Identifying Market Maker Trades as "Retail" from TAQ: No Shortage of False Negatives and False Positives

Robert Battalio
Professor of Finance
Mendoza College of Business
University of Notre Dame

Robert Jennings
Professor of Finance (Emeritus)
Kelley School of Business
Indiana University

Mehmet Saglam
Associate Professor of Finance
Carl Lindner College of Business
University of Cincinnati

Jun Wu
Research Support Director
Wharton Research Data Services

Abstract: Boehmer et al. (2021) propose a methodology to infer retail trades from publicly available NYSE Trade and Quote (TAQ) data. Their methodology relies on assumptions about what types of orders do and do not trade on non-mid-point sub-penny increments via the Trade Reporting Facility. We obtain proprietary data from one or more order flow wholesalers that are known to receive orders from retail brokers. We use these data to demonstrate that the Boehmer et al methodology identifies less than one-third of these trades generally assumed to be from retail investors. In addition, we obtain proprietary data on non-retail trades that demonstrate that a large number of such trades print on non-mid-point sub-penny prices in violation of the assumption by Boehmer et al. that institutional orders trade only on penny or half-penny increments. Thus, there are both Type I and Type II errors that affect the ability of Boehmer et al to identify retail and only retail trades from TAQ. We show that these errors result in low correlations between order imbalance measures that are computed using Boehmer et al methodology and order imbalance measures computed with known retail orders.

Current draft, November 23, 2022.

Acknowledgments: The authors received no outside funding for this study and have no conflicts of interests related to this study.

A considerable body of academic research investigates the trading behavior of retail investors. Key to all such pursuits is identifying the trades of retail investors. One line of research utilizes proprietary data to identify a subset of retail investors. The seminal paper by Odean (1998) generated a plethora of papers that utilize data from a discount stock broker to examine retail trading behavior. Kaniel et al. (2008) utilize a proprietary dataset provided by the New York Stock Exchange (NYSE) to study individual investor trading practices. Since the passage of Regulation NMS, which greatly reduced the NYSE's market share, it has become more difficult to characterize retail investor behavior using data from a single stock exchange. Kelley and Tetlock (2013) obtain a proprietary dataset that includes all retail orders executed by an order flow wholesaler between 2003 and 2007 to examine retail trading behavior. Unfortunately, retail brokers or the execution venues to which they send their orders rarely provide proprietary transaction level data.

A second approach uses algorithms intended to identify retail trades in the NYSE's Trade and Quote (TAQ) database, which has been publicly available since 1993. One of the first algorithms used to identify retail trades from TAQ data was trade size. Reilly (1979) asserts that trades of 1,000 or fewer shares are made primarily by individuals and uses this cutoff to identify institutional trades when analyzing transaction records obtained from Francis Emory Fitch, Inc. Lee (1992) uses a dollar-based threshold of $10,000 to distinguish between retail and institutional trades. Once identified as a retail trade, the marketable side of the transaction is typically inferred using some version of the Lee and Ready (1991) algorithm. Lee and Radhakrishna (2000) use the NYSE TORQ database to demonstrate that trade size can be effective at distinguishing between retail and institutional traders. Cready, et al. (2014) notes that a significant concern in studies that use trade size cutoffs to identify retail investors "is spurious effects attributable to misclassification of transactions, particularly those originating from large investors." As noted by Hvidkjaer (2008), the likelihood that a small trade was a piece of a large institutional order became much higher after the shift to decimal pricing in 2001 and Reg NMS in 2005. As a result, most researchers currently avoid using trade size cutoffs to identify retail and institutional trades in U.S. equity markets.

Boehmer, Jones, Zhang, and Zhang (2021), hereafter BJZZ, propose and develop a methodology to infer retail purchases and sales from publicly available data by identifying trades with sub-penny prices that are reported to FINRA's Trade Reporting Facility (TRF) rather than through an exchange. BJZZ assert that "in the United States, most marketable equity orders initiated by retail investors do not take place on one of the dozen or so registered exchanges." In May 2022, Rosenblatt finds that off-exchange trading was 40% of total U.S. equity trading volume. Further, Rosenblatt estimates that wholesalers account for 42.5% of off-exchange trading, dark pools about 23%, and capital commitment and manual crossing of institution trading interests another 22%.[1] To separate retail and institutional trades reported to FINRA's TRF, BJZZ's methodology relies on the logic that only retail trades print on sub-penny prices at any increment other than a price equal to the mid-point of the quoted spread on the TRF. As noted in BJZZ, most retail order flow is routed to so-called order flow wholesalers (often, but not always with the retail brokers charging payment for the order flow) and provided price improvement as the wholesalers compete to provide best execution for their clients' orders (see, e.g., Battalio and Jennings (2022)). Conversely, BJZZ state that institutional trades generally cannot receive non-midpoint sub-penny prices but that institutional trades "often occur at the midpoint of the prevailing bid and ask prices." BJZZ assert that their "approach is therefore likely to pick up a majority of the overall retail trading activity."

[1] See Trading Talk – Market Structure Analysis: An Update on Retail Market Share in US Equities, June 24, 2022.

Using proprietary retail and non-retail data provided by one or more wholesaler(s)[2] and institutional orders executed by a major investment bank, we examine the accuracy of the BJZZ algorithm in identifying retail trades and excluding institutional trades in the NYSE's TAQ database. We find that the methodology identifies less than one third of known retail trades and frequently could include known institutional trades as retail.

The methodology in the BJZZ paper has been widely employed to infer retail trading and to draw conclusions about retail trading.[3] Within several of these papers, however, are caveats regarding the ability of the BJZZ algorithm to correctly identify retail trades in the publicly available TAQ database. For example, Blankespoor et al. (2018) examines the market's response when the Associated Press began using algorithms to write articles about firms' earnings announcement. Using the BJZZ algorithm to identify retail trades in TAQ, the authors present evidence that retail trading increases around the release of these articles. The authors suggest in a footnote that this conclusion relies, in part, on the assumption that by excluding trades that executed at the round penny or around the half-penny they are eliminating institutional trades. Using the BJZZ algorithm to identify retail trades in TAQ, Barber et al. (2021) present evidence that Robinhood investors engage in more attention induced trading than other retail investors. In a footnote, however, they state that "this conclusion assumes that there is no bias in the Boehmer et al. (2021) methodology that would affect concentration measures." Bradley et al. (2022) write that the BJZZ methodology "is conservative in the sense that it has a low type I error (i.e., trades classified as retail are very likely to be retail)" but "does omit retail trades that occur on exchanges." Finally, Baradehi et al. (2022) utilize the Boehmer et al. algorithm to examine how

[2] For literary convenience and brevity, we refer to one or more wholesale(s) in the singular in the remainder of the paper.

[3] See, for example, Bonsall, Green, and Muller (2020), Bushee, Cedergren, and Michaels (2020), Farrell et al (2022), Guest (2021), and Israeli, Kasnik, and Sridharan (2021).

wholesalers internalize retail order flow to provide liquidity to institutional investors. The authors note that the BJZZ algorithm "has minimal, if any, misclassifications for transactions that correspond to non-binding penny quoted spreads."

In this draft of the paper, we analyze retail trades executed by a wholesaler during the first ten trading days of December 2020. We will expand the sample period of wholesaler retail trading data in future drafts of the paper. In addition, we look at two samples of non-retail trade executions that print in non-midpoint sub-penny increments. We first provide summary statistics on a sample of over two million such trades from our retail-order data provider during December 2021. Secondly, we analyze a sample of institutional orders routed through the smart order routing volume-weighted-average-price (VWAP) algorithms of a large broker dealer between January 2011 and March 2012.[4] Overall, we document the potential for many Type I errors and find that a majority of known retail trades are not identified as such (Type II errors).

We obtain all marketable retail orders routed to the cooperating wholesaler(s) during our sample period. Interestingly, and in stark contrast to the assertion made in BJZZ that retail orders are seldom routed to exchanges, 19.6% of the trades generated by the retail orders that are routed to the wholesaler execute using liquidity sourced outside the wholesaler (e.g., exchanges or other sources of liquidity) and thus interact with other (i.e. non-data-provider) order flow. We begin our analysis of the BJZZ methodology by matching the proprietary trades that are known to be retail to trades reported to the TRF obtained from the NYSE's TAQ database. Overall, despite a tight constraint on the time lag between the proprietary and TAQ timestamps and the aforementioned fact that nearly 20% of trades occur after external routing to venues including exchanges, we match more than 78% of the wholesaler retail trades to TAQ TRF trades. Nearly 40% of the matched

[4] These are the same data used by Battalio, Hatch, and Saglam (2022).

trades have execution prices that have no sub-penny prices and roughly 30% of the matched trades have trade prices that have a sub-penny in the interval [0.4, 0.6] and, thus, are not identified as retail trades by BJZZ. As a result, only about 30% of the sample of retail trades that are matched to TRF trades are classified as retail by the BJZZ methodology. Just over 94% of these trades have the correct inferred order side using the BJZZ methodology.

It is well-documented in the literature that institutional order flow can be informative (e.g., Hendershott, et al (2015)). One of the assumptions used by the BJZZ algorithm is that institutional order flow executed in Alternative Trading Systems (e.g., dark pools) and non-ATS execution venues (e.g. single-dealer platforms) and reported through the TRF does not frequently receive non-mid-point sub-penny price improvement. Evidence to the contrary would make it difficult to interpret the results of studies that examine whether retail trades identified using the BJZZ methodology are informed. Our data provider examines a sample of 19,802,471 non-retail trades and, after eliminating 2,741,318 trades that receive mid-point pricing, furnishes us a sample of over two million non-retail trades filled at non-mid-point sub-penny prices on venues that BJZZ argue do not receive such prints. We also use an alternative set of institutional trades from a major investment bank to examine whether this assertion by BJZZ is correct. Of the 166,266 sample institutional trades that obtain liquidity from electronic liquidity providers like Citadel Securities, Getco, and Knight between January 2011 and March 2012, we find that over 78% would be classified as retail trades by BJZZ. Looking separately at the 136,832 institutional orders executed in the broker's ATS, one-third of these trades is classified as retail by the BJZZ algorithm. Thus, a substantial portion of institutional trades with electronic liquidity providers or in an investment banking firm's ATS could be misidentified as retail trades. Whether or not these Type I errors are severe enough to alter inferences in studies of retail trading behavior of retail investors are questions that we cannot answer here. However, at the very least, these results suggest institutional order flow *can* receive price improvement measured in small fractions of a cent per share.

Do Type I (i.e., identifying non-retail trades as retail trades) and Type II errors (e.g., the failure to correctly identify retail trades) result in any substantive differences in the inferences researchers using the BJZZ methodology have made? BJZZ examine the ability of order imbalances constructed using inferred retail trades to anticipate future stock price movements to examine whether retail order flow is informed. Our data provider also produces the four BJZZ-defined order imbalance measures on a stock-day basis using all of their proprietary retail trades for the period of time from August 3, 2020 through July 26, 2022. We construct analogous order imbalance measures for BJZZ-identified trades for these stocks from TAQ during the same period and compare them to the order imbalance measures from data provider. We find the correlations between the actual and the BJZZ-inferred order imbalance measures are less than one-half as high as BJZZ found when comparing their order imbalance measures to order imbalance (OIB) measures created using a sample of proprietary retail trades provided by Nasdaq. Furthermore, we find systematic differences between the wholesaler order imbalance measures and the BJZZ-inferred metrics. Specifically, the BJZZ-inferred OIB measures are less sell/more buy oriented than the "true" OIB measures using the data-provider retail orders.

In the next section, we discuss the differences between our paper and the Barber et al (2022) paper. We then describe our data sets. In Section III, we evaluate how the BJZZ methodology performs using data that we know originates with either retail or institutional traders. In the subsequent section, we evaluate the implications of the differences in Type II errors in retail order classification for OIB measures. Finally, we conclude.

II. Literature Review.

Barber et al (2022) conduct an experiment to evaluate the effectiveness of the BJZZ methodology at identifying retail trades by placing over 85,000 trades in 128 stocks between December 21, 2021 and June 9, 2022 through six retail brokers. The 128 sample stocks are the result of a randomized sampling process after stratifying all stocks with a CRSP security code of 10 or 11 priced greater than $1.00 on market capitalization, liquidity (turnover), volatility, and stock price. They submit orders for a $100 notional amount (requiring an integer number of shares) or, if the share price exceeds $100 at least one share between 9:40am and 3:50pm. To mimic day trading activity, they initiate positions by buying and then selling 30 minutes later and carry no inventory overnight. They find that the BJZZ methodology identifies about 35% of their actual retail trades as retail and correctly signs (as buys or sells) about 72% of those. Furthermore, they conclude that, on a stock basis, 30% of BJZZ-constructed order imbalance measures are uninformative because the accuracy rate for signing trades does not differ from 50%.

We can distinguish our paper from Barber et al (2022) in several dimensions. We are the first, to our knowledge, to demonstrate that institutional trades reported on the TRF can print in sub-penny intervals other than the half-penny. Thus, the notion that BJZZ identify as retail trades only trades that are from retail investors is incorrect.

Second, our sample of known retail trades is much different from Barber et al (2022). Our sample of retail orders reflects overall retail trading interests – it contains orders from almost 9,600 trading symbols, some not CRSP security codes 10 or 11. Retail investors trade a number of securities (e.g., Exchange Traded Funds) that are not simple common stocks. It is important to preview our sample of BJZZ-matched actual retail orders with their sample. On a trade-weighted basis, our data's mean trade price is $123.94 and the mean trade size is 231 shares. This implies an average notional trade size of $28,630, much larger than the $118 average trade size in their sample. We demonstrate that the success of the BJZZ methodology is sensitive to order and trade size as well as stock price. Our order-weighted mean National Best Bid Offer spread is $0.09 versus their reported mean spread of $0.17. Less than 20% of their trades occur when the NBBO quoted spread is $0.01 but about 44% actual retail trades identified by the BJZZ methodology in our sample occur when the spread is at minimum. This difference might be at least partially explained by the fact that liquidity tends to concentrate at the open and the close – almost 10% of our sample trades occur before 9:40am or after 3:50pm.

Finally, as do Barber et al. (2022), we find that the correlation between BJZZ-inferred order imbalance measures and the actual order imbalance measures computed with known retail trades is considerably lower that computed by BJZZ. We believe that we are the first to ask whether a disagreement between order imbalance measures might lead to difference in conclusions about the informativeness of retail order flow.

III. Data.

To begin our analysis of the BJZZ methodology we obtain proprietary marketable order and trade data from a wholesaler for the month of December 2020. We receive all of the marketable retail orders handled by the wholesaler during this time period. The order data include: date and time of order entry, a unique order identification number, stock trading symbol, the type of order (market or marketable limit), the limit price if applicable, the order's side (buy or sell, with an indicator variable for short sell), and the order quantity. The trade data include: date and report time of trade, the unique order identification number mapping back to the order data, a unique trade reference number, the stock trading symbol, the number of shares filled by this execution, and the execution price. Using the order identification number, we can merge order and trade data.

We restrict our analysis of Type II errors to trades reported via FINRA's Trade Reporting Facility (TRF) and not to an exchange. In order to facilitate potential analyses by stock-day, we restrict our sample to stock symbols averaging at least 100 round-lot trades per day (2,200 trades for the sample month) without any days reporting zero TRF trades. Furthermore, we require that the end-of-month stock price be greater than one dollar to mitigate sub-penny limit order pricing. That provides us with a sample of 2,741 stock symbols (slightly less than 29% of the symbols in the original database). The sample stocks produce 85% of the trades contained in the proprietary data. We provide some descriptive statistics of our sample in Table 1.

[Insert Table 1 about here.]

End-of-month share price ranges from the designed minimum of $1.01 to over $3,000 and averages $64.57. Round-lot trades for the month range from the designed minimum of 2,201 to nearly two million and averages 23,328. Overall, there is substantial variation in the two variables used to restrict the sample stocks.

The BJZZ methodology uses publicly available TAQ data to infer retail trades and order sides. To assess the success of this methodology we match our set of proprietary retail trades to the corresponding TAQ trade (for which we will use the BJZZ methodology to infer its retail status). To begin this process, we gather all trades reported to the TRF (exchange code 'D') in the restricted-sample stocks for December 2020 from Daily TAQ. Following BJZZ, we exclude trades with non-normal condition codes and trades executed at prices less than $1. During the sample month, there were 327,542,261 trades for the sample stocks reported via the TRF. The mean trade price was $109.84 with a range of $0.30 to $3248.99.[5] There are a disproportionate number of

[5] Clearly, requiring the stock to have a trade price greater than $1 at month's end was insufficient to eliminate all trade prices less than $1. The BJZZ methodology we replicate later in the paper eliminates all trades priced less than $1.

trades in higher priced stocks as the trade-weighted trade price exceeds the equally-weighted share price. Mean trade size was 240 shares with a range of one share to 6,199,125 shares.

In this version of the paper, we restrict our analysis to the first ten trading days of December 2020. Future drafts will incorporate the remainder of the sample month. We match the proprietary data trades to TAQ TRF trades based on symbol, time, price and quantity for the 2,741 activity-restricted-sample stock. Symbol, price and quantity are unambiguous matching criteria. We have two times; one from the data provider(s) and another from TAQ. We choose TAQ's Participant Timestamp as the benchmark (the time the participant reported the trade to the SIP) because it is the earliest TAQ timestamp and allow a maximum of ten milliseconds difference between it and the proprietary data timestamp.[6] We match 78.16% of the proprietary trades with 81.25% of the volume to TAQ trades within the ten-millisecond window. Examining the trade matching success on a stock-level basis, the mean (median) matching success is 75.05% (75.92%) with a range from 22.21% to 95.88%. Given that nearly 20% of the wholesaler trades execute using liquidity sources external to the wholesaler (e.g., exchanges and ATSs), we did not expect to match all of the data-provider trades. Table 2 provides some descriptive statistics regarding the trades we were and were not able to match.

[Insert Table 2.]

Overall, the retail trades that we match to TAQ appear to be a representative sample of the original retail order data set. The average matched trade occurs slightly later in the trading day primarily due to proportionally fewer matched trades in the period close to the open. Because the open is the busiest time for trading, it likely results in a higher-than-average delay in the TAQ timestamp. The matched trade sample is somewhat more likely to be a buy order and be in a stock

[6] Participant timestamp in TAQ is measured in milliseconds instead of microseconds for the data provider(s).

with a higher transaction price. The mean order size for the matched orders is the only order/trade characteristic for which the matched sample differs remarkably from the original proprietary data set. The mean order size is substantially smaller for the TAQ-matched trades than the overall retail database. Not surprisingly, the larger orders are less frequently completely filled with wholesaler internalization so proportionally fewer of these trades make their way to the TRF (as noted previously, 19.6% of the trades are consummated by the data provider accessing external liquidity sources from venues such as exchanges and Alternative Trading Systems). Finally, the mean time difference between the proprietary data timestamp and the TAQ Participant Time is just over one millisecond (result not tabulated in Table 2).

IV. How does the BJZZ algorithm perform?

We use the BJZZ methodology to infer which of the TAQ-matched trades came from retail investors (all of which are known to be retail trades). To replicate BJZZ, we compute their classification variable $Z_{i,t}$ = 100 * mod($Price_{i,t}$ .01), where the subscript i,t denotes stock i at time t. Table 3 contains the sub-penny pricing distribution for matched retail trades by trade side conditional on the order-receipt time NBBO width.

[Insert Table 3 about here.]

We find that 39.66% of the matched retail trades have no sub-penny pricing and 29.90% have sub-penning pricing that BJZZ consider as mid-point pricing (and therefore not retail). Of trades with sub-penny increments in the range [0.4, 0.6], 91.74% of those trades are exactly at a sub-penny pricing increment of 0.5 (not tabulated on Table 3). Thus, only 30.44% of our sample of matched retail trades has $Z_{i,t}$ in the (0, .4) or (.6, 1) intervals and is classified as retail by the BJZZ methodology. This is slightly lower than the Barber et al (2022) identification rate of 35%.

When the quoted spread is \$0.01, the BJZZ algorithm does slightly better identifying just over 34% of the known retail trades as retail.

To better understand whether trades generated by institutional orders executed away from exchanges are as BJZZ state on their page 2,255 "usually in round pennies," we obtain two samples of non-retail orders. Firstly, we obtain a sample of 2,100,769 non-retail trades that occur at sub-penny but not half-penny prices during December 2021 from our data provider(s). These trades represent the wholesaler providing liquidity to non-retail orders at sub-penny prices other than a half cent. Thus, of the 19,802,471 non-retail trades examined, 24.45% occur at sub-penny prices and are reported to the TRF. Of those non-retail trades reported to the TRF that occur at non-penny prices, 43.4% occur at non-mid-point sub-penny intervals. Of the 2,100,769 trades occurring at non-mid-point sub-penny increments, 2,100,162 occur during regular trading hours and comprise the sample we analyze. The majority of these trades (52.38%) are sell orders. Almost 92% of these trades execute on the wholesaler's non-ATS single-dealer platforms (the wholesaler is the sole counterparty) and the remainder are roughly evenly split between Alternative Trading Systems (dark pools and ECNs) and Exchange RLP. (The latter, although representing a retail trade, is a retail trade occurring on and reported through an exchange.) We provide descriptive statistics on these trades in Table 4.

[Insert Table 4 about here.]

Although most of these trades do not result from retail orders, it would be difficult to distinguish them from retail trades described in Table 2 based on order size, trade price, or trade time. The mean trade size is 302 shares for the institutional orders versus 223 shares for the retail orders. Over 86% of the institutional trades are for fewer than 500 shares. There are increased tendencies for institutional trades to occur early and late in the trading day compared to the retail orders but there is only an eight second difference in mean execution time.

[Insert Table 5 about here.]

In Table 5, we describe our experience with the frequency of various sub-penny pricing intervals. Of the 2.1 million non-retail, non-half-penny sub-penny trades, only 17% would be eliminated using BJZZ's "near one-half penny" exclusion rule (recall the data provider removed all of the mid-point sub-penny trades). Thus, nearly 1.75 million institutional trades from our data provider(s) would be included as retail using BJZZ.

In addition, we obtain a sample of institutional parent orders and the corresponding child order executions processed by a large investment bank's (IB's) volume weighted average price (VWAP) algorithm between January 2011 and March 2012 used by Battalio et al. (2022). Unlike Battalio et al., we consider the entire set of child order executions and construct two samples of trades: a sample of 166,266 trades that executed in the investment bank's dark pool and a sample of 136,833 trades that source liquidity from electronic liquidity providers like Getco, Citadel Securities, and Knight Securities. Table 6 contains the sub-penny pricing distribution for each collection of trades.

[Insert Table 6 about here.]

Focusing first on the second column of Table 6, we see that 38.4% of the trades filled in the IB's dark pool executed in round pennies and 28.4% of the trades executed with sub-penny increments in the range [0.4, 0.6]. This implies that 33.2% of the institutional trades executed in the IB's dark pool have sub-penny pricing increments, which are classified as retail by the BJZZ algorithm. Moving to the third column of Table 6, we see that 78.3% of the trades executed by ELPs away from exchanges have sub-penny pricing increments that make the trades eligible to be classified as retail by the BJZZ algorithm. The results from both sets of institutional orders contradict the assertion that institutional trades executed in the dark "are usually in round pennies" and suggests that Type I errors may plague studies that use BJZZ-identified retail trades to examine ***retail*** trading behavior.

[Insert Table 7 about here.]

Finally, we turn to an analysis of Type II errors. In Table 7, we provide some descriptive statistics regarding our experience using the BJZZ methodology to classify our activity-restricted-sample of known retail trades as retail. From Panel A, we determine that the BJZZ methodology's inferred retail trade sample differs from the known retail trade universe in that the average trade (order) size is somewhat larger (smaller), the average trade-weighted trade price is slightly lower, and the average order receipt time is slightly later in the trading day. Panels B and C demonstrate that the methodology is least effective at identifying known retail orders throughout the opening half hour of trading. Interestingly, Panels D and E suggest that BJZZ's algorithm is less effective at identifying odd lot orders and trades than (most) round or partial-round orders and trades.[7] Panel F finds that the BJZZ approach is more effective at identifying retail trades for stocks with prices less than $100 per share, likely because these stocks are more frequently quoted at the minimum tick size of $0.01.[8]

We now turn to examining the success of BJZZ's methodology in correctly inferring trade side. For this analysis, we focus on the 7,349,520 trades the BJZZ methodology identifies as retail

[7] This is consistent with Barber et al (2022), which finds higher identification rates for 100 share orders than odd lots. Bartlett (2022) suggests that is due to SEC Rule 605, which requires venues to post execution quality statistics for round lots but not for odd lots.

[8] Finally, in results that are not tabulated in this version of the paper, we examine orders filled in a single execution versus orders filled in multiple executions. We find that about 93.6% of the retail orders are filled in a single trade. 99% of orders have four or fewer trades but the maximum number of trades in an order can be quite large (maximum of 753). For multiple-fill retail orders, BJZZ identify all trades associated with the order as retail for only about one of five orders.

trades. Overall, we find that just over 94% (compared to their robustness check of 98.2%) of BJZZ-identified retail trades have the correct inferred side.[9] This suggests that about 6% of the BJZZ-identified retail trades are getting substantial (i.e., better than mid-point) price improvement, which results in misclassification of order side by BJZZ.[10] In Table 8, we provide some detail on the trade side inference of BJZZ's approach.

[Insert Table 8 about here.]

From Panel A we find that it is about 1.5 times more likely for BJZZ to misclassify a buy order as a sell than it is to misclassify a sell order as a buy (3.59% versus 2.29%).[11] In Panel B, we examine BJZZ's side inference success by time of day. The BJZZ algorithm is slightly less successful in the first half hour of trading but there is not substantive variation in success rate across the day. In Panel C, we examine BJZZ's order side inference success rate by order and, more practically, trade size. Generally speaking, the methodology is slightly less accurate for smaller orders and trades than for large orders/trades suggesting that small orders/trades are more likely to receive greater price improvement from wholesalers than larger orders/trades. In Panel D, we examine the algorithm's success in identifying whether a retail trade is buyer- or seller-initiated conditional on whether the retail order generating the trade was filled with a single execution. For the 5,971,948 orders filled with a single execution that are identified as retail by the algorithm, the inferred order side is the actual order side 94.85% of the time. For retail orders filled with multiple trades, the overall success rate for inferring order side is less than 63%. The algorithm correctly infers the order side for all of the trades generated by an order requiring multiple trades only about 36% of the time (258,546 of 711,641 BJZZ-identified retail orders with

[9] This differs markedly from Barber et al (2022) likely due to a difference in sample stock selection.
[10] Note that executions on a full penny price might represent price improvement when quoted spreads exceed $0.01.
[11] The magnitude and bias in misidentification of order side with our data are greater than and opposite that documented in BJZZ (page 2261).

multiple fills). For a relatively small number of orders with multiple trades (65,905 of BJZZ-identified retail orders), BJZZ gets all of the inferred trade sides incorrect. Finally, for 388,190 (almost 55%) of BJZZ-inferred retail orders with multiple fills, BJZZ infers different sides for different fills and produces an overall correct side 48.19% of the trades.

In Panel E of Table 8, we consider what happens when we restrict the sample to instances where the quoted spread is a penny. This is important because at spreads wider than a penny, the BJZZ approach to inferring order side is more difficult to apply. Consider a stock with an NBB of $10.00 and an NBO of $10.01. A trade at $10.002 has a sub-penny increment of 0.2 and is (most likely) properly typed as a sell by BJZZ. Suppose that the NBO increases to $10.02 and a trade occurs at $10.012. Again, the sub-penny increment is 0.2 but it seems more likely that the trade is a buy. Without restricting the NBBO to a penny, the algorithm misclassifies the order side for 5.98% of the matched sample of retail trades. However, as shown in Panel E of Table 7, when we restrict our sample to matched retail trades received by the wholesaler when the width of the NBBO is $0.01, the percentage of trades for which the inferred order side is incorrect falls to 0.60%.

To summarize, we find that several of the assumptions made to derive the BJZZ algorithm are suspect. At least during our sample period, there is a substantial portion of the retail trades potentially executed on an exchange. One can imagine that the sample of retail orders sent to exchanges to be executed is different than the sample of retail orders that are executed in the dark (e.g., away from exchanges). This, coupled with the fact that a large percentage of our sample of retail trades are executed either with no sub-penny price increment or with a sub-penny increment in the interval [0.4, 0.6] means that the BJZZ algorithm only has a chance of identifying a fraction of the initial sample of retail trades; in our case less than one-third. For the minority of known retail trades that is typed as retail by BJZZ, most are assigned the correct order side by the algorithm – although it struggles with multiple-trade orders. Finally, the fact that Reg NMS prohibits "orders from having sub-penny limit prices" does not appear to restrict institutional trades from being executed with non-midpoint sub-penny price increments.[12]

V. Implications.

a. Order Imbalances

The analysis thus far indicates that the BJZZ methodology fails to identify as retail trades nearly 70% of the retail trades from our data provider(s) and frequently identifies non-retail trades as retail (in our limited non-retail samples). The next step in the inquiry is to address whether these failures result in any substantive differences in important inferences. One issue of importance is the measure of order imbalances as BJZZ go on to associate order imbalance measures constructed using their inferred retail trades with future stock price movements in an effort to judge the informativeness of retail order flow. On page 2262 of their paper, BJZZ find that their inferred order imbalance and the order imbalance derived from a dataset provided by Nasdaq with all trades and known order side in 117 stocks is 0.70. In order to assess whether this relatively high correlation persists in our data, we compute each of their four order imbalance measures using the BJZZ-inferred retail trades' inferred side and the data provider does the same using the entirety of their retail order data (regardless of execution venue) with the known order side on a daily basis. These calculations are done for all of the symbols traded by the data provider. $MROIBVOL_{i,t}$ is the signed difference between the retail buy volume and the retail sell volume normalized by the sum of the retail buy and sell volume for stock i on day t. $MROIBTRD_{i,t}$ is the signed difference between the number of retail buy trades and the number of retail sell trades normalized by the sum

[12] BJZZ write that "in the early part of our sample, a small number of dark pools allowed some sub-penny orders and provided non-midpoint sub-penny execution prices, but our results hold when we exclude this subperiod."

of retail buy and sell trades for stock i on day t. $ODDMROIBVOL_{i,t}$ and $ODDMROIBTRD_{i,t}$ are the imbalance measures computed using only odd lot trades in stock i on day t.

We then compute correlations between order imbalance measures on a stock-day basis and average those daily correlations over the 104-week sample period of August 3, 2020 to July 26, 2022. For any given stock-week observation, we eliminate all that have any of the eight (four based on the wholesaler's retail orders and four based on BJZZ's methodology of identifying the same set of stocks in TAQ) or stock-weeks with any extreme order imbalance values (0, +1, or -1) as all of these indicate that there is a paucity of observations when computing the order imbalance measures. We begin with 1,021,091 observations. After eliminating stock-weeks where there is insufficient trading to obtain a reliable imbalance metric, we retain 823,621 observations. In Table 9, we report our experience with each of their order imbalance measures.

[Insert Table 9 about here.]

In Panel A, we see that our overall correlations between the BJZZ-inferred order imbalance and the comprehensive wholesaler retail order imbalance is less than one-half of what BJZZ report. Volume-based measures are more highly correlated than trade-based metrics, but none exceed 35% correlation. Furthermore, we find that there is a bias in the BJZZ measures – they tend to over-estimate the buy imbalance for all trades and under-estimate the buy imbalance for odd lot trades.

In Panel B, we examine some descriptive statistics regarding the order imbalance (OIB) measures for the wholesaler sample of all trades and the BJZZ-inferred measure. In column two (three), we report the mean (median) difference (wholesaler minus BJZZ-inferred) in order imbalance measures with stock-week as the unit of observation. Consistent with Panel A, we find that the BJZZ-inferred measure is, on average, more (less) buy imbalance oriented than the actual wholesaler measure for the imbalance measures focusing on total (odd lot) volume and trades. In columns four and five, we count the number of sample weeks for which the BJZZ measure is less than the measure computed using all wholesaler retail trades in our sample and conduct a binomial test that the true proportion equals one-half (57 weeks). For the total volume- and trade-based measures, there is a clear bias of the BJZZ-inferred measure to be more buy oriented than the wholesaler measure. For the odd lot volume-based measure there is the opposite bias but for the odd lot trade-based measure there appears to be no bias in a binomial test.

b. Regressions of future stock returns.

In progress.

V. Conclusion.

Given the wide-ranging interest in the activities of retail investors in the academic literature and the professional and regulatory spheres, it is important that researchers properly identify retail trades in publicly available data in order to draw correct conclusions for policy formation. Boehmer et al (2021) provide a recently-popular methodology to infer retail trades from TAQ data that has been used in numerous academic and practitioner studies. We undertake a study of the accuracy of their assumptions in designing the methodology and conduct one analysis of the implications of any documented inaccuracies. Using proprietary data, we demonstrate that the BJZZ methodology correctly identifies as retail less than one-third of trades that are commonly thought of as retail. Furthermore, we document that it is frequently the case that known institutional trades that BJZZ assume will not be identified as retail by their methodology are indeed retail trades. Do these Type I and Type II errors lead to important incorrect inferences? We follow BJZZ to examine one such research path. Specifically, we compare the order imbalance measures computed with a known set of retail orders and compare that to the order imbalance measures computed using BJZZ methodology to infer retail trades. The order imbalance measures demonstrate a relatively low correlation with each other – less than one-half the correlation that BJZZ find with the limited sample of Nasdaq proprietary data.

References

Baradehi, Yashar, Dan Bernhardt, Zhi Da, and Mitch Warachka, 2022. Uncovering the liquidity premium in stock returns using retail liquidity provision, Mendoza College of Business working paper.

Terrance Odean, 1998, Are investors reluctant to realize their losses?, *Journal of Finance* 53, 1775-1798.

Barber, Brad, Xing Huang, Phillipe Jorion, Terrance Odean and Christopher Schwartz, 2022. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ, working paper.

Barber, Brad, Xing Huang, Terrance Odean, and Christopher Schwarz, 2021. Attention induced trading and returns: Evidence from Robinhood users, forthcoming in the *Journal of Finance*.

Bartlett, 2022, Modernizing odd lot trading, *Columbia Business Law Review*, forthcoming.

Battalio, Robert, and Robert Jennings, 2022. Why do brokers who do not charge payment for order flow route marketable to Wholesalers, Mendoza College of Business, working paper.

Battalio, Robert, Brian Hatch, and Mehmet Saglam, 2022. The cost of exposing large institutional orders to electronic liquidity providers, Mendoza College of Business working paper.

Blankespoor, Elizabeth, Ed deHaan, and Christina Zhu, 2018. Capital market effects of media synthesis and dissemination: Evidence from robo-journalism, *Review of Accounting Studies* 23, 1 – 36.

Boehmer, Ekkehart, Charles Jones, Xiaoyan Zhang, and Xinran Zhang, 2021. Tracking retail investor activity, *Journal of Finance* 2249 – 2305.

Bonsall IV, S. B., Green, J., & Muller III, K. A. (2020). Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics*, *69*(1), 101264.

Bushee, B., Cedergren, M., & Michels, J. (2020). Does the media help or hurt retail investors during the IPO quiet period? *Journal of Accounting and Economics*, *69*(1), 101261.

Bradley, Daniel, Russell James, and Jared Williams, 2022. Non-deal roadshows, informed trading, and analyst conflicts of interest, *Journal of Finance* 77, 265 – 315.

Cready, William, 1988. Information value and investor wealth: The case of earnings announcements, *Journal of Accounting Research*, 1 – 27.

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

Farrell, M., Green, T. C., Jame, R., & Markov, S. (2022). The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, *145*(2), 616-641.

Guest, N. M. (2021). The information role of the media in earnings news. *Journal of Accounting Research*, *59*(3), 1021-1076.

Israeli, D., Kasznik, R., & Sridharan, S. A. (2022). Unexpected distractions and investor attention to corporate announcements. *Review of Accounting Studies*, *27*(2), 477-518.

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

Hvidkjaer, Soeren, 2008. Small trades and the cross-section of stock returns, *Review of Financial Studies* 21, 1123 – 1151.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008. Individual investor trading and stock returns, *Journal of Finance*, 273 – 310.

Kelley, Eric and Paul Tetlock, 2013. How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229 – 1265.

Lee, Charles, 1992. Earnings news and small traders, *Journal of Accounting and Economics* 15, 265 – 302.

Lee, Charles and Mark Ready, 1991. Inferring trade direction from intraday data, *Journal of Finance* 46, 733 – 746.

Lee, Charles and Balkrishna Radhakrishna. 2000. "Inferring Investor Behavior: Evidence from TORQ Data." *Journal of Financial Markets* 3, 83–111.

Odean, Terrance, 1998. Are investors reluctant to realize their losses?, *Journal of Finance,* 1775 - 1798.

Reilly, Frank, 1979. Investment Analysis & Portfolio Management, The Dryden Press, Hinsdale, Illinois.

Trading Talk – Market Structure Analysis: An Update on Retail Market Share in US Equities, June 24, 2022.

Welch, Ivo, 2022. The wisdom of the Robinhood crowd, *Journal of Finance* 77, 1489 – 1527.

Table 1. Descriptive statistics for the 2,741 sample securities.

From the 9,584 unique security symbols traded by our data provider(s) for the month of December 2020, we select symbols that average at least 100 trades per trading day and have no days without trades. Finally, to reduce the frequency with which securities can be quoted in sub-penny increments, we require a closing security price of greater than $1.00 at month's end. Statistics in the table are equally-weighted across sample stocks.

Statistic	End-of-Month Price	Round-Lot Trades in Month
Mean	$64.57	23,328
Median	$25.86	6,274
Minimum	$1.01	2,201
Maximum	$3,255.63	1,967,137

Table 2. Descriptive statistics for the retail trades in our proprietary dataset reported to FINRA's TRF and for those retail trades we could match to TAQ TRF trades for the ten trading days from December 1 through December 14, 2020.

Dark Retail Trades are retail trades from our proprietary data reported to FINRA's TRF. We match to TAQ TRF trade data based on security symbol, trade price and size, and trade time. Symbol, trade size, and trade price require exact matches. We allow ten milliseconds difference between the TAQ Participant Timestamp and the data provider(s)' timestamp. The average time difference between the TAQ participant time of the matched TAQ trade and the execution time of the retail trade is 0.00121 seconds. We are able to match 78.16% of our proprietary data retail trades to TRF trades in the TAQ database.

Panel A. Descriptive statistics for the original and matched samples of retail trades.

Variable	Mean	
	Dark Retail Trades (N = 30,881,022)	Matched Retail Trades (N = 24,135,132)
Execution Time	12:13:50	12:15:00
Trade Quantity	223 shares	231 shares
Trade Price (trade-weighted)	$113.19	$123.94
Mean Share Price (symbol-weighted)	$62.49	$62.47
Order Quantity	2,307 shares	784 shares
% Buy Orders	55.84%	56.82%
% Short Sell Orders	3.17%	2.84%

Panel B. Distribution of trade times throughout the trading day for the original and matched sample of retail trades. There are 812 trades occurring at exactly 16:00:00 in the proprietary data that we do not match to TAQ.

Hour	Dark Retail Trades (N = 30,881,022)	Matched Retail Trades (N = 24,135,132)
9:30 to 10:00	15.96%	15.57%
10:00 to 11:00	21.46%	21.36%
11:00 to 12:00	14.34%	14.39%
12:00 to 1:00	11.95%	12.07%
1:00 to 2:00	11.43%	11.53%
2:00 to 3:00	10.86%	10.98%
3:00 to 4:00	14.00%	14.11%

Table 3. Sub-penny pricing distribution for matched retail trades by trade side conditional on the order receipt time NBBO width. We match proprietary wholesaler(s)' trades to TAQ TRF trade data based on security symbol, trade price and size, and trade time. Symbol, trade size, and trade price require exact matches. We allow ten milliseconds difference between the TAQ Participant Timestamp and the data provider(s)' timestamp.

	All Spreads		NBBO Width = $0.01	
100*mod(Price,0.01)	Buy Orders (N=13,714,104)	Sell Orders (N=10,421,028)	Buy Orders (N=5,977,765)	Sell Orders (N=4,556,201)
0.00	21.42%	18.24%	13.06%	12.82%
0.00<mod≤0.10	0.34%	9.98%	0.15%	11.33%
0.10<mod≤0.20	0.29%	0.99%	0.00%	1.46%
0.20<mod≤0.30	0.33%	0.63%	0.03%	0.79%
0.30<mod<0.40	0.19%	0.63%	0.00%	0.73%
0.40≤mod≤0.60	17.60%	12.30%	23.80%	16.09%
0.60<mod≤0.70	0.83%	0.20%	0.97%	0.00%
0.70<mod≤0.80	0.95%	0.21%	1.05%	0.02%
0.80<mod≤0.90	1.52%	0.22%	2.01%	0.00%
0.90<mod<1.00	13.34%	0.11%	15.68%	0.01%
0.00<mod<0.40	1.15%	12.23%	0.18%	14.31%
0.60<mod<1.00	16.64%	0.74%	19.71%	0.01%

	All Spreads		NBBO Width = $0.01	
100*mod(Price,0.01)	Buy Orders (N=13,714,104)	Sell Orders (N=10,421,028)	Buy Orders (N=5,977,765)	Sell Orders (N=4,556,201)
0.00	17.71%	18.98%	13.85%	15.55%
0.00<mod≤0.10	0.22%	16.87%	0.12%	17.99%
0.10<mod≤0.20	0.14%	1.01%	0.00%	1.31%
0.20<mod≤0.30	0.15%	0.52%	0.02%	0.57%
0.30<mod<0.40	0.11%	0.46%	0.00%	0.50%
0.40≤mod≤0.60	12.52%	11.30%	15.23%	13.67%
0.60<mod≤0.70	0.59%	0.13%	0.58%	0.00%
0.70<mod≤0.80	0.61%	0.11%	0.66%	0.01%
0.80<mod≤0.90	1.34%	0.12%	1.71%	0.00%

0.90<mod<1.00	17.03%	0.07%	18.21%	0.01%

Table 4 – Non-retail Order Flow Trading on Sub-penny Prices for Sample Stocks in December 2021. The data provider identified 2,100,769 non-retail, sub-penny executions. Requiring that the trades occur in regular market hours reduced the sample to 2,100,162 observations.

Panel A. Descriptive Statistics

Variable	Mean	Median	Minimum	Maximum
Execution Time	12:21:42	11:55:49	9:30:00	16:00:00
Shares Filled	302	100	1	1,600,000
Trade Price	$112.28	$38.61	$1.0005	$5850.0005

Panel B. Percent of Trades by Time of Trading Day

Time Interval	Percent of Trades
9:30:00 to 9:59:59.999999	17.85%
10:00:00 to 11:59:59.999999	19.54%
11:00:00 to 11:59:59.999999	13.43%
12:00:00 to 12:59:59.999999	10.24%
13:00:00 to 13:59:59.999999	9.42%
14:00:00 to 14:59:59.999999	10.50%
15:00:00 to 16:00:00.000000	19.01%

Panel C. Percent of Trades by Trade Size

Trade Size Interval	Percent of Trades
1 – 99 shares	37.17%
100 – 499 shares	49.25
500 – 999 shares	7.08%
1000 – 1999 shares	3.86%
2000 – 4999 shares	2.01%
5000 – 9999 shares	0.51%
> 9999 shares	0.14%

Panel D. Percent of Trades by Trade Price

Trade Price Interval	Percent of Trades
≥ $1 and < $10.00	20.84%
≥ $10 and < $50	35.29%
≥ $50 and < $100	13.73%
≥ $100 and < $250	18.62%
≥ $250 and < $500	8.88%
$500+	2.65%

Table 5 – Distribution of Sub-penny Increments for a Sample of 2,100,162 Non-Retail Trades in December 2021 by Trade Side. Sub-penny increments are defined by BJZZ as Z = 100*mod(price,.01).

Interval	Buy Orders	Sell Orders
$0 < Z \leq .1$	39.43%	2.66%
$.1 < Z \leq .2$	21.75%	3.73%
$.2 < Z \leq .3$	6.86%	3.02%
$.3 < Z \leq .4$	9.14%	9.09%
$.4 < Z \leq .5$	1.61%	1.03%
$.5 < Z \leq .6$	6.46%	6.60%
$.6 < Z \leq .7$	4.93%	4.42%
$.7 < Z \leq .8$	7.02%	15.48%
$.8 < Z \leq .9$	2.34%	20.43%
$.9 < Z < 1.00$	0.46%	33.54%

Table 6. Sub-penny pricing distribution for a sample of institutional trades executed away from exchanges between January 2011 and March 2012.

We obtain a sample of trades generated by institutional orders executed by a large investment bank (IB) between January 2011 and March 2012 used by Battalio et al. (2022). From this sample of trades, we extract all trades executed by an electronic liquidity provider (ELP) like Getco, Citadel, and Knight Securities and all trades executed in the IB's dark pool.

	% of Trades	
100*mod(Price,0.01)	Trades Executed in IB's Dark Pool (N = 166,266)	Trades Executed by ELP (N = 136,822)
0.00	38.4%	21.2%
0.00 < mod < 0.40	17.1%	38.7%
0.40 ≤ mod ≤ 0.60	28.4%	0.5%
0.60 < mod < 1.00	16.1%	39.6%

Table 7. Descriptive statistics comparing BJZZ-identified retail trades to the sample of retail trades matched to TAQ trades.

BJZZ require that the trade price be on one of two sub-penny intervals; greater than zero and less than \$0.004 or greater the \$0.006 and less than \$1.000. In addition, BJZZ require that the trade price exceed \$1.00 and not be subject to any non-normal trade condition. The BJZZ methodology results in identifying 7,349,520 of 24,135,132 (30.45%) of TAQ-matched proprietary data known retail trades.

Panel A. Descriptive Statistics of BJZZ-Identified retail trades and all matched retail trades.

Variable	Mean	
	Matched Retail Trades (N = 24,135,132)	BJZZ-Identified Retail Trades (N = 7,349,520)
Execution Time	12:15:00	12:19:27
Execution Size	231 shares	296 shares
Trade Price (trade-weighted)	\$123.14	\$108.88
Trade Price (stock-weighted)	\$62.47	\$63.18
Order Quantity	784 shares	582 shares
Order Side – Percent Buys	56.87%	57.96%
Percent Short Sells	2.84%	3.94%

Panel B. Distribution of trade times throughout the trading day for the BJZZ-Identified retail trades and all matched retail trades.

Hour	Matched Retail Trades (N = 24,135,132)	BJZZ-Identified Retail Trades (N = 7,349,520)
9:30 to 10:00	15.61%	12.72%
10:00 to 11:00	21.15%	23.68%
11:00 to 12:00	14.43%	14.37%
12:00 to 1:00	12.10%	11.59%
1:00 to 2:00	11.56%	11.04%
2:00 to 3:00	11.01%	10.75%
3:00 to 4:00	14.15%	15.85%

Panel C. BJZZ-Identified retail trades a percentage of the matched retail trades by time of day.

Hour	BJZZ Success Rate
9:30 to 10:00	24.98%
10:00 to 11:00	33.75%
11:00 to 12:00	30.41%
12:00 to 1:00	29.25%
1:00 to 2:00	29.16%
2:00 to 3:00	29.83%
3:00 to 4:00	34.20%

Panel D. BJZZ-Identified retail trades as a percentage of matched retail trades by order size.

Order Size (shares)	BJZZ Success Rate
1 – 99	28.13%
100 – 499	32.49%
500 – 999	33.31%
1000 – 1999	34.56%
2000 – 4999	45.74%
> 4999	26.52%

Panel E. BJZZ-Identified retail trades as a percentage of matched retail trades by trade size.

Trade Size (shares)	BJZZ Success Rate
1 – 99	28.45%
100 – 499	32.88%
500 – 999	32.17%
1000 – 1999	32.64%
2000 – 4999	48.74%
> 4999	48.34%

Panel F. BJZZ-Identified retail trades as a percentage of matched retail trades by trade price.

Execution Price	BJZZ Success Rate
$1.00 – $9.9999	33.93%
$10.00 – $49.9999	30.70%
$50.00 – $99.9999	39.95%
$100.00 – $249.9999	28.75%
$250.00 – $499.9999	28.89%

$500.00+	25.59%

Table 8. Descriptive statistics regarding the success of BJZZ trade side inference.

BJZZ infer trade side from the sub-penny pricing increment. Those trades with sub-penny pricing greater than $0.000 and less than $0.004 are inferred sells and those trades with sub-penny pricing greater than $0.006 and less than $0.01 are inferred buys. We compare the BJZZ-inferred side to the known order side in the proprietary retail data.

Panel A. Overall mix of the actual trade side to BJZZ inferred trade side.

	Known Order Side from Matched Retail Trade	
BJZZ Inferred Side	Buy	Sell
Buy	54.37%	2.29%
Sell	3.59%	39.74%

Panel B. Percentage of matched retail trades with the correct BJZZ-inferred trade side by time of trading day.

Hour	Correct Trade Side
9:30 to 10:00	93.77%
10:00 to 11:00	95.50%
11:00 to 12:00	94.54%
12:00 to 1:00	94.12%
1:00 to 2:00	93.96%
2:00 to 3:00	93.99%
3:00 to 4:00	95.11%

Panel C. Percentage of matched retail trades with the correct BJZZ-inferred trade side by order size and trade size.

Size	Order Size (shares)	Trade Size (shares)
1 – 99	94.24%	93.71%
100 – 499	93.72%	94.56%
500 – 999	94.09%	97.42%
1000 – 1999	94.56%	98.56%
2000 – 4999	99.13%	99.61%
> 4999	99.28%	98.82%

Panel D. Percentage of matched retail trades with the correct BJZZ-inferred trade side for matched retail orders filled with one and with multiple executions.

	Number of Orders	Success Rate
Retail orders filled with one trade	5,971,948	94.85%
Retail orders filled with multiple trades	711,641	62.62%

Panel E. Overall mix of the actual trade side to BJZZ inferred trade side when order placement time quote is $0.01.

	Known Order Side from Matched Retail Trade	
BJZZ Inferred Side	Buy	Sell
Buy	57.56%	0.07%
Sell	0.53%	41.84%

Table 9. Order imbalance descriptive statistics.

$MROIBVOL_{i,t}$ is the signed difference between the retail buy volume and the retail sell volume normalized by the sum of the retail buy and sell volume for stock i on day t. $MROIBTRD_{i,t}$ is the signed difference between the number of retail buy trades and the number of retail sell trades normalized by the sum of retail buy and sell trades for stock i on day t. $ODDMROIBVOL_{i,t}$ is the signed difference between the retail buy volume and the retail sell volume normalized by the sum of the retail buy and sell volume for odd lot trades in stock i on day t. $ODDMROIBTRD_{i,t}$ is the signed difference between the number of retail buy trades and the number of retail sell trades normalized by the sum of retail buy and sell trades for odd lot trades in stock i on day t. We construct each measure separately for the entire sample of proprietary retail trades and for the retail trades identified in the TAQ database by the BJZZ methodology. Our sample period is August 3, 2020 through July 26, 2022. We characterize the daily across stock correlations between the statistics computed using the inferred and the actual retail trading data in the table below.

Panel A. Overall Correlation Statistics

OIB Measure	Correlation	Mean Differences (Wholesaler data – BJZZ inferences)*
MROIBVOL	.3415	-0.0129
MROIBTRD	.3021	-0.0557
ODDMROIBVOL	.3181	+0.0324
ODDMROIBTRC	.2937	+0.0327

* All differences are statistically significant at beyond the .0001 level with a standard t-test.

Panel B. Weekly Order Imbalance (OIB) Statistics

Measure	Weekly Across-Stock Wholesaler – Inferred OIB		# Weeks with Mean Across-Stock Wholesaler OIB > Mean Inferred OIB	# Weeks where Binomial Test Indicates Wholesaler OIB > Inferred OIB Significance level in parenthesis (Hypothesis is ρ = .5)
	Mean	Median		
MROIBVOL	-0.0094	-0.0122	27 out of 104	< 0.0001
MROIBTRD	-0.0082	-0.0105	0 out of 104	< 0.0001
ODDMROIBVOL	0.0764	0.0766	102 out of 104	< 0.0001
ODDMROIBTRD	s.0767	0.0770	56 out of 104	.1634