Assessing the Reliability of Retail Trade Classification: Evidence from the Tick Size Pilot

Wan Soo Choi

University at Buffalo

September 30, 2025

Abstract

This paper uses the SEC's Tick Size Pilot (TSP) to evaluate retail trade classification errors. Group 2, with both quoting and trading restrictions and limited exemptions for sub-tick trading, provides a clean benchmark. By contrast, Group 1 restricted only quoting and the Control imposed no constraints, so retail classifications in both groups include institutional trades mislabeled as retail. This misclassification is most pronounced off-exchange. Excess retail classifications relative to Group 2 are about 10pp for Group 1 under both BJZZ and QMP (FDR $\approx$ 65%). For the Control group, BJZZ shows 12pp excess (FDR $\approx$ 66.3%) versus 7.6pp for QMP (FDR $\approx$ 39%), indicating a clear improvement. On-exchange, by contrast, sub-tick executions occur mainly through Retail Liquidity Programs, yielding precise classifications with negligible mislabeling. Overall, retail classifications are highly reliable on-exchange but substantially overstated off-exchange.

Keywords: Retail Trade Classification, Tick Size Pilot, False Discovery Rate

Email address: `wansooch@buffalo.edu` (Wan Soo Choi)

1. Introduction

Accurately measuring retail trading is central to understanding investor behavior and the functioning of equity markets. While there is mixed evidence on whether retail investors are informed,[1] these differences partly reflect the underlying data. Early studies relied on proprietary brokerage records, typically covering one or a few discount brokers, while subsequent research broadened the scope by using exchange audit-trail data with trader identifiers or wholesaler datasets. For example, Kaniel et al. (2008), Boehmer et al. (2008), and Kaniel et al. (2012) analyze NYSE account-type data; Kelley and Tetlock (2013) study a large U.S. wholesaler dataset; and Fong et al. (2014) examine the Australian Securities Exchange.

Despite these insights, brokerage and account-type datasets are rarely accessible and often capture only a slice of total retail activity. To address coverage and access, Boehmer, Jones, Zhang, and Zhang (2021) propose a rule based on sub-penny price improvements in off-exchange trades to identify retail activity. Applied to consolidated trades-and-quotes (TAQ) data, this rule enables researchers to estimate the retail sample at scale, and its influence is so substantial that Wharton Research Data Services (WRDS) incorporates it to provide daily retail-trading measures. However, its growing influence has spurred discussion about its scope and accuracy. For instance, Bradley, Jame, and Williams (2022) characterize the BJZZ measure as conservative—with low Type I error for identified trades—while noting it omits retail executions that occur on exchanges. Complementing this, Anand, Irvine, Puckett, and Venkataraman (2021) use FINRA OATS to show that 26.18% of trades generated by institutional parent orders in October 2016 executed in ATSs, SDPs, or by wholesalers (and thus reported to TRFs); when such prints occur at non-midpoint subpennies, they can be mechanically swept into BJZZ's retail category.

Recent work directly evaluates measurement error in retail classifications using either controlled experiments or proprietary datasets. Barber, Huang, Jorion, Odean, and Schwarz

[1] Early brokerage-account studies suggest that individuals trade excessively, incur high transaction costs, underperform benchmarks, and are often guided by attention rather than fundamentals (Odean, 1998; Barber and Odean, 2000, 2008), whereas other work finds that retail order flow can forecast returns and contribute to price discovery (Barrot, Kaniel, and Sraer, 2016; Kaniel, Saar, and Titman, 2008; Boehmer, Jones, and Zhang, 2008; Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013; Fong, Gallagher, and Lee, 2014).

(2024) conduct a large field experiment—placing over 85,000 orders across six retail brokers—and find that the BJZZ algorithm identifies about 35% of their actual retail trades and signs roughly 72% of those correctly (evidence primarily on Type II error). Battalio, Jennings, Saglam, and Wu (2024) use wholesaler, investment-bank, and pension-fund order and execution records to document frequent false negatives (Type II error) and false positives (Type I error) when institutional trades transact at non-midpoint subpennies and are reported to TRFs.

Recent papers evaluating retail classifications tend to emphasize Type I and Type II errors. While these measures are important, additional metrics such as the False Discovery Rate (FDR) (the proportion of institutional trades within the set of trades classified as retail) or the Precision of the test (the share of correctly classified retail trades among retail-classified trades) can also provide useful perspectives, particularly when the goal is to build and assess a classifier.[2] The distinction is subtle but important: Type I error, in the retail classification setting, measures the fraction of true institutional trades that are mistakenly classified as retail (false positives relative to all institutional trades). Type I error is typically expressed as a *rate*—the false positive rate ($1-$ specificity). By contrast, the False Discovery Rate (FDR) captures the proportion of institutional trades within the set of trades classified as retail (false positives relative to all predicted positives), making it the complement of precision ($1-$ precision). Table A.1 illustrates these relationships by mapping actual and predicted classes to the corresponding error types and performance metrics. Incorporating complementary metrics such as FDR or precision provides a fuller assessment of BJZZ's performance. What matters is not only the overall rate at which institutional trades are misclassified (the Type I error rate), but also their prevalence within the set of trades classified as retail. The presence of institutional trades in this predicted retail sample directly increases the FDR (or equivalently lowers precision), thereby reducing its reliability. These measures are particularly important when the classifier's outputs feed

[2]The emphasis on FDR has strong precedent in statistics and machine learning. The idea of controlling the proportion of false discoveries was first introduced by Soric (1989) and later formalized as the False Discovery Rate by Benjamini and Hochberg (1995). Subsequent works have extended this perspective beyond multiple testing; for example, Scott, Bellala, and Willett (2009) demonstrated its usefulness in classification, where FDR provides a natural way to assess the reliability of predicted positives.

into subsequent analyses of market quality. While Type I and Type II errors, defined relative to the true classes, capture overall misclassification, FDR and precision directly evaluate the trustworthiness of the estimated retail sample itself.

There has also been effort to adapt or extend retail classification methods to the Tick Size Pilot. Barardehi, Bernhardt, Da, and Warachka (2022) show that the algorithm, though originally designed for sub-penny executions in a one-cent tick environment, can be scaled to detect sub-tick executions under the Pilot's five-cent regime. Although their analysis applies BJZZ in this setting, their focus is not on retail classification itself but rather on how wholesaler internalization decisions shape retail order flow imbalances. Thus, while theirs is an early trial of retail classification during the TSP, my study is the first to use the Pilot Program to evaluate the classification errors of retail classification algorithms.

In parallel, recent work develops methods to classify on-exchange retail activity. Chung, Rösch, and Zhang (2025) use the NYSE's Retail Liquidity Program (RLP) to identify retail trades on exchange via the RPI flag, complementing BJZZ's off-exchange focus. Their analysis shows that on-exchange retail trades receive larger price improvements and narrower effective spreads than both off-exchange retail and non-retail trades, highlighting execution-quality differences across venues. While their contribution is primarily to measure execution quality rather than classification error, their method for identifying on-exchange retail trades provides a promising avenue for future research.

This paper provides a public-data assessment of retail classification errors by exploiting the SEC's Tick Size Pilot (TSP). Implemented in 2016, the TSP offers a natural setting to evaluate the reliability of retail classifications. The pilot imposed asymmetric rules across small-cap stocks: five-cent quoting only (Group 1); five-cent quoting and trading with explicit exemptions for midpoint executions, retail price improvement of at least \$0.005, and negotiated trades (Group 2); the Trade-At Rule layered on top of Group 2 (Group 3); and an unconstrained Control. These structural differences create a setting where sub-tick executions reflect a mixture of retail and institutional trades in the Control group and Group 1 off-exchange, but map cleanly to retail activity in Group 2. Exploiting this contrast, the analysis uses a difference-in-differences design for identification, a dynamic DiD specification to assess parallel trends, and sale-condition filters to refine the trade sample, thereby yielding a measure of the false discovery rate.

From FINRA's pilot lists, I obtain 1,912 eligible securities. To construct this sample, I exclude temporary test tickers, issues without a `permno`, non–common shares, group switchers, and ticker changes. The analysis uses TAQ trades and quotes beginning on November 1, 2016, after the pilot rules were fully phased in, to study the post-TSP environment. The pre-TSP period is May 1 to September 1, 2016. I match trades to quotes under standard quality filters (Holden and Jacobsen, 2014) to construct consistent measures of spreads, price improvement, and retail classifications.

Retail classifications are aligned with the pilot's tick-size regime. For Control and Group 1, I apply the Boehmer et al. (2021) sub-penny rule in combination with the quote-midpoint rule of Barber et al. (2024). For Groups 2 and 3, where trading is constrained to five-cent increments, I construct a tick-symmetric cent mapping that captures deviations from the \$0.05 grid, similar to the approach of Barardehi et al. (2022). I then additionally enforce the minimum retail price improvement requirement ($\geq$\$0.005) before applying the BJZZ and quote-midpoint logic to classify sub-tick retail executions. To further refine the measure, I exclude trades carrying sale-condition codes that signal institutional activity: `.F` for intermarket sweep orders (ISOs) and `.B`/`.W` for weighted-average price transactions (with `.B` used in CTA and `.W` in UTP).

Descriptive patterns show why the pilot program's design matters. Table 1 illustrates how price positions differ sharply between on- and off-exchange trading. On exchanges, activity is overwhelmingly concentrated at round ticks and at the midpoint: for example, in Group 2 more than 93% of trades occur at whole-5 cent prices and another 6.8% exactly at the midpoint. Sub-tick executions inside the grid are extremely rare, and when they do appear they almost never carry institutional sale-condition flags. This pattern is consistent with sub-penny executions on exchanges arising through Retail Liquidity Programs, which (via Retail Price Improvement orders and related Rule 612 exemptions) are specifically targeted at retail orders.

Off-exchange distributions look different. In Control stocks, only about 62% of trades occur at whole pennies, 17% at midpoints, and the remaining 20% fall into price-improvement bins. Within these bins, a substantial fraction of trades—often 10–20% depending on the interval—are flagged with institutional-like conditions (`.F`, `.W`/`.B`). Similar patterns appear in Groups 1–3, where much of the sub-tick activity occurs and about 10–20% of these trades carry institutional-like flags. This contrast between on- and off-exchange price distributions highlights why the pilot's design matters: on exchanges, sub-tick executions map cleanly to retail participation, while off-exchange they are readily mixed with institutional activity—especially in the Control group and Group 1, which lack trading restrictions or retail-specific exemptions.

Difference-in-differences estimates reinforce this contrast. Group 2, which restricts both quoting and trading to the five-cent grid while allowing explicit exemptions for retail orders, serves as the cleanest benchmark for measuring retail activity. Group 1 and the Control group instead reflect the "usual" trading environment—Group 1 with five-cent quotes but no constraint in trading, and Control with no constraints at all. Against the benchmark Group 2, measured off-exchange retail activity rises sharply more in Group 1 once the pilot begins. Under BJZZ, the difference-in-differences estimates show that off-exchange retail share in Group 1 increases by 10.2 percentage points relative to Group 2 at the daily level ($t = 21.7$), while the number of trades rises by about 116% (0.77 in logs, $t = 17.4$). Excluding institutional sale-condition codes (`F`, `B`, `W`) narrows the share gap slightly to 9.9 points ($t = 21.0$) but strengthens the count effect to 0.82 ($t = 17.9$), or roughly 127%. The QMP measure yields nearly identical magnitudes: retail share rises by 9.4 points ($t = 19.8$) and log counts by 0.63 ($t = 14.9$). Hourly specifications produce very similar results, with BJZZ showing a 9.5 percentage point increase in retail share ($t = 23.0$) and a 0.79 rise in log counts ($t = 18.3$), while QMP delivers an 8.8 point increase ($t = 20.8$) and a 0.65 log-count effect ($t = 15.4$). Taken together, these DiD estimates indicate that, relative to Group 2, measured off-exchange retail activity in Group 1 expands sharply under both classification methods and across time aggregations.

Interpreting these magnitudes in terms of classification accuracy, the roughly 10 percentage point gap in off-exchange retail classifications between Group 1 and Group 2 can be viewed as false positives. Relative to the overall off-exchange retail-classified share of Group 1 (13.68%), this implies a False Discovery Rate of $10.02/13.68 \approx 74.6\%$. In other words, nearly three-quarters of trades classified as retail in Group 1 are likely institutional.

While this calculation provides only an estimate of the FDR, it aligns closely with direct evidence on high Type I error rates reported by Battalio et al. (2024), who find that 86.7% of wholesaler institutional trades were misclassified as retail, and that 33.2% of ATS trades and 78.3% of ELP trades were misclassified in an investment bank dataset (Table 5 of their paper).

In addition to the Group 1–Group 2 comparison, the Control–Group 2 contrast reveals a similar pattern. To ensure consistency, I aligned Control with Group 1 off-exchange, so that both serve as comparable baselines relative to the Group 2 benchmark. In each case, the excess retail classifications can be interpreted as false discoveries, where institutional trades enter the retail bins and lower the reliability of the predicted retail sample. Results remain highly stable across both daily and hourly aggregation, with hourly specifications delivering similar effect sizes and higher t-statistics.

By contrast, no systematic differences appear on exchanges. Group 2 again serves as the benchmark for retail activity, since its sub-tick executions are tied to explicit exemptions under Rule 612. Group 1 and the Control group represent the "usual" trading environment, but on exchanges their classifications behave almost identically to Group 2 despite the differences in quoting and trading rules. In Group 1 versus Group 2 regressions, coefficients on retail share are statistically indistinguishable from zero, with estimates effectively at 0 percentage points and t-statistics below 1. Log-retail count coefficients are likewise negligible, around –0.03 with $t \approx -1.30$. Control–Group 2 comparisons yield similarly small effects, with the retail share differences again near zero and log-retail count estimates mildly negative but economically trivial. Dynamic difference-in-differences estimates following Rambachan and Roth (2023) show pre-treatment coefficients fluctuating slightly above zero (less than 5×10^{-4}) but economically negligible, consistent with parallel trends. Post-treatment coefficients remain near zero with wide 95% confidence intervals, indicating no meaningful effect for Group 1 relative to Group 2. The vertical axis is scaled at 10^{-3}, so the estimated differences are effectively indistinguishable from zero (Figure 4).

These findings align with the evidence in Table 1, showing that on-exchange sub-tick executions in the treatment groups occur almost exclusively on venues operating Retail Liquidity Programs (via Rule 612 exemptions), with institutional sale-condition codes largely absent. By contrast, Control-group executions are more dispersed across venues without such facilities, and a nontrivial share carry institutional-like flags. Even so, the difference-in-differences estimates reveal no meaningful gap between Control and Group 2 on exchanges, suggesting that these noisier Control trades do not materially affect retail classifications in the on-exchange setting.

The remainder of the paper is organized as follows. Section 2 describes the Tick Size Pilot Program, outlines the construction of the sample, and explains how trades and quotes are identified. Section 2.3 and Section 2.4 set out the retail trade classification rules, including tick-grid adjustments and exceptions based on sale-condition codes. Section 3 documents descriptive price distributions across groups and exchange types and shows how institutional flags cluster in specific bins. Section 4 examines off-exchange trading. It first exploits the asymmetric tick-size rules between Group 1 and Group 2 to assess the False Discovery Rate in retail classification, then establishes the alignment between the Control group and Group 1 as consistent comparison groups, and finally contrasts the unconstrained Control group with Group 2. Section 5 turns to on-exchange trading, showing how retail classifications behave on venues with RLPs and Rule 612 exemptions. Finally, Section 6 concludes that off-exchange retail classifications are substantially overstated, with high False Discovery Rates reflecting institutional mislabeling, whereas on-exchange retail classifications remain reliable.

2. Data Description

2.1. Tick Size Pilot Program

The U.S. Securities and Exchange Commission (SEC) launched the Tick Size Pilot Program (TSP) on October 3, 2016 to examine the effects of wider quoting increments on market quality in small and illiquid stocks. The program targeted securities trading above $2 with a market capitalization below $3 billion. Using a stratified random sampling procedure, eligible stocks were assigned to one of three test groups (TG) —TG1, TG2, and TG3—each containing approximately 400 stocks, or to a control group of about 1,200 stocks.

The TSP assigned stocks to four groups. The control group was not subject to the pilot and continued to be quoted and traded in $0.01 increments. Test Group 1 was subject to a $0.05 quoting increment but continued to trade at the standard $0.01 price increment.

Test Group 2 required both quoting and trading at \$0.05 minimum increments, but allowed exemptions for midpoint executions (midpoint of the NBBO or the protected best bid-offer), retail investor executions (retail investor orders may receive price improvement of at least \$0.005 better than the best-protected bid or offer), and negotiated trades.[3] Test Group 3 adopted all rules from Test Group 2 and, in addition, implemented a *Trade-At* prohibition. This rule restricted off-exchange trading centers from executing orders at the NBBO unless they were also displaying at that price or qualified under specific exemptions.[4] In practice, unlike in Group 2, wholesalers in Group 3 could not simply match the NBBO; they had to either provide at least \$0.005 of price improvement or route the order to the venue posting the protected quote. This made Group 3 the most restrictive setting in the Pilot, with the goal of reinforcing the role of displayed liquidity.

2.2. TSP Stocks and Identification of Their Trades and Quotes

The sample of stocks in TSP is based on the date March 2, 2017. I begin with a list of 2,236 stocks included in the Tick Size Pilot Program, obtained from the Financial Industry Regulatory Authority website.[5] Following the filtering strategy of Rindi and Werner (2017), I apply several sample restrictions. First, I drop all test stocks—defined as securities with “pilot” or “test” in their names (12 stocks). I then exclude stocks with missing `permno` (35 stocks removed). I also restrict the sample to common stocks, identified as those with a CRSP share code of 10 or 11 (325 stocks removed).

Additionally, I exclude stocks that changed groups during the Tick Size Pilot Program, based on “The Pilot Change List” available from the same source. One of the fields in the change file is the effective date, and it indicates that many stocks in the treatment group did not enter treatment on the program’s official start date (October 3, 2016), but instead transitioned in weekly batches. A majority of the stocks (1,179 stocks) became part of the

[3] Details on how negotiated trades are defined can be found at `https://www.federalregister.gov/documents/2015/05/13/2015-11425/joint-industry-plans-order-approving-the-national-market-system-plan-to-implement-a-tick-size-pilot#p-57`

[4] These exemptions included, among others, block-size trades, midpoint executions, retail investor orders with at least \$0.005 of price improvement, Intermarket Sweep Orders, negotiated trades, and system malfunctions.

[5] `https://www.finra.org/rules-guidance/key-topics/tick-size-pilot-program/data-collection-securities-and-pilot-securities-files`.

treatment group only after October 3, 2016. For example, 187 stocks became effective on October 10, 602 on October 17, 95 on October 24, and 295 on October 31. As a result, I restrict the later TAQ-based analysis to begin on November 1, 2016. Finally, I drop stocks that changed ticker symbols during the pilot period (52 stocks removed).

However, I do not apply several additional filters used in Rindi and Werner (2017)[6], as my objective is to retain trade-level data rather than construct a balanced panel. After applying above filtering, the resulting sample includes 1,912 stocks: 977 in the control group, 318 in Group 1, 314 in Group 2, and 303 in Group 3.

To identify trades and quotes associated with Tick Size Pilot stocks, I match stock tickers from the pilot list to the TAQ data. Following Holden and Jacobsen (2014), I filter quotes by keeping those with positive bid and ask prices and sizes. The ask price must be greater than or equal to the bid. I exclude quotes with non-normal quote conditions.[7] Quotes with a spread above $5 are also dropped. For trades, I keep only those with positive prices. Trades and quotes are merged without any time lag. The sample is restricted to regular trading hours, from 9:30 a.m. to 4:00 p.m.

2.3. Retail Trade Identification Using the Tick Size Pilot Program

To identify retail trades under the Tick Size Pilot Program (TSP), I use different classification rules depending on the treatment group. For the control group and Group 1 stocks, where transaction price follow a $0.01 tick size, I follow the approach in Boehmer et al. (2021) and classify trades based on their position within the penny. For Group 2 and Group 3 stocks, which are quoted and traded in $0.05 increments but allow sub-penny price improvement, I adjust the rule to account for the $0.05 grid and classify trades based on how far they deviate from the nearest tick. The specific methods for each group are described below.

[6]Such as excluding stocks added after program initiation, those removed before the sample period, stock-days with fewer than 20 trades, or stocks for which more than half of daily observations have fewer than 20 trades

[7]Specifically, I exclude conditions A, B, H, O, R, and W in the Daily Trade and Quote (DTAQ) dataset, and conditions 4, 7, 9, 11, 13, 14, 15, 19, 20, 27, and 28 in the Monthly Trade and Quote (MTAQ) dataset.

2.3.1. Retail Classification for Control Group Stocks

For the Control group stocks, retail trades are identified using a rule adapted from Boehmer et al. (2021). Each trade is classified based on its sub-penny position within the \$0.01 tick. Specifically, for stock i at time t, let

$$Z_{it} \equiv 100 \times \text{mod}(P_{it}, 0.01),$$

where P_{it} is the transaction price and $Z_{it} \in [0, 1)$ represents the fraction of a penny associated with that trade.

Following the classification rule in Boehmer et al. (2021), trades with $Z_{it} \in (0, 0.4)$ are labeled as retail sells, while trades with $Z_{it} \in (0.6, 1)$ are classified as retail buys. To be conservative, trades with $Z_{it} = 0$ (i.e., executed at a full penny) and those near the half-penny mark ($0.4 \leq Z_{it} \leq 0.6$) are left unclassified and not assigned to the retail category.

2.3.2. Retail Classification for Treatment Group Stocks

For Group 2 and Group 3 stocks, identifying retail trades requires a different approach because these stocks are quoted and traded in \$0.05 increments and have a different minimum requirement for retail price improvement.

To capture deviations from the \$0.05 tick grid, I first calculate the value of `Cent`, which reflects the residual cent and sub-cent component of a trade price.

Formally, for a trade price P and for stock i at time t, I define:

$$\texttt{Cent}_{it} = \text{round}\left(((\text{round}(P_{it} \bmod 1,\ 6) \times 10) \bmod 1) \times 10,\ 6\right)$$

This expression[8] isolates the second decimal digit and beyond. For example, a price of \$5.10 yields $\texttt{Cent} = 0$, while \$5.2193 returns 1.93—capturing the sub-cent variation relevant to our tick-size-based classification. This is a necessary step to ensure that prices such as \$5.11 and \$5.16 are treated symmetrically in the analysis, as they are equivalent modulo \$0.05 under the tick-size regime.

[8] An alternative expression could be $\texttt{Cent}_{it} = \text{round}\left((P_{it} \bmod 0.01),\ 6\right) \times 100$ which works in most cases but fails when the price ends in a round penny (e.g., for \$12.90, it incorrectly gives 10 instead of 0).

To reflect this equivalence modulo \$0.05, I apply a *tick-symmetric cent mapping* that folds all values of `Cent` onto the $[0, 5)$ interval using the modulo operator. Specifically, for stock i at time t, define the *symmetric cent*, S_{it}, as

$$S_{it} \equiv Cent_{it} \bmod 5,$$

so that $S_{it} \in [0, 5)$ reflects the distance of the transaction price from the nearest \$0.05 tick multiple. For example, if the transaction price is \$1.5385, then $Cent_{it} = 3.85$ and $S_{it} = 3.85$; if the price is \$1.5885, then $Cent_{it} = 8.85$ and $S_{it} = 3.85$.

Similar to the method in Boehmer et al. (2021), I classify trades based on the value of S_{it}, but the definition of retail differs between Group 1 and Groups 2 and 3. For Group 1, where no minimum price improvement threshold was explicitly defined, I classify as retail those trades inside the 5 cent grid but away from the half-tick zone. Specifically, I treat as retail all trades with $S_{it} \in (0, 2) \cup (3, 5)$, while, to be conservative, I exclude the bins immediately surrounding the half-tick boundary ($2.0 \leq S_{it} \leq 3.0$). For Groups 2 and 3, by contrast, the SEC required a minimum price improvement of half a cent for retail trades. Thus, price bins with less than half a cent of improvement are not considered retail. In these groups, retail trades are defined as $S_{it} \in (0.5, 2.0) \cup (3.0, 4.5)$. To be conservative, trades with $S_{it} \in [0, 0.5)$, those near the half-tick boundary ($2.0 \leq S_{it} \leq 3.0$), and those with $S_{it} \in [4.5, 5.0)$ are excluded from the retail category.

2.4. Exceptions in Retail Classification Algorithm

The BJZZ algorithm assumes that institutional order flow—when executed in an Alternative Trading System (ATS) or a Single Dealer Platform (SDP) and reported to a Trade Reporting Facility (TRF)—does not receive non-midpoint sub-penny price improvement. This assumption is convenient, but it may overlook signals embedded in the trade data.

One particularly useful feature is the *Sale Condition* field. As explained in the Consolidated Tape System (CTS) Specification[9], the *Sale Condition* field is a 4-character code in each trade message that describes how the trade was executed. Each position reflects a

[9] See pages 82-87 of `https://www.ctaplan.com/publicdocs/ctaplan/CTS_Pillar_Output_Specification.pdf`

different category, with maximum one code per category; unused positions are left blank. Most of the case, only one or two character codes are observed. The first character (Category 1) indicates the trade's settlement type; the second (Category 2) flags trade-through exemptions or special pricing conditions; the third (Category 3) captures extended hours or sequence-related characteristics; and the fourth (Category 4) provides Self-Regulatory Organization (SRO)-specific trade details.

In addition to the BJZZ and QMP methods, I extend the classification by excluding trades marked with certain sale conditions when identifying retail activity. Specifically, I classify trades flagged as Intermarket Sweep Orders (`.F`) and Weighted Average Price trades (`.W`) as institutional rather than retail, and I report results with and without this filter separately in the later analysis.

2.4.1. Intermarket Sweep Order

In the additional version, I classify all off-exchange trades marked with the `.F` sale condition (indicating Intermarket Sweep Orders) as institutional rather than retail. This classification is supported by both the regulatory structure and empirical evidence on ISO usage patterns.

An Intermarket Sweep Order (ISO) is a special order type defined under Regulation NMS Rule 600(b)(30), designed to enable rapid and aggressive execution across fragmented markets. Under the normal requirements of the Order Protection Rule (Rule 611), a trading venue must check for and route to any better-priced protected quotations elsewhere before executing an incoming order. By contrast, an ISO allows a broker-dealer to execute immediately on a specific trading venue without that venue performing this check. This is because the broker submitting the ISO certifies that it has already routed orders to all other venues displaying superior prices, thereby satisfying Rule 611 on its own. For example, suppose a broker receives a buy order for 1,000 shares at $10.04. The national market shows the best offer at $10.01 on Exchange A (400 shares available) and $10.02 on Exchange B (200 shares available), with the remaining liquidity at $10.04 on Exchange C. Without ISO, if the broker routes this order to Exchange C first, C must check the NBBO and see that better offers exist on A and B. C would either reject the order or route portions to A and B before filling at \$10.04. This sequential process can delay execution and increase the risk that available liquidity is taken by other market participants. With ISO, the broker can simultaneously route to A and B to clear the better-priced quotes and execute the rest on C, ensuring that all available liquidity is accessed in a single action.

ISO usage is overwhelmingly institutional across both on-exchange and off-exchange venues. Empirical research demonstrates that "informed institutions are the main users of ISO trades," Chakravarty, Jain, Upson, and Wood (2012). Sophisticated market participants—such as algorithmic trading desks, high-frequency traders, and large institutional brokers—dominate ISO activity because they possess the advanced infrastructure and regulatory capabilities required for simultaneous multi-venue routing and real-time NBBO compliance. This institutional dominance is particularly pronounced in off-exchange markets, where ISOs are typically used for cross-venue arbitrage, dark pool sweeps, and complex execution strategies requiring professional-grade market access. While retail orders can occasionally become ISOs when routed to public exchanges, off-exchange retail ISOs remain extremely unlikely due to structural factors: retail flow predominantly executes through internalization or Payment for Order Flow arrangements that provide price improvement without requiring multi-venue sweeps.

2.4.2. Weighted Average Price

The weighted-average price modifier in TAQ—coded as `B` for CTA trades and `W` for UTP trades—is a strong indicator of institutional activity. According to FINRA guidance (FAQ §404.1), the weighted-average price modifier must be used when a broker-dealer executes multiple trades to fill a customer order and then sells to the customer at a price equal to the volume-weighted average cost of those trades plus a net difference under a net trading agreement.[10,11] Because the BJZZ algorithm relies heavily on execution price characteristics,

[10] As described in `https://www.finra.org/filing-reporting/market-transparency-reporting/trade-reporting-faq#404`, for example, Member BD1 receives an order from a customer to buy 5,000 shares of ABCD, accumulates the shares through five separate trades, and reports each of these trades to the tape. BD1 then sells the 5,000 shares to the customer at its weighted-average cost with a net difference to reflect the compensation arrangement, and reports this customer-leg trade to the tape with the `.W` modifier.

[11] However, if the broker-dealer acts as a riskless principal, by having no markup or markdown, the customer leg is reported off-tape and without the weighted-average price modifier (FAQ §404.3).

VWAP-based trades—especially those resulting in sub-penny prices—could be misclassified as retail if the sale condition is ignored. The presence of the `.W` (or CTA `B`) flag therefore provides a clear, rule-based signal for filtering out such non-retail trades.

Retail trades can also be executed at a weighted-average price, for example when a customer order is filled in multiple parts and the broker passes the average price to the customer. However, in typical retail handling, this is done in a riskless principal capacity with no markup or markdown, so the customer leg is reported off-tape and without the weighted-average price modifier. Consequently, it is rare for a retail trade to carry the `.W` (or `B`) flag—such as in a negotiated principal transaction—and the flag remains a strong proxy for non-retail executions.

The SEC's Federal Register notice defining negotiated trades under the Tick Size Pilot interprets weighted-average price transactions as a form of *benchmark trade*, and thus part of the "negotiated trade" exemption available to Groups 2 and 3.[12] Since negotiated trades constitute the third explicit exception to the TSP's trade-at and increment rules, excluding trades with the `.W/B` flag when counting retail activity provides a more accurate classification by filtering out institutional benchmark executions.

3. Price Distribution by Tick Grid and Institutional Flags

To examine how trade prices are distributed within the penny or nickel grid, I construct frequency tables by Tick Size Pilot group and exchange type (Table 1). The sample includes all eligible trades tagged with their pilot group (`C`, `G1`, `G2`, `G3`). For each group, I split trades into *On-Exchange* (venues without the "D" off-exchange flag) and *Off-Exchange* (venues with the "D" flag). Within each subset, trades are assigned to bins according to their price location within the quoting increment, following the rules in Section 2.3.

For each bin, I report two counts: the total number of trades and the number carrying a specific set of sale condition codes. As described in Section 2.4, these codes reflect characteristics typical of institutional trades: `F` (Intermarket Sweep Order), `B`, or `W` (Weighted Average Price). The "Percent" column reports the share of trades in the bin relative to

[12] See `https://www.federalregister.gov/documents/2015/05/06/2015-10514/order-approving-nms-plan-to-implement-a-tick-size-pilot-program`, NMS Plan Sections (I)(P), (I)(C).

the total number of trades in the corresponding group–exchange. The "Percent$_{\text{ITF}}$" column reports the share of trades in the bin with one of these institutional flags, expressed as a percentage of the bin's total count.

Table 1 contains four panels. Panel A reports the price distribution for the control group, where the price bin is divided into 0.1-cent units. Panels B–D report the treatment groups—Group 1, Group 2, and Group 3—where the bin width is 0.5 cents, corresponding to one-tenth of the five-cent quoting increment imposed under the Tick Size Pilot.

3.1. Control Group

In the control group, where no Tick Size Pilot (TSP) rules apply, the vast majority of on-exchange trades occur at round pennies (96.43%), with another 3.22% at the midpoint (half-penny). While BJZZ is not designed for on-exchange activity, the distribution of sub-penny executions provides a proxy for price improvement, such as through Retail Liquidity Programs. Excluding the midpoint and adjacent 0.1-cent bins, only 0.3% of trades fall into this category, which serves as the on-exchange analogue to BJZZ's retail measure. Table 2 shows that, under BJZZ, 0.3% of on-exchange trades are classified as retail, while the quote-midpoint rule (QMP) yields a nearly identical estimate of 0.29%. Panel B of the Table 2 applies a stricter definition by excluding trades with institutional-like sale condition codes—most notably intermarket sweep orders (`F`) and weighted average price transactions (`W` or `B`, depending on whether reported under CTA or UTP). Under this definition, the on-exchange retail share falls to 0.19% under both BJZZ and QMP.

Off-exchange in the control group, 61.85% of trades occur at round pennies and 17.38% at the midpoint, with the remainder distributed across price-improved bins. These patterns reflect the prevalence of off-exchange price improvement, often facilitated by wholesalers providing executions at prices better than the NBBO. Aggregating across all bins yields the retail shares in Table 2, where BJZZ classifies 18.11% of off-exchange trades as retail, compared with 16.45% under QMP. Applying the stricter Panel B definition reduces these figures to 16.79% and 15.22%, respectively. Taken together, these results show that meaningful retail price improvement is concentrated off-exchange, while on-exchange price improvement is largely confined to programs such as the Retail Liquidity Program.

Overall, the on-exchange retail share is low across both control and treatment groups, in contrast to the higher prevalence off-exchange. As Ernst, Spatt, and Sun (2024) emphasize, Retail Liquidity Programs (RLPs) have largely failed to attract volume, accounting for well under one percent of total equity trading, mainly due to their lack of transparency, absence of quote protection, and wholesalers retaining the most attractive retail flow. Because RLPs are the main channel for retail flow on exchanges, their limited role leaves only a small share of retail trading on-exchange.

3.2. Group 1

In Group 1, stocks were required to be quoted in five-cent increments but could still trade in one-cent increments. Table 1 shows that, on-exchange, 93.03% of trades occur at the tick size (5 cents) and 6.86% at the midpoint (2.5 cents), with relatively little activity in sub-tick price-improved bins. Compared with the Control group, the share of midpoint trades more than doubled, while sub-tick activity declined. Notably, when price improvement did occur, it was almost always close to the quote— relatively large price improvements are rare once the tick size increases. Table 2 further shows that only 0.11% or 0.33% of on-exchange trades are classified as retail under BJZZ or QMP respectively. Applying the stricter definition in Panel B, which excludes institutional-like sale condition codes, leaves the BJZZ retail share unchanged at 0.11%, but QMP retail share slightly shrinks to 0.26%. This stability contrasts with the Control group, where nearly half of sub-penny price-improved trades carried institutional-like flags. For Group 1, such institutional-like trades are virtually absent, and it is likely that ISOs and weighted-average pricing were largely eliminated once the quoting increment widened. This pattern is consistent across all treatment groups. At the same time, the data highlight the potential role of the Retail Liquidity Program: there remain roughly 15,000 trades with less than one-cent price improvement, almost none of which bear institutional-like sale condition codes.

Off-exchange in Group 1, Table 1 shows a wider distribution than on-exchange: 61.48% of trades occur at the tick size and 23.04% at the midpoint (2.5 cents). The remainder are spread across sub-penny bins, reflecting the continued ability of wholesalers to offer price improvement that, while proportional to the tick size, is larger in absolute terms when quoting is constrained to five cents. Table 2 reports that BJZZ classifies 13.68% of off-exchange trades as retail, compared with 15.27% under QMP. The bins subject to retail classification in Group 1 are given by $S_{it} \in (0, 2.0) \cup (3.0, 5.0)$ where S_{it} is previously defined as the symmetric-cent measure of a trade's price location within the tick grid.[13] Applying the stricter definition in Panel B lowers these shares modestly, to 12.10% under BJZZ and 13.66% under QMP. Taken together, these results indicate that although Group 1's quoting constraints limit displayed liquidity to five-cent increments, off-exchange venues continue to deliver sub-penny price improvement, in contrast to the negligible retail shares observed on-exchange.

3.3. Group 2

Group 2 stocks are subject to both quoting and trading constraints at five-cent increments. The rule permitted three exemptions: (i) midpoint trades at the NBBO, (ii) retail investor orders that received at least \$0.005 of price improvement relative to the best protected quote, and (iii) negotiated trades. As a result, on-exchange trades with less than five cents in price improvement are virtually absent (only a single trade appears in Group 2), in contrast to Group 1 where several thousand such trades occurred. Instead, most sub-tick executions in Group 2 take place either at the midpoint or within bins showing price improvement of at least half a cent but less than one cent. Consistent with the SEC rule, retail trades in Group 2 and 3 are identified as $S_{it} \in (0.5, 2.0) \cup (3.0, 4.5)$ while trades with $S_{it} \in [0, 0.5)$, those near the midpoint boundary ($2.0 \leq S_{it} \leq 3.0$), and those with $S_{it} \in [4.5, 5.0)$ are excluded from the retail category. Among the remaining sub-tick price-improved trades, institutional-like sale condition codes are rare on-exchange, suggesting that these executions are natural candidates to be treated as retail-liquidity program trades. Notably, almost none of these trades carry institutional-like sale condition codes.

93.06% of on-exchange trades occur at the round tick, 6.81% at the midpoint, and 0.13% are executed with price improvement greater than half a cent but less than one cent. Table 2 shows that for Group 2, 0.13% of on-exchange trades are classified as retail by both BJZZ

[13] Unlike Groups 2 and 3, I include trades with price improvement below \$0.005 because Group 1 was not subject to a rule requiring a minimum of \$0.005 price improvement for retail orders.

and QMP. Applying the stricter rule, as in Panel B, does not affect the classification rate of BJZZ, but does mildly affect QMP retail share.

Off-exchange, 66.97% of trades are executed at the round tick, 23.59% at the midpoint, and the share of sub-tick price-improved trades is lower than in Group 1. Trades with less than half-cent price improvement do occur—unlike on-exchange—but more than half of these carry institutional-like sale condition codes, a pattern also observed in Group 3 but not in the control group or Group 1. As a result, BJZZ classifies 6.72% of trades as retail and QMP 8.63%, as reported in Table 2. After removing institutional-like trades, the retail share falls to 5.64% under BJZZ and 7.46% under QMP, as shown in Panel B.

One distinctive feature of Group 2 off-exchange trading is that the rule allowed three exemptions to the five-cent tick size: retail investor orders, midpoint trades, and negotiated trades. The latter two can largely be filtered out—midpoint executions can be easily identified, and negotiated trades can be partially removed by excluding those marked with the weighted-average price flag. Consequently, the remaining sub-tick executions in Group 2 provide a much clearer representation of retail activity, a feature unique to the TSP structure. By contrast, Group 1 had no such exemptions, so the retail classifications there may also capture institutional activity. As will be revisited in Section 4.1, the excess in Group 1 relative to Group 2 therefore provides an estimate of the False Discovery Rate.

3.4. Group 3

Test Group 3 adopted all rules from Test Group 2 and additionally implemented a *Trade-At* prohibition, which restricted trading centers from executing trades at the NBBO unless they were also displaying at that price or provided price improvement. As intended, the share of on-exchange trading is higher in Group 3 compared with the other groups: 83.65% of trades occur on-exchange, versus 77.89% in the control group, 72.36% in Group 1, and 73.00% in Group 2 (these figures are not reported in the tables).

On-exchange in Group 3, similar to Group 2, 92.33% of trades occur at the round tick and 7.55% at the midpoint. Sub-tick price improvement is virtually absent, except for trades with more than half a cent but less than one cent of improvement. Overall, the on-exchange price distributions across the treatment groups are quite similar, with the main exception that Group 1 includes trades with less than \$0.005 of price improvement. Group 3 also shows a slightly higher share of midpoint executions, likely reflecting the effect of the Trade-At rule. Similar to Group 2, most sub-tick price-improved trades fall within the range of greater than half a cent but less than one cent of price improvement. Moreover, none of these trades carry institutional-like sale condition codes. For BJZZ and QMP, approximately 0.12% and 0.33% of the trades are classified as retail (Table 2). Similar to all other treatment groups, applying the stricter rule does not impact these classification.

Off-exchange, only 23.37% of trades occur at the round tick, a much smaller share than in the other groups. This reflects the direct effect of the Trade-At rule, which prevented trading centers from executing at the NBBO. Instead, the proportion of midpoint trades surged to nearly 60%. About 1.5% of trades occurred with less than half-cent price improvement, which is not considered retail, and more than half of these carry institutional-like sale condition codes. Summing the trades with more than half-cent price improvement (except the midpont) yields retail classifications of 12.18% under BJZZ and 16.01% under QMP. Applying the stricter rule reduces these shares to 10.37% and 14.07%, respectively.

Table 1 around here.

Table 2 around here.

4. Asymmetric Tick-Size Rules and Off-Exchange Retail Classification

4.1. Comparison Between Off-Exchange Group 1 vs Group 2

A key structural distinction between Group 1 and Group 2 is how the five-cent rule was implemented. In Group 1 it applied only to quoting, so sub-tick executions could reflect both retail and institutional activity. In Group 2 it extended to both quoting and trading but included three explicit exemptions. Since midpoint and many negotiated trades can be filtered out, the remaining sub-tick executions provide a clearer window into retail activity. For this analysis, I concentrate on off-exchange trading, where the way wholesalers provide sub-tick price improvement interacts directly with these regulatory differences. This structural distinction provides a unique opportunity to compare how BJZZ and QMP classify retail trades: in Group 1, the retail-classified set is mixed with institutional executions, whereas Group 2 offers a sharper benchmark for isolating retail participation. Figure 1 illustrates this idea, similar to the conceptual diagrams used by Aghbabali, Choi, Im, Rösch, and Roy (2025) to explain external validity.

Utilizing this structural distinction between Group 1 and Group 2, I implement a difference-in-differences design to assess how the classification of retail trading changes in Group 1 relative to Group 2 once the Tick Size Pilot (TSP) begins. Formally, the specification is given by

$$\text{RetailClassification}_{s,t} = \beta \cdot \big(\text{OFF\_G1}_{s,t} \times \text{TSP}_{s,t}\big) + \text{FE} + \epsilon_{s,t}, \tag{1}$$

where $\text{RetailClassification}_{s,t}$ denotes the off-exchange retail classification outcome for stock s at time t. $\text{OFF\_G1}_{s,t}$ is an indicator equal to one if stock s belongs to Group 1 and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The coefficient of interest, β, on the interaction term $\text{OFF\_G1}_{s,t} \times \text{TSP}_{s,t}$ captures whether the retail classification of Group 1 stocks shifted differentially relative to Group 2 stocks once the Tick Size Pilot (TSP) began. A simple post-TSP comparison of Groups 1 and 2 would only reflect level differences, which may partly stem from pre-existing gaps or market-wide shocks. By instead comparing the change in retail classification for Group 1 before versus after the TSP to the corresponding change for Group 2, the difference-in-differences approach nets out both baseline differences between groups and common time effects, isolating the incremental impact of the pilot's asymmetric trading rules on measured retail activity. The fixed effects, denoted by FE, absorb stock-by-day variation. In the hourly specification, I additionally include one-hour time interval fixed effects.

To assess the validity of the identifying assumption underlying equation (1), I estimate a dynamic difference-in-differences model following Rambachan and Roth (2023). The specification is

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{OFF\_G1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}, \tag{2}$$

where RetailShare_{it} denotes the share of off-exchange trades classified as retail for stock i on day t, measured using the BJZZ algorithm. The indicator OFF_G1_i equals one if stock i belongs to Group 1 and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. The event-time indicator $\mathbf{1}\{\text{event\_time}_t = k\}$ maps each day t into an event week k relative to the defined start of the post-TSP period, with $k = 0$ denoting the first observed post-treatment week.[14] The fixed effects, FE, include stock and date, and standard errors are clustered by stock and date.

Figure 2 plots the estimated coefficients $\hat{\beta}_k$ along with 95% confidence intervals. Each $\hat{\beta}_k$ measures the difference in retail share between Group 1 and Group 2 in event week k, relative to the omitted reference period $k = -4$.[15] The pre-TSP coefficients are close to zero with 95% confidence intervals including zero, so the null of no difference cannot be rejected, consistent with the parallel trends assumption in equation (1). In contrast, the post-TSP coefficients rise sharply and remain persistently positive, with 95% confidence intervals excluding zero, indicating a stable and statistically significant increase in off-exchange retail classification for Group 1 relative to Group 2. This dynamic pattern supports interpreting β in equation (1) as the causal effect of the asymmetric tick-size rules on retail classification.

Table 3 presents the results from this regression framework. The table is structured to compare two classification approaches—BJZZ Boehmer et al. (2021) (columns 1 to 4) and the quote midpoint rule (QMP) Barber et al. (2024) (columns 5 to 8) —applied at two levels of time aggregation. Panel A reports daily regressions, while Panel B uses hourly data to increase precision. Each panel reports results for four dependent variables. `Share` measures the proportion of off-exchange trading volume classified as retail. `LogCount` records the number of trades classified as retail, expressed as the natural logarithm of one plus this count. The `Excluding ITF` specifications add a filter that excludes trades with institutional-like sale condition codes (`F`, `B`, or `W`) in counting retail trades. For the QMP retail classification method, the algorithm is refined by excluding trades that fall within the 40%–60% range of the NBBO that are close to the midpoint.

Post-TSP, Group 1 exhibits a systematically larger increase in retail-classified trades

[14] Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis.

[15] Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

relative to Group 2. At the daily level (Panel A), Column 1 shows that the post-TSP increase in retail share for Group 1 is 10.2 percentage points larger ($t = 21.67$) than the corresponding change for Group 2. In terms of the number of trades, Column 2 reports a coefficient of 0.768 ($t = 17.38$), which implies that the post-TSP increase in classified trades for Group 1 is $\exp(0.768) - 1 \approx 116\%$ larger than the relative increase for Group 2. When trades with institutional-like sale conditions are excluded from the retail classification (Excluding ITF), the coefficients move in opposite directions. Column 3 shows that the differential increase in retail share drops slightly to 9.9 percentage points ($t = 20.96$). By contrast, Column 4 shows that the differential increase in LogCount rises to 0.819 ($t = 17.86$), corresponding to $\exp(0.819) - 1 \approx 127\%$. This pattern reflects that a nontrivial portion of Group 1's retail-classified trades were associated with institutional-like sale condition codes. Once these trades are filtered out, the retail share gap narrows, but the relative increase in the number of retail-classified trades becomes even more pronounced. Assuming Group 2 provides a more accurate benchmark for retail classification, the observed gap can be interpreted as an estimate of the False Discovery Rate in Group 1, since the retail-classified set there includes institutional executions. The estimate is about 9–10 percentage points, so roughly one in ten trades labeled as retail in Group 1 are actually institutional executions that would not appear in a pure retail benchmark (Group 2). The QMP results in Columns 5–8 of Panel A mirror the same qualitative pattern, though the magnitudes are slightly smaller. The coefficient on retail share is 9.4 percentage points ($t = 19.84$), while the coefficient on LogCount is 0.633 ($t = 14.91$), implying a relative increase of $\exp(0.633) - 1 \approx 88.3\%$. After applying the Excluding ITF filter, the retail share differential falls to 9.2 points ($t = 21.08$), while the LogCount coefficient increases to 0.816 ($t = 17.62$), or about 127%.

Turning to Panel B, which uses hourly rather than daily aggregation, the results remain robust when finer time variation is considered. These hourly models include stock-by-day and one-hour fixed effects, and all standard errors are clustered by stock and date. For BJZZ, the retail share differential is 9.5 percentage points ($t = 22.97$), and the LogCount coefficient is 0.796 ($t = 18.34$), which translates into a relative increase of $\exp(0.796) - 1 \approx 122\%$. With the Excluding ITF filter, the share differential falls slightly to 9.1 percentage points ($t = 22.21$), while the LogCount coefficient increases to 0.852 ($t = 18.81$), or about 134%.

The QMP method again yields slightly smaller magnitudes but an identical qualitative pattern: the retail share differential is 8.8 percentage points (t = 20.81), and the LogCount coefficient is 0.652 (t = 15.41), corresponding to $\exp(0.652) - 1 \approx 91.9\%$. Under the Excluding ITF filter, the share effect narrows to 8.5 percentage points (t = 20.57), while the LogCount coefficient rises to 0.684 ($t = 16.20$), or roughly 98.18%. The consistency of these results across BJZZ and QMP and across both daily and hourly specifications indicates that Group 1's elevated retail classifications are structurally induced. They stem from the absence of retail exemptions in Group 1's tick-size rule, so the retail-classified set includes a substantial share of institutional off-exchange trades, thereby inflating the False Discovery Rate.

For example, consider the roughly 10 percentage point gap in off-exchange retail classifications between Group 1 and Group 2. This gap arises from the structural differences in Tick Size Pilot rules and exemptions, and thus represents excess retail share relative to the clean Group 2 benchmark. Taking this gap as false positives, dividing by the overall off-exchange retail-classified share of Group 1 (13.68%) yields a False Discovery Rate of $10.02/13.68 \approx 73.25\%$. In other words, nearly three-quarters of trades classified as retail in Group 1 are likely institutional, or equivalently, the precision of the classification is about 26.75%. While this is an estimate of the False Discovery Rate, it aligns with the direct evidence on high Type I error rates reported by Battalio et al. (2024). Using proprietary data, they find that 86.7% of wholesaler institutional trades were misclassified as retail, and in an investment bank dataset 33.2% of ATS trades and 78.3% of ELP trades were misclassified (Table 5 of their paper).

Table 3 around here.

Figure 2 around here.

4.2. Off-Exchange Control Group vs. Group 1: Comparison Group Alignment

In Section 4.3, I examine the comparison of Control versus Group 2. Before doing so, it is important to establish that the Control group and Group 1 behave similarly in off-exchange trading. Although both serve as comparison groups, they do so in different ways. Group 1 reflects a "within-pilot" reference: stocks that faced quoting constraints but remained free to trade as usual. The Control group, by contrast, was explicitly designed as the benchmark of the Tick Size Pilot, left unconstrained in both quoting and trading. If the Control group and Group 1 are broadly similar, then differences with Group 2 can be interpreted as the effect of the asymmetric trading rules rather than artifacts of which comparison group is chosen.

To test this, I estimate a difference-in-differences model of the form

$$\text{RetailClassification}_{s,t} = \beta \cdot \left(\text{OFF\_C}_{s,t} \times \text{TSP}_{s,t}\right) + \text{FE} + \epsilon_{s,t}, \tag{3}$$

where $\text{RetailClassification}_{s,t}$ denotes off-exchange retail share outcomes for stock s on day or hour t. The indicator $\text{OFF\_C}_{s,t}$ is an indicator equal to one if stock s belongs to Control Group and trades off-exchange, and zero if it belongs to Group 1 and trades off-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The specification includes stock-by-day fixed effects, and in hourly models also one-hour interval fixed effects, with standard errors clustered by stock and date.

Table A.7 reports the results. Across both BJZZ and QMP classifications, and at both daily and hourly aggregation, the coefficient is small but varies: around 1–2 percentage points for retail share, and negative in trade-count specifications. This difference in sign simply arises from whether retail activity is measured by volume share or by trade frequency, and in both cases the effects are modest relative to the 9–10 percentage point gaps observed in the Group 1–Group 2 comparison.

To assess the identifying assumption, I also estimate a dynamic specification following Rambachan and Roth (2023) as before:

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{OFF\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}. \tag{4}$$

Figure A.1 plots the event-study coefficients. The pre-treatment estimates hover around zero with 95% confidence intervals generally including zero, consistent with parallel trends between Control and Group 1 prior to the pilot. After the TSP begins, the coefficients shift modestly upward, in the range of 2–4 percentage points, and remain persistently positive. Although statistically distinguishable from zero, these effects are small relative to the much larger differences observed in the Group 1–Group 2 comparisons.

Overall, the evidence suggests that Control and Group 1 can be regarded as broadly similar comparison groups. While small differences exist, they are negligible relative to the much larger differences observed once Group 2 is introduced. This alignment supports using Control and Group 1 as reference points when interpreting the sharper contrasts with Group 2.

4.3. Comparison Between Off-Exchange Control Group vs Group 2

Having established that Control and Group 1 behave similarly in off-exchange trading (Section 4.2), I now turn to the comparison of Control versus Group 2.

In addition to the Group 1 versus Group 2 analysis, comparing the Control group with Group 2 adds an additional dimension, showing how unconstrained stocks behave relative to the benchmark. Control stocks faced no change in quoting or trading rules, whereas Group 2 operated under the five-cent constraint with limited exemptions. As discussed earlier, these exemptions ensure that most sub-tick trades in Group 2 map more directly to retail activity. Thus, contrasting Control with Group 2 highlights how the retail classification methods perform in a standard unconstrained setting versus one where regulatory design yields a cleaner benchmark for retail participation. As before, the analysis is restricted to off-exchange trading.

Leveraging the contrast between the unconstrained Control group and the tightly regulated Group 2, I implement a difference-in-differences design to evaluate how retail trade classifications evolve under the Tick Size Pilot, similar to Section 4.1. Formally, the specification is

$$\text{RetailClassification}_{s,t} = \beta \cdot \left(\text{OFF\_C}_{s,t} \times \text{TSP}_{s,t}\right) + \text{FE} + \epsilon_{s,t}, \tag{5}$$

where $\text{RetailClassification}_{s,t}$ denotes the off-exchange retail classification outcome for stock s at time t. The indicator $\text{OFF\_C}_{s,t}$ is an indicator equal to one if stock s belongs to Control Group and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The interaction term $\text{OFF\_C}_{s,t} \times \text{TSP}_{s,t}$ therefore captures the differential shift in retail classifications of Control stocks relative to

Group 2 following the onset of the pilot program. The fixed effects, denoted by FE, account for stock-by-day variation, with the hourly specification further including one-hour interval fixed effects.

To evaluate the identifying assumption in equation (5), I estimate a dynamic difference-in-differences model (Rambachan and Roth, 2023):

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{OFF\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}, \tag{6}$$

where RetailShare_{it} denotes the share of off-exchange trades classified as retail for stock i on day t using the BJZZ algorithm. The indicator OFF_C_i equals one if stock i belongs to Control Group and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. $\mathbf{1}\{\text{event\_time}_t = k\}$ maps each day into an event week k relative to the start of the post-TSP period, with $k = 0$ denoting the first post-treatment week. The specification includes stock and date fixed effects, and standard errors are clustered by stock and date.

Figure 3 plots the coefficients $\hat{\beta}_k$ with 95% confidence intervals. Each coefficient measures the difference in retail share between Control and Group 2 stocks in event week k, relative to the omitted reference week $k = -4$. The pre-treatment coefficients stay close to zero with confidence intervals including zero, consistent with the parallel trends assumption. In contrast, the post-treatment coefficients rise sharply and remain persistently positive, with confidence intervals excluding zero, indicating a stable and statistically significant increase in off-exchange retail classification for Control stocks relative to Group 2 once the TSP began.

Table 4 presents the results of the difference-in-differences regressions comparing the Control group to Group 2. The structure mirrors Table 3: columns 1–4 report results using BJZZ Boehmer et al. (2021), while columns 5–8 use the quote midpoint rule (QMP) Barber et al. (2024). Panel A uses daily aggregation, and Panel B uses hourly aggregation. Within each panel, `Share` measures the fraction of off-exchange trading volume classified as retail, `LogCount` is the log of one plus the retail-classified trade count, and the `COND` specifications exclude trades flagged with institutional-like sale condition codes (`F`, `B`, or `W`).

Post-TSP, Control stocks display a consistently larger increase in retail-classified trades relative to Group 2. In Panel A, BJZZ results show that retail share rises by 12.0 percentage points more for Control than for Group 2 (Column 1, $t = 30.61$). The number of retail-classified trades increases by $\exp(0.507) - 1 \approx 66\%$ more for Control relative to Group 2 (Column 2, $t = 13.15$). Excluding the institutional like flags reduces the share effect slightly to 11.7 percentage points (Column 3, $t = 29.76$), while the relative trade count effect increases to $\exp(0.584) - 1 \approx 79\%$ (Column 4, $t = 14.38$). The QMP method in Columns 5–8 produces qualitatively similar results, though smaller in magnitudes: share differentials of about 7.6 percentage points ($t = 20.21$) and relative trade count increases of $\exp(0.178) - 1 \approx 19.5\%$ (Column 6, $t = 20.21$).

Taken together, the Control–Group 2 results show that, with no constraints, Control stocks record excess retail classifications of about 7–12 percentage points, depending on the classification methods. Because Group 2 provides a clean benchmark with explicit retail exemptions, the 7–12 percentage point excess observed in Control can be interpreted as institutional activity entering the retail-classified set. Put differently, the percentage point gaps reflect false positives as a share of all Control group trades. Since the overall off-exchange retail-classified share in the Control group is 18.11%, taking the ratio $12/18.11 \approx 66.3\%$ for BJZZ and $7/18.11 \approx 39\%$ for QMP shows that a substantial fraction of trades labeled as retail are in fact institutional executions that would not appear under a pure retail benchmark. Although this is an estimate of the False Discovery Rate, the relatively high value is consistent with the evidence on high Type I error reported by Battalio et al. (2024).

Table 4 around here.

Figure 3 around here.

5. On-Exchange Retail Classification under TSP

5.1. On-Exchange Retail Activity: RLPs and RPIs

The Tick Size Pilot (TSP) provides a unique setting to study on-exchange retail executions. While the pilot constrained quoting and trading increments, exemptions for retail orders remained in place. During this period, sub-penny price improvements on exchanges arose through exchange-operated Retail Liquidity Programs (RLPs). Under these programs, approved Retail Member Organizations (RMOs) routed qualified retail orders to interact with hidden Retail Price Improvement (RPI) orders supplied by designated Retail Liquidity

Providers (also abbreviated RLPs). The NYSE family of exchanges (NYSE, NYSE Arca, NYSE American) ran formal Retail Liquidiy Program with dedicated RPI orders, while Nasdaq BX and Cboe BYX implemented analogous arrangements branded as "Retail Price Improvement Programs" (RPI Programs), both operating under limited SEC exemptions from Rule 612(c).[16] Importantly, both BX and BYX operated under an inverted maker–taker fee model, where liquidity takers earned rebates and liquidity providers paid fees. This structure created strong incentives for brokers to route marketable retail orders to these venues, as it generated rebates in addition to price improvement. Accordingly, the sub-penny retail-classified executions observed on BX and BYX reflect both the SEC's sub-penny exemptions and the economic incentives embedded in the inverted fee model.

As shown in Table A.4, on-exchange trades classified as retail are heavily concentrated on venues operating Retail Liquidity Programs, which function under limited Rule 612 exemptions that authorize sub-penny executions for retail orders. This concentration, however, does not extend uniformly to the control group. For instance, Nasdaq Tape C—which terminated its Retail Price Improvement Program in 2015 and did not operate any sub-penny exemption during the TSP[17]—still accounts for the largest single share of retail-classified trades (26.55%) in the post-TSP period. As further evidence, Table 1 shows that, on-exchange, control-group trades carry institutional-like sale condition codes (COND) in roughly half of sub-penny executions across bins, whereas such flags are virtually absent for the treatment groups. This also reinforces the interpretation that sub-tick executions in Groups 1 and 2 are overwhelmingly retail in origin, rather than reflecting institutional trading activity.

[16] For Nasdaq BX, see SEC Release No. 34-84961 (Dec. 20, 2018), https://www.federalregister.gov/documents/2018/12/26/2018-27821/self-regulatory-organizations-nasdaq-bx-inc-order-granting-an-extension-to-limited-exemptions-from. For Cboe BYX, see SEC Release No. 34-81308 (Aug. 9, 2017), https://www.federalregister.gov/documents/2017/08/15/2017-17162/self-regulatory-organization-bats-byx-exchange-inc-order-granting-an-extension-to-limited-exemption or in general, https://www.cboe.com/us/equities/trading/offerings/retail_price_improvement/ and https://www.nasdaqtrader.com/content/BXRPIfs.pdf.

[17] Nasdaq terminated its Retail Price Improvement Program (Rule 4780) in 2015. See SEC Release No. 34-75228, June 17, 2015, https://www.sec.gov/rules/sro/nasdaq/2015/34-75228.pdf.

5.2. Comparison Between On-Exchange Group 1 vs Group 2

While off-exchange Group 1 trades can include institutional activity slipping into the sub-tick category, this does not appear to be the case on-exchange, because executions in both treatment groups occur on venues authorized to provide retail price improvement. The distinction is that Group 2 routes such trades through explicit exemptions tied to retail orders, whereas Group 1 permits sub-tick (and in practice sub-penny) executions to arise under its one-cent trading increment. Any gap that emerges between the two groups on exchange would therefore reflect these structural differences and indicate institutional trades entering the retail-classified set, raising the False Discovery Rate.

To test for differences between Group 1 and Group 2 on-exchange, I estimate a difference-in-differences model of the form

$$\text{RetailClassification}_{s,t} = \beta \cdot \big(\text{ON\_G1}_{s,t} \times \text{TSP}_{s,t}\big) + \text{FE} + \epsilon_{s,t}, \tag{7}$$

where $\text{RetailClassification}_{s,t}$ captures on-exchange retail classifications for stock s at time t, and the indicator ON_G1_i equals one if stock i belongs to Group 1 and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange.

Table 5 reports results using both BJZZ and quote midpoint classifications under daily and hourly aggregation. Across all specifications, the ON_G1 $\times$ TSP interaction is indistinguishable from zero, indicating that in the difference-in-differences setting, Group 1 and Group 2 exhibit no meaningful differences in retail classification outcomes. This contrasts with the off-exchange setting, where institutional trades are likely to be classified within Group 1 sub-tick executions. On exchange, sub-tick executions in both groups are more likely to represent genuine retail trades, with little evidence of institutional activity in the retail-classified set.

To further assess the identifying assumption in equation (7), I estimate a dynamic difference-in-differences model (Rambachan and Roth, 2023):

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_G1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}, \tag{8}$$

where RetailShare_{it} denotes the on-exchange retail share for stock i in event week t, classified using the BJZZ algorithm. The indicator ON_G1_i equals one for Group 1 stocks that trade on-exchange and zero for Group 2 stocks that trade on-exchange. $\mathbf{1}\{\text{event\_time}_t = k\}$ maps each day into an event week k relative to the start of the post-TSP period, with $k = 0$ denoting the first post-treatment week. The specification includes stock and date fixed effects, and standard errors are clustered by stock and date.

Figure 4 plots the estimated coefficients $\hat{\beta}_k$ with 95% confidence intervals. Each coefficient measures the difference in on-exchange retail share between Group 1 and Group 2 in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates fluctuate slightly above zero (less than 5×10^{-4}), but the magnitudes are economically negligible, supporting the parallel trends assumption. Post-treatment coefficients remain near zero with wide confidence intervals, indicating no statistically or economically meaningful effect for Group 1 relative to Group 2. The vertical axis is scaled at 10^{-3}, so the estimated differences are effectively indistinguishable. This pattern is consistent with the difference-in-differences results and suggests that on-exchange, sub-tick executions in Group 1 and Group 2 behave similarly, with little evidence of institutional trades in the retail-classified set.

Table 5 around here.

Figure 4 around here.

5.3. On-Exchange Control Group vs. Group 1: Comparison Group Alignment

Similar to Section 4.2, I begin by examining how the Control group and Group 1 behave on exchange. Both serve as comparison groups but in different ways: Group 1 provides a "within-pilot" reference with quoting constraints but unrestricted trading, while the Control group represents the fully unconstrained benchmark of the pilot. Establishing their similarity ensures that any subsequent differences with Group 2 can be attributed to the trading-rule asymmetries rather than to the choice of comparison group.

To test this, I estimate a difference-in-differences model of the form

$$\text{RetailClassification}_{s,t} = \beta \cdot \left(\text{ON\_C}_{s,t} \times \text{TSP}_{s,t}\right) + \text{FE} + \epsilon_{s,t}, \tag{9}$$

where $\text{RetailClassification}_{s,t}$ denotes on-exchange retail share outcomes for stock s at time

t. The indicator $\text{ON\_C}_{s,t}$ equals one for Control stocks that trade on-exchange and zero for Group 1 stocks on-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The specification includes stock-by-day fixed effects, and in hourly models also one-hour interval fixed effects, with standard errors clustered by stock and date.

Table A.8 reports the on-exchange results. Across both BJZZ and QMP classifications, the C $\times$ TSP coefficients are effectively zero for the share of trades classified as retail and modestly negative for the number of trades. In Panel A (daily aggregation), the estimated differential in retail share is only about 0.2 percentage points and statistically insignificant, while the trade-count coefficients are around –0.065 in logs ($t \approx -3.55$), corresponding to a 6–7% decline in the number of trades classified as retail. Panel B (hourly aggregation) yields nearly identical results: retail share differentials remain close to zero and only marginally significant, while trade-count effects again indicate a 6–7% decline. The similarity of results across BJZZ and QMP (though QMP log count shows stronger difference), and across daily and hourly specifications, shows that Control and Group 1 behave almost identically in their on-exchange retail classifications, with virtually no differences—far stronger similarity than in off-exchange.

To assess further the identifying assumption, I also estimate a dynamic specification following Rambachan and Roth (2023):

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}. \tag{10}$$

Figure A.2 plots the event-study coefficients. The pre-TSP estimates remain close to zero with confidence intervals including zero, consistent with the parallel trends assumption. Post-TSP coefficients drift modestly upward but remain economically small (around 0.2–0.4 percentage points) and statistically indistinguishable from zero, with confidence intervals generally including zero. Overall, the evidence indicates no statistically or economically meaningful change in on-exchange retail classification for Control stocks relative to Group 1 after the pilot, and thus they can be treated as an aligned comparison group.

5.4. Comparison Between On-Exchange Control Group vs Group 2

The on-exchange comparison of Control versus Group 2 parallels the earlier Group 1–Group 2 analysis, with the Control group serving as the comparison group in the same way that Group 1 did previously. As summarized in Table A.4, however, the on-exchange context differs: While treatment-group trades are overwhelmingly concentrated on venues operating Retail Liquidity Programs under limited Rule 612 exemptions, for the Control group a substantial share of on-exchange retail-classified trades occur on venues without such facilities—most notably Nasdaq Tape C, which terminated its Retail Price Improvement Program in 2015. In addition, nearly half of these on-exchange Control-group trades carry sale-condition codes typically associated with institutional activity, making this comparison less clean than within the treatment groups.

To estimate the effect, I use the same difference-in-differences model as before:

$$\text{RetailClassification}_{s,t} = \beta \cdot \big(\text{ON\_C}_{s,t} \times \text{TSP}_{s,t}\big) + \text{FE} + \epsilon_{s,t}, \tag{11}$$

where $\text{RetailClassification}_{s,t}$ denotes on-exchange retail classifications for stock s at time t, and $\text{ON\_C}_{s,t}$ equals one if stock s belongs to Control Group and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. The specification includes stock-by-day fixed effects, and in hourly models also one-hour interval fixed effects, with standard errors clustered by stock and date.

Table 6 reports the results. The ON_C×TSP coefficients are essentially zero in retail share specifications (about 0–0.2 percentage points and statistically insignificant), but consistently negative and highly significant in trade-count specifications, a decline of roughly 9% in the number of trades classified as retail for Control stocks relative to Group 2 under BJZZ, and about a 38% decline under QMP. The pattern is robust across both BJZZ and QMP methods and across daily and hourly aggregation.

In sum, the difference-in-differences results indicate that Control and Group 2 behave broadly similarly on exchange. Despite differences in venue composition and the presence of institutional-like sale-condition codes in Control, the estimates show no statistically or economically meaningful divergence in retail classifications. This suggests that, on exchange, sub-tick executions in both groups primarily capture genuine retail activity, with little evidence that institutional trades are included in the retail-classified set.

To assess the identifying assumption for the Control–Group 1 on-exchange comparison, I estimate a dynamic difference-in-differences specification following Rambachan and Roth (2023):

$$\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}, \tag{12}$$

where RetailShare_{it} is the share of on-exchange trades classified as retail for stock i in event week t, and ON_C_i equals one for Control stocks traded on-exchange and zero for Group 2 stocks traded on-exchange. The specification includes stock and date fixed effects, with standard errors clustered by stock.

Figure 5 plots the dynamic DiD estimates $\hat{\beta}_k$ with 95% confidence intervals, relative to the reference week $k = -4$. The pre-treatment estimates stay near zero with confidence intervals including zero, so the null of no difference cannot be rejected, consistent with parallel trends prior to the pilot. Post-treatment coefficients fluctuate slightly above zero but remain economically negligible (well under one percentage point), with wide confidence intervals that include zero. This supports the regression results: on exchange, Control and Group 2 exhibit no statistically or economically meaningful change in retail classifications after the pilot.

Taken together, the on-exchange analyses show little evidence of systematic differences. Group 1 and Group 2 exhibit no meaningful gap in retail classifications. Control and Group 2 also show no difference, while Control and Group 1 behave similarly and can be regarded as aligned comparison groups. Overall, the evidence indicates that on exchange, sub-tick executions in all groups largely reflect genuine retail trades, and the False Discovery Rate is minimal—unlike off exchange, where excess classifications reveal a sizable FDR.

Table 6 around here.

Figure 5 around here.

6. Conclusion

This paper examines the reliability of retail trade classification algorithms using the SEC's Tick Size Pilot (TSP), focusing on the False Discovery Rate (FDR)—the share of institutional trades mislabeled as retail among all trades classified as retail. Alongside Type I and Type II errors, the FDR is crucial for evaluating classification errors, since it captures the degree to which reported retail activity may in fact reflect institutional trading.

The Tick Size Pilot imposed distinct rules across groups: no change in the Control group, quoting-only restrictions in Group 1, and full quoting-and-trading restrictions in Groups 2 and 3, with the latter also subject to Trade-at rule. Exemptions permitted limited sub-penny executions—most notably for retail trades—while others, such as midpoint and negotiated trades, can be filtered out as non-retail. Thus, sub-tick executions across groups provide a setting where they can be linked directly to retail activity in some groups but may also capture institutional trades in others.

Trades classified as retail in Group 2 provide a clean benchmark for retail activity, since its wider tick size applies to both quoting and trading and only limited exemptions permit sub-penny executions. Group 1, by contrast, restricts only quoting, and both institutional and retail trades can occur in any price bin. I compare Group 1 with Group 2 to test whether the absence of sub-tick trading restrictions in Group 1 leads to inflated retail classifications, with the magnitude of this inflation capturing the extent of false positives relative to the cleaner Group 2 baseline. In addition to this comparison, I also contrast the Control group with Group 2, where the Control group is fully unconstrained. I align the Control group with Group 1 to confirm that observed differences are not simply driven by baseline gaps.

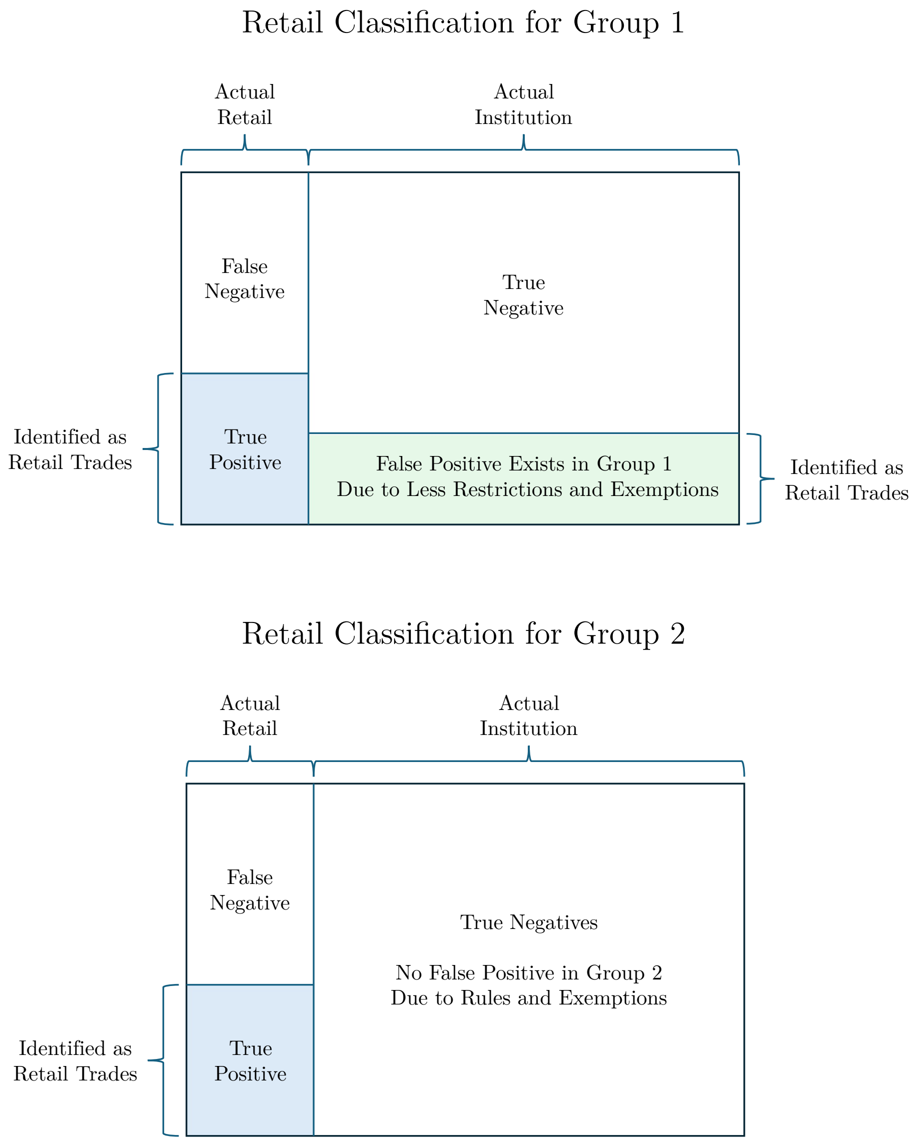
The difference-in-differences analysis—using Group 2 as the retail benchmark—reveals a sharp divergence between on- and off-exchange trading. Off-exchange, retail classifications overstate true retail activity because institutional trades enter the retail-classified sample. The estimates indicate that, relative to Group 2, the post-TSP share of trades classified as retail increased by about 9–10 percentage points in Group 1 and by about 7–12 percentage points in the Control group. These excesses represent false positives, arising in the groups where institutional trades can be misclassified as retail (Group 1 and the Control), in contrast to Group 2 where retail classifications remain clean. This implies a False Discovery Rate

(FDR) of about 73.25% in Group 1 and 66.3% for BJZZ and 39% for QMP in the Control group.

On-exchange, by contrast, sub-tick executions in Groups 1 and 2 almost entirely reflect retail activity. Both static and dynamic difference-in-differences estimates show no meaningful gap between the two groups, and comparisons with the Control group likewise show virtually no difference. The evidence suggests that misclassification under the retail classification algorithm is minimal on-exchange, as reflected in low false positives, with small and insignificant differentials, since sub-tick executions are mainly channeled through Retail Liquidity Programs (RLPs).

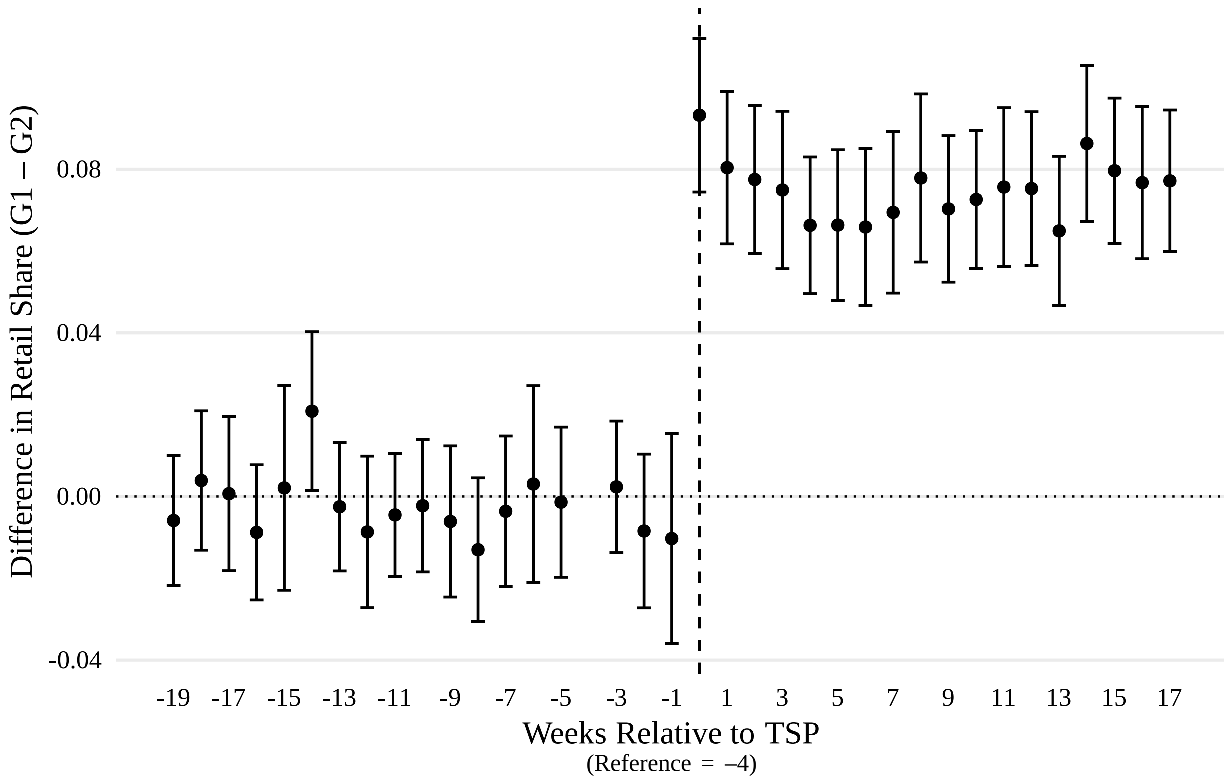
In sum, retail classifications are high-precision on-exchange but much less reliable off-exchange, where high FDR reveal substantial mislabeling of institutional trades as retail. QMP reflects a clear improvement over BJZZ, exhibiting lower FDR and misclassifying fewer institutional trades as retail.

Figure 1. Illustration of the Comparison of Group 1 vs Group 2 (Off-Exchange)



This figure presents a stacked illustration of off-exchange retail classifications for Group 1 (top panel) and Group 2 (bottom panel). In Group 2, the strict rules of the Tick Size Pilot (a five-cent quoting and trading increment) combined with exemptions for retail trades (e.g., executions with meaningful price improvement and other recognizable carve-outs) imply that retail-classified trades form a clean sample of retail activity and are unlikely to contain institutional trades (false positives). By contrast, Group 1 lacks such restrictions and exemptions, so institutional trades may occur in any price bin; as a result, the set of trades classified as retail in Group 1 plausibly mixes both retail and institutional orders. The difference between the two groups provides the intuition for the False Discovery Rate (FDR) that I estimate, although this illustration does not directly depict the difference-in-differences framework used in the panel regressions.

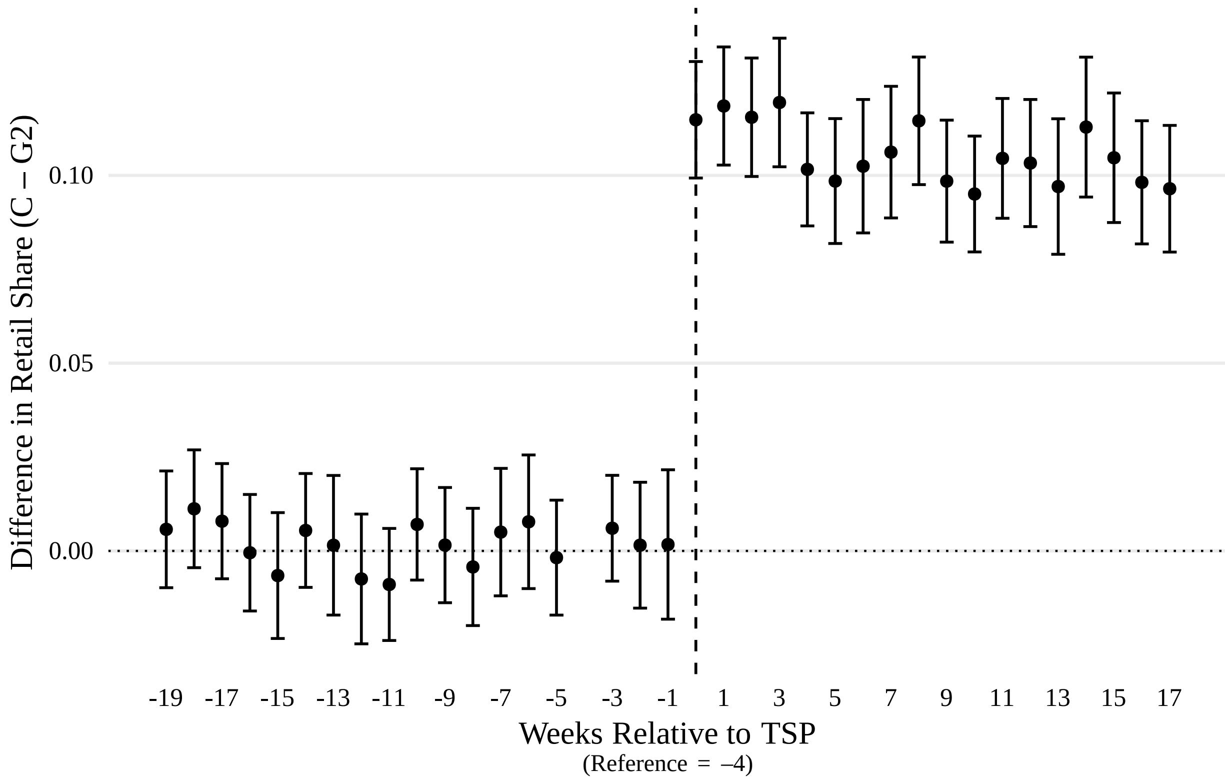
Figure 2. Dynamic DiD Estimates of Retail Share: Group 1 vs. Group 2 (Off-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k\neq-4} \beta_k \cdot \text{OFF\_G1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of off-exchange trades for stock i on day t classified as retail by BJZZ. The variable OFF_G1_i equals one if a stock belongs to Group 1 and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. Each coefficient $\hat{\beta}_k$ measures the G1–G2 difference in off-exchange retail share in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates are close to zero with 95% confidence intervals including zero, so the null of no difference cannot be rejected, supporting the parallel trends assumption. In contrast, post-treatment coefficients rise sharply and remain persistently positive, with 95% confidence intervals excluding zero throughout, indicating a stable and statistically significant increase in off-exchange retail classification for Group 1 stocks relative to Group 2 after the TSP.

[a]Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

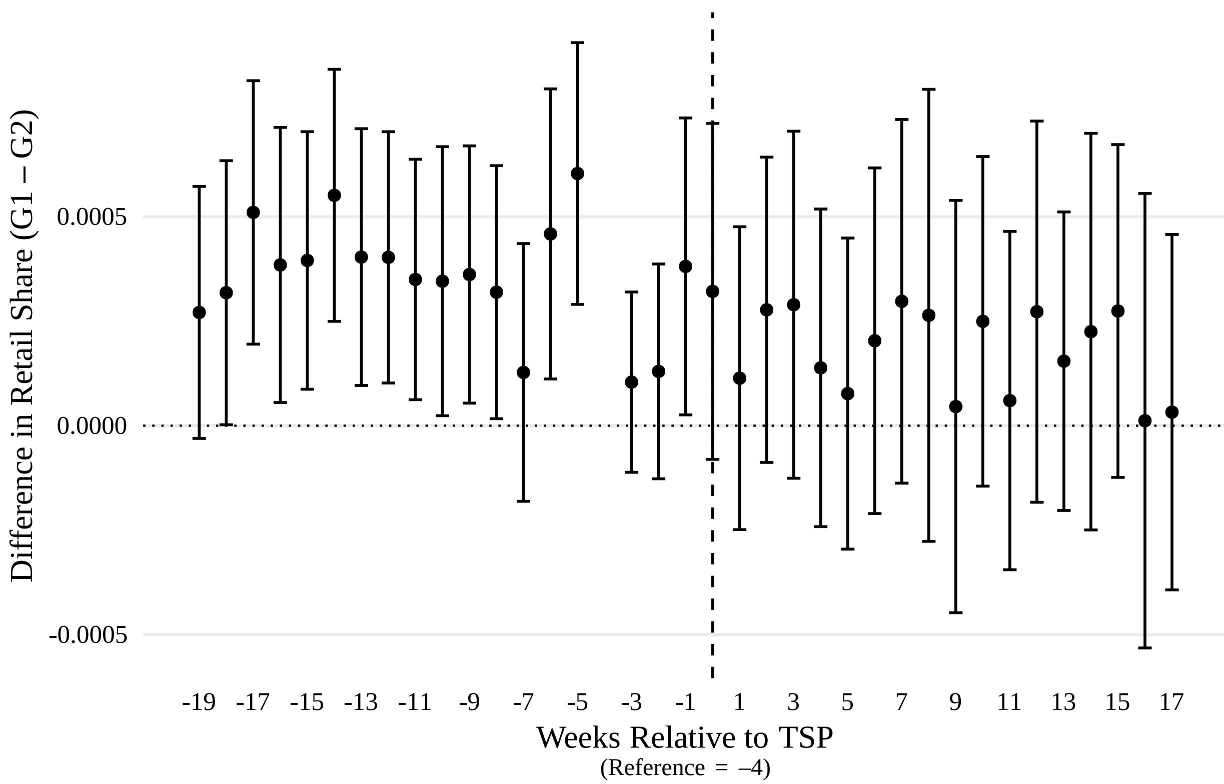
Figure 3. Dynamic DiD Estimates of Retail Share: Control vs. Group 2 (Off-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k\neq -4} \beta_k \cdot \text{OFF\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of off-exchange trades for stock i on day t classified as retail by BJZZ. The variable OFF_C_i equals one if a stock belongs to the Control group and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. Each coefficient $\hat{\beta}_k$ measures the Control–G2 difference in off-exchange retail share in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates stay close to zero with 95% confidence intervals including zero, so the null of no difference cannot be rejected, supporting the parallel trends assumption. In contrast, post-treatment coefficients rise sharply and remain persistently positive, with 95% confidence intervals excluding zero throughout, indicating a stable and statistically significant increase in off-exchange retail classification for Control stocks relative to Group 2 after the TSP.

[a]Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

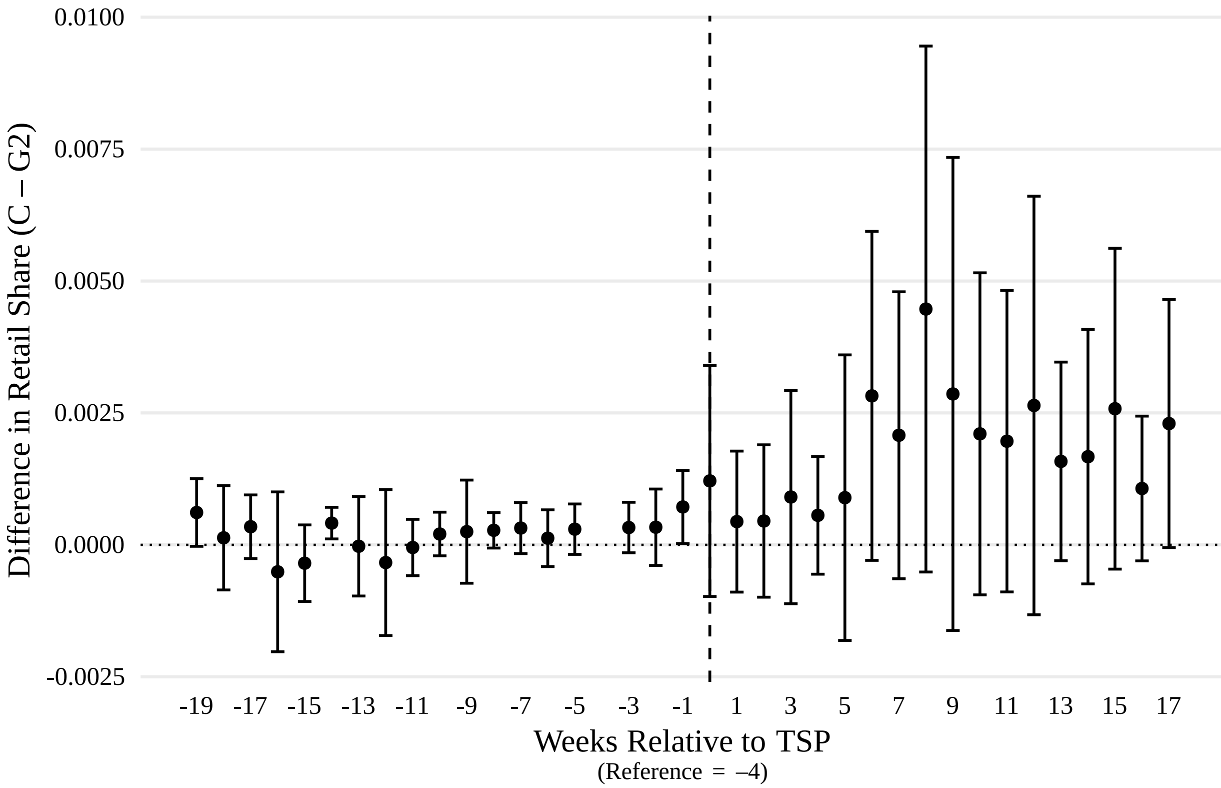
Figure 4. Dynamic DiD Estimates of Retail Share: Group 1 vs. Group 2 (On-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected on-exchange retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_G1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of on-exchange trades for stock i on day t classified as retail. The variable ON_G1_i equals one if a stock belongs to Group 1 and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. Each coefficient $\hat{\beta}_k$ measures the G1–G2 difference in on-exchange retail share in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates fluctuate slightly above zero (a little less than 5×10^{-4}), but the magnitudes are economically negligible, supporting the parallel trends assumption. Post-treatment coefficients remain near zero with wide 95% confidence intervals, indicating no statistically or economically meaningful effect for Group 1 relative to Group 2. The vertical axis is scaled at 10^{-3}, so the estimated differences are effectively indistinguishable from zero.

[a] Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

Figure 5. Dynamic DiD Estimates of Retail Share: Control Group vs. Group 2 (On-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected on-exchange retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_C}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of on-exchange trades for stock i on day t classified as retail. The variable ON_C_i equals one if a stock belongs to the Control group and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. The pre-treatment estimates stay near zero with 95% confidence intervals including zero, so the null of no difference cannot be rejected, supporting the parallel trends assumption. Post-treatment coefficients fluctuate slightly above zero but remain economically negligible (well under one percentage point), with wide 95% intervals that include zero throughout. This indicates no statistically or economically meaningful change in on-exchange retail classification for Control stocks relative to Group 2 after the TSP.

[a]Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

Table 1. Trade Price Distributions within Tick Increments

This table reports the distribution of trade prices across sub-tick increment bins for on-exchange and off-exchange executions, separated by Tick Size Pilot (TSP) groups. Panels A–D correspond to the Control group (Panel A), Test Group 1 (Panel B), Test Group 2 (Panel C), and Test Group 3 (Panel D). Within each panel, the left half shows on-exchange trades and the right half shows off-exchange trades. For each bin, the table reports the number of trades (`Count`), their percentage share of total trades within the group–exchange (`Percent`), and the subset flagged by institutional-type sale conditions ($\texttt{Count}_{\texttt{ITF}}$ and $\%_{\text{ITF}}$). Institutional-type flags (`ITF`) exclude trades marked with sale condition codes `.F` (Intermarket Sweep Orders) and `.B` [CTA] or `.W` [UTP] (Weighted Average Price). The statistic $\%_{\text{ITF}}$ is calculated relative to the number of trades in that specific price bin. The sample covers the post-TSP period from November 1, 2016, to March 2, 2017.

	On-Exchange					**Off-Exchange**			
Bin	Count	Percent	Count_{ITF}	$\%_{ITF}$	Bin	Count	Percent	Count_{ITF}	$\%_{ITF}$
Panel A: Group C									
0	30,498,173	96.43	15,330,277	50.27	0	5,553,823	61.85	11,652	0.21
(0.0–0.1)	13,060	0.04	7,573	57.99	(0.0–0.1)	388,788	4.33	14,160	3.64
[0.1–0.2)	25,917	0.08	4,596	17.73	[0.1–0.2)	143,735	1.60	15,211	10.58
[0.2–0.3)	8,446	0.03	4,421	52.34	[0.2–0.3)	172,390	1.92	14,992	8.70
[0.3–0.4)	6,775	0.02	3,682	54.35	[0.3–0.4)	114,734	1.28	14,816	12.91
[0.4–0.5)	6,474	0.02	3,436	53.07	[0.4–0.5)	119,128	1.33	14,759	12.39
0.5	1,018,707	3.22	204,910	20.11	0.5	1,560,683	17.38	2,324	0.15
(0.5–0.6]	8,388	0.03	4,588	54.70	(0.5–0.6]	118,846	1.32	14,747	12.41
(0.6–0.7]	5,598	0.02	2,814	50.27	(0.6–0.7]	115,103	1.28	14,921	12.96
(0.7–0.8]	6,276	0.02	3,385	53.94	(0.7–0.8]	171,385	1.91	14,948	8.72
(0.8–0.9]	21,301	0.07	3,417	16.04	(0.8–0.9]	144,470	1.61	15,286	10.58
(0.9–1.0)	8,424	0.03	4,567	54.21	(0.9–1.0)	375,876	4.19	14,645	3.90
Total	31,627,539	100.00	15,577,666		Total	8,978,961	100.00	162,461	
Panel B: Group G1									
0	12,768,880	93.03	6,521,676	51.07	0	3,222,954	61.48	7,866	0.24
(0–0.5)	2,999	0.02	0	0.00	(0–0.5)	175,246	3.34	12,435	7.10
[0.5–1.0)	4,861	0.04	0	0.00	[0.5–1.0)	60,924	1.16	10,276	16.87
[1.0–1.5)	69	0.00	14	20.29	[1.0–1.5)	71,170	1.36	9,472	13.31
[1.5–2.0)	0	0.00	0	0.00	[1.5–2.0)	61,489	1.17	9,164	14.90
[2.0–2.5)	2	0.00	0	0.00	[2.0–2.5)	46,741	0.89	9,016	19.29
2.5	941,174	6.86	125,085	13.29	2.5	1,207,756	23.04	1,325	0.11
(2.5–3.0]	7	0.00	2	28.57	(2.5–3.0]	47,538	0.91	8,985	18.90
(3.0–3.5]	3	0.00	0	0.00	(3.0–3.5]	60,696	1.16	9,171	15.11
(3.5–4.0]	33	0.00	11	33.33	(3.5–4.0]	69,504	1.33	9,378	13.49
(4.0–4.5]	3,823	0.03	1	0.03	(4.0–4.5]	59,816	1.14	10,198	17.05
(4.5–5.0)	3,049	0.02	0	0.00	(4.5–5.0)	158,141	3.02	12,363	7.82
Total	13,724,900	100.00	6,646,789		Total	5,241,975	100.00	109,649	
Panel C: Group G2									
0	12,043,892	93.06	6,174,632	51.27	0	3,203,724	66.97	7,421	0.23
(0–0.5)	1	0.00	1	100.00	(0–0.5)	18,478	0.39	10,820	58.56
[0.5–1.0)	9,907	0.08	1	0.01	[0.5–1.0)	84,793	1.77	8,749	10.32

Continued on next page

Table 1. (continued)

	On-Exchange					Off-Exchange			
Bin	Count	Percent	Count_{ITF}	$\%_{ITF}$	Bin	Count	Percent	Count_{ITF}	$\%_{ITF}$
[1.0–1.5)	11	0.00	1	9.09	[1.0–1.5)	40,380	0.84	8,306	20.57
[1.5–2.0)	0	0.00	0	0.00	[1.5–2.0)	39,430	0.82	8,252	20.93
[2.0–2.5)	3	0.00	3	100.00	[2.0–2.5)	45,847	0.96	7,960	17.36
2.5	881,314	6.81	119,575	13.57	2.5	1,128,491	23.59	1,228	0.11
(2.5–3.0]	2	0.00	0	0.00	(2.5–3.0]	47,164	0.99	7,986	16.93
(3.0–3.5]	2	0.00	1	50.00	(3.0–3.5]	38,047	0.80	8,475	22.28
(3.5–4.0]	5	0.00	0	0.00	(3.5–4.0]	40,260	0.84	8,637	21.45
(4.0–4.5]	7,058	0.05	0	0.00	(4.0–4.5]	78,699	1.65	9,212	11.71
(4.5–5.0)	0	0.00	0	0.00	(4.5–5.0)	18,729	0.39	11,167	59.62
Total	12,942,195	100.00	6,294,214		Total	4,784,042	100.00	98,213	
Panel D: Group G3									
0	13,347,163	92.33	7,439,754	55.74	0	660,001	23.37	6,101	0.92
(0–0.5)	0	0.00	0	0.00	(0–0.5)	21,167	0.75	10,955	51.76
[0.5–1.0)	9,109	0.06	0	0.00	[0.5–1.0)	67,973	2.41	8,851	13.02
[1.0–1.5)	6	0.00	0	0.00	[1.0–1.5)	67,087	2.38	8,418	12.55
[1.5–2.0)	1	0.00	0	0.00	[1.5–2.0)	40,702	1.44	8,277	20.34
[2.0–2.5)	1	0.00	1	100.00	[2.0–2.5)	43,937	1.56	8,204	18.67
2.5	1,091,303	7.55	186,342	17.08	2.5	1,688,044	59.76	1,686	0.10
(2.5–3.0]	0	0.00	0	0.00	(2.5–3.0]	46,373	1.64	8,276	17.85
(3.0–3.5]	0	0.00	0	0.00	(3.0–3.5]	39,913	1.41	8,219	20.59
(3.5–4.0]	9	0.00	3	33.33	(3.5–4.0]	65,508	2.32	8,344	12.74
(4.0–4.5]	7,626	0.05	0	0.00	(4.0–4.5]	62,869	2.23	9,094	14.46
(4.5–5.0)	0	0.00	0	0.00	(4.5–5.0)	21,082	0.75	10,989	52.13
Total	14,455,218	100.00	7,626,100		Total	2,824,656	100.00	97,414	

Table 2

Retail Classification Rates across Tick Size Pilot Groups and Venues

This table reports the percentage of trades classified as retail across Tick Size Pilot (TSP) groups and execution venues, using two classification methods: BJZZ (Boehmer et al., 2021) and the quote midpoint (QMP) method (Barber et al., 2024). Panel A presents raw retail classification shares, measured as the fraction of total trades within each group–venue identified as retail by the respective method. Panel B reports retail classification shares after excluding trades flagged with institutional-type sale condition codes (`.F` for Intermarket Sweep Orders and `.B` [CTA] or `.W` [UTP] for Weighted Average Price). On-exchange and off-exchange results are shown separately. The sample covers the post-TSP period from November 1, 2016, to March 2, 2017.

	On-Exchange		**Off-Exchange**	
TSP_Group	BJZZ (%)	QMP (%)	BJZZ (%)	QMP (%)
Panel A: Retail Share (%)				
Control Group	0.30	0.29	18.11	16.45
Group 1	0.11	0.33	13.68	15.27
Group 2	0.13	0.35	6.72	8.63
Group 3	0.12	0.33	12.18	16.01
Panel B: Retail Share, COND (%)				
Control Group	0.19	0.19	16.79	15.22
Group 1	0.11	0.26	12.10	13.66
Group 2	0.13	0.28	5.64	7.46
Group 3	0.12	0.26	10.37	14.07

Table 3
DiD Estimates of Off-Exchange Retail Classification under Asymmetric Tick Constraints: G1 vs. G2

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how off-exchange retail trade classifications for Group 1 stocks changed relative to Group 2 after the Tick Size Pilot (TSP) began. Group 2 stocks were constrained to a $0.05 tick size for both quoting and trading, while Group 1 faced only quoting constraints, allowing trades to execute at finer price increments. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \left(\text{OFF\_G1}_{s,t} \times \text{TSP}_{s,t}\right) + \text{FE} + \epsilon_{s,t}$, where $\text{OFF\_G1}_{s,t}$ equals one if stock s belongs to Group 1 and trades off-exchange, and zero if it belongs to Group 2 and trades off-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The dependent variables capture various measures of off-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of off-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
OFF_G1 × TSP	0.102***	0.768***	0.099***	0.819***	0.094***	0.633***	0.091***	0.662***
	(21.67)	(17.38)	(20.96)	(17.86)	(19.84)	(14.91)	(19.55)	(15.65)
Num. Obs.	57,471	57,471	57,471	57,471	57,471	57,471	57,471	57,471
Adj. R^2	0.377	0.641	0.399	0.626	0.333	0.646	0.356	0.628
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
OFF_G1 × TSP	0.095***	0.796***	0.091***	0.852***	0.088***	0.652***	0.085***	0.684***
	(23.04)	(18.34)	(22.21)	(18.81)	(20.81)	(15.41)	(20.57)	(16.20)
Num. Obs.	333,771	333,771	333,771	333,771	333,771	333,771	333,771	333,771
Adj. R^2	0.433	0.599	0.459	0.582	0.383	0.595	0.411	0.574
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4
DiD Estimates of Off-Exchange Retail Classification under Asymmetric Tick Constraints: Control Group vs. G2

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how off-exchange retail trade classifications for Control Group stocks changed relative to Group 2 after the Tick Size Pilot (TSP) began. Group 2 stocks were constrained to a $0.05 tick size for both quoting and trading, while Control Group faced no constraints. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \text{OFF\_C}_{s,t} \times \text{TSP}_{s,t} + \text{FE} + \epsilon_{s,t}$ where $\text{OFF\_C}_{s,t} = 1$ for Control Group stocks, and $\text{TSP}_{s,t} = 1$ in the post-TSP period. The dependent variables capture various measures of off-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of off-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
OFF_C × TSP	0.120***	0.507***	0.117***	0.584***	0.076***	0.178***	0.074***	0.219***
	(30.61)	(13.15)	(29.76)	(14.38)	(20.21)	(4.83)	(20.08)	(5.88)
Num. Obs.	116,350	116,350	116,350	116,350	116,350	116,350	116,350	116,350
Adj. R^2	0.344	0.597	0.366	0.583	0.297	0.595	0.319	0.580
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
OFF_C × TSP	0.114***	0.535***	0.110***	0.617***	0.072***	0.192***	0.071***	0.234***
	(33.38)	(14.62)	(32.40)	(15.92)	(21.61)	(5.38)	(21.53)	(6.50)
Num. Obs.	652,979	652,979	652,979	652,979	652,979	652,979	652,979	652,979
Adj. R^2	0.397	0.563	0.422	0.548	0.347	0.558	0.375	0.539
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5
DiD Estimates of On-Exchange Retail Classification under Asymmetric Tick Constraints: Group 1 vs. Group 2

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how on-exchange retail trade classifications for Group 1 stocks changed relative to Group 2 after the Tick Size Pilot (TSP) began. Group 2 stocks were constrained to a \$0.05 tick size for both quoting and trading, while Group 1 faced only quoting constraints, allowing trades to execute at finer price increments. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \left(\text{ON\_G1}_{s,t} \times \text{TSP}_{s,t}\right) + \text{FE} + \epsilon_{s,t}$, where $\text{ON\_G1}_{s,t}$ equals one if stock s belongs to Group 1 and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The dependent variables capture various measures of on-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of on-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
ON_G1 × TSP	-0.000	-0.026	-0.000	-0.027	-0.000	-0.001	-0.000	-0.004
	(-0.64)	(-1.30)	(-0.66)	(-1.34)	(-0.73)	(-0.02)	(-0.73)	(-0.11)
Num. Obs.	57,511	57,511	57,511	57,511	57,511	57,511	57,511	57,511
Adj. R^2	0.022	0.230	0.022	0.230	0.024	0.380	0.023	0.351
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
ON_G1 × TSP	-0.000	-0.029	-0.000	-0.029	-0.000	-0.002	-0.000	-0.005
	(-1.20)	(-1.28)	(-1.22)	(-1.32)	(-1.37)	(-0.05)	(-1.38)	(-0.14)
Num. Obs.	338,431	338,431	338,431	338,431	338,431	338,431	338,431	338,431
Adj. R^2	0.040	0.197	0.040	0.197	0.043	0.333	0.041	0.303
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

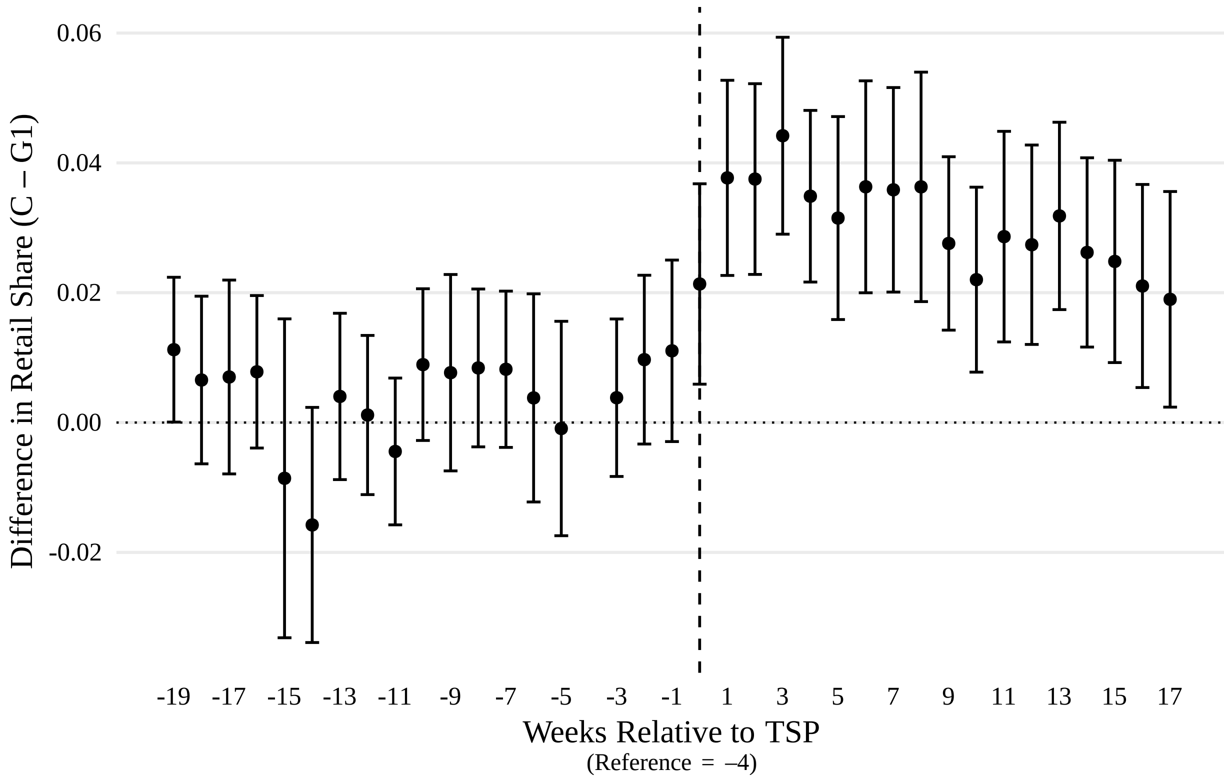
Table 6
DiD Estimates of On-Exchange Retail Classification under Asymmetric Tick Constraints: Control Group vs. Group 2

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how on-exchange retail trade classifications for Control Group stocks changed relative to Group 2 after the Tick Size Pilot (TSP) began. Group 2 stocks were constrained to a $0.05 tick size for both quoting and trading, while Control Group faced no constraints. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \text{ON\_C}_{s,t} \times \text{TSP}_{s,t} + \text{FE} + \epsilon_{s,t}$ where $\text{ON\_C}_{s,t}$ equals one if stock s belongs to Control Group and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The dependent variables capture various measures of on-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of on-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
ON_C × TSP	0.002	-0.091***	0.000	-0.095***	-0.000	-0.489***	-0.001**	-0.409***
	(1.42)	(-4.75)	(0.55)	(-5.13)	(-0.21)	(-15.32)	(-2.06)	(-14.81)
Num. Obs.	116,491	116,491	116,491	116,491	116,491	116,491	116,491	116,491
Adj. R^2	0.192	0.230	0.092	0.227	0.183	0.300	0.084	0.280
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
ON_C × TSP	0.002	-0.093***	0.001	-0.097***	0.000	-0.530***	-0.001	-0.442***
	(1.60)	(-4.36)	(0.97)	(-4.75)	(0.11)	(-15.50)	(-1.53)	(-14.91)
Num. Obs.	667,341	667,341	667,341	667,341	667,341	667,341	667,341	667,341
Adj. R^2	0.247	0.211	0.155	0.208	0.239	0.275	0.144	0.255
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A.1. Supplementary Materials

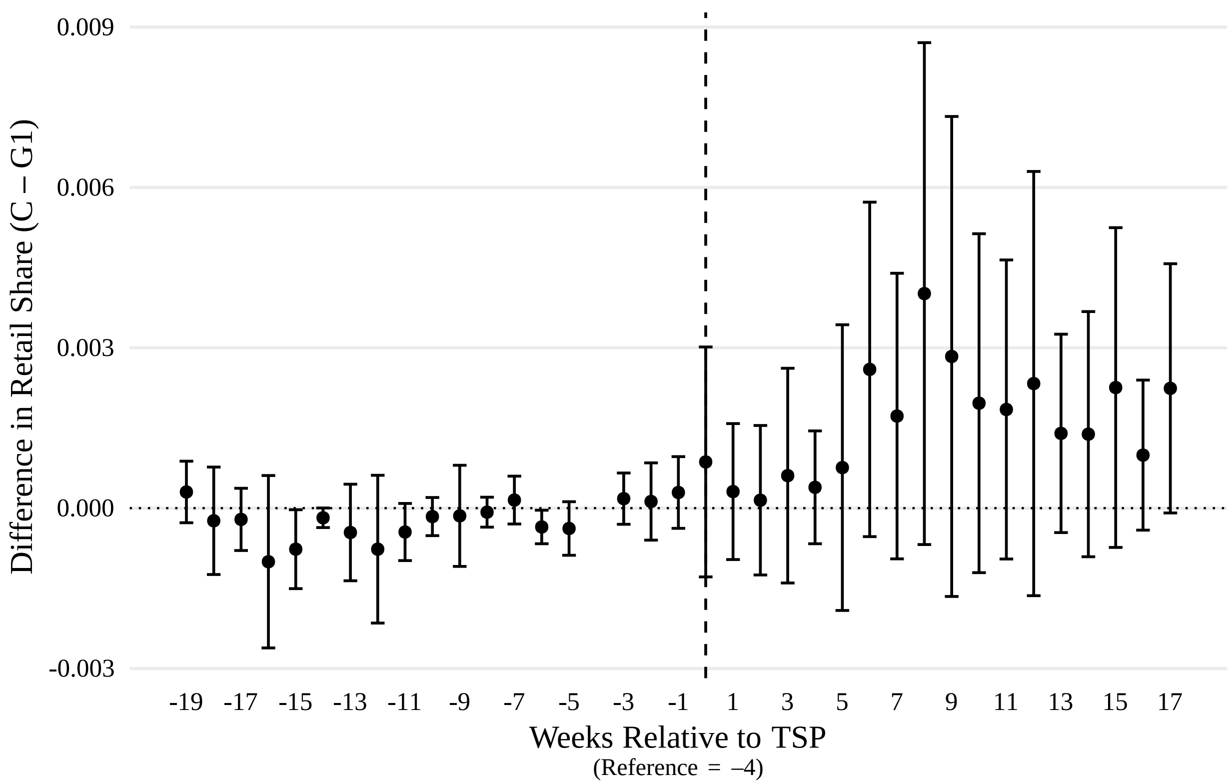
Figure A.1. Dynamic DiD Estimates of Retail Share: Control vs. Group 1 (Off-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected off-exchange retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{OFF\_C1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of off-exchange trades for stock i on day t classified as retail by BJZZ. The variable OFF_C1_i equals one if a stock belongs to the Control group and trades off-exchange, and zero if it belongs to Group 1 and trades off-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. Each coefficient $\hat{\beta}_k$ measures the Control–G1 difference in off-exchange retail share in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates stay close to zero with 95% confidence intervals generally including zero, supporting the parallel trends assumption. Post-treatment coefficients rise and remain positive, with 95% intervals excluding zero, indicating a statistically significant increase in off-exchange retail classification for Control stocks relative to Group 1 after the TSP. However, the magnitude of this effect is smaller than the corresponding difference observed between Group 1 and Group 2.

[a] Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

Figure A.2. Dynamic DiD Estimates of Retail Share: Control vs. Group 1 (On-Exchange)



The figure shows dynamic difference-in-differences estimates (Rambachan and Roth, 2023) of how the Tick Size Pilot (TSP) affected on-exchange retail trade classification. The regression estimated is $\text{RetailShare}_{it} = \sum_{k \neq -4} \beta_k \cdot \text{ON\_C1}_i \cdot \mathbf{1}\{\text{event\_time}_t = k\} + \text{FE} + \varepsilon_{it}$, where RetailShare_{it} is the share of on-exchange trades for stock i on day t classified as retail using exchange-provided RPI flags. The variable ON_C1_i equals one if a stock belongs to the Control group and trades on-exchange, and zero if it belongs to Group 1 and trades on-exchange. The variable event_time_t maps each day t into an event week k, with $k = 0$ denoting the first post-TSP week.[a] The term FE denotes stock and date fixed effects, and standard errors are clustered by stock and date. Each coefficient $\hat{\beta}_k$ measures the Control–G1 difference in on-exchange retail share in event week k, relative to the omitted reference week $k = -4$. The pre-treatment estimates remain close to zero with 95% confidence intervals including zero, supporting the parallel trends assumption. Post-treatment coefficients drift modestly upward but remain economically small (around 0.2–0.4 percentage points) and imprecisely estimated, with 95% intervals that generally include zero. This indicates no statistically or economically meaningful change in on-exchange retail classification for Control stocks relative to Group 1 after the TSP.

[a] Although the official implementation date of the TSP was October 3, 2016, the post-TSP sample begins on November 1 to ensure that all TSP rules were fully in effect. Because the rollout occurred gradually over October, that month is omitted from the analysis. Results are shown relative to event week $k = -4$, chosen as a clean baseline prior to the October 2016 staggered rollout of pilot stocks. Selecting $k = -4$ (early September) provides a pre-period reference that is safely before treatment while leaving enough pre-TSP weeks for estimation.

Table A.1

Confusion Matrix with Error Types and Common Classification Metrics

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{TP+FN}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	**Negative Predictive Value** $\frac{TN}{TN+FN}$	**Accuracy** $\frac{TP+TN}{TP+TN+FP+FN}$

Notes: The confusion matrix summarizes classification outcomes by comparing predicted and actual classes. It defines Type I errors (false positives) and Type II errors (false negatives), and reports common metrics such as sensitivity, specificity, precision, and accuracy, thereby clarifying how these measures relate to one another. For Type I and Type II errors, the conditioning is on the *actual class* (e.g., given an institutional trade or a retail trade, how many of them are incorrectly identified). When expressed as rates, these correspond to the false positive rate (1 − specificity) and the false negative rate (1 − sensitivity), respectively. By contrast, Precision and the False Discovery Rate (FDR) condition on the *predicted class*: given a trade classified as retail (a positive prediction), Precision measures the share that are truly retail, while the FDR measures the share that are actually institutional. Importantly, the False Discovery Rate is given by FDR = 1 − Precision, which is the central measure of misclassification emphasized in this paper. FDR can also be expressed as

$$\mathrm{FDR} \;=\; \frac{FP}{TP+FP} \;=\; \frac{(1-p)\alpha}{p(1-\beta)+(1-p)\alpha}.$$

where p denote the base rate of retail trades, α the Type I error rate, and β the Type II error rate.

Table A.2
Sale Condition Codes Used in CTA and UTP Trades

This table lists the sale condition codes defined in the Daily TAQ (Trades and Quotes) data specification, which records every trade reported to the consolidated tape by the CTA (Consolidated Tape Association, Tapes A/B) and UTP (Unlisted Trading Privileges, Tape C) Securities Information Processors (SIPs). Sale condition codes indicate how a trade was executed or reported (e.g., regular trade, average price trade, intermarket sweep order, odd lot, cross, sold last). These codes are embedded in the TAQ trades file as part of the SIP feed and are critical for interpreting whether trades qualify as regular executions or carry special reporting conditions.

CTA Code	**Description**	**UTP Code**	**Description**
blank	Regular Trade (no associated conditions)	@	Regular Trade
B	Average Price Trade	A	Acquisition
C	Cash Trade (Same Day Clearing)	B	Bunched Trade
E	Automatic Execution	C	Cash Sale
F	Intermarket Sweep Order	D	Distribution
H	Price Variation Trade	E	Placeholder for future use
I	Odd Lot Trade	F	Intermarket Sweep
K	Rule 127 (NYSE) or Rule 155 Trade (NYSE American)	G	Bunched Sold Trade
L	Sold Last (Late Reporting)	H	Price Variation Trade
M	Market Center Official Close	I	Odd Lot Trade
N	Next Day Trade (Next Day Clearing)	K	Rule 155 Trade (NYSE American only)
O	Market Center Opening Trade	L	Sold Last
P	Prior Reference Price	M	Market Center Close Price
Q	Market Center Official Open	N	Next Day
R	Seller	O	Opening Prints
T	Extended Hours Trade	P	Prior Reference Price
U	Extended Hours (Sold Out of Sequence)	Q	Market Center Open Price
V	Contingent Trade	R	Seller
X	Cross Trade	S	Split Trade
Z	Sold (Out of Sequence)	T	Form-T Trade
4	Derivatively Priced	U	Extended Trading Hours
5	Market Center Re-Opening Trade	V	Stock-Option Trade
6	Market Center Closing Trade	W	Average Price Trade
7	Qualified Contingent Trade (from Aug 3, 2015)	X	Cross Trade
9	Corrected Consolidated Close Price	Y	Yellow Flag
		Z	Sold (Out of Sequence)
		1	Stopped Stock - Regular Trade
		4	Derivatively Priced
		5	Re-opening Prints
		6	Closing Prints
		7	Exempt Qualified Contingent Trade (QCT)
		8	Placeholder for 611 Exempt
		9	Corrected Consolidated Close (per listing market)

Note: Sale condition codes indicate how a trade was executed or reported. CTA refers to the Consolidated Tape Association (Tape A/B, e.g., NYSE/Arca), while UTP refers to the Unlisted Trading Privileges Plan (Tape C, e.g., Nasdaq). A space in CTA (shown as [space]) indicates a regular trade with no special condition.

Table A.3

Trade Sale Condition Codes and Frequencies by Reporting SIP

This table reports the distribution of trade sale condition codes across the two Securities Information Processors (SIPs): CTA (Consolidated Tape Association, SOURCE = C) and UTP (Unlisted Trading Privileges, SOURCE = N). For each SIP, the table lists sale condition codes alongside their observed trade counts. For retail-classification purposes, the following codes are treated as institutional-type flags: `.F` (Intermarket Sweep Orders), and `.B` [CTA] or `.W` [UTP] (Weighted Average Price). All other codes are reported for completeness. The frequencies highlight which conditions are most prevalent in each feed, providing context for how classification methods (such as BJZZ or QMP) should treat or exclude particular conditions.

CTA (SOURCE = C)						**UTP (SOURCE = N)**					
Cond	**Count**	**Cond**	**Count**	**Cond**	**Count**	**Cond**	**Count**	**Cond**	**Count**	**Cond**	**Count**
@	18,095,400	B	474	UB	29	@	24,660,998	N W	513	@O	10
F	9,517,081	4ZI	469	N T	25	@F	12,975,904	@FZ	345	N ZI	10
I	6,096,641	FZ	429	P	24	@ I	8,796,026	@ T	341	N T	9
F I	5,518,929	N B	351	4 P	19	@F I	8,141,067	N4 W	233	@ V	8
4 B	205,945	6	300	6 I	18	@4 W	248,890	R W	200	@4ZW	6
Q	56,135	O X	275	TB	9	@ Q	103,869	@ M	199	@7ZV	5
O	27,335	TI	186	7ZV	5	@O X	47,872	@FZI	168	N Z	3
4 I	6,673	FZI	163	V	5	@4 I	10,568	@7 W	122	N7 V	3
Z	4,805	N4 B	158	4ZB	4	@ Z	8,080	@6 X	97	R4 I	3
N	4,045	R B	145	C ZI	2	@4	5,463	@FT	87	@4ZP	2
4	3,044	N TI	102	N Z	2	N	4,998	C4 W	75	@6 I	2
R	2,572	R4 B	79	N ZI	2	@7 V	4,121	@F W	70	R4 P	2
7 V	2,234	7 B	64	R4 I	2	R	2,720	@ TI	68	@ UI	1
C	1,916	FT	63	ZB	2	@O I	2,717	@FTI	55	@ UP	1
N I	1,559	C4 B	61	7 I	1	@ ZI	2,442	@ P	44	@ ZP	1
ZI	1,366	M	48	C Z	1	C	2,337	R4 W	39	@6	1
R I	929	FALSE	41	N7 V	1	N I	2,293	@7 I	26	@F P	1
T	813	F B	40	NF	1	@4Z	1,088	N TI	21	C Z	1
C I	731	O I	36	UI	1	R I	937	@4 P	20	N4 I	1
4Z	485	FTI	31			C I	690	@ UW	14	R Z	1
						@4ZI	673	@ ZW	12	R ZI	1
						@ W	594	@ TW	11	RFZ	1

Table A.4

Exchange Distribution by Group and Time for On-Exchange Retail Classified Trades

This table reports the distribution of on-exchange retail-classified trades across exchanges before (Pre-TSP) and after (Post-TSP) the start of the Tick Size Pilot (TSP). Panels A–D correspond to the Control group (Panel A), Test Group 1 (Panel B), Test Group 2 (Panel C), and Test Group 3 (Panel D). For each panel, the left half of the table shows the pre-TSP distribution of retail-classified trades across exchanges, and the right half shows the post-TSP distribution. Exchange codes are the standard CTA/UTP identifiers (e.g., `Y` = Cboe BYX, `B` = Nasdaq BX, `N` = NYSE). Percentages are calculated relative to the total number of on-exchange retail-classified trades within each group and period.

Pre-TSP			**Post-TSP**		
Code	Exchange	%	Code	Exchange	%
Panel A: Group C					
Y	Cboe BYX	47.94	Q	NASDAQ (Tape C)	26.55
B	NASDAQ BX	21.31	P	NYSE Arca	11.19
N	NYSE	16.64	N	NYSE	10.06
P	NYSE Arca	13.51	K	Cboe EDGX	9.63
A	NYSE American	0.59	Z	Cboe BZX	9.39
M	NYSE Chicago	0.01	T	NASDAQ (Tape A,B)	9.08
Q	NASDAQ (Tape C)	0.01	Y	Cboe BYX	8.80
			B	NASDAQ BX	6.63
			J	Cboe EDGA	3.72
			V	IEX	3.42
			X	NASDAQ PSX	1.33
			A	NYSE American	0.17
			C	NYSE National	0.01
			M	NYSE Chicago	0.01
Panel B: Group G1					
Y	Cboe BYX	45.68	B	NASDAQ BX	38.84
N	NYSE	23.24	Y	Cboe BYX	37.47
B	NASDAQ BX	20.43	N	NYSE	15.89
P	NYSE Arca	9.41	P	NYSE Arca	7.15
A	NYSE American	1.24	A	NYSE American	0.40
M	NYSE Chicago	0.01	Z	Cboe BZX	0.12
			K	Cboe EDGX	0.09
			J	Cboe EDGA	0.04
			M	NYSE Chicago	0.01
Panel C: Group G2					
Y	Cboe BYX	46.42	B	NASDAQ BX	42.11
B	NASDAQ BX	23.23	Y	Cboe BYX	37.26
N	NYSE	17.25	N	NYSE	13.35
P	NYSE Arca	12.22	P	NYSE Arca	6.90
A	NYSE American	0.87	A	NYSE American	0.36
			M	NYSE Chicago	0.02
Panel D: Group G3					
Y	Cboe BYX	47.80	Y	Cboe BYX	44.48
B	NASDAQ BX	20.03	B	NASDAQ BX	35.82

Continued on next page

Table A.4
(continued)

Pre-TSP			Post-TSP		
Code	Exchange	%	Code	Exchange	%
P	NYSE Arca	16.69	N	NYSE	11.01
N	NYSE	15.06	P	NYSE Arca	8.44
A	NYSE American	0.42	A	NYSE American	0.23
			M	NYSE Chicago	0.02

Table A.5. Trade Price Distributions within Tick Increments (Pre-TSP)

This table summarizes the pre–Tick Size Pilot (pre-TSP) distribution of trade prices across sub-penny bins for on-exchange and off-exchange executions, separated by the four TSP groups. Because the sample is pre-TSP, all stocks quote and trade in one-cent increments; the bins record each execution's location within the penny. Panels A–D correspond to the Control group (Panel A), Test Group 1 (Panel B), Test Group 2 (Panel C), and Test Group 3 (Panel D). Within each panel, the left half shows on-exchange trades and the right half shows off-exchange trades. For each bin, the table reports the number of trades (`Count`), their percentage share of total trades within the group–exchange cell (`Percent`), and the subset flagged by institutional-type sale conditions ($\texttt{Count}_{\texttt{COND}}$ and $\%_{\text{COND}}$). Institutional-type flags (`COND`) mark trades carrying sale condition codes `.F` (Intermarket Sweep Order), `.B` [CTA], or `.W` [UTP] (Weighted Average Price), which are typically associated with institutional activity. The statistic $\%_{\text{COND}}$ is computed as $\texttt{Count}_{\texttt{COND}}$ divided by `Count` in that bin. The sample covers the pre-TSP period from May 1 to September 1, 2016.

	On-Exchange					**Off-Exchange**			
Bin	Count	Percent	$Count_{COND}$	$\%_{COND}$	Bin	Count	Percent	$Count_{COND}$	$\%_{COND}$
Panel A: Group C									
0	37,141,919	97.73	19,871,035	53.50	0	7,067,588	62.19	67,722	0.96
(0.0–0.1)	0	0.00	0	0.00	(0.0–0.1)	539,182	4.74	16,683	3.09
[0.1–0.2)	14,100	0.04	0	0.00	[0.1–0.2)	163,515	1.44	18,201	11.13
[0.2–0.3)	1,401	0.00	0	0.00	[0.2–0.3)	195,640	1.72	20,013	10.23
[0.3–0.4)	2,209	0.01	0	0.00	[0.3–0.4)	115,885	1.02	17,636	15.22
[0.4–0.5)	700	0.00	0	0.00	[0.4–0.5)	104,240	0.92	17,217	16.52
0.5	829,784	2.18	237,203	28.59	0.5	2,083,509	18.33	10,508	0.50
(0.5–0.6]	632	0.00	0	0.00	(0.5–0.6]	104,434	0.92	17,684	16.93
(0.6–0.7]	1,657	0.00	0	0.00	(0.6–0.7]	117,038	1.03	17,904	15.30
(0.7–0.8]	1,131	0.00	0	0.00	(0.7–0.8]	191,759	1.69	19,936	10.40
(0.8–0.9]	11,285	0.03	0	0.00	(0.8–0.9]	156,453	1.38	18,331	11.72
(0.9–1.0)	1	0.00	0	0.00	(0.9–1.0)	525,635	4.63	17,041	3.24
Total	38,004,819	100.00	20,108,238		Total	11,364,878	100.00	258,876	
Panel B: Group G1									
0	13,048,947	97.14	6,878,456	52.71	0	2,513,199	61.66	26,516	1.06
(0.0–0.1)	0	0.00	0	0.00	(0.0–0.1)	180,841	4.44	6,610	3.66
[0.1–0.2)	4,215	0.03	0	0.00	[0.1–0.2)	54,595	1.34	6,979	12.78
[0.2–0.3)	448	0.00	0	0.00	[0.2–0.3)	66,057	1.62	7,392	11.19
[0.3–0.4)	779	0.01	0	0.00	[0.3–0.4)	38,588	0.95	6,730	17.44
[0.4–0.5)	226	0.00	0	0.00	[0.4–0.5)	34,412	0.84	6,564	19.07
0.5	374,137	2.79	89,271	23.86	0.5	818,501	20.08	3,719	0.45
(0.5–0.6]	210	0.00	0	0.00	(0.5–0.6]	34,756	0.85	6,692	19.25
(0.6–0.7]	597	0.00	0	0.00	(0.6–0.7]	38,682	0.95	6,785	17.54
(0.7–0.8]	382	0.00	0	0.00	(0.7–0.8]	66,311	1.63	7,995	12.06
(0.8–0.9]	3,537	0.03	0	0.00	(0.8–0.9]	53,027	1.30	6,955	13.12
(0.9–1.0)	0	0.00	0	0.00	(0.9–1.0)	176,776	4.34	6,625	3.75
Total	13,433,478	100.00	6,967,727		Total	4,075,745	100.00	99,562	
Panel C: Group G2									
0	12,201,142	97.54	6,526,605	53.49	0	2,318,076	62.10	23,975	1.03
(0.0–0.1)	0	0.00	0	0.00	(0.0–0.1)	167,642	4.49	5,998	3.58

Continued on next page

Table A.5. (continued)

	On-Exchange					Off-Exchange			
Bin	Count	Percent	Count_{COND}	$\%_{COND}$	Bin	Count	Percent	Count_{COND}	$\%_{COND}$
[0.1–0.2)	4,044	0.03	0	0.00	[0.1–0.2)	52,583	1.41	6,217	11.82
[0.2–0.3)	463	0.00	0	0.00	[0.2–0.3)	61,588	1.65	6,966	11.31
[0.3–0.4)	677	0.01	0	0.00	[0.3–0.4)	37,043	0.99	5,869	15.84
[0.4–0.5)	246	0.00	0	0.00	[0.4–0.5)	32,332	0.87	5,878	18.18
0.5	297,091	2.38	80,660	27.15	0.5	713,731	19.12	3,758	0.53
(0.5–0.6]	242	0.00	0	0.00	(0.5–0.6]	32,276	0.86	5,937	18.39
(0.6–0.7]	615	0.00	0	0.00	(0.6–0.7]	37,779	1.01	5,964	15.79
(0.7–0.8]	378	0.00	0	0.00	(0.7–0.8]	60,759	1.63	6,784	11.17
(0.8–0.9]	3,804	0.03	0	0.00	(0.8–0.9]	50,417	1.35	6,396	12.69
(0.9–1.0)	0	0.00	0	0.00	(0.9–1.0)	168,661	4.52	5,962	3.53
Total	12,508,702	100.00	6,607,265		Total	3,732,887	100.00	89,704	
Panel D: Group G3									
0	11,580,924	97.52	6,139,767	53.02	0	2,237,331	61.89	22,689	1.01
(0.0–0.1)	0	0.00	0	0.00	(0.0–0.1)	169,199	4.68	5,136	3.04
[0.1–0.2)	3,896	0.03	0	0.00	[0.1–0.2)	52,001	1.44	5,624	10.82
[0.2–0.3)	469	0.00	0	0.00	[0.2–0.3)	61,959	1.71	6,088	9.83
[0.3–0.4)	736	0.01	0	0.00	[0.3–0.4)	36,966	1.02	5,504	14.89
[0.4–0.5)	198	0.00	0	0.00	[0.4–0.5)	31,686	0.88	5,309	16.76
0.5	284,286	2.39	80,685	28.38	0.5	684,896	18.94	3,174	0.46
(0.5–0.6]	181	0.00	0	0.00	(0.5–0.6]	31,530	0.87	5,151	16.34
(0.6–0.7]	618	0.01	0	0.00	(0.6–0.7]	37,909	1.05	5,347	14.10
(0.7–0.8]	332	0.00	0	0.00	(0.7–0.8]	59,806	1.65	5,980	10.00
(0.8–0.9]	3,620	0.03	0	0.00	(0.8–0.9]	48,844	1.35	5,674	11.62
(0.9–1.0)	0	0.00	0	0.00	(0.9–1.0)	163,133	4.51	5,165	3.17
Total	11,875,260	100.00	6,220,452		Total	3,615,260	100.00	80,841	

Table A.6

Retail Classification Rates across Tick Size Pilot Groups and Venues (Pre-TSP)

This table reports the share of trades classified as retail across Tick Size Pilot (TSP) groups and execution venues, using pre-TSP data (before the Tick Size Pilot). Two classification methods are applied: BJZZ (Boehmer et al., 2021) and the quote midpoint (QMP) method (Barber et al., 2024). Panel A presents raw retail classification shares, measured as the fraction of trades in each group–venue classified as retail by the respective method. Panel B reports retail classification shares after excluding trades flagged with institutional-type sale condition codes (`.F` for Intermarket Sweep Orders and `.B` [CTA] or `.W` [UTP] for Weighted Average Price). On-exchange and off-exchange results are shown separately. The sample covers the pre-TSP period from May 1 to September 1, 2016.

	On-Exchange		**Off-Exchange**	
TSP_Group	BJZZ (%)	QMP (%)	BJZZ (%)	QMP (%)
Panel A: Retail Share (%)				
Control Group	0.08%	0.08%	17.64%	16.45%
Group 1	0.07%	0.07%	16.56%	15.55%
Group 2	0.08%	0.08%	17.05%	15.95%
Group 3	0.08%	0.08%	17.42%	16.33%
Panel B: Retail Share, COND (%)				
Control Group	0.08%	0.08%	16.36%	15.26%
Group 1	0.07%	0.07%	15.18%	14.25%
Group 2	0.08%	0.08%	15.71%	14.69%
Group 3	0.08%	0.08%	16.19%	15.19%

Table A.7
DiD Estimates of Off-Exchange Retail Classification under Asymmetric Tick Constraints: Control Group vs. Group 1

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how off-exchange retail trade classifications for Control Group stocks changed relative to Group 1 after the Tick Size Pilot (TSP) began. Group 1 stocks were constrained to a $0.05 tick size for quoting, although not trading, while Control Group faced no constraints. The comparison is designed to assess whether Control and Group 1 exhibited similar or divergent behavior in retail classification following the pilot. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \text{OFF\_C}_{s,t} \times \text{TSP}_{s,t} + \text{FE} + \epsilon_{s,t}$ where $\text{OFF\_C}_{s,t}$ equals one if stock s belongs to Control Group and trades off-exchange, and zero if it belongs to Group 1 and trades off-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The dependent variables capture various measures of off-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of off-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
OFF_C × TSP	0.018***	-0.261***	0.018***	-0.235***	-0.018***	-0.454***	-0.017***	-0.443***
	(4.80)	(-7.56)	(5.01)	(-6.85)	(-4.79)	(-13.03)	(-4.43)	(-12.88)
Num. Obs.	116,779	116,779	116,779	116,779	116,779	116,779	116,779	116,779
Adj. R^2	0.321	0.589	0.347	0.572	0.291	0.597	0.316	0.580
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
OFF_C × TSP	0.019***	-0.260***	0.019***	-0.235***	-0.016***	-0.460***	-0.014***	-0.450***
	(5.37)	(-7.46)	(5.54)	(-6.73)	(-4.47)	(-13.05)	(-4.13)	(-12.89)
Num. Obs.	656,082	656,082	656,082	656,082	656,082	656,082	656,082	656,082
Adj. R^2	0.357	0.545	0.389	0.523	0.328	0.555	0.358	0.534
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table A.8
DiD Estimates of On-Exchange Retail Classification under Asymmetric Tick Constraints: Control Group vs. Group 1

This table reports fixed-effects panel regressions in a difference-in-differences design, comparing how on-exchange retail trade classifications for Control Group stocks changed relative to Group 1 after the Tick Size Pilot (TSP) began. Group 1 stocks were constrained to a \$0.05 tick size for quoting, although not trading, while Control Group faced no constraints. The comparison is designed to assess whether Control and Group 1 exhibited similar or divergent behavior in retail classification following the pilot. The pre-TSP period is defined as May 1 to September 1, 2016, and the post-TSP period spans November 01, 2016, to March 2, 2017. Two classification methods—BJZZ Boehmer et al. (2021) and quote midpoint (QMP) Barber et al. (2024)—are compared across two levels of time aggregation (daily and hourly). The regression model is specified as: $\text{RetailClassification}_{s,t} = \beta \cdot \text{ON\_C}_{s,t} \times \text{TSP}_{s,t} + \text{FE} + \epsilon_{s,t}$ where $\text{ON\_G1}_{s,t}$ equals one if stock s belongs to Control Group and trades on-exchange, and zero if it belongs to Group 2 and trades on-exchange. $\text{TSP}_{s,t}$ equals one in the post-TSP period. The dependent variables capture various measures of on-exchange retail trading identified by either BJZZ or QMP. `Share` refers to the proportion of on-exchange volume classified as retail; `LogCount` is the natural logarithm of one plus the number of classified trades; `COND` excludes trades with sale condition codes `F`, `B`, or `W` from being classified as retail, as these conditions are typically associated with institutional orders. `QMP` excludes trades priced within the 40% to 60% range of the NBBO, in addition to excluding trades without meaningful price improvement. Panel A uses daily aggregation, while Panel B uses hourly aggregation. All regressions include stock-by-day fixed effects; hourly models additionally include fixed effects for one-hour time intervals. Standard errors are double clustered at the stock and date. Coefficients are reported with t-statistics in parentheses. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	BJZZ				**Quote Midpoint**			
	Including ITF		Excluding ITF		Including ITF		Excluding ITF	
	Share	LogCount	Share	LogCount	Share	LogCount	Share	LogCount
Panel A: Daily Aggregation								
ON_C × TSP	0.002	-0.065***	0.001	-0.068***	0.000	-0.488***	-0.001	-0.405***
	(1.60)	(-3.55)	(0.92)	(-3.88)	(0.04)	(-14.54)	(-1.39)	(-14.01)
Num. Obs.	116,946	116,946	116,946	116,946	116,946	116,946	116,946	116,946
Adj. R^2	0.187	0.221	0.086	0.218	0.177	0.298	0.078	0.276
StockDayFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Hourly Aggregation								
ON_C × TSP	0.002*	-0.065***	0.001	-0.068***	0.000	-0.528***	-0.001	-0.437***
	(1.79)	(-3.11)	(1.40)	(-3.43)	(0.34)	(-14.88)	(-1.02)	(-14.22)
Num. Obs.	670,304	670,304	670,304	670,304	670,304	670,304	670,304	670,304
Adj. R^2	0.245	0.203	0.152	0.199	0.236	0.273	0.139	0.252
StockDayTimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

References

Aghbabali, T., W. S. Choi, T. Im, D. Rösch, and P. Roy (2025). How to: Dynamic reports and reproducible research. *Working Paper*.

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman (2021). Institutional parent and child orders and off-exchange executions: Evidence from finra oats. Technical report, Working Paper.

Barardehi, Y., D. Bernhardt, Z. Da, and M. Warachka (2022). Institutional liquidity costs, internalized retail trade imbalances, and the cross-section of stock returns. *Working Paper*.

Barber, B. M., X. Huang, P. Jorion, T. Odean, and C. Schwarz (2024). (sub)penny for your thoughts: Tracking retail investor activity in taq. *Journal of Finance*. Forthcoming.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance 55*(2), 773–806.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Barrot, J.-N., R. Kaniel, and D. Sraer (2016). Are retail traders compensated for providing liquidity? *Journal of Finance 120*, 146–168.

Battalio, R., R. Jennings, M. Saglam, and J. Wu (2024). Difficulties in obtaining a representative sample of retail trades from public data sources. Technical report, Working Paper. Available at SSRN: 4579159.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 289–300.

Boehmer, E., C. M. Jones, and X. Zhang (2008). Which shorts are informed? evidence from short interest changes. *Journal of Finance 63*(2), 491–527.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2021). Tracking retail investor activity. *The Journal of Finance 76*(5), 2249–2305.

Bradley, D., R. Jame, and J. R. Williams (2022). Non-deal roadshows, informed trading, and analyst conflicts of interest. *Journal of Finance 77*(6), 2633–2671.

Chakravarty, S., P. Jain, J. Upson, and R. Wood (2012). Clean sweep: Informed trading through intermarket sweep orders. *Journal of Financial and Quantitative Analysis 47*(2), 415–435.

Chung, K. H., D. Rösch, and Y. Zhang (2025). Competition and execution quality in the market for retail trading. Working paper, last revised June 11 2025.

Ernst, T., C. S. Spatt, and J. Sun (2024). Why Did Retail Liquidity Programs Fail? *working paper*.

Fong, K. Y. L., D. R. Gallagher, and A. C. Lee (2014). Individual investors and broker types. *Journal of Financial and Quantitative Analysis 49*(2), 431–451.

Holden, C. W. and S. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance 69*(4), 1747–1785.

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

Kaniel, R., G. Saar, and S. Titman (2008). Individual investor trading and stock returns. *Journal of Finance 63*(1), 273–310.

Kelley, E. K. and P. C. Tetlock (2013). How wise are crowds? insights from retail orders and stock returns. *Journal of Finance 68*(3), 1229–1265.

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance 53*(5), 1775–1798.

Rambachan, A. and J. Roth (2023, 02). A more credible approach to parallel trends. *The Review of Economic Studies 90*(5), 2555–2591.

Rindi, B. and I. M. Werner (2017). U.s. tick size pilot. *Working Paper*.

Scott, C., G. Bellala, and R. Willett (2009). The false discovery rate for statistical pattern recognition. *Electronic Journal of Statistics 3*, 651–677.

Soric, B. (1989). Statistical discoveries and effect-size estimation. *Journal of the American Statistical Association 84*(406), 608–610.