Difficulties in obtaining a representative sample of retail trades from public data sources

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
Senior Research Support Director
Wharton Research Data Services

Abstract: Researchers using the Boehmer et al. (2021) methodology implicitly assume that the sample of retail trades identified as retail has the same characteristics as retail trades that are not identified as retail. Moreover, researchers must also assume the algorithm does not falsely identify institutional trades to be retail trades. We demonstrate that neither of these assumptions are valid in practice. We also show that the subset of known retail trades identified by the algorithm to be retail differs significantly along many dimensions from the subset of known retail trades that are not classified as retail.

Current draft, July 23, 2024.

Acknowledgments: The authors received no outside funding for this study and have no conflicts of interests related to this study. The authors thank Zhi Da, Tom Ernst, John Shim, Andriy Shkilko (discussant) and participants in the University of Maryland finance department's workshop, the Microstructure Exchange, and the 3rd TCU Finance conference for helpful comments on previous versions of the paper. Finally, we thank one or more wholesalers for providing some of the data used in this paper. An earlier version of this paper was circulated under the title, "Identifying Market Maker Trades as 'Retail' from TAQ: No Shortage of False Negatives and False Positives."

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 study trades placed by a subset of retail investors. For example, the seminal paper by Odean (1998) generated a plethora of papers that utilize data from a discount stockbroker to examine retail trading behavior. A second approach uses algorithms intended to identify retail trades in publicly available trade and quote data. 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. Once identified as a retail trade in this manner, the marketable side of the transaction is typically inferred using some version of the Lee and Ready (1991) algorithm. Cready, et al. (2014), however, note that a significant concern in studies using trade size cutoffs to identify retail investors "is spurious effects attributable to misclassification of transactions, particularly those originating from large investors." Hvidkjaer (2008) posits that the likelihood that a small trade is a piece of a large institutional order became much higher after the shift to decimal pricing in 2001 and the passage of Regulation NMS in 2005. As a result, most researchers currently avoid using trade size cutoffs to distinguish between 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 one of Financial Industry Regulatory Authority's (FINRA's) Trade Reporting Facilities (TRFs) rather than to 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." Consistent with this assertion, in May 2022, Rosenblatt finds that off-exchange trading was 40% of total U.S. equity trading volume. Further, Rosenblatt estimates that wholesalers, who execute marketable retail orders, account for 42.5% of off-exchange trading, Alternative Trading Systems (ATSs) and Single Dealer Platforms (SDPs) account for about 23%, and capital commitment and manual crossing of institution trading interests another 22%.[1] To separate retail and institutional trades reported to a TRF, BJZZ assume that only retail trades print on sub-penny prices at any increment other than the midpoint of the quoted spread. As noted in BJZZ, most retail orders are routed to so-called order flow wholesalers (often, but not always with the retail brokers charging payment for order flow) and are provided price improvement relative to the National Best Bid and Offer (NBBO) as the wholesalers compete to provide best execution for their clients' orders (see, e.g., Battalio and Jennings (2023a)). Conversely, BJZZ state that while institutional trades generally cannot execute at sub-penny prices, they "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."

The BJZZ paper had an immediate impact on the literature; we identify 28 papers applying their methodology that have been published or are forthcoming in top accounting and finance journals as of February 2024.[2] Of these papers, half provide some sort of caveat with respect to the ability of the BJZZ algorithm to identify retail trades. For example, 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." In this paper, we utilize both proprietary and publicly available data to evaluate the accuracy of the BJZZ algorithm and provide one analysis for which inaccuracies in the methodology affect conclusions drawn.

---

[1] See Trading Talk – Market Structure Analysis: An Update on Retail Market Share in US Equities, June 24, 2022.
[2] Appendix Table A1 contains a list of the 28 papers, as well as any caveat that is made regarding the ability of the BJZZ algorithm to cleanly identify retail trades. The table also includes Barber et al., 2023, which evaluates the ability of the BJZZ algorithm to identify actual retail trades. As of the date of this draft, Google Scholar lists 430 citations.

Barber et al. (2023) conduct an experiment to evaluate the effectiveness of the BJZZ methodology at identifying retail trades by placing over 85,000 orders in 128 stocks through six retail brokers. 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. We utilize a proprietary dataset of actual retail orders in over 6,800 securities to evaluate the generalizability of the conclusions in Barber et al. Given the nature of their dataset, Barber et al. cannot document the frequency with which the BJZZ algorithm incorrectly identifies institutional trades that execute off exchanges as retail trades. Anand et al. (2021) access the proprietary FINRA Order Audit Trail System database and examine institutional parent and child orders executed in a size-stratified group of 300 stocks in October 2016. They find that 26.18% of the trades generated by these parent orders are executed either in ATSs, SDPs, or by wholesalers, each of whom reports trades to a TRF. If these trades are executed on subpennies other than the spread midpoint, they have the potential of being identified as retail by the BJZZ algorithm. Our paper contributes to the literature by utilizing proprietary data obtained from several sources that allows us to identify the frequency that the BJZZ algorithm incorrectly identifies institutional trades reported to a TRF as retail trades.

We begin by examining whether the findings regarding Type II errors of Barber et al. (2023), obtained using their sample of very small (most trading only $100 worth of stock) marketable orders placed in a limited number of stocks, are representative of the BJZZ algorithm's success for a comprehensive sample of trades placed by actual retail investors. To do this, we obtain a dataset containing over 53 million trades generated by over 40 million retail marketable orders executed by one or more wholesalers in May 2022 and document the ability of the BJZZ algorithm to successfully identify the trades generated from these orders as retail.[3] Using the New York Stock Exchange's Trade and Quote (TAQ) database, we match the proprietary trades that are known to be retail to trades reported to a TRF and use this sample of matched trades to evaluate the likelihood of Type II errors (i.e., that the BJZZ methodology fails to identify them as retail trades). Almost 40.5% of the matched trades have execution prices without sub-penny prices and roughly 32% of the matched trades have trades that have a sub-penny price in the interval [0.4, 0.6] and, thus, are not identified as retail trades by BJZZ. As a result, less than 28% of the sample of known 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. These results suggest that the findings of Barber et al. (2023a) regarding the frequency of Type II errors are somewhat understated compared to datasets containing trades placed by actual retail investors and that their findings regarding the correct inference of order side might be specific to the orders they submitted. The results of a multivariate analysis indicate that the success rate of the BJZZ algorithm regarding identifying actual retail trades in a stock on a given day is higher for less volatile, high-priced stocks with spreads wider than $0.01. Thus, the Barber et al. sample of BJZZ-identified retail trades in a set of randomly selected stocks may not be representative of actual retail trades.

Another assumption of BJZZ is that institutional order flow executed in an ATS or a SDP and reported to a TRF does not receive non-midpoint sub-penny price improvement, i.e., that there is a low frequency of Type I errors. Evidence to the contrary would make it difficult to interpret the results of studies that examine whether "retail" trades identified using the BJZZ methodology

[3] For literary convenience and brevity, we refer to "one or more wholesaler(s)" with simply wholesaler(s) in the remainder of the paper when referring to the data-providing wholesaler(s). Our dataset contains more than 30% of all orders and shares routed to and executed by the six largest wholesalers in May 2022 as reported in SEC Rule 605 reports.

are informed. As our second contribution, we examine the extent to which institutional trades reported to a TRF potentially are misclassified as retail trades by the BJZZ algorithm.

In stark contrast to the assertions made by BJZZ, we find that a nontrivial percentage of institutional trades reported to a TRF in three separate datasets are executed in subpennies other than the spread midpoint. We obtain a sample of 2,408,465 institutional trades executed with sub-penny prices other than 0 or 5 in May 2022 from our data-providing wholesaler(s) and find that 2,087,284 are misclassified as retail by the BJZZ algorithm. We also use an alternative set of institutional trades from a major investment bank to evaluate the BJZZ algorithm.[4] 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 trades executed in the broker's ATS, one-third are 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. Finally, we examine trades executed on behalf of the Colorado Public Employee's Retirement Association (CoPERA) in 2016 and 2017.[5] Of the 363,459 (6,203) pension fund's trades in an ATS or SDP (with electronic liquidity providers), we find that about 6.1% (26.4%) would be classified as retail using BJZZ. Together, these results suggest Type I errors are not insignificant and call into further question whether the samples of BJZZ-identified retail trades used in prior academic analyses are representative of true samples of retail trades.

To better understand the potential for Type I errors to affect inferences in studies of retail investor trading behavior, we offer two additional analyses. First, we compare the volume of retail

[4] These are the same data used by Battalio, Hatch, and Saglam (2023).
[5] Data obtained via a Freedom of Information Act filing associated with a different research project.

marketable orders executed by the eight wholesalers operating in the U.S. equity markets as reported in their Rule 605 reports to the volume of BJZZ-identified retail trades in the NYSE's TAQ database in May 2022 on a stock-by-stock basis. The wholesaler 605 reports may contain trades reported to both exchanges and to a TRF. Thus, if the BJZZ algorithm works well, the volume of BJZZ-identified retail TAQ trades in a given stock cannot be larger than the volume of retail trades executed in that stock by the eight wholesalers.[6] We identify several stocks with a consolidated trading volume of over 100 million shares in May 2022 for which the volume of BJZZ-identified retail trades ***exceeds*** the volume of retail trades executed by the eight wholesalers. One of these stocks is Zynga, which is the 13th most actively traded stock in May 2022. These results indicate that Type I errors have the potential to significantly bias inferences made using BJZZ-identified retail trades to investigate retail trading in individual stocks.

Second, using the proprietary sample of institutional orders obtained from a large investment bank, we show that there is a significant positive correlation between the direction of institutional trading activity and an order imbalance measure created using BJZZ-inferred retail trades, suggesting that the "retail" order imbalance measure in Boehmer et al. is biased on days with institutional trading. Together with our analyses of individual institutional trades, these results suggest that the prevalence of Type I errors makes it difficult to assess the extent to which a sample of BJZZ-identified retail trades is representative of actual retail trades.

[6] As the Rule 605 reports only provide information for marketable orders seeking to trade 100 shares or more, we compare the volume of BJZZ-identified retail TAQ trades for 100 shares or more to the volume of retail trades executed by the eight wholesalers. Since an order for 100 shares can be executed in two trades of 50 shares, the trading volume of the eight wholesalers in a given stock should exceed the BJZZ-implied retail TAQ trading volume even if all trades are executed in subpennies. However, our work and that of Barber et al. (2023a) indicate the BJZZ algorithm only identifies about 30% of retail trades as being retail. Thus, if Type I errors are not an issue (e.g., institutional trades are rarely identified by the BJZZ algorithm as retail trades), the volume of BJZZ-identified retail TAQ trades in a given stock should, on average, be around 30% of the volume of trades executed in that stock by the eight wholesalers.

We conclude by examining whether the typical process wholesalers use to execute marketable retail order flow induces a bias with respect to the market conditions in which marketable retail orders execute on a sub-penny interval and thus are recognized as retail by the BJZZ algorithm. We first hypothesize why this might be the case and then we produce evidence that execution quality metrics such as price improvement, effective and realized spreads, and potential price impact (i.e., the price concession an order would have to make to fully execute against displayed market-wide liquidity) are correlated with the time elapsing between order arrival and trade execution. Next, we examine the ability of the BJZZ methodology to identify known retail trades as retail across time-to-trade deciles and find that the proportion of trades identified as retail increases (almost) monotonically across deciles until falling dramatically in the slowest decile. We find that these biases in identification rates result in BJZZ-identified retail trades having a higher price improvement rate, higher effective and realized spreads, and less price impact than those retail trades not identified as retail by the methodology. Our evidence suggests that Type II errors potentially explain why the cost estimates produced by the U.S. Securities and Exchange Commission (SEC) in the proposed Order Competition Rule[7] are significantly higher when actual retail trades rather than BJZZ-identified retail trades are analyzed.[8] We expect this bias may plague other studies of retail trading behavior that utilize BJZZ-identified retail trades.

The remainder of the paper has four sections. In the following two sections, we evaluate the BJZZ algorithm's propensity to commit Type II (not identifying known retail trades as retail) and Type I (identifying known institutional trades as retail), respectively. Both types of errors are

[7] https://www.sec.gov/rules/2022/12/order-competition-rule
[8] https://www.cato.org/regulation/winter-2023-2024/examining-secs-proposed-order-competition-rule#:~:text=In%20late%202022%20it%20proposed,such%20as%20a%20registered%20exchange.

alarmingly frequent. We then conduct an analysis demonstrating that, for at least one application of the BJZZ methodology, conclusions are affected by Type II errors. The final section concludes.

I. Type II Errors.

A. Data.

To examine the BJZZ algorithm's ability to correctly identify actual retail trades, we obtain proprietary marketable order and trade data from the data-providing wholesaler(s) for the month of May 2022.[9] We receive all the marketable retail orders handled by the wholesaler(s) during this 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.

The BJZZ methodology uses publicly available TAQ data to infer retail trades and order sides. Thus, to assess the success of the BJZZ algorithm for identifying a sample of known retail trades, we match our set of proprietary retail trades to trades reported to a TRF in May 2022. For a retail trade to be matched to a TRF trade, we require the symbol, trade size, trade price, and trade date to be the same. Further, we require that the timestamps of the trades (one from the proprietary database and one from TAQ) are within one second of each other. When multiple TRF trades can be matched to one retail trade, we choose the TRF trade that minimizes the difference in execution

[9] These are the same data are described in detail in Battalio and Jennings (2023a), who use wholesaler Rule 605 reports to show their dataset is not an outlier with regards to execution quality statistics reported by the top six wholesalers in in May 2022.

times. Finally, a TRF trade can only be matched to one retail trade. We match 98.59% of the proprietary retail trades to trades reported to the TRF.

We restrict our analysis to symbols that trade each day in May 2022. Since securities priced below a dollar can quote in subpennies, we eliminate securities with trades occurring at prices below a dollar from our sample. This reduces the set of securities in our sample from 11,901 unique symbols to 6,823 unique symbols. We provide descriptive statistics of our sample in Table 1.

[Insert Table 1 about here.]

Looking first at Panel A, across sample securities, the beginning-of-month share price ranges from the designed minimum of $1.01 to over $4,300 and averages $47.71. Round-lot trades for the month range from 0 to 757,119 and the average sample security has 2,306 round-lot trades. Odd-lot trades range from 2 to 1,077,137 with the average security having 4,226 odd-lot trades.

Panel B provides information on the time elapsing between order receipt and trade time from the wholesaler(s) perspective for sample trades. Over half of the trades occur less than seven milliseconds after retail marketable orders are received by the wholesaler(s) and the wholesaler(s) executed 95% of the sample trades within 49 milliseconds after they receive the orders. Some orders take some time to execute, as evidenced by the fact that 1% of the sample trades are executed more than thirteen seconds after orders are received.[10]

B. Failing to identify retail trades as retail trades.

We use BJZZ's methodology to infer which of the TRF-matched known retail trades came from retail investors. If the algorithm does not have Type II errors, it will identify all of the TRF-matched trades as 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 2 contains the sub-

[10] These tend to be limit orders that are initially marketable with a market price dynamic causing the order to become non-marketable for some period of time.

penny pricing distribution for matched retail trades by trade side conditional on the order-receipt-time (ORT) NBBO width.

[Insert Table 2 about here.]

We find that 40.47% of the matched retail trades have no sub-penny pricing and 31.70% have sub-penning pricing that BJZZ consider as midpoint pricing (and therefore not classified as retail). Of trades with sub-penny increments in the range [0.4, 0.6], 91% of those trades are exactly at a sub-penny pricing increment of 0.5 (not tabulated in Table 2). Thus, only 27.83% of our sample of matched retail trades has $Z_{i,t}$ in the (0, .4) or (.6, 1) intervals and are classified as retail by the BJZZ methodology. This is lower than the Barber et al. (2023) identification rate of 35%. When the quoted spread is $0.01, the BJZZ algorithm does slightly less well, identifying 27.11% of the known retail trades as retail, with trades more frequently occurring around the spread midpoint.

[Insert Table 3 about here.]

In Table 3, we provide some descriptive statistics regarding the frequency with which BJZZ-identified retail trades are signed correctly as buys or sells by the BJZZ methodology. Panel A reveals that the algorithm misclassifies the initiator of 9% of the trades in our sample of actual retail trades, which is much lower than the algorithm's misclassification rate of 28% for the sample of trades used in Barber et al. (2023). In Panel B, 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 error prone. 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. Panel B indicates that when we restrict our sample to matched retail trades received by the wholesaler(s) when the width of the NBBO is \$0.01, the percentage of trades for which the inferred order side is incorrect falls to 0.33%. The fact that the 25th percentile trade in Barber et al. (2023) has an execution time NBBO spread width of a penny compared to 40 percent of the trades in our sample suggests that the differences in error rates across the two studies are related to the percentage of sample trades executed when the NBBO is a penny.[11] Overall, our results question whether the Barber et al. suggestion of using Lee and Ready (1991) to type trades might be unnecessary with a sample that is more representative of actual retail trading.

C. Determinants of Type II errors.

In this section, we examine whether Type II errors within a stock on a given day are associated with characteristics of the trade, the stock, or the market conditions prevailing when the trade occurs. For our sample of retail trades in May 2022, we use ordinary least squares to estimate the following equation:

$$\begin{aligned} Success\ Rate_{i,t} &= \alpha + \beta_1\ Penny\ Spread_{i,t} + \beta_2\ Log(Avg.\ Trade\ Price)_{i,t} \\ &+ \beta_3\ Log(Avg.\ Trade\ Size)_{i,t} + \beta_4\ Stock\ Volatility_{i,t} + \beta_5\ Log(Market\ Cap)_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{1}$$

where $Success\ Rate_{i,t}$ is the percentage of retail trades in stock i on day t that are correctly identified as retail by the BJZZ algorithm, $Penny\ Spread_{i,t}$ is an indicator variable taking a value of 1 if all the retail trades in stock i on day t executed when the width of the NBBO was a penny, $Log(Avg.\ Trade\ Price)_{i,t}$ is the log of the average trade price for sample trades in stock i on day t, $Log(Avg.\ Trade\ Size)_{i,t}$ is the log of the average trade size for sample trades in stock i on day t, $Stock\ Volatility_{i,t}$ is the standard deviation of stock i's daily returns over the past 20 trading days prior to

[11] Table I of Barber et al. reports that the execution time NBBO spread width for the 50th, 75th, and 90th trades are \$0.05, \$0.15, and \$0.35 respectively. In our sample of known retail trades, the execution time NBBO spread width for the 50th, 75th, and 90th trades are \$0.02, \$0.06, and \$0.19 respectively.

day t, and *Log(Market Cap)*$_{i,t}$ is the logarithm of the market capitalization of stock *i* on day *t*. We cluster standard errors at the stock and day level.

The univariate results in Table 2 indicate that sample retail trades executed when the width of the NBBO is a penny are more likely to receive midpoint price improvement and are less likely to trade on the round penny than sample retail trades executed when the NBBO spread width exceeds a penny. Overall, the results in Table 2 suggest the algorithm's success rate for a given stock on a given day will be slightly lower if all the sample retail trades in that stock executed when the spread width is a penny. It follows that the algorithm's success rate will be higher for stocks with higher average trade prices, as those stocks tend to have wider bid ask spreads. Since liquidity provision is generally more costly for larger trades and trades executed in more volatile stocks, we posit price improvement is less frequent in these stocks. As result, we expect the algorithm's success rate to be negatively related to a stock's average trade size and stock price volatility. Finally, assuming inventory management costs are lower in stocks with higher market capitalizations, we expect price improvement rates and the algorithm's success rate to be higher in stocks with larger market capitalizations.

[Insert Table 4 about here.]

The results of our analysis are presented in Table 4. Consistent with our expectations, our multivariate analysis suggests that in a controlled setting, the ability of the algorithm to correctly identify sample retail trades in the average stock on a given day is slightly lower when each of the retail trades in the stock occur when the width of the NBBO is a penny. As expected, the average success rate of the BJZZ algorithm is lower for stocks that have higher prices, larger average trade sizes, and higher volatility. We find little evidence that the algorithm's success rate is related to the stock's market capitalization.

To summarize, we find that several of the assumptions made to derive the BJZZ algorithm are suspect. A substantial majority 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]. Thus, the BJZZ algorithm only has a chance of identifying a fraction of the initial sample of retail trades. In our case, less than 28% of the retail trades in our sample are identified as retail. For the minority of known retail trades typed as retail by BJZZ, most are assigned the correct order side. Finally, our multivariate analysis suggests researchers using BJZZ-identified retail trades to study retail trading behavior in a randomly selected sample of stocks will over-sample less volatile, higher priced stocks that have bid ask spreads that exceed a penny.

II. Type I Errors.

A. Falsely identifying institutional trades as retail trades.

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 three samples of non-retail trades. As discussed previously, the institutional investor order/trade data used to examine Type I errors (institutional trades falsely identified as retail trades) come from the data-providing wholesaler(s), a major investment bank, and a public pension fund.

Firstly, we obtain a sample of 2,408,465 institutional trades reported to a TRF during May 2022 from the same data provider(s) who supplied our sample of retail orders. As the point of this exercise is to examine the frequency with which the BJZZ algorithm falsely classifies institutional trades as retail, our data provider(s) only provided trades that were not executed with sub-penny increments of 0 or 5. The mean and median trade prices are $95.06 and $45.07 respectively. Regarding trade size, 38% of this sample of institutional trades executed by our data provider(s) in the dark are odd lots, 50% are for 100 to 499 shares, and 6% are for 500 to 999 shares. Only 0.52% of the sample trades are for 5,000 shares or more. Thus, while these trades do not result from retail orders, it would be difficult to distinguish them from retail trades based on price and size.

[Insert Table 5 about here.]

In Panel A of Table 5, we summarize the frequency of various sub-penny pricing intervals for these institutional trades. Of the 2.4 million institutional trades not executed with sub-penny increments of 0 or 5, only 13% would be eliminated using BJZZ's "near one-half penny" exclusion rule. Thus, 2,087,284 institutional trades from our data provider(s) would be included as retail using BJZZ's algorithm in May 2022. To put this into context, the BJZZ algorithm identified 7,048,164 of the actual retail trades executed by the same data provider(s) in May 2022. Given that other wholesalers also operate single dealer platforms, the prevalence of Type I errors calls into question whether BJZZ-identified retail TRF trades are representative of actual retail trades.

Secondly, 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 algorithm.[12] This data set was also used by Battalio et al. (2022) and covers large institutional orders in S&P 500 stocks between January 2011 and March 2012. Here, we consider the entire set of 166,266 child order executions in the IB's dark pool and 136,833 child order executions that source liquidity from electronic liquidity providers like Getco, Citadel Securities, and Knight Securities. Panel B of Table 5 contains the sub-penny pricing distribution for each collection of trades. Focusing first on the second column of Panel B, 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 classified as retail by the BJZZ algorithm.

[12] Volume weighted average price algorithms are one of the most commonly used trading algorithms that seek to match the volume-weighted average price realized during the trading horizon.

Moving to the third column of Panel B, we see that 78.3% of the trades executed by electronic liquidity providers away from exchanges have sub-penny pricing increments that make the trades eligible to be classified as retail by the BJZZ algorithm.

Thirdly, we employ the dark pool and electronic liquidity provider trades of the Public Employee's Pension Association of Colorado (CoPERA) during 2016-2017. Panel C of Table 5 provides an analysis of the sub-penny trade prices of CoPERA. Column 2 contains the results of trades in the ATS and/or SDP operated by the broker CoPERA used for the parent orders and Column 3 presents results for trades executed in ATSs and SDPs not operated by the broker handling CoPERA's order.[13] Finally, Column 4 reports the results of trades with electronic liquidity providers (again, such as wholesalers). In the broker's own dark pool trades about 15.6% of the trades occur at sub-penny prices BJZZ would classify as retail trades. In dark venues not operated by the broker handling the order, the percentage of BJZZ-inferred retail trades is much smaller, approximately 3.8%. Finally, 26.4% of the pension fund's trades executed by electronic liquidity providers are classified as retail by BJZZ due to the price improvement provided.

The results from all three sets of institutional trades contradict the assertion that institutional trades executed in the dark "are usually in round pennies" and suggest that Type I errors may plague studies that use BJZZ-identified trades to examine ***retail*** trading behavior. Indeed, this assertion seems (marginally) appropriate only for the CoPERA's trades where, overall, "only" 6.5% of the trades are classified as retail by BJZZ. There is the potential for these magnitudes of Type I errors to produce many incorrect retail attributions. As noted above, Rosenblatt estimates that wholesalers account for about 17% of consolidated tape trading volume and ATSs and SDPs account for approximately another 9% during May 2022. Across our three

[13] The distinction between trades internalized by CoPERA's routing broker and trades conducted in liquidity pools external to the routing broker were in the dataset we received from CoPERA.

samples of electronic liquidity provider trades, the estimates of non-half-penny sub-penny prints range from 11% to 78%, and across our two samples of dark pool trades, the analogous range is 6% to 33%. Given that the NYSE's TAQ database reports there 1,919,062,354 trades in the consolidated tape in May 2022, this implies 46.3 (311.5) million potentially mis-identified trades per month at the low (high) end of the range. 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 and, therefore, be identified as a retail trade.[14]

B. Actual versus inferred retail trading volume.

SEC (2022) finds that "broker-dealers route more than 90% of marketable orders of individual investors in NMS (National Market System) stocks to a small group of six off-exchange dealers, often referred to as 'wholesalers.'" Dyherberg et al. (2023) identify eight wholesalers operating in U.S. equity markets in 2022.[15] Battalio and Jennings (2023b) find that even brokers who do not accept order flow inducements route their marketable orders to wholesalers. If we define marketable retail orders as orders executed by wholesalers, then we can use the SEC-mandated Rule 605 execution quality reports published by the eight wholesalers to quantify retail trading volume in a given stock in a given month. The one caveat to this claim is that the Rule 605 reports do not provide information for orders seeking to trade odd lots (usually fewer than 100 shares). Here, we use Rule 605 data to gain insight as to the severity of Type I errors encountered when using the BJZZ algorithm by comparing the aggregate volume of BJZZ-identified trades for

[14] 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."
[15] The eight wholesalers by market share are Citadel, Virtu, G1, Jane Street, Two Sigma, UBS, Merrill Lynch, and Morgan Stanley. As some of these wholesalers report trades under multiple market maker ids, we obtain the market maker ids that correspond to the execution of retail order flow from Andriy Shkilko.

100 or more shares in TAQ to the aggregate volume of marketable wholesaler orders for 100 shares or more on a stock-by-stock basis in May 2022.

If each marketable retail order in May 2022 is executed in one fill (e.g., a 100 share market order produces a 100 share trade and not two 50 share trades), if each retail order is sent to one of the eight wholesalers, if all retail trades are reported to a TRF, and if the BJZZ algorithm works perfectly (e.g., there are not Type I or Type II errors), the ratio of BJZZ-inferred retail trading volume to wholesaler Rule 605 trading volume should be equal to 1.00 for stocks that retail investors trade. The results presented in Barber et al. (2023) and in Section I of this paper suggest that on average, the BJZZ algorithm correctly identifies only a minority of actual retail trades. Thus, if there are no Type I errors, the average across-stock ratio of inferred-to-actual retail trading volume should be around 30%. Given that marketable retail orders are often executed in multiple trades, some of which may be odd lots, this ratio should never exceed 100% for a given stock/month.[16]

For each sample common stock in May 2022, we compute the ratio of BJZZ-inferred retail trading volume to aggregate wholesaler Rule 605 trading volume, hereafter referred to as the inferred-to-actual retail trading ratio, as follows. We begin by using the BJZZ algorithm to identify retail trades within the entire universe of market-wide trades reported to a TRF in May 2022 for each sample stock. On a stock-by-stock basis, we then aggregate all BJZZ-identified retail trades that are for 100 or more shares. Finally, we compute the aggregate wholesaler Rule 605 trading volume as the sum of all marketable retail trades executed by the eight wholesalers.[17] This will

[16] 94.32% of our sample of marketable retail orders fully execute with one trade (results not reported).

[17] We drop all tickers that are not able to be matched across the TAQ database, the Rule 605 database, and the CRSP database. We also drop stocks that experience a stock split or a share repurchase as the Rule 605 data do not adjust for these corporate actions. We also drop stocks whose average daily closing price in CRSP is less than $1.00 in May 2022. A spreadsheet containing the resulting database is available from the authors.

include fully internalized retail orders (e.g., orders that fully source wholesaler liquidity) and orders that source external liquidity from competing wholesalers, in ATSs and SDPs, or on stock exchanges.[18] Finally, to assess the extent of inferred and/or actual retail trading volume in a given stock/month, we compute the aggregate trading volume excluding odd lot trades in each stock for May 2022. To ensure sufficient trading interest, we restrict our analysis to the 3,704 stocks with at least (arbitrarily set) 100,000 shares of consolidated trading volume. Table 6 provides descriptive statistics by decile of consolidated trading volume in May 2022.

[Insert Table 6 about here.]

The average consolidated volume in the bottom (top) decile is 0.3 (247.3) million shares. The average percentage of actual retail trading volume for stocks in the bottom decile is 26.8%, while the average amount of inferred retail trading volume for these stocks is 10.6%. Average retail trading volume as a percentage of total trading volume is smallest, 7.6%, in decile 9. The difference in the concentration of actual retail trading across stock-deciles suggests selecting a random sample of stocks across volume deciles (as was done by Barber et al. (2023)) is unlikely to produce a random sample of the stocks traded by actual retail investors. For the least actively traded decile of stocks, the average inferred-to-actual retail trading ratio is about 45%. As a percentage of consolidated volume, actual retail and BJZZ-identified retail trading volume generally fall as consolidated trading volume increases except for the most active decile. The average percentage of actual retail trading volume for stocks in the top decile is 11.07%, while the average amount of inferred retail trading volume for these stocks is 6.75%. This implies that the average inferred-to-actual retail trading ratio for the most actively traded decile of stocks is 58.9%.

[18] As the BJZZ algorithm can only identify retail trades reported to a TRF but the wholesaler 605 retail volume consists of retail trades reported to the TRF and exchange-executed retail trades, the implied to actual retail trading ratio should never exceed one even if each marketable retail order in the proprietary sample is fully executed with a single trade.

Overall, each of the ten deciles have average inferred-to-actual retail trading ratios more than 40% and four of the deciles have average inferred-to-actual retail trading ratios in excess of 50%. As noted earlier, given the results in Barber et al. and our Table 2, one would expect that on average these ratios would be close to 30% if Type I errors are inconsequential.

[Insert Figure 1 about here.]

For another perspective, in Figure 1, we plot the inferred-to-actual retail trading ratios for each percentile of stocks, where stocks are ranked from low to high based on their inferred-to-actual retail trading ratios in May 2022. We present separate plots for the 100 stocks with the highest consolidated trading volume and for all 3,740 that stocks that survive our screens. Focusing on the 100 stocks with the highest volume, the dashed line in Figure 1 illustrates that the inferred-to-actual retail trading ratio is greater than 30% for each of the 100 most actively traded sample stocks. Of these stocks, 58 have inferred-to-actual retail trading ratios exceeding 60%, 17 have ratios in excess of 80%, and 5 stocks have ratios in excess of 100%. Zynga, the sample stock with the 13th highest trading volume (861 million shares excluding odd lots) has Rule 605 retail trading volume of 20.5 million shares, BJZZ-implied retail trading volume of 30.5 million shares, and an inferred-to-actual retail trading ratio of 148%. Thus, for Zynga in May 2022, non-retail trades comprise *at least* one-third of the sample of BJZZ-identified retail trades. These results suggest that Type I errors pose a problem for actively traded stocks that are likely of interest to institutions.

The solid black line in Figure 1 presents inferred-to-actual retail trading ratio by percentile for all stocks. Seventy-five stocks have ratios that exceed 100%, which suggest that for these stocks non-retail trades are likely to make up a large percentage of the sample of BJZZ-identified retail trades. Approximately 19% of sample stocks have ratios exceeding 60%, 41% of sample stocks have ratios exceeding 50%, and 89% of sample stocks have ratios of more than 30%. Interestingly, only around 11% of the sample stocks have inferred-to-actual retail trading ratios that are less than 30%. Together, these results suggest that Type I errors can be an issue for the more actively traded stocks that are more likely to be traded by institutional investors. At the very least, the results presented in Table 6 and illustrated in Figure 1 demonstrate that Type I errors are likely prevalent in most samples of BJZZ-identified retail trades.

C. Determinants of Type I errors.

If a large institutional parent order includes many sub-penny trades in a stock on a trading day with low retail trading activity, the BJZZ algorithm can perform poorly. In our investment bank data set, we find some extreme examples of poor performance. For example, the BJZZ algorithm reports 37,777 shares traded by retail investors on July 11th, 2011, in Vulcan Materials Company (NYSE:VMC). However, examining the sample of institutional trades provided by the IB, we find that 32,800 of these shares (corresponding to 87%) are traded by institutional investors with sub-penny increments BJZZ classifies as retail on that day.

We use data from the IB to examine whether Type I errors are associated with characteristics of the parent order and/or the market conditions in which parent orders are placed. We first compute *% Parent Misclassified*$_i$ as the fraction of a parent order *i*'s trades that the BJZZ algorithm incorrectly identifies as retail. Across all parent orders, *% Parent Misclassified*$_i$ ranges from 0% to 100% with an average value of 10.9%. We estimate the following regression with various parent-order and stock-level characteristics to examine the first-order determinants of the percentage of a parent order's trades that are misclassified as retail:

$$\begin{aligned} \%\ \textit{Parent Misclassified}_i = {} & \alpha + \beta_1\, \textit{Participation Rate}_i + \beta_2\, \textit{Penny Spread}_i \\ & + \beta_3\, \textit{Stock Volatility}_i + \beta_4\, \textit{Turnover}_i + \beta_5\, \textit{Log(Market Cap}_i\textit{)} + \varepsilon_i, \qquad (2) \end{aligned}$$

where *Participation Rate*$_i$ is the ratio of the shares in the i^{th} executed parent order to the trading volume of the underlying stock during the parent order *i*'s life, *Penny Spread*$_i$ is an indicator variable that equals one if each of the parent's child trades are executed when the when the width of the NBBO is a penny and equals 0 otherwise, *Stock Volatility*$_i$ is the volatility of the stock price over the life of the i^{th} parent order, *Turnover*$_i$ is the executed shares of stock during the lifetime of the parent order life divided by the stock's shares outstanding, and *Log(Market Cap)*$_i$ is the logarithm of the market capitalization of the underlying stock on the day parent order *i* is executed. We cluster standard errors at the stock-day level.

All else equal, a higher parent order participation rate increases the likelihood that the order executes on exchanges at displayed prices because the IB is more aggressively sourcing liquidity from all available sources to fill a large order. Similarly, when the underlying stock price or the market is more volatile, price improvement of any amount is less likely as it is risky to provide better prices in unsettled markets. This suggests the percentage of parent order's trades that are misclassified as retail will be inversely related to the aggressiveness of the parent order and to the volatility of the underlying stock price. All else equal, when the width of the stock's bid ask spread is equal to a penny, price improvement must be in subpennies. Thus, we expect a positive relationship between the percentage of a parent order's share volume that is misclassified as retail by the BJZZ algorithm and the binary variable *Penny Spread*. Finally, we posit that inventory management costs are lower and price improvement rates higher in stocks with larger market capitalizations. This leads us to expect a positive relationship between the percentage of a parent order's trades that are misclassified and the underlying stock's market capitalization.

[Insert Table 7 about here.]

Columns (1) to (5) in Table 7 report the regression results in various specifications of equation (2). We find consistent patterns across these models. As predicted, the percentage share volume of a parent order that is misidentified as retail by the BJZZ algorithm is, on average, lower when the parent order's participation rate or the underlying stock's volatility is higher. Furthermore, as hypothesized, the underlying stock's market capitalization and having all trades executed when the spread is a penny are each positively correlated with the percentage of a parent order's trades that are misclassified. Overall, the analysis summarized in Table 7 suggests the bias associated with Type I errors will be higher for stocks with significant institutional trading.

D. Type I errors and BJZZ-implied retail order imbalances.

BJZZ use their algorithm to construct (supposedly) retail order imbalance measures. Most of the articles mentioned in Appendix Table A1 use this measure to address the impact of retail trading. In this section, we formally test whether Type I errors introduce a bias in order imbalance measures created using BJZZ-identified retail trades. Despite the misclassification of institutional trades as retail by the BJZZ algorithm as documented earlier, it is possible that there is no statistically significant bias in this aggregate retail trading proxy. However, our analysis thus far suggests that institutional trades that are misidentified as retail will be executed for larger orders in less volatile market conditions and when the underlying stock has a lower share price and minimum spread (e.g., a higher relative spread).

$Mroibvol_{i,t}$ is formally defined in Boehmer et al. as the signed difference between the BJZZ-identified retail buy volume and the BJZZ-identified retail sell volume normalized by the sum of the retail buy and sell volume for stock i on day t. As the institutional trading data obtained from the IB are for S&P500 stocks, we compute the *Mroibvol* measure for all sample securities in the S&P 500 from January 2011 through March 2012. Note that only a small fraction of these stock-day combinations is observed in the sample IB's institutional data set (approximately 12%). If these aggregate stock-level proxies were driven by solely retail trading demand, we should find no significant correlation to statistics generated by the institutional large order data set, especially considering the weak overlap between the data sets.

We run the following regression to determine whether the BJZZ-implied *Mroibvol* is significantly correlated with some basic characteristics of the institutional order:

$$Mroibvol_{i,t} = \alpha + \beta_1\, Buy_{i,t} + \beta_2\, Sell_{i,t} + \beta_3\, InstSubPenny_{i,t} + \varepsilon_{i,t}, \tag{3}$$

where $Buy_{i,t}$ takes a value 1 if there is an institutional buy order in our dataset on stock i and day t, $Sell_{i,t}$ takes a value 1 if there is an institutional sell order in our dataset on stock i and day t, and $InstSubPenny_{i,t}$ is the ratio of the signed sub-penny institutional volume (excluding the midpoint range) to total volume during the execution period for stock i on day t. We again cluster our standard errors at the stock-day level. Table 8 reports the regression results.

[Insert Table 8 about here.]

We find a significant positive (negative) correlation between *Mroibvol* and the presence of an institutional buy (sell) order in our IB data. This finding implies that the direction of institutional trading is informative for the net "retail" order imbalance. The coefficients are sizeable economically when compared to the estimated constant. This is quite surprising given that only 12% of the stock-day combinations are covered in the sample IB's institutional data. Further, the sub-penny institutional trading activity as proxied by *InstSubPenny* is also significantly positively correlated with *Mroibvol.* This finding implies that sub-penny fills of institutional orders are biasing the estimates of net retail order imbalance with high statistical significance.

About 37% of overall 2023 trading volume in U.S. equity markets was executed through algorithms and/or smart routers and one of the most popular algorithms is the VWAP algorithm, which the IB used to execute our sample of institutional orders.[19] Based on these facts, our findings suggest that Type I errors have the potential to bias inferences as to the behavior of retail investors that are made using order imbalance measures constructed from BJZZ-identified retail trades. This is especially true in stocks primarily traded by institutional (and not retail) investors.

III. Why the presence of Type II errors may lead to faulty inferences.

Below, we employ our proprietary marketable retail order data to determine if Type II errors in the BJZZ methodology (failure to identify all trades from retail orders as retail trading) affect inferences regarding retail trading. The setting we choose is the cost analysis conducted by the SEC in support of the Order Competition Rule (OCR), which was proposed in December 2022. As written, the OCR would essentially require that most marketable retail orders be submitted to a qualified auction lasting between 100 and 300 milliseconds before the order could be internalized by a wholesaler. Acknowledging that failed auctions can impose costs on retail investors, the SEC estimated this cost by examining the frequency with which the market moves against the investor (the offer increases following a buy or the bid falls following a sale) in the 100 to 300 milliseconds following a trade. However, rather than utilize actual retail orders contained in the Consolidated Audit Trail data in their analysis, the Commission used BJZZ-identified retail trades from TAQ to estimate the costs associated with failed auctions. Using the same data as employed in Section I of this paper, Battalio and Jennings (2023b) demonstrate that the use of BJZZ-identified retail trades led to cost estimates that are lower than those obtained using actual retail trades. In this section, we present evidence that suggests that the approach the wholesaler(s) use to provide execution

[19] These statistics are referenced in the article, "Electronic platforms capture growing share of US equity trading volume," by Alex Pugh. This article was published by Best Execution on the internet on January 16, 2024. The article attributes the statistics regarding the prevalence of algorithmic trading to the Coalition Greenwich report, US Equity Markets 2024: Trends and Opportunities. See: https://www.bestexecution.net/electronic-platforms-capture-growing-share-of-us-equity-trading-volume/

services to marketable retail order flow imposes a bias as to market conditions in which actual retail trades fail to be identified as retail by the BJZZ algorithm. This bias offers one explanation as to why the sample of BJZZ-identified retail trades produced lower cost estimates than actual retail trades.

A. Hypothesis.

To understand how the methodology used by wholesalers to execute marketable retail orders could impose a bias in the set of trades the BJZZ algorithm identifies as retail, consider the following example. A wholesaler will first quickly execute (at least partially) any order that can fill at the midpoint of the ORT NBBO (the benchmark used to evaluate execution quality). Likewise, should the wholesaler fear that quotes are likely to fade (e.g., move against the investor's trading interest), the wholesaler might immediately fill an order at the far touch.[20] These orders will trade at prices that either have a sub-penny increment of 0 or 5 and thus, will not be identified as retail by the BJZZ algorithm. Next, the wholesaler will seek to execute the orders at non-midpoint prices that are both within the ORT NBBO and allow the wholesaler to earn a normal return on capital. These trades occur a few additional milliseconds after orders are received by the wholesaler compared to those filled immediately because of midpoint liquidity availability or fear of quote fading. This second group of orders is likely to have sub-penny trade prices in increments that the BJZZ methodology classifies as retail. Finally, if market conditions dictate an additional delay in execution, the wholesaler will fill remaining shares at the far touch, again at a zero sub-penny increment and not classified as retail by the BJZZ approach. As a result, the BJZZ methodology potentially excludes retail trades that occur most quickly, includes a higher proportion of trades that take slightly longer to execute, and excludes more of the trades that take

[20] For marketable buy (sell) orders, the far touch is the NBO (NBB).

the longest time to fill. Thus, the BJZZ-algorithm might not identify the trades with the shortest time between order arrival and trade execution nor does it identify the trades with the longest time to execution. Between those extremes, it is hypothesized that the BJZZ identification rate is higher.

To the extent that our description of the process used by wholesalers to execute marketable retail order flow is accurate, we posit that the likelihood of a known retail trade being identified by BJZZ as retail is an inverted U-shaped curve plotted with time to execution on the horizontal axis. As the likelihood of a quote changing also is likely to be a function of time passage, the difference between the order-receipt-time and the execution time NBBOs for retail trades that are not identified as retail by the BJZZ algorithm are at one of two extremes. Orders executed immediately have minimal differences in the order receipt time and execution time quote midpoints while orders executed with delay at the far touch have the greatest differences. The price impact of orders executed at sub-penny prices that the BJZZ algorithm uses to identify trades as retail will be between these two extremes. Thus, if the share-weighted price impact of trades executed at the NBBO midpoint and trades executed at the far touch is not equal to share-weighted price impact of trades executed at subpennies (and not at the midpoint), the set of trades identified as retail by the BJZZ algorithm will, on average, have a price impact that differs from the average price impact of all retail orders. In this section, we examine whether the execution process used by wholesalers could at least partially explain why the Commission's use of BJZZ-identified trades to estimate the potential costs associated with its recently proposed OCR results in cost estimates that are lower than those obtained using actual retail trades.

B. Descriptive Statistics.

For this analysis, we begin with the data utilized in Section I, which are the same data used in Battalio and Jennings (2023a, 2023b). We use the BJZZ algorithm to partition our dataset of retail trades into two subsamples: a sample of 13,754,704 BJZZ-identified retail trades and a sample of 35,672,623 BJZZ-unidentified retail trades. In Table 9 we provide some descriptive statistics for each sample of trades.

[Insert Table 9 about here.]

As seen in Panel A of Table 9, while the average trade size is roughly the same for each set of trades, the average size of an order generating a BJZZ-unidentified retail trade is around six times larger than the average size of a sample order that generates a BJZZ-identified retail trade (4,338 shares versus 710 shares).[21] This is consistent with the notion that larger order sizes are more likely to exhaust liquidity at the NBBO and subsequently walk up or down the order book and trade at decimal prices. Panel A also suggests that the algorithm is not biased with respect to identifying retail short sales or order side.

Panel B of Table 9 illustrates that, although BJZZ is slightly more likely to identify trades early and late in the trading day, there are not large differences in the distribution of BJZZ-identified and BJZZ-unidentified retail trades throughout the trading day. Panel C reiterates that the success rate of the BJZZ algorithm in identifying sample retail trades is highest at the beginning and end of the day. Panel D presents the algorithm's success rate in identifying actual retail trades by order size. Consistent with the evidence in Panel A, the algorithm is most successful in identifying retail trades generated from smaller orders as retail. For example, BJZZ correctly identify 30.45% of the trades generated by retail odd lot orders as retail but it correctly identifies only 11.60% of the trades generated by retail orders seeking to trade 5,000 or more shares as retail. Evidence presented in Panel E suggests that this relationship does not carry over to trade size. The algorithm has a 29.38% success rate in identifying odd lot retail trades as retail versus a 39.62%

[21] For orders executed with a single trade, the average order sizes for BJZZ-identified and for BJZZ-unidentified trades are respectively 244 shares and 237 shares.

success rate in identifying retail trades of 5,000 or more shares as retail. Finally, suggesting that it is easier to provide price improvement to high-priced stocks with wider bid-ask spreads, Panel F notes the BJZZ-algorithm performs best for stocks trading at prices of $500 or higher.

In Table 10, we provide some execution quality statistics of retail trades arranged by deciles of the time elapsing between order arrival and trade execution for our sample of retail trades.

[Insert Table 10 about here.]

Panel A provides trade characteristics and ORT market conditions in each decile. Almost all trades happen quickly. The 90th percentile time to trade is 0.02647 seconds. Trade size is largest in the slowest-to-trade decile, consistent with the notion that these are the larger, more difficult to fill orders. Deciles 1, 8, and 9 have the second largest trade sizes. Deciles 2 through 7 have similar mean and median trade sizes. The median trade is an odd lot in deciles 1 through 7. Based on the difference in mean and median trade price, the different time-to-trade deciles appear to contain different securities. Trade price increases through decile 7 and then falls such that deciles 9 and 10 have the lowest mean and median priced securities. Consistent with rising mean price levels, mean dollar quoted spreads generally increase through decile 6 and then fall (median quoted spreads are relatively unchanged across deciles at one or two cents). Consistent with the median trade being an odd lot, the mean potential price impact is negative (indicating the existence of displayed odd lot liquidity supplying orders bettering the official NBBO quoted prices) through decile 7.[22] Consistent with the fact that the median trade size in deciles 8, 9 and 10 is a round lot (e.g., between

[22] Recall that we define potential price impact as the difference between the weighted-average execution price of an order that fills against displayed liquidity at the top and depth of a consolidated limit order book across all markets and the current NBO (for buy orders) or NBB (for sell orders. Thus, if a 200 share order to buy arrives when there are 100 offered at $20.00 and another 100 shares offered at $20.01, the potential price impact is $0.50; the average share-weighted trade price of $20.015 minus the current NBO of $20.00.

two to five times larger than the median trade size in the other deciles), the potential price impact becomes positive and is substantial in decile 10 (about one percent of the mean trade price).

Panel B provides execution quality statistics of the trades in each time-to-trade decile. The trade-weighted price improvement rate increases through decile 6 and then falls dramatically in deciles 9 and 10. Likewise, effective and realized spreads increase throughout deciles 2 to 6, fall in deciles 7 to 9, and rise again in decile 10. With the exceptions of the extreme deciles, the likelihood of an adverse price move in the 100 milliseconds after order arrival increases monotonically. The average adverse quote movement is largest in deciles 1 and 10. Overall, the correlations between time-to-trade and execution quality statistics are consistent with more difficult retail orders taking longer to trade – less price improvement, higher effective and realized spreads, and more frequent and larger adverse quote moves.

Finally, as hypothesized, the fraction of the trades printed on a sub-penny interval other than a half-penny are lowest in the extreme deciles of time to trade. Deciles 6, and 9 have the highest fractions of BJZZ-identified retail trades (recall all of our trades are presumably retail). Decile 10 trades, the trades taking the longest time to trade and typically exhibiting the lowest quality execution statistics are the most underrepresented in the sample of BJZZ-identified trades. Conversely, decile 9 trades, which are overrepresented in the BJZZ-identified sample, have low effective and realized spreads and (mean) adverse price moves. The relevant question is whether these biases in the likelihood of a known retail trade being identified as retail by the BJZZ methodology result in a bias in important execution quality statistics. Does under-weighting the decile 1 trades with average execution quality and decile 10 trades with poor execution quality result in stark differences between BJZZ-identified and BJZZ-unidentified subsamples?

[Insert Table 11 about here.]

In Table 11, we provide execution quality statistics conditional on whether a given retail trade was identified by BJZZ as a retail trade. Panel A describes the order/trade characteristics and market conditions for the two samples of retail trades. BJZZ-identified orders tend to be very slightly larger, are more likely to be marketable limit orders, and are somewhat less likely to be fully internalized by the wholesaler(s). Median trade sizes and prices are similar between the two samples (although the mean trade size is smaller and mean trade price larger for the identified sample). Median ORT quoted spreads are identical (to four decimal points), although the mean quoted spread for the identified sample is somewhat wider. On average, BJZZ-identified orders require considerably less time to print a trade (although median times-to-trade do not differ much).

Panel B provides trade-weighted execution quality statistics. The analysis is at the trade level, with each trade being equally weighted. In each cell, we report mean/median (means only for binary variables). A few statistics require elaboration. The price improvement rate is the percent of trades with an execution price better than the relevant ORT quoted price (NBB for sell orders and NBO for buys). The realized spread is taken at the 100 millisecond level (the SEC's OCR envisions auctions that last between 100 and 300 milliseconds) but are not remarkably different from those computed with a 300 millisecond lag. Potential Price Impact is the difference between the order-receipt time far touch price (NBB for sell orders and NBO for buy orders) and the price that the order would walk (up for buy orders or down for sell orders) a consolidated limit order book across all exchanges including displayed odd lot orders in exchange data feeds. Thus, because odd lots are excluded in the official NBBO, it is possible to have executions better than the NBBO and negative Potential Price Impact. An adverse quote movement is a movement in the quote over a 100 millisecond period post order receipt time that moves against the investor. Thus, an adverse quote movement is a decrease (increase) in the NBB (NBO) for a sell (buy) order. % with Adverse Quote Moves is the fraction of trades that execute when an adverse quote movement occurs.[23]

Panel B of Table 11 indicates that 97% of the BJZZ-identified retail trades are price improved while only 66% of the BJZZ-unidentified retail trades receive price improvement.[24] The differences in price improvement statistics for BJZZ-identified and BJZZ-unidentified are a byproduct of the methodology used by the algorithm to identify retail trades. For example, the median amounts of price improvement for BJZZ-identified and BJZZ-unidentified retail trades are 0.0001 and $0.0050 respectively.[25] Consistent with the assertion that BJZZ-identified and BJZZ-unidentified retail trades are executed in different market conditions and with different amounts of price improvement, the mean effective spread, the average realized spread, the average potential price impact, and the average adverse price impact are each larger for BJZZ-identified retail trades than for BJZZ-unidentified trades. Additionally, 18.47% of the BJZZ-unidentified retail trades experience an adverse price impact, i.e., for buy (sell) orders the offer increases (bid decreases) in the subsequent 100 milliseconds post trade, versus 15.9% for BJZZ-identified retail trades.[26]

Overall, we find that the association between time to execution and the likelihood of a trade being identified as retail by the BJZZ methodology appear to produce a non-random sample of retail trades for studies of execution quality and short-term price impact. BJZZ-identified retail trades have wider quoted, effective, and realized spreads than (known) retail trades that are not

[23] Again, we report statistics only for quote moves in the 100 milliseconds following a trade. The equivalent numbers using a 300 millisecond time window are not substantially different – a slightly higher % with Adverse Quote Moves and the amount of the average adverse quote move. We report the 100 millisecond numbers to be conservative.
[24] The BJZZ-identified retail trades that did not receive price improvement were either price disimproved or were assigned the incorrect sign by the algorithm.
[25] Dollar price improvement is the amount by which the trade price differs from the relevant quoted price (positive numbers indicating price improvement and negative numbers indicating price disimprovement.)
[26] The differences in the percent of trades with adverse quote moves and the average dollar amount of an adverse quote move between the BJZZ-identified and BJZZ-unidentified are consistent with the Battalio and Jennings (2023b) discussion of the SEC's cost analysis of the proposed OCR.

identified as retail. In addition, identified trades have higher price improvement rates, but lower mean price improvement amounts than unidentified trades. Finally, the retail trades that are not identified as such by BJZZ are also less likely to experience adverse quote moves in the 100 milliseconds following a trade than are BJZZ-unidentified retail trades.

IV. Conclusions.

Since the introduction of penny pricing to U.S. stock markets in 2001, researchers seeking to study the behavior of known retail investors have been forced to use proprietary data or data from foreign markets. First posted to SSRN in August 2016, the methodology introduced by Boehmer et al. (BJZZ) seemingly makes it possible for individuals to pursue this line of research using the NYSE's publicly available TAQ data base. Blankespoor et al. (2018) was among the first published papers to utilize BJZZ-identified retail TRF trades to investigate retail investing behavior. In their paper, the authors note that "this approach has a low type I error rate but high type II error rate; that is, trades identified as retail are likely correct, but many retail trades are unidentified" and conclude that "while results showing a change in our measure of retail trading can be reliably interpreted as evidence of retail trading, results of no change in retail trading should be interpreted with caution." They were half right – Type II error rates are high but Type I errors also pose a significant problem.

Using a proprietary dataset of 85,000 small marketable orders in a stratified random sample of 128 securities, which the authors placed with retail brokers, Barber et al. (2023) confirm the conjecture that the BJZZ methodology has a high Type II error rate in their sample. Using a proprietary dataset of over 40 million orders resulting in 53 million trades across 6,800 securities that contains all marketable retail orders executed by one or more retail order wholesalers in May 2022, we find the BJZZ algorithm identifies 28% of the wholesaler's retail trades that were reported to a TRF, which is lower than the 35% accuracy rate documented by Barber et al. for their sample of trades. This ignores the fact that some of the retail trades in our dataset are executed on exchanges, which cannot be identified as retail by the BJZZ algorithm. Our multivariate analysis suggests that the frequency of Type II errors is a function of a stock's trading characteristics such as price, volatility, and bid ask spread, implying that any bias associated with Type II errors will vary across stocks and that a random sample of listed securities is not likely to best capture the securities that retail investors trade.

We utilize three proprietary datasets of institutional trades to demonstrate that, inconsistent with the assumptions of BJZZ, institutional orders produce trades at sub-penny prices other than whole- and half-penny that are reported to a TRF. As a result, Type I errors are a potential problem for researchers using the BJZZ algorithm to investigate retail trading. From the data providing wholesaler about 2.4 million institutional trades from an original 7.5 million institutional trades from a sample month are from retail traders according to BJZZ. Using proprietary institutional trading data from an investment bank, we present evidence that institutional trading interest in a stock is associated with increases in a commonly used retail order imbalance measure created using BJZZ-identified TRF retail trades. To better understand how big of a potential problem Type I errors are, we compare BJZZ-identified TRF retail trading volume to actual retail trading volume as reported by wholesalers' 605 execution quality reports across stocks in May 2022. Despite the fact that on average, the BJZZ algorithm only identifies around 30% of retail trades, we find that for several actively traded stocks, the volume of BJZZ-identified TRF retail trades is *greater* than the volume of actual retail trades. BJZZ-identified TRF retail trading volume is more than 30% of actual retail trading volume for 89 of the 100 most actively traded stocks in May 2022. Depending upon the level of institutional trading interest in a given stock, these results suggest that samples of BJZZ-identified TRF trades in randomly selected stocks are likely to contain a significant number of institutional trades. Thus, changes in BJZZ-identified retail trading behavior cannot be reliably attributed to changes in actual retail trading behavior.

We conclude by providing evidence that the process used by wholesalers when executing marketable retail orders imposes a bias as to market conditions in which retail trades are identified as retail by the BJZZ algorithm. We find that BJZZ-identified retail trades receive more frequent price improvement but in lower amounts, have higher effective and realized spreads, and lower potential price impact than BJZZ-unidentified retail trades. As a result, quotes are less likely to fade in the 100 seconds following order receipt for BJZZ-identified retail trades. We believe this provides an explanation for the results of Battalio and Jennings (2023b), who find that fade costs estimated using BJZZ-identified retail trades are, on average, lower than fade costs estimated using BJZZ-unidentified retail trades.

Overall, our results suggest that the prevalence of Type I and Type II errors have the potential to alter the inferences made from empirical analyses that utilize BJZZ-identified TRF retail trades. We believe there is no simple modification that can be made to improve the algorithm's ability to classify retail trades as retail and to avoid classifying institutional trades as retail. One could use the monthly rule 605 data to identify the stocks with the highest percentage of retail trading and use the BJZZ algorithm to investigate retail trading activity in these stocks. While doing this would reduce Type I errors, Type II errors would remain and, as demonstrated in this paper, could potentially bias any inferences made using these data. An alternative approach for researchers is to frame research questions in terms of monthly trading activity and use the monthly rule 605 data that will soon be produced by individual retail brokers to investigate the impact of retail trading activity on markets.[27]

[27] In March 2024 the SEC adopted the Disclosure of Order Execution Information rule, which requires brokers to produce monthly 605 execution quality reports on a stock-by-stock basis. Thus, retail brokers will have to provide details as to the volume of marketable and nonmarketable orders each month.

References

Anand, Amber, Mehrdad Samadi, Jonathan Sokobin, and Kumar Venkataraman, 2021. Institutional order handling and broker-affiliated trading venues, *Review of Financial Studies* 34, 3364-3402.

Barber, Brad, Xing Huang, Phillipe Jorion, Terrance Odean and Christopher Schwartz, 2023. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ, forthcoming in the *Journal of Finance*.

Battalio, Robert, Brian Hatch, and Mehmet Saglam, 2024. The cost of exposing large institutional orders to electronic liquidity providers, *Management Science* 70, 3597-3618.

Battalio, Robert and Robert Jennings, 2023a. Wholesaler execution quality, SSRN working paper number 4304124.

Battalio, Robert and Robert Jennings, 2024. On the potential cost of mandating qualified auctions for marketable retail orders, *Journal of Investing* 33, 69-99.

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.

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, 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.

Dyhrberg, Anne, Andriy Shkilko, and Ingrid Werner, 2023. The retail execution quality landscape, SSRN working paper number 4313095.

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

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

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.

Securities and Exchange Commission Release 34-96495, Order Competition Rule, Proposed on December 14, 2022.

Securities and Exchange Commission Release 34-99679, Disclosure of Order Execution Information, Final Rule issued on March 6, 2024.

Table 1. Descriptive statistics for the sample of retail marketable orders executed by one or more wholesalers in May 2022.

From the 11,901 unique security symbols traded by our data provider(s) for the month of May 2022, we select symbols that have no days without trades. 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 the end of April 2022. Trades at prices below $1.00 are also excluded in the sample. Statistics in the table are equally-weighted across the 6,823 sample stocks. There are 21 trading days in May 2022.

Panel A. Across-stock characterization of the price-level and size of sample trades.

Statistic	Month-Start Price	Round-Lot Trades	Odd-Lot Trades
Mean	$47.71	2,306	4,226
Median	$23.67	341	643
Minimum	$1.01	0	2
Maximum	$4,376.21	757,119	1,077,137

Panel B. Time elapsing between order receipt and trade execution as seen from the wholesaler(s) perspective.

	Milliseconds
10^{th} Percentile	5.46
25^{th} Percentile	5.92
50^{th} Percentile	6.88
75^{th} Percentile	12.99
90^{th} Percentile	26.51
95^{th} Percentile	48.69
99^{th} Percentile	13,182

Table 2. Sub-penny pricing distribution for TAQ-matched proprietary known retail trades by trade side conditional on the width of the order receipt time National Best Bid and Offer.

Sub-penny increments are defined by BJZZ as Z = 100*mod(price, 0.01).

Z	All Spreads		NBBO Width = $0.01	
	Buy Orders (N=27,064,919)	Sell Orders (N=22,362,408)	Buy Orders (N=12,859,851)	Sell Orders (N=10,530,935)
Z = 0.00	20.68%	19.79%	14.29%	15.39%
0 < Z ≤ .1	0.24%	9.47%	0.01%	9.82%
.1 < Z ≤ .2	0.45%	0.59%	0.00%	0.58%
.2 < Z ≤ .3	0.41%	0.46%	0.02%	0.35%
.3 < Z ≤ .4	0.32%	0.63%	0.00%	0.89%
.4 < Z ≤ .5	16.05%	12.46%	21.55%	17.07%
.5 < Z ≤ .6	2.36%	0.83%	3.72%	0.88%
.6 < Z ≤ .7	0.95%	0.23%	1.43%	0.00%
.7 < Z ≤ .8	0.66%	0.27%	0.53%	0.01%
.8 < Z ≤ .9	0.79%	0.33%	0.74%	0.01%
.9 < Z < 1.00	11.84%	0.19%	12.67%	0.03%
0.00 < Z < 0.40	1.42%	11.15%	0.04%	11.64%
0.60 < Z < 1.00	14.25%	1.01%	15.38%	0.05%

Table 3. 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	
BJZZ Inferred Side	Buy	Sell
Buy	51.20%	3.65%
Sell	5.09%	40.06%

Panel B. Overall mix of the actual trade side to BJZZ inferred trade side when order placement time spread is \$0.01.

	Known Order Side	
BJZZ Inferred Side	Buy	Sell
Buy	56.74%	0.18%
Sell	0.15%	42.93%

Table 4. Determinants of the success rate of the BJZZ algorithm in correctly identifying actual retail trades in a given stock-day as retail trades.

For our sample of retail trades in May 2022, we use ordinary least squares to estimate the following equation,

$$\begin{aligned} \textit{Success Rate}_{i,t} &= \alpha + \beta_1 \textit{ Penny Spread}_{i,t} + \beta_2 \textit{ Log(Avg. Trade Price)}_{i,t} \\ &+ \beta_3 \textit{ Log(Avg. Trade Size)}_{i,t} + \beta_4 \textit{ Stock Volatility}_{i,t} + \beta_5 \textit{ Log(Market Cap)}_{i,t} + \varepsilon_{i,t}, \end{aligned}$$

where $\textit{Success Rate}_{i,t}$ is the percentage of retail trades in stock i on day t that are correctly identified as retail by the BJZZ algorithm, $\textit{Penny Spread}_{i,t}$ is an indicator variable taking a value of 1 if all the retail trades in stock i on day t executed when the width of the NBBO was a penny, $\textit{Log(Avg. Trade Price)}_{i,t}$ is the log of the average trade price for retail trades in stock i on day t, $\textit{Log(Avg. Trade Size)}_{i,t}$ is the log of the average trade size for retail trades in stock i on day t, $\textit{Stock Volatility}_{i,t}$ is the standard deviation of stock i's returns over the past 20 trading days prior to day t, and $\textit{Log(Market Cap)}_{i,t}$ is the logarithm of the market capitalization of stock i on day t. We cluster standard errors at the stock and day level.

	Dependent Variable: $\textit{Success Rate}_{i,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Penny Spread	-0.03***					-0.02***
	(0.003)					(0.003)
Log(Avg. Trade Price)		-0.003***				-0.01***
		(0.001)				(0.001)
Log(Avg. Trade Size)			-0.01***			-0.02***
			(0.001)			(0.001)
Stock Volatility				-0.13***		-0.88***
				(0.03)		(0.18)
Log(MktCap)					-0.004***	-0.0000
					(0.001)	(0.001)
Constant	0.33***	0.33***	0.38***	0.33***	0.37***	0.46***
	(0.002)	(0.004)	(0.01)	(0.002)	(0.01)	(0.01)
Observations	142,097	142,097	142,097	141,761	141,672	141,672
Adjusted R^2	0.003	0.005	0.01	0.001	0.002	0.02

Table 5. Frequency with which the BJZZ algorithm falsely identifies non-retail trades as retail trades.

Panel A. Sub-penny pricing distribution for a sample of non-retail trades executed by one or more wholesalers in May 2022 by trade side. The data providing wholesaler(s) eliminated about five million trades that receive whole-penny or half-penny pricing from the dataset. Sub-penny increments are defined by BJZZ as Z = 100*mod(price, .01).

Z	% of Trades		
	Buys (N = 1,130,784)	Sells (N = 1,277,544)	Total (N = 2,408,328)
Z = 0.00	n.a.	n.a.	n.a.
0.00 < Z < 0.40	69.14%	17.94%	41.98%
0.40 ≤ Z ≤ 0.60	14.02%	12.72%	13.33%
0.60 < Z < 1.00	16.85%	69.33%	44.68%

Panel B. Sub-penny pricing distribution for a sample of institutional trades executed away from exchanges between January 2011 and March 2012. This sample includes institutional trades executed by an electronic liquidity provider (ELP) like Getco, Citadel, and Knight Securities and all trades executed in the IB's dark pool. Sub-penny increments are defined by BJZZ as Z = 100*mod(price, .01).

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

Panel C. Sub-penny pricing distribution for a sample of institutional trades executed for the Public Employees' Retirement Association of Colorado in 2016 and 2017. Own dark pool refers to trades executed in the facilitating broker's dark pool. Sub-penny increments are defined by BJZZ as Z = 100*mod(price, .01).

Z	% of Trades		
	Own Dark Pool (N = 70,739)	All Other Dark Pool (N = 292,720)	All ELP Trades (N = 6,203)
Z = 0.00	35.8%	54.3%	50.4%
0.00 < Z < 0.40	7.2%	1.9%	9.0%
0.40 ≤ Z ≤ 0.60	48.8%	42.0%	24.1%
0.60 < Z < 1.00	8.4%	1.9%	17.4%

Table 6. Across-stock volume of actual wholesaler trades and BJZZ-identified retail trades by volume decile in May 2022.

We include 3,704 common stocks with at least 100,000 shares of consolidated volume in May 2022. Actual retail trading volume in a given stock is the volume of the market orders and the marketable limit orders for 100 or more shares executed by eight wholesalers in May 2022. BJZZ-identified retail trading volume is the volume of BJZZ-identified trades of 100 or more shares in the stock in May 2022. Consolidated volume is the total volume of shares traded in the stock in May 2022. Within each decile, the statistics presented below are across-stock averages.

Consolidated Volume Decile	Consolidated Volume (MM)	Ratio of Retail to Consolidated Volume	Ratio of BJZZ-Identified Retail to Consolidated Volume	Ratio of BJZZ-Identified Retail to Actual Retail Volume
1	0.320	26.80%	10.60%	45.47%
2	0.979	17.86%	7.26%	43.73%
3	2.000	15.41%	6.35%	44.75%
4	3.666	13.19%	5.73%	46.73%
5	6.417	9.54%	4.28%	47.57%
6	10.001	8.12%	3.77%	49.46%
7	16.682	7.90%	3.95%	51.12%
8	27.721	8.20%	4.40%	54.14%
9	52.309	7.56%	4.11%	55.64%
10	247.308	11.07%	6.75%	58.91%

Table 7. Determinants of the percentage of an institutional parent order misclassified as retail by the BJZZ algorithm.

For each parent order in our proprietary sample of institutional parent orders executed by a volume-weighted average price algorithm, we use ordinary least squares to estimate the following equation,

$$\textit{\% Parent Misclassified}_i = \alpha + \beta_1 \textit{ Participation Rate}_i + \beta_2 \textit{ Penny Spread}_i + \beta_3 \textit{ Stock Volatility}_i + \beta_4 \textit{ Turnover}_i + \beta_5 \textit{ Log(Market Cap)}_i + \varepsilon_i,$$

where *% of Parent Misclassified*$_i$ is the percentage of the trades of the i^{th} parent order (e.g., child trades) that are identified as retail by the BJZZ algorithm, *Participation Rate*$_i$ is the ratio of the shares in the i^{th} executed parent order to the trading volume of the underlying stock during the parent order *i*'s life, *Penny Spread*$_i$ is an indicator variable that equals one if each of the parent's child trades are executed when the when the width of the NBBO is a penny and equals 0 otherwise, *Stock Volatility*$_i$ is the volatility of the stock price over the life of the i^{th} parent order, *Turnover*$_i$ is the turnover of the underlying stock during the lifetime of the parent order execution, and *Log(Market Cap)*$_i$ is the logarithm of the market capitalization of the underlying stock on the day parent order *i* is executed. We cluster standard errors at the stock-day level.

	Dependent Variable: % of Parent Order Trades Misclassified as Retail				
	(1)	(2)	(3)	(4)	(5)
Participation Rate	-0.54***	-0.52***	-0.53***	-0.53***	-0.45***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Penny Spread		0.05***	0.04***	0.04***	0.04***
		(0.005)	(0.005)	(0.005)	(0.004)
Stock Volatility			-0.52***	-0.51***	-0.44***
			(0.18)	(0.18)	(0.18)
Turnover				-0.0000	0.0002
				(0.0002)	(0.0002)
Log(MktCap)					0.01***
					(0.002)
Constant	0.10***	0.09***	0.10***	0.10***	-0.05
	(0.004)	(0.004)	(0.01)	(0.01)	(0.04)
Observations	20,335	20,335	20,335	20,335	20,335
Adjusted R^2	0.01	0.04	0.04	0.04	0.04

Table 8. Does the presence of an institutional order impact estimates of order imbalances computed following the BJZZ methodology?

We use ordinary least squares to estimate the following equation,

$$Mroibvol_{i,t} = \alpha + \beta_1\ Buy_{i,t} + \beta_2\ Sell_{i,t} + \beta_3\ InstSubPenny_{i,t} + \varepsilon_{i,t},$$

where we compute $Mroibvol_{i,t}$ for stock i on day t is computed as the volume of BJZZ-identified retail buy trades minus the volume of BJZZ-identified retail sell trades normalized by the sum of the BJZZ-identified retail buy and retail sell trades, $Buy_{i,t}$ is an indicator variable that takes the value of 1 if the IB receives one or more institutional buy orders in stock i on day t and zero otherwise, $Sell_{i,t}$ is an indicator variable that takes the value of 1 if the IB receives one or more institutional sell orders in stock i on day t and zero otherwise, and $InstSubPenny_{i,t}$ is the ratio of the signed sub-penny institutional volume (excluding the midpoint range) to total volume during the execution period for stock i on day t. We cluster standard errors at the stock-day level. Our sample consists of S&P500 stocks from January 2011 through March 2012.

	Dependent variable:			
	Mroibvol			
	(1)	(2)	(3)	(4)
Buy	0.01**			0.01**
	(0.003)			(0.003)
Sell		-0.01***		-0.01***
		(0.004)		(0.004)
InstSubPenny			1.92***	1.29***
			(0.47)	(0.48)
Constant	-0.03***	-0.03***	-0.03***	-0.03***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	147,933	147,933	147,933	147,933
Adjusted R^2	0.001	0.001	0.001	0.001

Table 9. Descriptive statistics for BJZZ-identified and BJZZ-unidentified retail trades from our proprietary data consisting of known retail 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. Sub-penny increments are defined by BJZZ as Z = 100*mod(price, .01).

Panel A. Descriptive statistics of BJZZ-identified retail trades and BJZZ-unidentified retail trades. Average order quantity (all) is the trade-weighted average order size for sample trades. So, orders with multiple executions are counted multiple times.

Variable	Trade-Weighted Average	
	BJZZ-Identified (N = 13,754,704)	BJZZ-Unidentified (N = 35,672,623)
Execution Time	12:33:17	12:21:06
Trade Size	256 shares	270 shares
Trade Price	$141.29	$129.83
Order Quantity (single fills)	244 shares	237 shares
Order Quantity (all)	710 shares	4338 shares
Order Side – Percent Buys	56.3%	54.2%
Percent Short Sells	5.81%	5.21%

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

Hour	BJZZ-Identified (N = 13,754,704)	BJZZ-Unidentified (N = 35,672,623)
9:30 to 10:00	14.11%	11.92%
10:00 to 11:00	17.60%	21.88%
11:00 to 12:00	14.00%	16.40%
12:00 to 1:00	11.88%	12.33%
1:00 to 2:00	11.29%	10.94%
2:00 to 3:00	11.81%	11.58%
3:00 to 4:00	19.30%	14.95%

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

Hour	BJZZ Success Rate
9:30 to 10:00	31.35%
10:00 to 11:00	23.67%
11:00 to 12:00	24.77%
12:00 to 1:00	27.09%
1:00 to 2:00	28.47%
2:00 to 3:00	28.23%
3:00 to 4:00	33.24%

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

Order Size (shares)	BJZZ Success Rate
1 to 99	30.45%
100 to 499	29.07%
500 to 999	27.25%
1,000 to 1,999	27.19%
2,000 to 4,999	27.31%
5,000 & higher	11.60%

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

Trade Size (shares)	BJZZ Success Rate
1 to 99	29.38%
100 to 499	25.39%
500 to 999	25.38%
1000 to 1999	24.98%
2000 to 4999	26.29%
5000 & higher	39.62%

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

Execution Price	BJZZ Success Rate
$\geq$ \$1 and < \$10	28.63%
$\geq$ \$10 and < \$50	27.88%
$\geq$ \$50 and < \$100	26.49%
$\geq$ \$100 and < \$250	27.08%
$\geq$ \$250 and < \$500	27.42%
\$500 and higher	33.89%

Table 10. Equal-weighted (by trade) execution quality statistics by time-to-trade deciles.

We place the 49,201,043 retail trades with complete data into deciles based on the time elapsing between the receipt of the order and the execution of the trade. Decile 1 has the trades that execute the quickest. All statistics use a trade as the unit of observation. We report means and medians for non-binary variables and means for binary variables. A trade is price improved if the execution price is better than the relevant NBBO quoted price. Realized spreads are computed at 100 milliseconds after order receipt. Quoted Spread is the width of the National Best Bid and Offer prevailing when the order is received by a wholesaler. Potential Price Impact is the difference between the NBB (NBO) and the price at which the trade would have occurred if it walked down (up) the consolidated limit order book, including displayed odd lot orders, across all exchanges for a sell (buy) order. As is evident in the table, the Potential Price Impact can be negative if there is an odd lot quote that betters the NBB (NBO) for sell (buy) orders. Consistent with the negative potential price impact through decile seven, the median trade size is less than 100 shares. A buy (sell) trade is associated with an adverse quote move if the NBO (NBB) prevailing when the trade is executed is higher (lower) than the NBO (NBB) prevailing when the order was received. The dollar amount of adverse price moves is the change in the relevant side (NBB for sell orders and NBO for buy orders) of the quote over 100 milliseconds.

Panel A. Statistics that are mostly observable at the time an order is executed. As trade size is determined by the passive orders resting on an exchange, which may be displayed or hidden, the trade size for orders routed to exchanges may be different than expected. Similarly, if there is hidden liquidity or the displayed liquidity changes between the time an order is routed to an exchange and the time that it trades, the trade price for orders routed to exchange may be different than expected. In results not reported, we find that the BJZZ-identified subsample's decile breaks are greater than the BJZZ-unidentified subsample for deciles 1 through 7 and less for deciles 8 and 9.

Time-to-Trade Decile	Maximum Time-to-Trade in Decile (Milliseconds)*	Trade Size (shrs.)		Trade Price		Quoted Spread		Potential Price Impact	
		Average	Median	Average	Median	Average	Median	Average	Median
1	5.47	300	15	$98.12	$45.47	$0.1115	$0.0100	-$0.0111	$0.0000
2	5.89	227	17	$97.37	$46.56	$0.1112	$0.0200	-$0.0118	$0.0000
3	6.06	237	19	$104.02	$47.79	$0.1142	$0.0200	-$0.0118	$0.0000
4	6.38	245	20	$117.69	$50.16	$0.1220	$0.0200	-$0.0126	$0.0000
5	6.88	248	20	$149.61	$63.32	$0.1387	$0.0200	-$0.0148	$0.0000
6	7.99	239	25	$211.39	$90.57	$0.1658	$0.0200	-$0.0187	$0.0000
7	10.03	204	36	$226.38	$95.81	$0.1449	$0.0200	-$0.0136	$0.0000
8	19.36	292	100	$166.91	$58.43	$0.1035	$0.0100	$0.0069	$0.0000
9	26.47	297	100	$63.98	$21.83	$0.0422	$0.0100	$0.0376	$0.0010
10	> 26.47	377	100	$91.39	$31.75	$0.0682	$0.0100	$0.9353	$0.0145

Table 10 (continued)

Panel B. Execution quality statistics and the success rate of the BJZZ algorithm.

Time-to-Trade Decile	% of Price Improved Trades	Effective Spread		Realized Spread		% with Adverse Quote Moves	Amount of Adverse Quote Move		% of Retail Trades Identified by BJZZ as Retail Trades
		Average	Median	Average	Median		Average	Median	
1	73.55	$0.0544	$0.0016	$0.0514	$0.0052	6.85	$0.0685	$0.01	23.91
2	82.17	$0.0489	$0.0000	$0.0485	$0.0000	4.55	$0.0333	$0.01	27.62
3	83.68	$0.0501	$0.0000	$0.0497	$0.0000	4.89	$0.0325	$0.01	28.96
4	85.00	$0.0501	$0.0000	$0.0525	$0.0000	5.44	$0.0329	$0.01	30.47
5	86.24	$0.0597	$0.0000	$0.0591	$0.0012	6.67	$0.0341	$0.01	33.55
6	87.71	$0.0713	$0.0000	$0.0703	$0.0072	9.24	$0.0355	$0.01	35.19
7	87.27	$0.0623	$0.0000	$0.0593	$0.0000	13.64	$0.0341	$0.01	33.00
8	81.68	$0.0460	$0.0062	$0.0336	$0.0014	23.42	$0.0414	$0.01	30.72
9	54.72	$0.0283	$0.0100	$0.0004	$0.0000	53.21	$0.0311	$0.01	36.41
10	29.38	$0.0963	$0.0300	$0.0392	$0.0100	49.66	$0.0603	$0.02	17.44

Table 11. Execution quality statistics for BJZZ-identified and BJZZ-unidentified retail trades.

Our sample consists of 49,201,043 trades with complete data. In Panels A and B, statistics are equal weighted by order (% marketable limit orders, % fully internalized, and order size) or equally weighted by trade (all other statistics). In Panels A and B, we report mean and median for continuous variables and means for binary variables. A trade is price improved if the execution price is better than the relevant NBBO quoted price. For BJZZ-identified trades, we report price improvement rates and amounts based on the actual order side in the table and footnote the equivalent numbers using the BJZZ-inferred order side. Realized spreads are computed at 100 milliseconds after order receipt. Quoted Spread is the width of the National Best Bid and Offer prevailing when the order is received by a wholesaler. Potential Price Impact is the difference between the NBB (NBO) and the price at which the trade would have occurred if it walked down (up) the consolidated limit order book, including displayed odd lot orders, across all exchanges for a sell (buy) order. Potential Price Impact can be negative if there is an odd lot quote that betters the NBB (NBO) for sell (buy) orders. A buy (sell) trade is associated with an adverse quote move if the NBO (NBB) prevailing when the trade is executed is higher (lower) than the NBO (NBB) prevailing when the order was received. The dollar amount of adverse price moves is the change in the relevant side (NBB for sell orders and NBO for buy orders) of the quote over 100 milliseconds.

Panel A. Order and trade characteristics.

		BJZZ-Identified Retail Trades N=13,653,983	BJZZ-Unidentified Retail Trades N=35,547,060
% Marketable Limit Orders		43.32%	23.21%
% Fully Internalized		83.47%	87.87%
Order Size	Average	386	377
	Median	27	25
Trade Size	Average	257	270
	Median	40	50
Trade Price	Average	$140.50	$129.70
	Median	$51.58	$54.13
Time-to-Trade	Average	125.72	2,742.59
	Median	7.14	7.98
Quoted Spread	Average	$0.1200	$0.1093
	Median	$0.0200	$0.0200

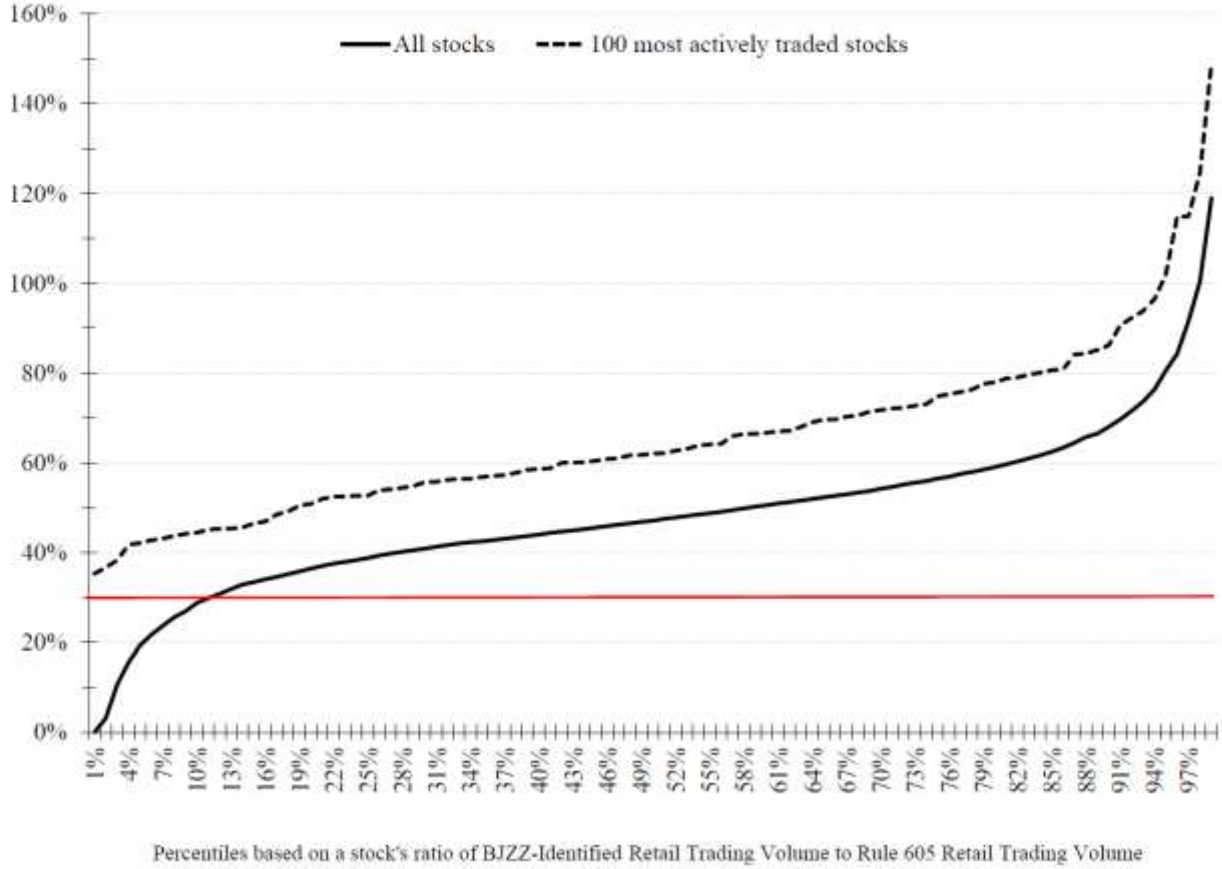
Table 11 (continued)

Panel B. Execution quality statistics.

		BJZZ-Identified Retail Trades	BJZZ-Unidentified Retail Trades
Price Improvement ($)	Average	$0.0177[1]	$0.0337
	Median	$0.0001	$0.0050
Price Improved (%)		97.35%[2]	65.91%
Effective Spread ($)	Average	$0.0964	$0.0419
	Median	$0.0098	$0.0000
Realized Spread ($)	Average	$0.0895	$0.0298
	Median	$0.0098	$0.0000
Potential Price Impact ($)	Average	-$0.0039	$0.1241
	Median	$0.0000	$0.0000
Adverse Price Impact (%)		15.90%	18.47%
Adverse Price Impact ($)	Average	$0.0362	$0.0465
	Median	$0.0100	$0.0100

[1]Using the BJZZ-inferred order side, mean (median) dollar price improvement for the identified subsample is $0.0118 ($0.0001).
[2]Using the BJZZ-inferred order side, 99.19% of trades are price improved.

Figure 1. Ratio of BJZZ-identified retail trading volume to Rule 605 retail trading volume in May 2022.



Appendix Table A1

Papers that utilize BJZZ-identified retail trades in their analysis published in the *Journal of Finance*, the *Review of Financial Studies*, the *Journal of Financial Economics*, the *Journal of Financial and Quantitative Analysis*, *Management Science*, the *Accounting Review*, the *Review of Accounting Studies*, the *Accounting Review*, and the *Journal of Accounting and Economics*.

Paper	**Comments made regarding the methodology.**
Barber et al., 2023a	"Our discussions with SEC staff, FINRA staff, and traders at financial institutions suggest that institutional orders are not typically sold to wholesalers. Institutions do trade off exchanges through crossing networks and alternative trading systems (ATS). However, most of those we spoke with thought that institutional trades are more likely to execute in round pennies." "In summary, an open question is whether using the sub-penny digit in off-exchange trades to identify retail trades leads to a large rate of false positive errors."
Barber et al., 2023b	"Finally, non-retail trades with exchange code "D" may be misidentified as retail. These issues are discussed in detail in Barber et al. (2022)."
Birru et al., 2023	None.
Blankespoor et al., 2019	"These trades do not include nonmarketable limit orders or any orders fulfilled on an exchange, so they have a low Type I error rate but a high Type II rate (i.e., many individuals' trades are unidentified)."
Blankespoor et al., 2018	"This approach has a low type I error rate but high type II error rate; that is, trades identified as retail are likely correct, but many retail trades are unidentified." "Thus, while results showing a change in our measure of retail trading can be reliably interpreted as evidence of retail trading, results of no change in retail trading should be interpreted with caution."
Bonsall IV et al., 2020	None.
Bradley et al., 2023	"This classification is conservative in the sense that it has a low type 1 error (i.e., trades classified as retail are very likely to be retail). However, this classification omits retail trades that occur on exchanges as well as limit orders that are not immediately executable."
Bradley et al., 2022	"This classification is conservative in the sense that is has a low type 1 error (i.e., trades classified as retail are very likely to be retail)."
Bushee et al., 2020	None.
Callan et al., 2023	None.

Appendix Table A1 continued

Paper	Comments made regarding the methodology.
Campbell et al., 2023	None.
Castellani et al., 2024	None.
Chen et al., 2023	None.
Cookson et al., 2023	None.
Dambra et al., 2023	None.
Dottling and Kim, 2022	None.
Drake et al., 2023	"This method identifies retail trades using a regulatory restriction that retail orders can receive price improvements (measured in small fractions of a cent per share) but institutional orders cannot."
Eaton et al. 2022	None.
Farrell et al., 2022	"The BJZZ approach is conservative in the sense that it has a low type 1 error (i.e., trades classified as retail are very likely to be retail). While this approach does omit some retail trading, including nonmarketable limit orders and retail traders that take place on registered exchanges, it "probably picks up a majority of overall retail trading activity.""
Fedyk and Hodson, 2023	"We present the analysis as suggestive, due to the caveat that there is no perfect measure of retail and institutional trades…"
Greenwood et al., 2023	"The Boehmer et al. (2021) algorithm must be treated with a bit of caution because it will identify, for example, trading days with retail buys or sells exceeding the total market capitalization. Where we use the data directly, we winsorize it at the 99% and 1% levels to account for such outliers."
Guest, 2021	None.
Holzman et al., 2023	None.
Huang et al., 2021	None.

Appendix Table A1 continued

Paper	**Comments made regarding the methodology.**
Israeli et al., 2022	“Blankespoor et al. (2018) note that this approach offers low type I and high type II errors. In other words, using this approach we are unlikely to misclassify trades as retail but probably are not capturing the full extent of retail trading.
Kim and Kim, 2023	None.
Liaukonyte and Zaldokas, 2022	“According to Boehmer et al. (2021), this approach can identify most overall retail trading activity.”
Moss et al., 2023	“As a result, the Boehmer et al. (2021) approach provides a good way of identifying retail trades (low type I error), but it does so at the cost of leaving many retail trades unidentified because many transactions occur at nonfractional prices (high type II error) (e.g., Blankespoor et al. 2020).”
Zhongjin et al., 2023	“While the BJZZ order imbalance is not a perfect measure of retail or- der flows (e.g., it does not capture limit retail order flows), it has quickly become the standard proxy for retail order flows due to the lack of alter- native measures that cover a broad panel of stocks.” “Furthermore, given that the BJZZ order imbalances are known to be filled by market makers (i.e., the fast arbitrageurs in our model), they are not contaminated by nonmarket makers’ liquidity provision.”

References for Appendix Table A1

Barber, Brad, Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz, 2023. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ, forthcoming in the *Journal of Finance*.

Barber, Brad, Lin Shengle, and Terrance Odean, 2023. Resolving a paradox: Retail trades positively predict returns but are not profitable, *Journal of Financial and Quantitative Analysis*, 1-34.

Birru, Justin, Sinan Gokkaya, Xi Liu, and Rene Stulz, 2023. Are analyst "top picks" informative?, forthcoming in the *Review of Financial Studies*.

Blankespoor, Elizabeth, Ed deHann, John Wertz, and Christina Zhu, 2019. Why do individual investors disregard accounting information? The roles of information awareness and acquisition costs, *Journal of Accounting Research* 57, 53-84.

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.

Bonsall IV, Samuel, Jeremiah Green, and Karl Muller III, 2020. Market uncertainty and the importance of media coverage at earnings announcements, *Journal of Accounting and Economics* 69, 1-22.

Bradley, Daniel, Jan Janousek Jr., Russell Jame, and Zicheng Xiao, 2023. Place your bets? The value of investment research on Reddit's Wallstreetbets, forthcoming in the *Review of Financial Studies*.

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

Bushee, Brian, Matthew Cedergren, and Jeremy Michels, 2020. Does the media help or hurt retail investors during the IPO quite period?, *Journal of Accounting and Economics* 69, 1-18.

Callan, Jeffrey, Ron Kaniel, and Dan Segal, 2023. Filing speed, information leakage, and price formation, *Review of Accounting Studies* 28, 1618-1656.

Campbell, Brett, Michael Drake, Jacob Thornock, and Brady Twedt, 2023. Earnings Virality, *Journal of Accounting and Economics* 75, 1-30.

Castellani, Vincent, Karl Muller III, and K.J. Park, 2024. Investor reaction to SPACs' voluntary disclosure, The *Accounting Review* 99, 105-137.

Chen, Zilin, Zhi Da, Dashan Huang, and Liyao Wang, 2023. Presidential economic approval rating and the cross-section of stock returns, *Journal of Financial Economics* 147, 106-131.

Cookson, J. Anthony, Joseph Engelberg, and William Mullins, 2023. Echo chambers, *Review of Financial Studies* 36, 450-500.

Dambra, Michael, Omri Even-Tov, and Kimberlyn Munevar, 2023. Are SPAC revenue forecasts informative? The *Accounting Review* 98, 121-152.

Dottling, Robin and Sehoon Kim, 2022. Sustainability preferences under stress: Evidence from COVID-19, *Journal of Financial and Quantitative Analysis*, 1-39.

Drake, Michael, James Moon Jr., Brady Twedt, and James Warren, 2023. Social media analysts and sell-side analyst research, *Review of Accounting Studies* 28, 385-420.

Eaton, Gregory, T. Clifton Green, Brian Roseman, and Yanbin Wu, 2022. Retail trader sophistication and stock market quality: Evidence from brokerage outages, *Journal of Financial Economics* 146, 502-528.

Farrell, Michael, T. Clifton Green, Russell Jame, and Stanimir Markov, 2022. The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145, 616-641.

Fedyk, Anastassia and James Hodson, 2023. When can the market identify old news?, *Journal of Financial Economics* 149, 92-113.

Greenwood, Robin, Toomas Laarits, and Jeffrey Wurgler, 2023. Stock market stimulus, *Review of Financial Studies* 36, 4082-4112.

Guest, Nicholas, 2021. The information role of the media in earnings news, *Journal of Accounting Research* 59, 1021-1076.

Holzman, Eric, Brian Miller, and Brady Twedt, 2023. Curbing enthusiasm: Media sentiment and the disciplining role of quarterly earnings announcements, The *Accounting Review* 98, 315-345.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhang, 2021. Innovation and informed trading: Evidence from Industry ETFs, *Review of Financial Studies* 34, 1280-1316.

Israeli, Dorn, Ron Kasznik, and Suhas Sridharan, 2022. Unexpected distractions and investor attention to corporate announcements, *Review of Accounting Studies* 27, 477-518.

Kim, Sehwa and Seil Kim, 2023. Fragmented securities regulation, information-processing costs, and insider trading, forthcoming in *Management Science*.

Liaukonyte, Jura and Alminas Zaldokas, 2022. Background noise? TV advertising affects real-time investor behavior, *Management Science* 68, 2465-2484.

Moss, Austin, James Naughton, and Clare Wang, 2023. The irrelevance of environmental, social, and governance disclosure to retail investors, forthcoming in *Management Science*.

Zhongjin, Lu, Steven Malliaris, and Zhongling Quin, 2023. Heterogenous liquidity providers and night-minus-day return predictability, *Journal of Financial Economics* 148, 175-200.