,448,368	−0.357	0.301	−0.265	−0.522	−0.115
$Mroibtrd > 0$	2,088,865	0.321	0.282	0.232	0.111	0.435
$Mroibtrd < 0$	2,329,910	−0.347	0.290	−0.261	−0.500	−0.123

Table 3: **Internalized Retail Order Flow and the Cross-section of Next Week's Returns.** This table presents estimates of the association between internalized retail order flow and the cross-section of the next week's returns (in percentage points). Daily returns are calculated based on the mid-points of best bid and ask prices at close as well as open prices, decomposing each day's close-to-close returns into intraday (open-to-close) and overnight (close-to-open) before aggregating each return type into weekly observations. Each of the three return cross-sections is decomposed based on the sign of previous week's internalized order flow to form a total of nine samples. According to equation (1), week w returns in each sample are regressed on week $w-1$'s internalized order flows ($Mroibvol_{w-1}$) and control variables including last week's return (R_{w-1}), last month's return (RET_{-1}), the return over the preceding five months ($RET_{(-7,-2)}$), volatility (VOLAT), and natural logs of turnover (ln(TO)), market capitalization (ln(Size)), and book-to-market ratio (ln(BM)). Estimates are based Fama-Macbeth regressions, featuring Newey-West corrected standard errors with 6 lags. Sample includes NMS common shares from Jan 2010 – Dec 2014, excluding observations with previous month-end's closing price below $1. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

Dependent Variable	Close-to-close return			Overnight return			Intraday return		
	$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
Constant	0.0063	−0.37	0.28	0.58***	0.59***	1.05***	−0.57**	−0.96***	−0.77***
	[0.02]	[−1.08]	[0.83]	[4.58]	[4.48]	[6.88]	[−2.10]	[−3.33]	[−2.66]
$Mroibvol_{w-1}$	0.087***	0.056***	0.12***	0.12***	0.16***	−0.0050	−0.029***	−0.10***	0.13***
	[13.73]	[3.75]	[7.37]	[25.53]	[20.54]	[−0.55]	[−4.41]	[−6.90]	[8.49]
R_{w-1}	−0.021***	−0.018***	−0.022***	0.00090	−0.0031	0.0038*	−0.022***	−0.015***	−0.026***
	[−5.86]	[−4.93]	[−5.78]	[0.50]	[−1.61]	[1.89]	[−7.07]	[−4.59]	[−7.64]
$RET_{(-1)}$	0.21	0.39**	0.014	−0.19**	−0.15*	−0.18*	0.40**	0.54***	0.20
	[1.14]	[2.12]	[0.07]	[−2.30]	[−1.84]	[−1.84]	[2.47]	[3.43]	[1.07]
$RET_{(-7,-2)}$	0.063	0.091	0.044	0.061**	0.047*	0.058*	0.0024	0.044	−0.014
	[0.84]	[1.18]	[0.53]	[2.45]	[1.83]	[1.83]	[0.03]	[0.63]	[−0.18]
ln(TO)	−0.037***	−0.032***	−0.047***	0.036***	0.030***	0.036***	−0.073***	−0.063***	−0.083***
	[−3.60]	[−3.08]	[−3.95]	[8.89]	[7.43]	[6.79]	[−8.16]	[−6.64]	[−8.05]
VOLAT	−6.44***	−6.55***	−6.17***	9.68***	8.20***	11.5***	−16.1***	−14.7***	−17.7***
	[−3.55]	[−3.73]	[−2.89]	[11.02]	[10.05]	[9.39]	[−10.03]	[−9.21]	[−9.33]
ln(Size)	0.020	0.036**	0.0065	−0.033***	−0.030***	−0.054***	0.053***	0.065***	0.061***
	[1.47]	[2.49]	[0.45]	[−5.31]	[−4.65]	[−7.57]	[4.39]	[5.22]	[4.75]
ln(BM)	0.058***	0.045**	0.064***	−0.038***	−0.025***	−0.046***	0.096***	0.070***	0.11***
	[2.73]	[2.12]	[2.66]	[−6.10]	[−3.88]	[−5.26]	[4.75]	[3.43]	[5.06]
Observations	3,330,408	1,875,061	1,448,395	3,330,408	1,875,061	1,448,395	3,330,408	1,875,061	1,448,395

Table 4: **Internalized Order Flow and the Cross-sections of Future Weeks' Returns.** This table presents estimates of the associations between internalized retail order flow and the cross-sections of future week $w+i$ returns (in percentage points), with $i \in \{0, 1, 2, 3, 6, 9, 12, 24, 36, 39, 42, 45, 48, 51, 54, 57, 60\}$. Daily returns are calculated based on the mid-points of best bid and ask prices at close as well as open prices, decomposing each day's close-to-close returns into intraday (open-to-close) and overnight (close-to-open) before aggregating each return type into weekly observations. Each of the three return cross-sections for a given week $w+i$ is decomposed based on the sign of week $w-1$'s internalized order flow to form a total of nine samples. According to equation (1), week $w+i$ returns in each sample are regressed on week $w-1$'s internalized order flows ($Mroibvol_{w-1}$), whose loadings are reported in the table, and control variables including last week's return (R_{w-1}), last month's return (RET_{-1}), the return over the preceding five months ($RET_{(-7,-2)}$), volatility (VOLAT), and natural logs of turnover (ln(TO)), market capitalization (ln(Size)), and book-to-market ratio (ln(BM)). Estimates are based Fama-Macbeth regressions, featuring Newey-West corrected standard errors with 6 lags. Sample includes NMS common shares from Jan 2010 – Dec 2014, excluding observations with previous month-end's closing price below $1. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

Dep. Var. =	Close-to-close return			Overnight return			Intraday return		
	$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
w	0.087***	0.056***	0.12***	0.12***	0.16***	−0.0050	−0.029***	−0.10***	0.13***
	[13.73]	[3.75]	[7.37]	[25.53]	[20.54]	[−0.55]	[−4.41]	[−6.90]	[8.49]
$w+1$	0.053***	0.019	0.093***	0.090***	0.14***	−0.015	−0.037***	−0.12***	0.11***
	[8.54]	[1.26]	[5.32]	[25.16]	[20.52]	[−1.56]	[−6.16]	[8.86]	[6.55]
$w+2$	0.045***	0.017	0.093***	0.077***	0.12***	−0.025***	−0.032***	−0.11***	0.12***
	[7.31]	[1.21]	[4.95]	[21.24]	[17.30]	[−2.81]	[−5.07]	[7.55]	[6.76]
$w+3$	0.034***	0.0052	0.080***	0.067***	0.12***	−0.031***	−0.033***	−0.11***	0.11***
	[6.04]	[0.38]	[4.71]	[20.56]	[16.00]	[−3.56]	[−5.76]	[8.05]	[6.87]
$w+6$	0.016***	−0.018	0.058***	0.050***	0.11***	−0.033***	−0.034***	−0.13***	0.091***
	[2.62]	[−1.28]	[3.32]	[14.56]	[15.25]	[−3.77]	[−6.05]	[9.49]	[5.83]
$w+9$	0.011**	−0.012	0.058***	0.038***	0.087***	−0.052***	−0.026***	−0.099***	0.11***
	[1.98]	[−0.89]	[3.23]	[10.15]	[12.64]	[−6.18]	[−4.82]	[−7.67]	[6.76]
$w+12$	0.0028	−0.048***	0.063***	0.042***	0.077***	−0.016*	−0.039***	−0.12***	0.079***
	[0.53]	[−3.54]	[3.53]	[12.54]	[12.06]	[−1.86]	[−7.06]	[9.93]	[4.96]
$w+24$	−0.0027	−0.054***	0.048***	0.025***	0.067***	−0.037***	−0.028***	−0.12***	0.085***

Continued on next page

Table 4 – *continued from previous page*

Dep. Var. =	Close-to-close return			Overnight return			Intraday return		
	$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
	[−0.45]	[−3.69]	[2.76]	[7.83]	[10.76]	[−4.19]	[−4.43]	[8.66]	[4.99]
$w+36$	−0.0044	−0.046***	0.068***	0.030***	0.065***	−0.019**	−0.034***	−0.11***	0.087***
	[−0.68]	[−2.86]	[4.18]	[7.57]	[8.88]	[−2.40]	[−5.51]	[7.82]	[5.69]
$w+39$	−0.011*	−0.039***	0.044***	0.016***	0.043***	−0.043***	−0.027***	−0.082***	0.087***
	[−1.78]	[−2.60]	[2.76]	[4.41]	[6.42]	[−5.13]	[−4.79]	[−5.97]	[5.43]
$w+42$	−0.010*	−0.042***	0.072***	0.019***	0.058***	−0.042***	−0.029***	−0.10***	0.11***
	[−1.85]	[−3.09]	[3.99]	[4.96]	[8.01]	[−4.70]	[−5.29]	[7.75]	[6.97]
$w+45$	−0.014**	−0.048***	0.055***	0.015***	0.053***	−0.045***	−0.029***	−0.10***	0.100***
	[−2.37]	[−3.21]	[3.04]	[3.65]	[6.79]	[−4.82]	[−5.14]	[7.62]	[6.12]
$w+48$	−0.011*	−0.051***	0.053***	0.023***	0.056***	−0.036***	−0.033***	−0.11***	0.088***
	[−1.82]	[−3.17]	[3.04]	[6.48]	[8.33]	[−4.21]	[−5.64]	[7.46]	[5.78]
$w+51$	−0.023***	−0.070***	0.054***	0.015***	0.047***	−0.036***	−0.038***	−0.12***	0.091***
	[−3.61]	[−4.95]	[3.18]	[3.81]	[5.77]	[−3.90]	[−6.00]	[8.32]	[5.60]
$w+54$	−0.021***	−0.050***	0.059***	0.011***	0.043***	−0.037***	−0.032***	−0.093***	0.096***
	[−3.43]	[−3.31]	[3.59]	[2.97]	[5.90]	[−3.69]	[−5.40]	[−6.74]	[6.19]
$w+57$	−0.0013	−0.051***	0.064***	0.011***	0.036***	−0.037***	−0.012**	−0.087***	0.10***
	[−0.21]	[−3.30]	[3.38]	[2.76]	[5.11]	[−4.16]	[−1.98]	[−6.17]	[5.35]
$w+60$	−0.022***	−0.071***	0.040**	0.014***	0.040***	−0.038***	−0.036***	−0.11***	0.078***
	[−3.54]	[−4.67]	[2.27]	[3.53]	[5.03]	[−4.08]	[−6.16]	[8.11]	[4.60]

Table 5: **Retail Order Internalization and Tick Size Pilot *Quote Rule.*** This table reports OLS and quantile (median) regression estimates of equation (6.2), comparing stocks in Test Group 1 ($g = 1$) to control stocks. Panels A and C report results for stocks whose average quoted spread in during August, 2016 was below sample median; and Panels B and D report results for stocks with above-median spreads. Sample periods spans the 10 trading day prior to implementation of TSP on 10/03/2016 as well as the 10 trading days following the full implementation of TSP on 10/17/2016 for Test Group 1 stocks. Outcome variables are constructed using trade and quote information of sub-penny-executed off-exchange transactions, and they include (1) the absolute value of $Mroibtrd$; (2) the absolute value of $Mroibvol$; (3) size-weighted average relative % price improvement, defined as the difference between the closer best quoted price and the transaction price, divided by the mid-point of best bid and ask; (4) total price improvement, defined as the sum, in dollars of, dollar price improvements, with respect to the closer best quoted price, across all sub-penny-executed transactions; (5) the total share volume, in round lots, of trades receiving price improvement; and (6) the size-weighted average sup-tick (sub-penny) fraction of execution prices of trades receiving price improvement. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

	Panel A: Low-spresd stocks, OLS						Panel B: High-spresd stocks, OLS					
	Outcome variable						Outcome variable					
	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI
Intercept	0.31***	0.39***	0.19***	367.2***	145.1***	0.30***	0.31***	0.39***	0.19***	367.2***	145.1***	0.30***
	(198.21)	(225.90)	(40.11)	(34.55)	(74.77)	(169.97)	(173.36)	(208.29)	(45.46)	(34.87)	(89.14)	(141.10)
Post	−0.047***	−0.047***	0.015*	31.3*	62.8***	−0.11***	0.10***	0.12***	0.19***	153.5***	−89.6***	0.029***
	(−17.74)	(−16.08)	(1.92)	(1.76)	(18.90)	(−38.56)	(32.11)	(34.77)	(24.48)	(7.83)	(−32.32)	(7.49)
Treat	−0.012***	−0.01**	−0.001	7.60	4.62	0.015***	−0.012***	−0.01**	−0.001	7.60	4.62	0.015***
	(−3.15)	(−2.33)	(−0.09)	(0.29)	(0.97)	(3.48)	(−2.76)	(−2.15)	(−0.10)	(0.30)	(1.16)	(2.89)
Post*Treat	0.0034	0.0015	0.0045	86.9**	13.6*	−0.025***	−0.019**	−0.010	−0.047**	−24.9	−3.34	0.0085
	(0.54)	(0.21)	(0.24)	(2.05)	(1.70)	(−3.54)	(−2.46)	(−1.25)	(−2.51)	(−0.53)	(−0.49)	(0.91)

	Panel C: Low-spresd stocks, Quantile regression						Panel D: High-spresd stocks, Quantile regression					
	Outcome variable						Outcome variable					
	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI
Intercept	0.23***	0.32***	0.068***	84.8***	48.9***	0.22***	0.23***	0.32***	0.068***	84.8***	48.9***	0.22***
	(132.83)	(136.29)	(99.87)	(73.16)	(71.81)	(193.39)	(102.92)	(112.02)	(79.44)	(72.70)	(107.96)	(165.43)
Post	−0.03***	−0.040***	0.019***	39.9***	43.9***	−0.068***	0.097***	0.14***	0.090***	47.5***	−35.1***	0.016***
	(−9.81)	(−10.07)	(16.64)	(20.60)	(37.65)	(−36.44)	(24.58)	(27.06)	(56.62)	(21.89)	(−45.42)	(6.72)
Treat	−0.014***	−0.015***	−0.001	4.16	−0.86	−0.00002	−0.014***	−0.015**	−0.001	4.16	−0.86	−0.00002
	(−3.40)	(−2.62)	(−0.60)	(1.47)	(−0.52)	(−0.01)	(−2.63)	(−2.15)	(−0.48)	(1.46)	(−0.78)	(−0.01)
Post*Treat	0.014**	0.011	0.040***	1.4***	30.6***	−0.004	−0.023**	−0.0075	−0.025***	19.8***	9.3***	0.016***
	(2.04)	(1.18)	(14.56)	(30.63)	(10.87)	(−0.88)	(−2.44)	(−0.61)	(−6.46)	(3.82)	(4.85)	(2.79)

Table 6: **Retail Order Internalization and Tick Size Pilot *Trade Rule*.** This table reports OLS and quantile (median) regression estimates of equation (6.2), comparing stocks in Test Group 2 ($g = 2$) to control stocks. Panels A and C report results for stocks whose average quoted spread in during August, 2016 was below sample median; and Panels B and D report results for stocks with above-median spreads. Sample periods spans the 10 trading day prior to implementation of TSP on 10/03/2016 as well as the 10 trading days following the full implementation of TSP on 10/17/2016 for Test Group 1 stocks. Outcome variables are constructed using trade and quote information of sub-penny-executed off-exchange transactions, and they include (1) the absolute value of $Mroibtrd$; (2) the absolute value of $Mroibvol$; (3) size-weighted average relative % price improvement, defined as the difference between the closer best quoted price and the transaction price, divided by the mid-point of best bid and ask; (4) total price improvement, defined as the sum, in dollars of, dollar price improvements, with respect to the closer best quoted price, across all sub-penny-executed transactions; (5) the total share volume, in round lots, of trades receiving price improvement; and (6) the size-weighted average sup-tick (sub-penny) fraction of execution prices of trades receiving price improvement. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

	Panel A: Low-spresd stocks, OLS						Panel B: High-spresd stocks, OLS					
	Outcome variable						Outcome variable					
	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI
Intercept	0.31***	0.39***	0.19***	369.0***	146.9***	0.30***	0.31***	0.39***	0.19***	369.0***	146.9***	0.30***
	(198.89)	(225.93)	(40.04)	(34.91)	(75.76)	(173.01)	(172.60)	(207.06)	(44.89)	(35.58)	(90.92)	(142.76)
Post	−0.056***	−0.065***	0.0089	62.1***	79.2***	−0.11***	0.087***	0.10***	0.18***	147.3***	−88.8***	0.046***
	(−21.80)	(−22.28)	(1.15)	(3.54)	(23.91)	(−37.93)	(27.91)	(31.63)	(23.53)	(7.76)	(−32.19)	(12.10)
Treat	0.0043	0.011**	0.015	−36.0	−1382.4***	0.0004	0.0043	0.011**	0.015	−36.0	−13.8***	0.0004
	(1.13)	(2.53)	(1.33)	(−1.38)	(−2.92)	(0.08)	(0.98)	(2.32)	(1.49)	(−1.41)	(−3.51)	(0.07)
Post*Treat	0.032***	0.076***	0.0063	−11.7	−32.8***	−0.058***	0.042***	0.052***	−0.0059	17.7	59.2	−0.099***
	(5.13)	(10.79)	(0.33)	(−0.27)	(−4.07)	(−8.20)	(5.44)	(6.27)	(−0.30)	(0.37)	(0.88)	(−10.19)

	Panel C: Low-spresd stocks, Quantile regression						Panel D: High-spresd stocks, Quantile regression					
	Outcome variable						Outcome variable					
	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI	Mroibtrd	Mroibvol	Relative %-PI	Relative $-PI	PI shr vol	Sub-tick PI
Intercept	0.22***	0.31***	0.068***	85.3***	49.5***	0.22***	0.22***	0.31***	0.068***	85.3***	49.5***	0.22***
	(125.61)	(131.66)	(102.76)	(70.72)	(71.84)	(196.96)	(97.95)	(111.11)	(79.70)	(72.43)	(109.61)	(165.95)
Post	−0.036***	−0.052***	0.015***	58.5***	57.9***	−0.06***	0.075***	0.12***	0.082***	46.8***	−32.9***	0.03***
	(−11.86)	(−13.06)	(13.95)	(29.23)	(49.29)	(−33.20)	(18.57)	(23.75)	(52.31)	(21.69)	(−42.81)	(12.62)
Treat	0.0058	0.0065	0.00064	−4.75	−54.6***	0.0018	0.0058	0.0065	0.0006	−4.75	−5.5***	0.0018
	(1.31)	(1.12)	(0.39)	(−1.60)	(−3.25)	(0.68)	(1.03)	(0.94)	(0.31)	(−1.64)	(−4.96)	(0.58)
Post*Treat	0.027***	0.091***	0.047***	26.1***	−23.3***	−0.059***	0.028***	0.092***	0.018***	15.9***	1.2	−0.086***
	(3.71)	(9.32)	(17.48)	(5.32)	(−8.13)	(−13.20)	(2.75)	(7.45)	(4.50)	(2.90)	(0.64)	(−14.25)

Table 7: **Portfolios of *Mroibvol*: Contemporaneous Return, Liquidity, Institutional Trading, and Short Interest.** The table presents the cross-sectional relationship between weekly *Mroibvol* and the contemporaneous return, institutional trade, and liquidity outcomes. Outcome variables include (1) returns (close-to-close, intraday, and overnight returns, with a version of overnight returns shifted by one day); (2) liquidity (dollar and relative quoted spreads, depth, in shares, and abnormal off-exchange midpoint executions of larger trades); (3) institutional trading (order flow and implementation shortfall, in bps/$1m); and (4) short interest (% change in bi-weeklys short interest). Each weekly cross-section is sorted into deciles of *Mroibvol*. The the average of an outcome variable Y is calculated by *Mroibvol* decile in each cross-section before the averages of mean-Y time-series are calculated. For short interest, bi-weekly relative % changes in short interest are constructed and *Mroibvol* is aggregated over two-week periods, before forming *Mroibvol* portfolios. Median short interest changes by *Mroibvol* and stock size tercile, before averaging the time-series of medians.

	Deciles of internalized retail order flow imbalance (*Mroibvol*)									
	1	2	3	4	5	6	7	8	9	10
Mroibvol	−2.043	−1.132	−0.745	−0.467	−0.238	−0.033	0.173	0.417	0.763	1.607
Returns (%)										
Close-to-close return	−0.019	0.091	0.135	0.179	0.219	0.249	0.269	0.290	0.267	0.321
Intraday return	0.098	0.053	0.019	−0.005	−0.063	−0.118	−0.176	−0.210	−0.237	−0.138
Overnight return	−0.116	0.038	0.117	0.184	0.283	0.367	0.445	0.500	0.505	0.459
Next-day overnight return	−0.134	0.019	0.100	0.166	0.257	0.340	0.423	0.490	0.488	0.456
Institutional Trading										
Order flow imbalance	0.336	0.299	0.282	0.266	0.251	0.246	0.227	0.226	0.231	0.211
Implementation shortfall	69.44	9.45	4.05	2.92	3.41	2.70	4.54	5.94	5.55	15.51
Change in Short Interest (%)										
Small stocks	−2.58	−1.90	−1.38	−0.87	−0.61	0.22	0.16	0.70	1.21	2.25
Mid-sized stocks	−0.70	−0.54	−0.39	−0.10	−0.01	0.29	0.26	0.37	0.63	0.41
Large stocks	−1.16	−0.58	−0.72	−0.33	−0.25	−0.27	0.06	0.04	0.20	0.80
Liquidity										
Dollar quoted spread	0.089	0.068	0.058	0.054	0.053	0.057	0.054	0.055	0.064	0.093
Relative quoted spread	0.0069	0.0046	0.0038	0.0033	0.0031	0.0032	0.0031	0.0034	0.0043	0.0070
Ask-side depth	972	1,288	1,409	1,557	1,738	1,857	1,893	1,751	1,500	905
Bid-side depth	972	1,306	1,449	1,602	1,790	1,935	2,000	1,864	1,618	960
Large midpoint executions	0.78	0.87	0.92	0.97	0.99	1.02	1.05	1.03	1.04	0.95

Table 8: **Asymmetric Persistence in Institutional Order Flow: Implications for Retail Flow Internalization.** This table presents estimates of the predictive power of past institutional order flow, conditional on it sign, for both current institutional order flow and current internalized retail order flow. Columns (1)–(4) report estimation results of equation (3) for $i \in \{3,4,5,6\}$ and $X = Inoibvol_w$. Columns (5)–(8) report estimation results of equation (3) for $i \in \{3,4,5,6\}$ and $X = Mroibvol_w$. Fama-MacBeth regressions are used with Newey-West-corrected standard errors using 6 lags. The sample contains stocks with average ANcerno-to-CRSP daily volume of 6.8% or higher. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

	Dependent variable: $Inoibvol_w$				Dependent variable: $Mroibvol_w$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.065**	0.038	0.023	0.0088	−0.16***	−0.15***	−0.15***	−0.14***
	[2.46]	[1.42]	[0.86]	[0.32]	[−14.54]	[−13.32]	[−12.50]	[−11.77]
$Inoibvol_{w-1}$	0.33***	0.33***	0.33***	0.33***	−0.016***	−0.016***	−0.016***	−0.016***
	[58.36]	[59.42]	[58.50]	[58.00]	[−7.43]	[−7.86]	[−7.66]	[−7.58]
$I(Inoibvol_{w-1} < 0) \times Inoibvol_{w-1}$	0.020***	0.022***	0.022***	0.023***	0.0083***	0.0085***	0.0081**	0.0085***
	[2.71]	[2.98]	[3.03]	[3.24]	[2.63]	[2.69]	[2.56]	[2.65]
$Inoibvol_{w-2}$	0.075***	0.072***	0.071***	0.069***	−0.0067***	−0.0062***	−0.0060***	−0.0059***
	[17.07]	[16.60]	[16.60]	[15.35]	[−3.41]	[−3.06]	[−2.94]	[−2.90]
$I(Inoibvol_{w-2} < 0) \times Inoibvol_{w-2}$	−0.023***	−0.020***	−0.020***	−0.018**	0.0059*	0.0051	0.0046	0.0044
	[−3.06]	[−2.70]	[−2.68]	[−2.46]	[1.85]	[1.57]	[1.38]	[1.31]
$Inoibvol_{w-3}$	0.062***	0.048***	0.045***	0.043***	−0.0069***	−0.0054***	−0.0052**	−0.0050**
	[13.26]	[10.52]	[9.90]	[9.65]	[−3.40]	[−2.64]	[−2.53]	[−2.41]
$I(Inoibvol_{w-3} < 0) \times Inoibvol_{w-3}$	−0.017***	−0.014**	−0.012*	−0.011*	0.0091***	0.0079***	0.0078***	0.0077**
	[−2.63]	[−2.14]	[−1.86]	[−1.79]	[3.09]	[2.66]	[2.63]	[2.54]
$Inoibvol_{w-4}$		0.052***	0.040***	0.037***		−0.0055***	−0.0048**	−0.0050**
		[12.29]	[9.65]	[8.77]		[−2.64]	[−2.30]	[−2.40]
$I(Inoibvol_{w-4} < 0) \times Inoibvol_{w-4}$		−0.023***	−0.021***	−0.019***		0.0078**	0.0080***	0.0078***
		[−3.51]	[−3.20]	[−2.90]		[2.58]	[2.69]	[2.60]
$Inoibvol_{w-5}$			0.041***	0.031***			−0.0041**	−0.0028
			[10.22]	[7.73]			[−2.11]	[−1.38]
$I(Inoibvol_{w-5} < 0) \times Inoibvol_{w-5}$			−0.029***	−0.025***			0.00047	0.000084
			[−4.14]	[−3.78]			[0.16]	[0.03]
$Inoibvol_{w-6}$				0.037***				−0.0044**
				[9.35]				[−2.15]
$I(Inoibvol_{w-6} < 0) \times Inoibvol_{w-6}$				−0.026***				0.0019
				[−3.79]				[0.63]
Observations	976,110	976,110	976,110	976,110	976,110	976,110	976,110	976,110

Table 9: **Predictability of Institutional Order Flow Using Internalized Retail Trading Imbalance.** This table presents estimates of the predictive power of past internalized order flow, conditional the sign the corresponding institutional order flow, for current institutional order flow. Equation (4) for $i \in \{3, 4, 5, 6\}$ and $X = Inoibvol_w$ is estimated using Fama-MacBeth regressions with Newey-West-corrected standard errors using 6 lags. The sample contains stocks with average ANcerno-to-CRSP daily volume of 6.8% or higher. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

	(1)	(2)	(3)	(4)
Constant	1.04***	1.09***	1.14***	1.17***
	[39.71]	[40.99]	[41.98]	[43.14]
$Mroibvol_{w-1}$	−0.020***	−0.021***	−0.021***	−0.020***
	[−3.69]	[−3.74]	[−3.74]	[−3.57]
$I(Inoibvol_{w-1} < 0) \times Mroibvol_{w-1}$	0.021***	0.020***	0.020***	0.021***
	[2.85]	[2.78]	[2.81]	[2.79]
$Mroibvol_{w-2}$	−0.013**	−0.014**	−0.013**	−0.013**
	[−2.43]	[−2.56]	[−2.43]	[−2.38]
$I(Inroibvol_{w-2} < 0) \times Mroibvol_{w-2}$	0.025***	0.025***	0.025***	0.024***
	[3.41]	[3.39]	[3.42]	[3.30]
$Mroibvol_{w-3}$	−0.0043	−0.0063	−0.0054	−0.0067
	[−0.72]	[−1.13]	[−0.93]	[−1.14]
$I(Inoibvol_{w-3} < 0) \times Mroibvol_{w-3}$	0.017**	0.018***	0.019***	0.020***
	[2.38]	[2.59]	[2.59]	[2.72]
$Mroibvol_{w-4}$		0.0047	0.0054	0.0035
		[0.70]	[0.87]	[0.57]
$I(Inoibvol_{w-4} < 0) \times Mroibvol_{w-4}$		0.0017	0.0038	0.0038
		[0.23]	[0.51]	[0.52]
$Mroibvol_{w-5}$			−0.0058	−0.0065
			[−1.08]	[−1.20]
$I(Inoibvol_{w-5} < 0) \times Mroibvol_{w-5}$			−0.0036	−0.0018
			[−0.45]	[−0.22]
$Mroibvol_{w-6}$				0.0025
				[0.42]
$I(Inoibvol_{w-6} < 0) \times Mroibvol_{w-6}$				0.0056
				[0.63]
Observations	976,110	976,110	976,110	976,110

Table 10: **Portfolios of *Mroibvol* and Future Weekly Returns.** The table presents the cross-sectional relationships between $Mroibvol$ and future weekly (%) returns. Each cross-section is sorted into portfolios (deciles) of $Mroibvol_{w-1}$ to calculate portfolio-specific averages of future close-to-close (CCR), intraday (IDR), and overnight (ONR) returns in week $w+i$, with $i \in \{0, 1, 2, 3, 6, 9, 12, 24, 36, 39, 42, 45, 48, 51, 54, 57, 60\}$. The means of the time-series of portfolio future returns are presented by $Mroibvol$ decile.

		Deciles of $Mroibvol_{w-1}$									
Week	Variable	1	2	3	4	5	6	7	8	9	10
w	CCR	0.07	0.14	0.14	0.16	0.17	0.15	0.16	0.18	0.28	0.42
	IDR	0.08	0.05	−0.03	−0.06	−0.12	−0.19	−0.22	−0.23	−0.14	0.03
	ONR	−0.01	0.09	0.17	0.22	0.29	0.34	0.38	0.41	0.42	0.40
$w+1$	CCR	0.13	0.15	0.14	0.15	0.15	0.14	0.17	0.16	0.21	0.34
	IDR	0.10	0.01	−0.06	−0.09	−0.14	−0.19	−0.21	−0.20	−0.16	−0.01
	ONR	0.03	0.14	0.21	0.24	0.29	0.33	0.38	0.36	0.37	0.35
$w+2$	CCR	0.14	0.16	0.17	0.16	0.16	0.15	0.16	0.17	0.21	0.31
	IDR	0.10	0.02	−0.03	−0.08	−0.13	−0.18	−0.20	−0.20	−0.15	−0.02
	ONR	0.04	0.14	0.20	0.24	0.29	0.34	0.36	0.37	0.35	0.33
$w+3$	CCR	0.17	0.20	0.18	0.18	0.17	0.17	0.17	0.18	0.23	0.29
	IDR	0.10	0.04	−0.03	−0.07	−0.12	−0.17	−0.18	−0.18	−0.13	−0.01
	ONR	0.07	0.16	0.22	0.25	0.29	0.33	0.35	0.36	0.35	0.30
$w+6$	CCR	0.19	0.17	0.19	0.18	0.16	0.16	0.18	0.18	0.21	0.26
	IDR	0.09	−0.01	−0.04	−0.08	−0.14	−0.17	−0.16	−0.16	−0.11	−0.03
	ONR	0.10	0.18	0.23	0.26	0.29	0.33	0.34	0.34	0.33	0.29

Continued on next page

Table 10 – *continued from previous page*

		Deciles of $Mroibvol_{w-1}$									
Week	Variable	1	2	3	4	5	6	7	8	9	10
$w+9$	CCR	0.14	0.16	0.16	0.13	0.13	0.12	0.10	0.11	0.15	0.19
	IDR	0.02	−0.02	−0.06	−0.12	−0.16	−0.19	−0.22	−0.22	−0.16	−0.07
	ONR	0.12	0.18	0.22	0.24	0.29	0.31	0.32	0.32	0.31	0.26
$w+12$	CCR	0.15	0.12	0.11	0.10	0.08	0.07	0.07	0.09	0.12	0.18
	IDR	0.04	−0.04	−0.09	−0.14	−0.18	−0.22	−0.23	−0.19	−0.17	−0.09
	ONR	0.11	0.16	0.19	0.24	0.25	0.29	0.30	0.28	0.29	0.27
$w+24$	CCR	0.21	0.18	0.19	0.15	0.14	0.13	0.13	0.15	0.16	0.22
	IDR	0.06	−0.02	−0.04	−0.10	−0.13	−0.16	−0.18	−0.15	−0.12	−0.02
	ONR	0.15	0.20	0.23	0.25	0.27	0.30	0.31	0.30	0.28	0.24
$w+36$	CCR	0.22	0.21	0.20	0.17	0.15	0.13	0.14	0.15	0.17	0.20
	IDR	0.09	0.03	−0.01	−0.06	−0.10	−0.13	−0.15	−0.13	−0.10	−0.04
	ONR	0.13	0.19	0.21	0.22	0.25	0.27	0.29	0.27	0.27	0.24
$w+39$	CCR	0.16	0.17	0.16	0.14	0.13	0.11	0.10	0.10	0.13	0.14
	IDR	0.03	−0.01	−0.04	−0.06	−0.10	−0.14	−0.16	−0.15	−0.11	−0.07
	ONR	0.13	0.18	0.20	0.20	0.23	0.26	0.26	0.25	0.24	0.21
$w+42$	CCR	0.18	0.15	0.13	0.13	0.12	0.11	0.09	0.08	0.12	0.15
	IDR	0.05	−0.01	−0.05	−0.07	−0.11	−0.15	−0.17	−0.16	−0.12	−0.05
	ONR	0.12	0.16	0.18	0.20	0.23	0.25	0.26	0.25	0.23	0.20

Continued on next page

Table 10 – *continued from previous page*

		Deciles of $Mroibvol_{w-1}$									
Week	Variable	1	2	3	4	5	6	7	8	9	10
$w+45$	CCR	0.19	0.17	0.15	0.14	0.12	0.10	0.09	0.10	0.12	0.14
	IDR	0.06	0.00	−0.05	−0.07	−0.12	−0.16	−0.17	−0.15	−0.11	−0.06
	ONR	0.13	0.18	0.20	0.21	0.24	0.26	0.26	0.25	0.23	0.20
$w+48$	CCR	0.14	0.13	0.11	0.09	0.07	0.05	0.06	0.04	0.06	0.10
	IDR	0.02	−0.02	−0.08	−0.11	−0.17	−0.19	−0.19	−0.20	−0.17	−0.10
	ONR	0.12	0.16	0.18	0.21	0.23	0.25	0.25	0.24	0.23	0.20
$w+51$	CCR	0.13	0.10	0.12	0.07	0.02	0.02	0.01	0.03	0.04	0.07
	IDR	0.01	−0.07	−0.08	−0.13	−0.19	−0.21	−0.23	−0.22	−0.18	−0.11
	ONR	0.11	0.17	0.20	0.21	0.22	0.23	0.24	0.24	0.22	0.18
$w+54$	CCR	0.08	0.10	0.08	0.08	0.04	0.01	0.00	−0.01	0.03	0.06
	IDR	−0.04	−0.05	−0.09	−0.12	−0.16	−0.21	−0.22	−0.23	−0.18	−0.11
	ONR	0.12	0.15	0.17	0.20	0.20	0.22	0.22	0.22	0.21	0.16
$w+57$	CCR	0.07	0.03	0.04	0.01	−0.01	−0.01	−0.01	0.02	0.03	0.05
	IDR	−0.07	−0.11	−0.13	−0.17	−0.20	−0.22	−0.23	−0.20	−0.18	−0.11
	ONR	0.13	0.14	0.18	0.19	0.20	0.21	0.22	0.21	0.21	0.15
$W+60$	CCR	0.08	0.07	0.04	0.01	0.00	0.00	−0.01	−0.02	0.00	0.00
	IDR	−0.04	−0.08	−0.13	−0.18	−0.21	−0.22	−0.24	−0.23	−0.21	−0.17
	ONR	0.12	0.15	0.17	0.19	0.21	0.22	0.23	0.22	0.21	0.17

Table 11: **Decomposition of *Mroibvol*'s Predictive Power for Future Weekly Returns (%).** This table presents a decomposition to the overall predictive power of $Mroibvol_{w-1}$ for future returns into those of persistence, contrarian trading, and residual components. Daily returns are calculated based on the mid-points of best bid and ask prices at close as well as open prices, decomposing each day's close-to-close returns into intraday (open-to-close) and overnight (close-to-open) before aggregating each return type into weekly observations. $Mroibvol_{w-1}$ is decomposed into Persistence, Contrarian, and Other components according to equation (5). Each of the three return cross-sections for a given week $w+i$, with $\in \{0,1,2,3,6,9,12,24,36,39,41,45,58,51,54,57,60\}$, is decomposed based on the sign of week $w-1$'s internalized order flow to form a total of nine samples. According to equation (9), week $w+i$ returns in each sample are regressed on week $w-1$'s Persistence, Contrarian, and Other components of the internalized order flow ($Mroibvol_{w-1}$), whose loadings are reported in the table, and control variables including last week's return (R_{w-1}), last month's return (RET_{-1}), the return over the preceding five months ($RET_{(-7,-2)}$), volatility (VOLAT), and natural logs of turnover (ln(TO)), market capitalization (ln(Size)), and book-to-market ratio (ln(BM)). Estimates are based Fama-Macbeth regressions, featuring Newey-West corrected standard errors with 6 lags. Sample includes NMS common shares from Jan 2010 – Dec 2014, excluding observations with previous month-end's closing price below $1. Numbers in brackets reflect t-statistics, and symbols ***, **, and * identify statistical significance at the 1%, 5%, and 10% type one errors, respectively.

Dependent Variable =		Close-to-close return			Overnight return			Intraday return		
		$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	Component	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
w	Persistence	0.32***	0.27***	0.43***	0.51***	0.47***	0.48***	−0.19***	−0.20***	−0.058
		[8.76]	[6.60]	[8.20]	[23.73]	[19.69]	[15.02]	[−5.53]	[−4.85]	[−1.21]
	Contrarian	1.22	0.86	1.56*	−0.74**	−0.78*	−0.42	1.97**	1.64	1.98**
		[1.48]	[0.88]	[1.79]	[−2.21]	[−1.80]	[−1.12]	[2.18]	[1.52]	[2.07]
	Other	0.079***	0.045***	0.12***	0.10***	0.15***	−0.013	−0.025***	−0.10***	0.13***
		[13.39]	[3.17]	[7.26]	[23.43]	[19.30]	[−1.46]	[−3.96]	[−6.97]	[8.71]
$w+1$	Persistence	0.27***	0.27***	0.31***	0.43***	0.43***	0.40***	−0.16***	−0.16***	−0.092*
		[7.43]	[5.99]	[6.21]	[20.64]	[17.88]	[14.18]	[−4.45]	[3.75]	[−1.77]
	Contrarian	−0.80	−1.40	−0.22	−0.39	−0.34	−0.31	−0.42	−1.06	0.094
		[−0.85]	[−1.34]	[−0.20]	[−1.02]	[−0.81]	[−0.57]	[−0.48]	[1.11]	[0.10]
	Other	0.046***	0.011	0.085***	0.078***	0.13***	−0.023**	−0.032***	−0.12***	0.11***
		[7.74]	[0.76]	[4.94]	[22.98]	[18.46]	[−2.48]	[−5.72]	[8.61]	[6.70]
$w+2$	Persistence	0.20***	0.19***	0.24***	0.39***	0.39***	0.34***	−0.19***	−0.20***	−0.098**
		[5.84]	[4.32]	[5.06]	[19.06]	[15.03]	[11.30]	[−5.46]	[−4.62]	[−1.98]
	Contrarian	−0.80	−0.80	−1.05	−0.78*	−0.41	−1.25	−0.027	−0.38	0.20
		[−1.04]	[−0.96]	[−0.88]	[−1.71]	[−0.93]	[−1.60]	[−0.04]	[−0.58]	[0.23]
	Other	0.042***	0.015	0.083***	0.067***	0.11***	−0.033***	−0.025***	−0.097***	0.12***
		[7.10]	[1.07]	[4.53]	[19.39]	[16.11]	[−3.67]	[−4.21]	[−7.18]	[6.87]

Continued on next page

Table 11 – *continued from previous page*

Dependent Variable =		Close-to-close return			Overnight return			Intraday return		
		$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	Component	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
$w+3$	Persistence	0.11***	0.076*	0.18***	0.33***	0.30***	0.32***	−0.22***	−0.22***	−0.14***
		[3.41]	[1.93]	[3.95]	[18.06]	[12.67]	[10.96]	[−6.96]	[5.74]	[−3.18]
	Contrarian	−2.06	−2.24*	−2.10	0.16	0.49	−0.29	−2.23*	−2.74*	−1.81
		[−1.60]	[−1.72]	[−1.46]	[0.41]	[1.04]	[−0.59]	[−1.68]	[1.88]	[−1.34]
	Other	0.031***	0.0033	0.075***	0.059***	0.11***	−0.039***	−0.027***	−0.11***	0.11***
		[5.97]	[0.24]	[4.43]	[18.98]	[15.66]	[−4.45]	[−5.01]	[7.82]	[7.06]
$w+6$	Persistence	0.12***	0.13***	0.13***	0.29***	0.29***	0.27***	−0.17***	−0.16***	−0.14***
		[3.48]	[3.09]	[2.64]	[14.27]	[12.46]	[8.25]	[−4.94]	[−4.03]	[−2.87]
	Contrarian	0.091	0.43	−0.067	−0.71	−0.083	−1.20*	0.80	0.51	1.13
		[0.09]	[0.37]	[−0.06]	[−1.33]	[−0.16]	[−1.78]	[0.95]	[0.54]	[1.03]
	Other	0.015***	−0.020	0.057***	0.043***	0.099***	−0.037***	−0.028***	−0.12***	0.093***
		[2.63]	[−1.56]	[3.28]	[13.28]	[14.39]	[−4.33]	[−5.24]	[−9.55]	[6.08]
$w+9$	Persistence	0.016	0.031	0.025	0.20***	0.24***	0.13***	−0.19***	−0.20***	−0.10**
		[0.48]	[0.76]	[0.57]	[9.45]	[9.64]	[3.60]	[−5.95]	[−5.19]	[−2.42]
	Contrarian	−1.23*	−1.09	−1.22*	−0.54	−0.45	−0.79	−0.69	−0.64	−0.43
		[−1.68]	[−1.08]	[−1.73]	[−1.06]	[−0.75]	[−1.40]	[−1.20]	[−0.90]	[−0.60]
	Other	0.011*	−0.012	0.055***	0.031***	0.079***	−0.056***	−0.020***	−0.092***	0.11***
		[1.94]	[−0.95]	[3.06]	[9.13]	[12.01]	[−6.85]	[−3.85]	[−7.40]	[6.90]
$w+12$	Persistence	0.018	0.0063	0.042	0.21***	0.19***	0.21***	−0.19***	−0.18***	−0.17***
		[0.45]	[0.13]	[0.83]	[9.67]	[7.59]	[6.36]	[−5.08]	[−4.03]	[−3.62]
	Contrarian	0.11	0.47	−0.72	−0.029	−0.032	−0.17	0.14	0.50	−0.56
		[0.19]	[0.67]	[−0.85]	[−0.06]	[−0.06]	[−0.28]	[0.19]	[0.66]	[−0.70]
	Other	0.0037	−0.050***	0.069***	0.035***	0.070***	−0.021**	−0.031***	−0.12***	0.090***
		[0.73]	[−3.82]	[3.88]	[10.78]	[11.36]	[−2.39]	[−5.96]	[−9.82]	[5.65]
$w+24$	Persistence	−0.028	−0.13***	0.11**	0.14***	0.13***	0.14***	−0.17***	−0.26***	−0.032
		[−0.84]	[−2.94]	[2.30]	[6.78]	[5.04]	[4.45]	[−5.14]	[−6.34]	[−0.68]
	Contrarian	−0.36	−1.11	0.20	−0.46	−0.59	−0.25	0.096	−0.52	0.45
		[−0.46]	[−1.08]	[0.29]	[−1.12]	[−1.29]	[−0.42]	[0.11]	[−0.47]	[0.53]
	Other	−0.0021	−0.049***	0.045***	0.022***	0.064***	−0.042***	−0.024***	−0.11***	0.087***
		[−0.37]	[−3.44]	[2.60]	[6.75]	[10.28]	[−4.71]	[−3.94]	[−8.39]	[5.04]

Continued on next page

Table 11 – *continued from previous page*

Dependent Variable =		Close-to-close return			Overnight return			Intraday return		
		$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	Component	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
$w+36$	Persistence	−0.00053	−0.019	0.052	0.12***	0.14***	0.094***	−0.12***	−0.16***	−0.042
		[−0.01]	[−0.40]	[1.12]	[5.92]	[5.06]	[2.99]	[−3.62]	[−3.47]	[−0.97]
	Contrarian	0.44	−0.40	1.56	0.42	0.30	0.69	0.018	−0.70	0.87
		[0.36]	[−0.35]	[1.00]	[0.81]	[0.41]	[1.28]	[0.02]	[−0.87]	[0.65]
	Other	−0.0027	−0.045***	0.069***	0.027***	0.061***	−0.020***	−0.029***	−0.11***	0.089***
		[−0.43]	[−2.84]	[4.21]	[7.16]	[8.52]	[−2.63]	[−4.94]	[−7.68]	[5.80]
$w+39$	Persistence	−0.052	−0.065	−0.019	0.075***	0.054**	0.057*	−0.13***	−0.12***	−0.077
		[−1.40]	[−1.38]	[−0.40]	[3.45]	[2.03]	[1.85]	[−3.53]	[−2.69]	[−1.54]
	Contrarian	0.41	0.60	−0.13	0.015	0.31	−0.18	0.39	0.29	0.049
		[0.47]	[0.74]	[−0.11]	[0.04]	[0.67]	[−0.39]	[0.44]	[0.35]	[0.04]
	Other	−0.0090	−0.037***	0.045***	0.014***	0.042***	−0.045***	−0.023***	−0.079***	0.090***
		[−1.55]	[−2.58]	[2.80]	[4.09]	[6.46]	[−5.41]	[−4.33]	[−5.94]	[5.65]
$w+42$	Persistence	−0.031	−0.094**	0.098*	0.073***	0.050*	0.085***	−0.10***	−0.14***	0.013
		[−0.80]	[−2.12]	[1.85]	[3.34]	[1.84]	[2.60]	[−2.95]	[−3.42]	[0.25]
	Contrarian	−0.72	−1.08	−0.20	−0.13	0.025	−0.15	−0.58	−1.10	−0.042
		[−0.93]	[−1.24]	[−0.21]	[−0.41]	[0.07]	[−0.19]	[−0.74]	[−1.38]	[−0.04]
	Other	−0.010*	−0.041***	0.069***	0.016***	0.056***	−0.045***	−0.026***	−0.097***	0.11***
		[−1.93]	[−3.07]	[3.78]	[4.36]	[7.92]	[−5.04]	[−5.03]	[−7.63]	[6.87]
$w+45$	Persistence	0.0032	0.015	0.016	0.12***	0.12***	0.096***	−0.11***	−0.10**	−0.080*
		[0.09]	[0.34]	[0.33]	[5.62]	[4.68]	[3.28]	[−3.37]	[−2.50]	[−1.81]
	Contrarian	−0.63	−0.59	−0.80	0.10	0.64	−0.61	−0.73	−1.23	−0.19
		[−0.49]	[−0.52]	[−0.51]	[0.23]	[1.15]	[−1.15]	[−0.57]	[−0.98]	[−0.14]
	Other	−0.014**	−0.049***	0.054***	0.011***	0.051***	−0.048***	−0.025***	−0.100***	0.10***
		[−2.30]	[−3.30]	[3.01]	[2.83]	[6.51]	[−5.17]	[−4.49]	[−7.59]	[6.32]
$w+48$	Persistence	−0.083**	−0.11**	−0.022	0.12***	0.12***	0.093***	−0.20***	−0.22***	−0.12**
		[−2.20]	[−2.43]	[−0.42]	[4.95]	[4.04]	[2.96]	[−5.77]	[−5.61]	[−2.35]
	Contrarian	−2.48***	−2.14**	−3.00***	−1.11**	−1.11	−1.03***	−1.37**	−1.03	−1.97***
		[−3.07]	[−2.16]	[−3.76]	[−2.40]	[−1.43]	[−2.92]	[−2.00]	[−1.30]	[−2.61]
	Other	−0.0049	−0.050***	0.061***	0.021***	0.054***	−0.037***	−0.025***	−0.10***	0.098***
		[−0.87]	[−3.15]	[3.48]	[6.12]	[8.13]	[−4.48]	[−4.52]	[−7.39]	[6.39]

Continued on next page

Table 11 – *continued from previous page*

Dependent Variable =		Close-to-close return			Overnight return			Intraday return		
		$Mroibvol_{w-1}$			$Mroibvol_{w-1}$			$Mroibvol_{w-1}$		
Week	Component	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
$w+51$	Persistence	−0.035	−0.074	0.052	0.10***	0.097***	0.096***	−0.14***	−0.17***	−0.044
		[−0.93]	[−1.56]	[1.19]	[4.60]	[3.12]	[2.97]	[−3.65]	[−3.54]	[−0.97]
	Contrarian	−1.86**	−1.79**	−2.21**	−0.63	−0.72	−0.55	−1.23**	−1.06	−1.66**
		[−2.42]	[−2.13]	[−2.40]	[−1.34]	[−1.13]	[−1.26]	[−2.13]	[−1.45]	[−2.26]
	Other	−0.022***	−0.069***	0.055***	0.011***	0.044***	−0.039***	−0.034***	−0.11***	0.094***
$w+54$	Persistence	−0.013	−0.0068	0.020	0.11***	0.10***	0.094***	−0.12***	−0.11**	−0.074
		[−0.35]	[−0.15]	[0.38]	[4.84]	[3.90]	[2.90]	[−3.29]	[−2.55]	[−1.56]
	Contrarian	−1.16*	−1.14	−1.32*	0.28	0.47	−0.013	−1.45**	−1.61**	−1.31*
		[−1.79]	[−1.36]	[−1.68]	[0.85]	[1.02]	[−0.03]	[−2.45]	[−2.28]	[−1.80]
	Other	−0.019***	−0.048***	0.060***	0.0077**	0.041***	−0.042***	−0.027***	−0.089***	0.10***
		[−3.24]	[−3.24]	[3.66]	[2.11]	[5.67]	[−4.13]	[−4.77]	[−6.61]	[6.56]
$w+57$	Persistence	−0.053	−0.094**	0.030	0.081***	0.082***	0.061*	−0.13***	−0.18***	−0.031
		[−1.35]	[−2.04]	[0.57]	[3.43]	[2.85]	[1.72]	[−3.80]	[−4.09]	[−0.67]
	Contrarian	0.67	0.41	0.99	1.02***	1.03**	1.00**	−0.35	−0.62	−0.017
		[0.79]	[0.48]	[1.01]	[2.82]	[2.46]	[2.26]	[−0.39]	[−0.62]	[−0.02]
	Other	−0.0015	−0.051***	0.065***	0.0071*	0.033***	−0.040***	−0.0086	−0.084***	0.10***
		[−0.25]	[−3.41]	[3.42]	[1.93]	[4.76]	[−4.58]	[−1.47]	[−6.10]	[5.56]
$w+60$	Persistence	−0.058	−0.046	−0.045	0.13***	0.14***	0.090***	−0.19***	−0.18***	−0.13**
		[−1.34]	[−0.90]	[−0.80]	[5.36]	[4.35]	[2.69]	[−4.64]	[−3.74]	[−2.47]
	Contrarian	−0.87	−0.82	−1.12	0.41	0.41	0.47	−1.28	−1.22	−1.59
		[−1.04]	[−0.72]	[−1.03]	[0.73]	[0.56]	[0.68]	[−1.23]	[−0.96]	[−1.39]
	Other	−0.020***	−0.071***	0.045**	0.012***	0.036***	−0.038***	−0.032***	−0.11***	0.083***
		[−3.34]	[−4.78]	[2.57]	[3.06]	[4.75]	[−4.17]	[−5.59]	[−8.11]	[4.90]

Figure 3: **Retail Order Types:** This figure uses buy orders to illustrate different retail order types from a wholesaler's perspective. When best bid and ask prices are \$9.97 and \$10.03 respectively, market and marketable limit buy orders seek execution at prices at or below the best ask of \$10.03. Non-marketable retail buy orders quoted above the midpoint, \$10.00, but below \$10.03 may be profitable to execute if internalized. Non-marketable retail buy orders quoted above the best bid, \$9.97, but below the midpoint, \$10.00, are unlikely to be profitable to execute if internalized and hence are likely routed to an exchange and added the order book. Non-marketable retail buy orders quoted at or below the best bid price of \$9.97 are most likely to be routed to an exchange and added to the order book.

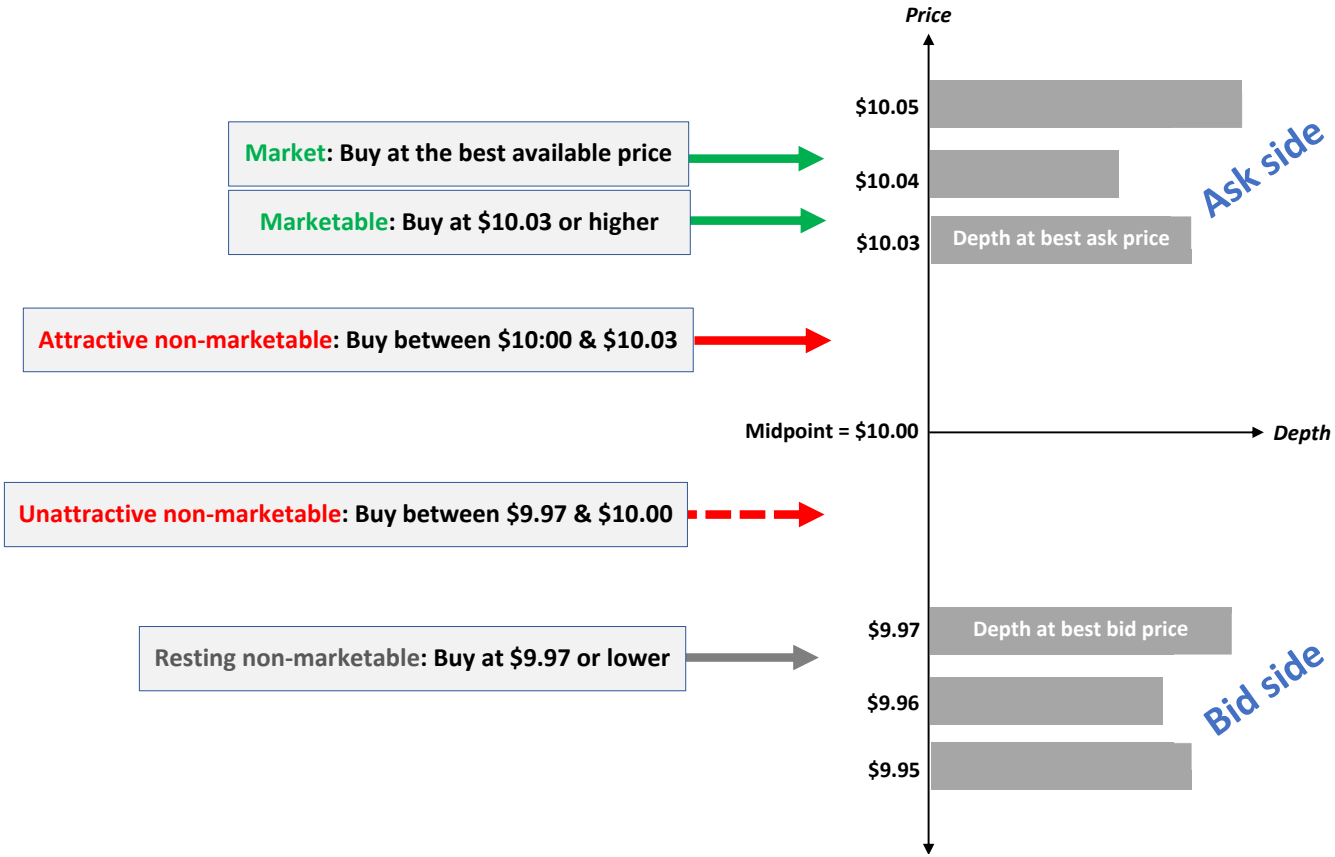


Figure 4: **Retail Order Flow Internalization and PFOF.**

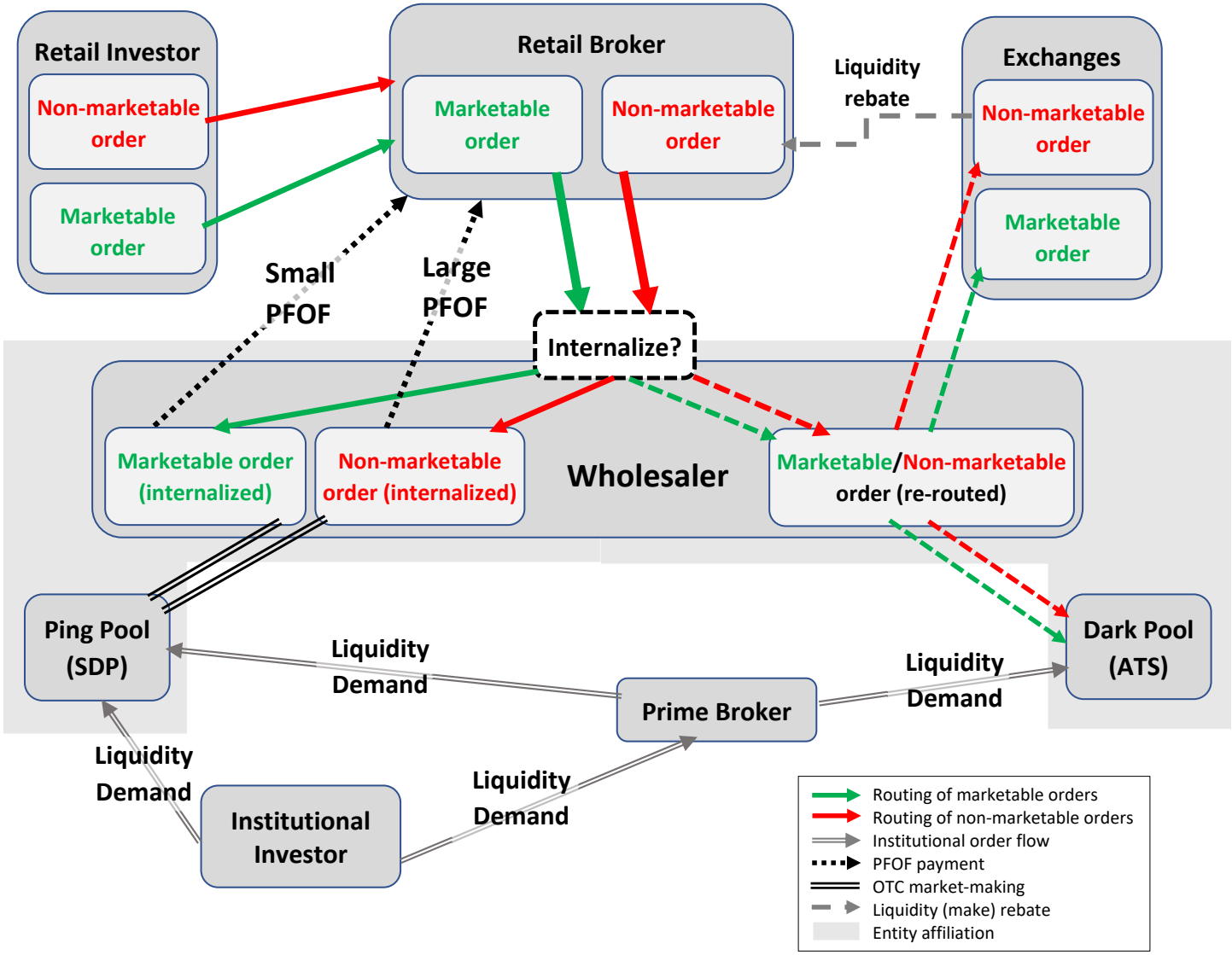


Figure 5: **Tick Size Pilot: *Quote Rule.*** This figure provides visual evidence associated with the results of the Difference-in-Difference specification in equation (6.2) for Test Group 1. The sample period spans the 10 trading days prior to the TSP's implementation on 10/03/2016 as well as the 10 trading days following its full implementation on 10/17/2016. The figure plots the daily medians for six outcome variables across the control and treatment groups. The outcome variables are constructed using trade and quote information for sub-penny-executed off-exchange transactions and include: (A) the absolute value of $Mroibtrd$; (B) the absolute value of $Mroibvol$; (C) size-weighted average relative % price improvement (difference between the relevant best quoted price and the transaction price, divided by the mid-point of the best bid and ask); (D) total price improvement (sum of dollar-denominated price improvements with respect to the relevant best quoted price across all sub-penny-executed transactions); (E) the total share volume of trades receiving price improvement; and (F) the size-weighted average sub-tick (sub-penny) fraction of trades receiving price improvement.

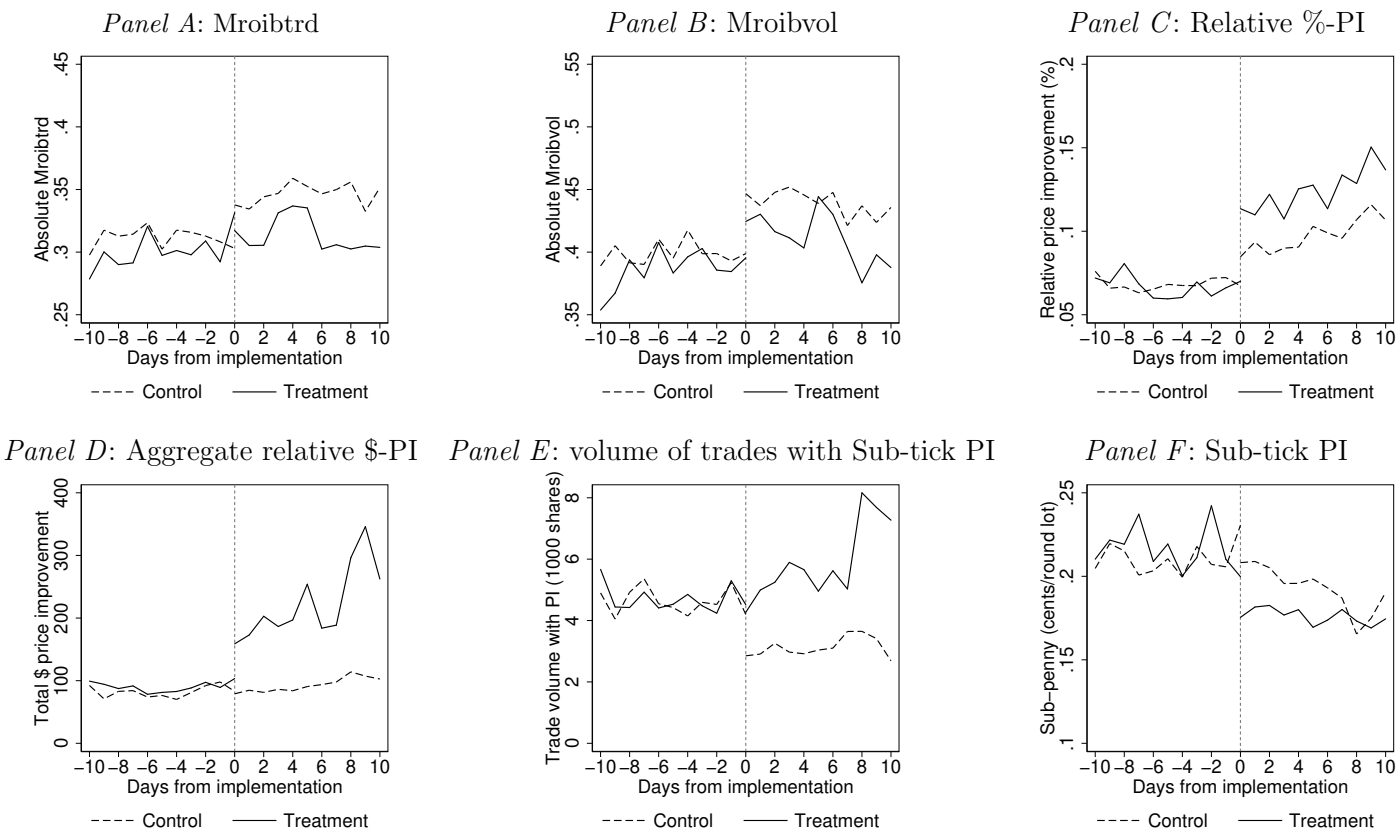


Figure 6: **Tick Size Pilot: *Trade Rule.*** This figure provides visual evidence associated with the results of the Difference-in-Difference specification in equation (6.2) for Test Group 2. The sample period spans the 10 trading days prior to the TSP's implementation on 10/03/2016 as well as the 10 trading days following its full implementation on 10/17/2016. The figure plots the daily medians for six outcome variables across the control and treatment groups. The outcome variables are constructed using trade and quote information for sub-penny-executed off-exchange transactions and include: (A) the absolute value of $Mroibtrd$; (B) the absolute value of $Mroibvol$; (C) size-weighted average relative % price improvement (difference between the relevant best quoted price and the transaction price, divided by the mid-point of the best bid and ask); (D) total price improvement (sum of dollar-denominated price improvements with respect to the relevant best quoted price across all sub-penny-executed transactions); (E) the total share volume of trades receiving price improvement; and (F) the size-weighted average sub-tick (sub-penny) fraction of trades receiving price improvement.

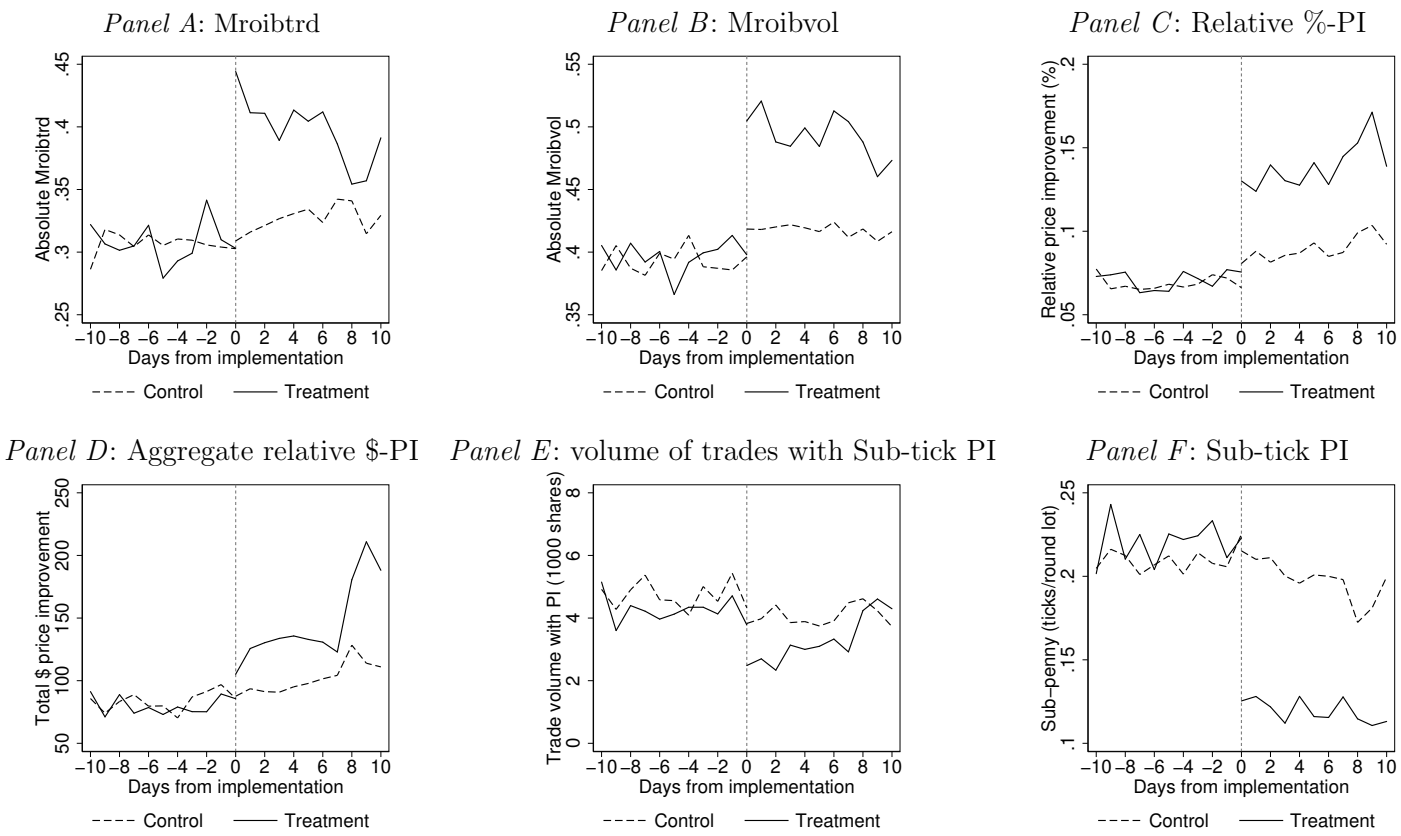
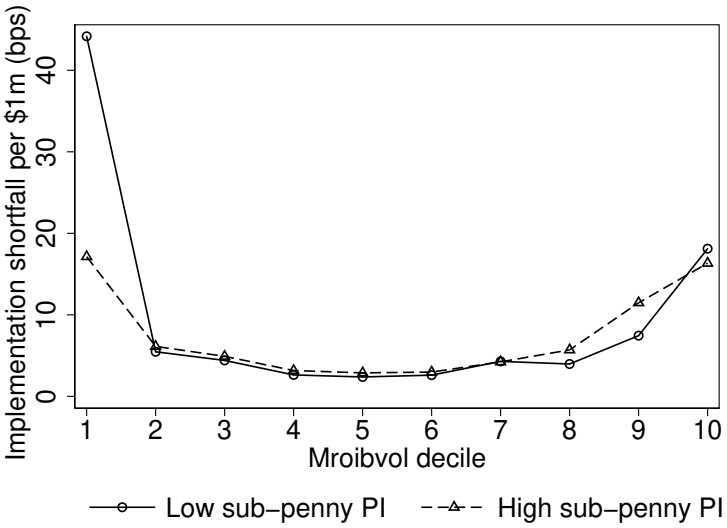


Figure 7: **Implementation Shortfalls and Future Returns Conditional on the Magnitude of Price Improvement.** This figure compares contemporaneous implementation shortfalls and week $w + 12$ returns when $Mroirbvol$ is constructed using retail trades with sub-penny price improvements that are low versus high. In Panel A, stocks are first sorted each day into deciles of low-sub-penny $Mroibvol$ and high-sub-penny $Mroibvol$. Panel A then plots median contemporaneous implementation shortfalls (in basis points per million dollars) across the deciles of both $Mroibvol$ measures. In Panel B, stocks are first sorted each day into deciles of low-sub-penny $Mroibvol_{w-1}$ and high-sub-penny $Mroibvol_{w-1}$. Panel B then plots the average week $w + 12$ return (in %) across the deciles of both $Mroibvol$ measures in week $w - 1$.

Panel A: Contemporaneous implementation shortfalls



Panel B: Week $w + 12$ returns