Who Is Minding the Store? Order Routing and Competition in Retail Trade Execution

Xing Huang, Philippe Jorion, Jeongmin Lee, and Christopher Schwarz*,†

July 22, 2024

Abstract

Using 150,000 actual trades, we study the U.S. equity retail broker-wholesaler market, focusing on brokers' order routing and competition among wholesalers. We document substantial and persistent dispersion in execution costs across wholesalers within brokers. Despite this, many brokers hardly change their routing and even consistently send more orders to the more expensive wholesalers, although there is considerable variation among brokers. We also document a case where, after a new wholesaler enters, existing wholesalers significantly reduce their execution costs. Overall, our findings and theoretical framework highlight the heterogeneity across brokers and are inconsistent with perfect competition in this market.

JEL Classifications: G12, G14, G50

Keywords: retail trading, execution quality, order routing, competition, market microstructure, regulation

*Christopher Schwarz (corresponding author, email: cschwarz@uci.edu) and Philippe Jorion are at the Paul Merage School of Business at the University of California at Irvine. Xing Huang is at the Olin Business School, Washington University in St. Louis. Jeongmin "Mina" Lee is at the Federal Reserve Board, Washington, DC. The views expressed here are the authors' own and do not reflect the views of the Federal Reserve System or its staff. We thank Alyssa Moncrief and Sanath Nair for research assistance. This paper benefited from helpful comments by numerous friends, colleagues, and conference participants at the 2023 ESSFM Asset Pricing Conference, 2024 MFA Conference, 2024 Northeastern University Finance Conference, SFS Cavalcade 2024, 9th Annual University of Connecticut Finance Conference, 10th "Women in Microstructure" Meeting (WiMM), as well as seminars at Baruch College, Columbia University, Cornell University, Iowa State University, Johns Hopkins University, the University of Maryland, Northeastern University, and Notre Dame University. In particular, we wanted to thank Amber Anand, Fatemeh Aramian, Thomas Ernst, Dermot Murphy, Mehrdad Samadi, Chester Spatt, and Mao Ye. All errors are our own.

†All brokerage accounts referenced were funded directly by the authors with personal money. No outside compensation was received from any broker or wholesaler for this study. This paper required no outside approval. Jeongmin Lee did not and will not contribute any funds or input to the design and execution of trading. Since Jeongmin Lee and her spouse and dependent children are not participating in any form to the trading process, the paper does not violate any ethics rules or Board policies.

1. Introduction

Retail investors' access to equity markets has vastly improved over the last few decades. In the U.S., the adoption of Regulation NMS (SEC (2005)) led most retail brokers to route customers' orders to specialized OTC market makers, called "wholesalers." In exchange for retail order flows, wholesalers provide direct compensation to most brokers, called payment for order flow (PFOF), and indirectly compensate investors through reduced effective bid-ask spreads relative to those quoted on exchanges, which is also known as "price improvement." Further, most retail brokers eliminated commissions in late 2019. These changes led to major reductions in trading costs. Even so, trading is not free, and the actual costs investors incur are determined by the execution quality provided to retail investors.

In this paper, we study brokers' order routing and wholesaler competition to provide insight into the execution quality of retail trades.[1] On the one hand, there are only four major players currently (Citadel, Virtu, G1X, and Jane Street), which raises a concern that the wholesaler market is too concentrated and has led to scrutiny from policy-makers.[2,3] On the other hand, brokers might be able to enforce price (Bertrand) competition among wholesalers. Although brokers do not have a direct economic stake in execution costs (which are entirely borne by investors), they do have a legal duty of "best execution." Brokers are

[1] Another element that can affect the execution quality is PFOF. However, the literature (e.g., Bartlett et al. (2023), Ernst et al. (2024), Levy (2022), Schwarz et al. (2023)) has largely found that PFOF is not related to execution quality for U.S. equities and that the economic magnitude of PFOF may be too small to explain large variations in execution quality. In addition, PFOF is the same across wholesalers within brokers, which should not influence routing decisions.

[2] Hu and Murphy (2022) argue that the two largest wholesalers in this market account for 70% of volume.

[3] In 2022, the SEC issued four proposals that would fundamentally revise the regulations governing the structure of U.S. equity markets. The proposed changes include: "Disclosure of Order Execution Information" (SEC (2022a)), which was recently adopted (SEC (2024)), "Order Competition Rule" (SEC (2022b)), "Regulation Best Execution" (SEC (2022c)), and "Regulation NMS: Minimum Pricing Increments, Access Fees, and Transparency of Better Priced Orders" (SEC (2022d)).

expected to monitor the quality of wholesaler execution to ensure that "order flow is directed to markets providing the most beneficial terms for their customers' orders" (FINRA (2014)).

Using over 150,000 independently-generated trades, we find results inconsistent with perfect price competition in this market. Measuring execution costs as effective spreads divided by quoted NBBO spreads (E/Q), we find significant and persistent differences in costs across wholesalers for each broker. Moreover, we find substantial heterogeneity in the responses of brokers to past execution quality, with the majority of brokers continuing to send more orders to more expensive wholesalers. Lastly, we document a case where existing wholesalers significantly lower execution costs following the entry of a new wholesaler. Our theoretical framework based on price (Bertrand) competition with frictions collectively explains our results.

Trading Experiment. Rather than relying on public disclosures that are aggregated at a high level, we provide a unique window into this debate by using data from a nearly two-year long trading experiment. We placed trades on randomly selected stocks at random times across multiple brokerage accounts.[4] To provide a controlled comparison across brokers, trades are placed in a parallel manner (i.e., trades in the same stock of the same order size at the same time). This setup is essential for the following reasons. First, randomization of our trades can mitigate the selection bias that could arise from potential strategic behaviors

[4]We generated a total of approximately 150,000 trades, equivalent to $22.4 million in notional, over the period from December 21, 2021, to May 31, 2023. Our trades are executed through E*Trade, Fidelity, Interactive Brokers (IBKR, with both their Pro and Lite account types), Robinhood, Schwab, and TD Ameritrade. We placed orders at different brokers that were identical in type (market orders), ticker/symbol (stock), size (dollars and shares traded), direction (buy or sell), and submission time (randomized across brokers.) All trades were intraday, i.e., we bought equities after the market opened and then sold them within 30 minutes, with trading spread out throughout the day.

of wholesalers.[5] Second, our approach allows us to directly conduct *within-broker* comparison of execution costs. Focusing on dispersion *within broker* is important since there are large variations of execution costs *across brokers* (Battalio et al. (2001) and Schwarz et al. (2023)). Third, because brokers may have heterogeneous criteria for making routing choices, having a wide range of brokers allows us to examine how different routing choices affect the competitiveness of the wholesaler market place.

Our trades are based on small market orders.[6] Although we show that our findings are robust to a limited sample of round-lot orders, we want to caution that our analysis may be more applicable to understanding how brokers handle small orders, rather than the entire order flow.[7] Nevertheless, small orders are becoming increasingly important.[8] We consider that this segment can still provide valuable insights for highlighting the rich heterogeneity in the behavior of brokers and wholesalers and uncovering the economic mechanisms in this important and growing market.[9]

Interestingly, we find that brokers employ two distinct routing styles. Approximately two-thirds of our brokers route stocks to wholesalers using what we call "proportional" routing. These brokers simply take a "slice" of their aggregate order flow, randomized across stocks,

[5]Wholesalers might tailor their execution quality to brokers' evaluation metrics that are very narrow. Consequently, using archival or public disclosure data may inadvertently bias the sample towards instances where strategic wholesalers apparently offer superior execution. Our approach, in contrast, should provide a more unbiased analysis.

[6]Our target order size is $100, with a minimum order size of a full share.

[7]More generally, we are only examining one segment of a broker's order flow. Different segments, whether sorted by spread, size, or other security characteristics, may look similar or different depending on the broker's exact routing criteria.

[8]As investors are spreading out large trades over time, and smaller-sized trades have become a natural response to the abolition of fixed commissions per order, odd lots (i.e., less than 100 shares) now account for 60% of orders and close to 20% of trading volume. This prevalence is likely even greater among retail trades and high-priced stocks (See `https://www.sec.gov/marketstructure/datavis/ma_overview.html`). Within our random sample of stocks, odd lots are also the second largest size bin in terms of total quoted spread dollar volume.

[9]In addition, a number of authors, such as O'Hara et al. (2014) and Bartlett et al. (2023), have emphasized the importance of odd lots in terms of hidden quotes, as well as price discovery informativeness.

and send it to each wholesaler; the only variation across wholesalers is the dollar size of each slice. In contrast, the remaining one-third of brokers use what we call "selective" routing ("smart-routing" in industry parlance), where the routing for each individual stock is selected from stock characteristics as well as the observed wholesaler execution quality.

Main Results. Our three main results test implications of perfect price (Bertrand) competition and find otherwise. First, under perfect competition, we would expect minimal dispersion in execution costs across wholesalers for any given broker since most orders should go to wholesalers with superior execution when brokers frictionlessly switch across wholesalers based on execution quality. Instead, we find a substantial dispersion. For a given broker, the gap between the maximum and minimum execution costs varies from 42% to 151% of the brokers' average. The dispersion is more pronounced among proportional brokers (70% to 151% of the broker's average) compared to selective brokers (42% to 58% of the broker's average). Moreover, this dispersion is persistent both at the aggregate and stock levels, implying that within-broker execution quality is highly predictable over time.

Second, under perfect competition, we would expect brokers to adjust their routing based on execution quality, shifting market share towards wholesalers with superior execution. Instead, despite significant and persistent dispersion, we find that proportional brokers hardly change their routing based on past execution, while selective brokers do. We also study how the level (not the change) of market share relates to past execution. Surprisingly, there is a positive relation between market share and past execution for proportional brokers, while the relation flips to negative for selective brokers. Thus, proportional brokers consistently allocate a larger fraction of our orders to more expensive wholesalers. To illustrate the economic impact of brokers' limited adjustments, we simulate dynamic routing strategies based on past execution, directing all orders to the lowest-cost wholesaler from the previous month. Our hypothetical routing lowers E/Q by 34% on average, with reductions ranging from 33% to 60% for proportional brokers and from 8% to 14% for selective brokers.

Third, if markets were perfectly competitive, we would expect minimal changes in execution costs following the entry of a new wholesaler, as the existing margins would not allow significant reductions. Instead, we document a case with significant reductions in execution costs among incumbent wholesalers. Jane Street entered as a new wholesaler for Robinhood, our most responsive selective broker, during our sample period in early 2022. Immediately after this entry, we find that incumbent wholesalers lowered E/Q by 14% on average, with reductions varying from 2% to 26%.[10]

Overall, our three main findings show that brokers' responses to execution quality vary substantially and are often limited when handling small, odd-lot orders. Additionally, it appears that there is insufficient competition among wholesalers, even when brokers are responsive. These results suggest that retail investors would benefit from more responsive broker routing and increased wholesaler competition.

Theoretical Framework. We develop a stylized model of price (Bertrand) competition with frictions to explain the puzzling result revealed by our empirical analysis: proportional brokers exhibit a positive relation between market share and past execution cost, consistently sending a larger fraction of their orders to more expensive wholesalers.

In the model, a broker optimally routes orders to two wholesalers. To represent the

[10] The large impact of entry suggests significant barriers to entry in the wholesaler market (e.g., Goolsbee and Syverson (2008)).

observed limited broker responses to execution differences, we assume that the broker faces quadratic switching costs when changing market shares across wholesalers. These switching costs may represent various frictions that limit the broker's ability to adjust its routing.[11] The puzzling positive relation between market share and execution costs, observed among proportional brokers, naturally emerges in equilibrium. The intuition is that the brokers' limited adjustments give wholesalers some scope for exercising market power, such that those with larger market shares also charge higher execution costs.[12]

Discussion. We discuss three main points. First, we show how broker priorities across different segments can be incorporated into our theoretical framework. If a broker prioritizes one segment over another, wholesalers respond by lowering execution costs for the prioritized segment and increasing them for the other. This results in an equilibrium where the broker effectively has higher switching costs for the non-prioritized segment. Thus, switching costs can be broadly interpreted to reflect broker priorities, as well as various real-world frictions.

Second, we explore how varying characteristics of broker order flows, such as order toxicity, may explain the heterogeneity in average execution across brokers. Under the uniform standard (NBBO) enforced by regulation, brokers with more toxic order flows and higher wholesalers' marginal costs face increased regulatory pressure, incentivizing them to choose smaller switching costs. The endogenous choice of switching costs helps explain a puzzling relation observed in the data: selective brokers with lower switching costs tend to have higher (rather than lower) overall execution costs.

[11] The frictions may include the time and costs associated with monitoring execution quality and implementing complex routing changes, managerial and organizational inertia, the desire to supply stable order flows to wholesalers, and so on.

[12] An extension accounting for heterogeneous marginal costs across wholesalers generates a negative relation between market share and execution costs, consistent with our findings for selective brokers.

Lastly, we examine how broker heterogeneity can complicate the interpretation of aggregate public disclosures and provide implications for disclosure policies. Our data show that wholesalers with lower execution costs attract more orders from brokers handling more toxic order flows, leading to higher overall execution costs for these brokers. This complicates comparisons of execution costs across wholesalers. Therefore, detailed, disaggregated data at the broker-wholesaler level is crucial for accurately assessing wholesaler performance. Implementing these public disclosures would likely increase pressure on brokers to improve monitoring and routing practices, potentially increasing competition.[13]

Related Literature. Our paper is most closely related to Dyhrberg et al. (2023) and Ernst et al. (2023).[14] These papers are complementary to ours and offer useful insights from different data sources that also contribute to a better understanding of this market.

Dyhrberg et al. (2023) study the competitiveness of the broker-wholesaler market place using public 605 forms by market centers, which provide execution costs at the wholesaler level. They find that market share and execution costs are negatively related, or that the largest wholesalers offer the lowest execution costs. They also claim that Jane Street's entry into the market did not cause competitors to change their pricing, when comparing two fixed windows around July 2021.

[13]Under the recently adopted disclosure requirements (SEC (2024)), the 605 forms (disclosures of order execution information by market centers) will now also be required for retail brokers, not just wholesalers, and extended to odd lots. However, the proposal still does not require disclosing execution costs for each broker-wholesaler pair. For details on Form 605, see "Disclosure of Order Execution Information" (SEC (2005)).

[14]Bartlett et al. (2023) compare wholesaler execution quality to that on exchanges and show that wholesaler execution is consistently superior to exchanges. Otherwise, previous work on execution quality has mostly focused on different market centers and types of trades. However, the market environment has drastically changed since the advent of zero commissions in 2019. For example, Battalio et al. (2016) and Battalio (2018) use data from 2012 and 2016, respectively, to examine fees and rebates on exchanges, where rebates apply to limit orders.

Despite using different data sources, we are able to replicate their main empirical findings. However, our interpretations are more nuanced based on more granular data. We highlight broker heterogeneity and find that the relation between market share and execution costs is, in fact, positive for proportional brokers. We also find that the lower-cost wholesalers attract more toxic order flows, complicating the comparison of execution quality based on aggregate execution costs due to Simpson's paradox. Finally, the entry of Jane Street occurred at different points in time for various brokers. Our detailed data allows us to isolate one specific (selective) broker, where we find significant reductions in execution costs after the entry.

Ernst et al. (2023) study whether brokers implement best execution using proprietary data provided by three anonymous brokers. They find that brokers respond to wholesaler execution and that wholesalers respond to the broker's change in priorities. Their findings on broker responsiveness generally align with our results for selective brokers but not for proportional brokers.

There are three main differences that may explain this. First, the measure of wholesaler execution is different. We consistently use equal-weighted E/Q of our independently generated trades for all brokers. They use different metrics for each broker based on their own criteria, including different security bins, size categories, and a proprietary score metric.

Second, the broker samples differ. We have six different brokers (four proportional and two selective). They have three anonymous brokers who voluntarily provide their data (only one of their brokers appears to be proportional). Given the stark heterogeneity across brokers, a larger sample of brokers provides a broader comparison of order routing practices across the industry. Additionally, relying on a voluntarily disclosed sample may bias results toward better responsiveness.

Third, our data focuses on small, odd-lot trades, while their data includes a broader range of order sizes. This implies the possibility that our non-responsive proportional brokers may still respond to wholesaler execution of larger, round-lot trades. In Section 6.1, we discuss how brokers prioritizing different segments of order flow can lead to varying responsiveness across segments and how this can be incorporated into our theoretical framework.

Our paper is also related to Ernst et al. (2024) who provide a systematic comparison of the current, status-quo broker-routing market system versus the SEC (2022b) proposed "Order Competition Rule" (OCR), which would rely on order-by-order auctions instead. The authors assume that competition is ensured in the status-quo model because brokers should monitor execution closely and route orders based on each market maker's aggregate performance. In practice, we find that this is not necessarily the case.

More generally, our paper is related to the industrial organization literature. Numerous studies show how market outcomes reflect the exercise of market power in a specific industry.[15] Here, we focus on the broker-wholesaler market. We use switching costs, which have been long used in the literature,[16] as a modeling device to represent brokers' limited adjustments and connect them to market power of wholesalers. We also highlight the role of regulations in providing heterogeneous and insufficient incentives for brokers to reduce switching costs, limiting competition. This connects to earlier studies like Daughety (1984), which show how setting the same price for all firms in an industry can lead to inefficiencies.

The rest of the paper is organized as follows. Section 2 reviews the institutional back-

[15]See for example Porter (1983), Bresnahan (1987), and Nevo (2001). Einav and Levin (2010) provide a comprehensive survey of the empirical industrial organization literature, including studies on market power.

[16]Prior literature mostly focuses on consumer behavior, e.g., Klemperer (1987), Farrell and Klemperer (2007).

ground, in particular interactions between brokers and wholesalers. Section 3 then describes our independently-generated data and patterns in the routing of individual stocks from brokers to various wholesalers. Next, Section 4 delves into our empirical analysis aimed at evaluating the degree of competitiveness in the wholesaler industry. Section 5 then presents a stylized model explaining our empirical findings, while Section 6 provides additional interpretation and implications of our analysis. Section 7 concludes.

2. Institutional Background

Starting in late 2019, almost all retail brokers went commission free, following Robinhood's example in 2015. This was made possible because brokers could route their retail trades off exchanges, directly to wholesalers that execute trades, with generally better pricing for clients. From the brokers' perspective, this setup can not only provide Payment for Order Flow (PFOF) revenues, but it also helps them fulfill their best execution requirements.[17]

Interestingly, the relation between brokers and wholesalers is rather loose, reflecting the nature of their private arrangements.[18] First, the broker selects a pool of wholesalers that satisfy its due diligence requirements. The broker then sets a level of payment for order flow, which can be zero. It is important to note that the broker typically sets "level", or identical, PFOF rates across wholesalers in order to avoid conflicts of interest in routing decisions. Next, the wholesaler can decide whether to accept or not the broker's orders. There are no other contractual obligations: brokers can route orders to any market centers,

[17] One concern is that PFOF could potentially lead to worse execution for retail traders due to its inherent conflict of interest. However, recent research (e.g., Battalio and Jennings (2023), Ernst et al. (2024), and Schwarz et al. (2023)) has largely shown that PFOF does not significantly impact price execution for equity trades. We should note that PFOF has been common for decades, although to a much lesser extent than nowadays.

[18] Schwab (2022), for example, provides an overview of order routing practices for U.S. equities.

and wholesalers do not commit to any set price improvement. Notably, brokers receive no indicative quotes, pre-trade, from wholesalers.

Both the originating broker and the wholesaler, who acts as "executing" broker, are subject to regulatory "best execution" requirements. In the U.S., the broker-dealer industry is overseen by the Financial Industry Regulatory Authority (FINRA), which has issued guidance on best execution practices. According to FINRA (2014)'s Rule 5310, a member firm

> "shall use reasonable diligence to ascertain the best market for the subject security and buy or sell in such market so that the resultant price to the customer is as favorable as possible under prevailing market conditions."

In practice, Section .09 requires the originating retail broker to periodically conduct regular and rigorous reviews of the quality of the executions, at least on a quarterly basis. FINRA (2015) also says that this must include both venues currently used by the broker as well as competing markets. Reviews are not sufficient, however. In addition:

> "In conducting its regular and rigorous review, a member must determine whether any material differences in execution quality exist among the markets trading the security and, if so, modify the member's routing arrangements or justify why it is not modifying its routing arrangements."

Thus, the best execution requirement is for both monitoring the quality of execution and taking action, i.e., changing routing if needed. It should be noted that brokers only observe directly the quality of execution of their *own* orders across wholesalers. Brokers can observe aggregate stock-level execution reported by market centers in their 605 forms, but these numbers are averaged across all trades for that center across all their customers; furthermore, they do not cover odd lots. Finally, it is important to note that using prior performance to change routing assumes that execution quality is persistent and that changes in order routing will not impact price improvement. We will test the first hypothesis.

In practice, best execution is defined very broadly with room for interpretation. Indeed, different brokers may have different objective functions underlying their routing decisions. Brokers may emphasize different aspects of execution, or may focus on special types of trades, e.g., small market orders. Brokers may have different clienteles (e.g., high net worth individuals versus small individual retail traders) that cause the broker to emphasize different types of orders in their routing decisions. In addition, the concept of "best execution" is more holistic than just price improvement (even though this is systematically listed first) and can include additional factors like execution time and fill rates.

From our discussions with the industry, brokers generally provide feedback ("scorecards") on how a wholesaler's price improvement compares to its competitors.[19] If execution is subpar, the broker can advise the wholesaler to provide better price improvement. Of course, brokers also have the option to route more of their orders to different wholesalers.

Practical considerations are also certainly important. For instance, it may be beneficial to keep small allocations to some venues to enable broader and continuous comparisons of execution information. Also, it would be unwise to route all trades to one single venue, even if it had the best execution, because this could lead to less competition in the long run. Reportedly, allocations above 50% would also attract the attention of regulators. In addition, brokers may not be willing to route a majority of their orders to one wholesaler in order to diversify against operational issues such as outages. Some wholesalers may not even

[19] Brokers provide anonymized rankings across their wholesalers. Reportedly, this is generally done across trade size segment, e.g., odd-lots, then in share "buckets" of 100-499, 500-1999, 2000-4999, and above 5,000. These round-lot buckets correspond to those in the 605 forms.

have the technical capabilities to handle multiples of their current trading volumes. More generally, the quality and breadth of services offered by the wholesaler are also important. Such considerations should slow down changes in allocations across wholesalers and can be modeled as switching costs.

3. Trading Experiment and Descriptive Evidence

To obtain granular data at the broker-wholesaler level and conduct within-broker comparisons of execution costs, we implement a trading experiment across multiple brokerage accounts that generates comparative pricing and routing data. This section begins by providing a high-level overview of the trading experiment. We then present the summary statistics and delve into different routing styles utilized by brokers.

3.1. Trading Experiment

Our main source of data is based on independently-generated trades. In summary, we placed trades on a randomly selected sample of stocks at random times across multiple brokerage accounts. Our brokerage accounts include E*Trade (ET), Fidelity (FD), Interactive Brokers (IBKR, with their Pro and Lite account types), Robinhood (RH), Schwab (SC), and TD Ameritrade (TD). All accounts except IBKR Pro charge zero commission.

After (before) the opening (closing) auction, we place round-trip "market" orders for 128 stocks, randomly selected from each of 128 bins sorted on four dimensions – market capitalization, daily volatility, daily share turnover, and stock price. We use market orders because they are the most widely used trade type by retail investors.[20] Our order target

[20] Schwab (2022), for example (p.10), indicates that about 75% of its equity trades are plain market orders.

size is \$100, with a minimum order size of a full share, so most of these trades are odd lots. We also generated a small sample of round-lot trades for comparative purposes.[21] A major feature of our setup is the placement of simultaneous identical trades (i.e., trades in the same stock of the same order size at the same time), which allows for a controlled comparison across brokers. Further details about the experiment can be found in Schwarz et al. (2023) and Barber et al. (2023), which utilize an early version of this dataset.

In total, we placed over 150,000 trades equivalent to \$19.3 million in notional. We supplement our trading data with the TAQ database, which has a complete record of all trades in U.S. equities. We identified each of our trades and retrieved the matching National Best Bid and Offer (NBBO) generated through WRDS.

To evaluate routing decisions and wholesaler execution, it is crucial for us to identify the actual venues to which our trades were sent. TAQ provides a broad classification of trade locations, with most of our trades executed off-exchange and coded simply as "D". This is insufficient, however. To obtain detailed routing information, we rely on SEC rule 606(b)(1), which requires brokers to provide clients with the exact routing of each of their trades over the last six months upon request.

Our data have three main advantages. Firstly, we have independent data for a wide range of major brokers in the industry. Secondly, the granularity in our data far exceeds what is available in public disclosures, which only report aggregate execution statistics and aggregate market shares across wholesalers on a monthly basis.[22] In contrast, we can precisely trace

[21]At the beginning, our trading experiment involved making larger \$1000 trades in parallel with our standard \$100 trades for Robinhood and TD Ameritrade. After six weeks of trading, we found no statistically significant differences in execution costs across these two trade sizes, leading us to close the extended experiment. These larger trades contain round lots, allowing us to compare parallel round-lot and odd-lot trades, each with about 100 round-trip observations. More details can be found in the Appendix.

[22]Specifically, brokers are required to file SEC Rule 606 reports that disclose the fraction of orders routed

the routing and execution for each one of our trades and therefore compare wholesalers at the same broker. In other words, we have the full joint distributions instead of only the marginal ones. Thirdly, the randomization design of our experiment ensures that our trades are uncorrelated to unobservable factors that may affect execution costs, such as the private information of retail investors, market conditions, stock characteristics, etc. These endogeneity concerns complicate analyses and interpretations when relying on endogenously placed trades in proprietary archival data or public disclosure data.

We also want to point out that since our experiment focuses on odd lot orders and only examines market orders for equities, caution should be exercised when generalizing our results to larger orders or other order types. However, despite odd lot trades not being covered by SEC Rule 605 filings, they constitute a majority of retail trades. We consider that this segment still provides a valuable perspective for exploring the behavior of brokers and wholesalers and understanding the economic mechanisms in the retail execution marketplace.[23]

3.2. Summary Statistics

In Table 1, we provide summary statistics on order routing for each of our brokerage accounts. The row totals give the total number of trades for each brokerage account, whereas the column totals give the total number of trades sent to each wholesaler across all accounts. Panel A shows the total number of trades across brokers and wholesalers, whereas Panel B shows the routing composition in percentages by brokers (also displayed in Figure 1).

to the venues used. However, there is no execution information. Likewise, venues are required to file SEC Rule 605 reports that display detailed execution statistics broken down by stock, but only aggregated across all of their clients.

[23] Odd lots currently are close to 60% of all trades. Based on our random sample of stocks, they are also the second largest size bin in terms of total dollar quoted spread, trailing only 100-499 share trades.

[Insert Table 1 about here]

Most of our commission-free trades are routed to four primary wholesalers: Citadel, Virtu, Jane Street, and G1X, which accounts for 94% of these trades. The remaining trades are mostly sent to Two Sigma (mainly by Robinhood), and to UBS, which saw a reduction in market share over this period, along with other venues. In contrast, the routing composition in the IBKR Pro account differs significantly, with the majority of trades sent to IBKR's own Alternative Trading System (ATS) and to exchanges.[24]

[Insert Figure 1 about here]

We use effective over quoted (E/Q) spread to measure execution cost. For buy trades, for example, the "effective" spread is defined as twice the difference between the execution price and the midpoint; this is then scaled by the (NBBO) quoted spread to give a unit-less ratio. E/Q is directly related to price improvement (PI), defined for buy trades as the ask quote minus the execution price, also scaled by NBBO. Mathematically, E/Q can be expressed as $E/Q = 1 - 2 \times PI$. Therefore, lower E/Q is equivalent to greater PI and indicates reduced execution costs.

Our data reveal substantial variation in order routing as well as execution costs, both over time and in the cross-section. To provide a visual representation of the details in our dataset, Figure 2 plots time-series data for Robinhood and Fidelity as an illustration. In Panels A and C, we report the percentage of our trades sent to each wholesaler over our

[24]ATSs are computerized systems such as Electronic Communication Networks (ECNs) that automatically match buyers and sellers of securities. A "dark pool" refers to an ATS that is not "lit", meaning that it does not publicly display pre-trade quotations. They are less regulated than exchanges but are still subject to the 1998 Regulation ATS. Both ATSs and wholesalers must also operate as broker-dealers, so are still subject to SEC and FINRA oversight. They generally charge no execution fees or fees that are lower than exchanges.

18-month trading period. In Panels B and D, we report the effective over quoted (E/Q) spread for each wholesaler, averaged across our trades.

[Insert Figure 2 about here]

3.3. Broker Routing Styles

Brokers can follow different approaches to routing. Our analysis of the routing data reveals two distinct routing styles. The first is to simply route a percentage of the broker's entire order flow to a wholesaler, which we refer to as "proportional" routing. With proportional routing, each wholesaler receives the same fraction of the broker's order flow across stocks. The second approach is for brokers to route orders on a stock-by-stock basis, which we call "selective" routing. With selective routing, however, different wholesalers would receive different order flow composition across stocks.

The differences between these two routing styles can be illustrated by the variation in the fraction of stock-level orders sent to a specific venue. For each broker, we first compute the percentage of our orders for each stock that are routed to Citadel, as an example, requiring at least 100 trades for a stock to be included. Next, we sort these percentages from lowest to highest. The results are displayed in Figure 3 for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, and TD Ameritrade in Panels A to F, respectively.

The horizontal axis corresponds to each of our stocks. Each vertical line summarizes the distribution of the time series of the percentage routed to the selected venue. The circle represents the historical average, in the middle of whiskers that show 95% confidence bands. Red lines indicate that the stock-level average is significantly different from the overall average for that broker.

[Insert Figure 3 about here]

If a broker uses "proportional" routing, we should observe that all stocks have essentially the same percentage of their orders routed to Citadel. Conversely, if many stocks deviate strongly from the average, the broker must be employing a "selective" routing method. The figure suggests that four of our brokers—E*Trade, Fidelity, Schwab, and TD Ameritrade—use a method close to proportional routing. In contrast, the graphs for IBKR (both Lite and Pro accounts) and Robinhood indicate that these two brokers use selective order routing. Indeed, these two brokers advertise this feature.[25]

To investigate factors that impact order routing, we perform a series of logistic regressions for each broker-wholesaler pair. The dependent variable is set to one if the broker's order was routed to that wholesaler, and zero otherwise. We include variables that may explain routing decisions and are directly observable by the broker, such as prior execution costs, stock characteristics, trade characteristics, etc.

We report results in Table 2 for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, TD Ameritrade, and IBKR Pro in Panels A to G, respectively.[26] Consistent with the patterns suggested by Figure 3, for the four brokers with proportional routing, the coefficients on

[25]Robinhood indicates that "[T]his algorithm, known as the smart order router, prioritizes sending your order to a market maker that's likely to give you the best execution, based on historical performance." IBKR also emphasizes its "SmartRouting" algorithm, which "searches for the best destination price in view of the displayed prices, sizes and accumulated statistical information about the behavior of market centers at the time an order is placed, then immediately seeks to execute that order electronically." See https://robinhood.com/us/en/support/articles/how-robinhood-makes-money/ and https://www.interactivebrokers.com/lib/cstools/faq/#/content/38448530/

[26]Note that not all wholesalers are present in the panels. This is either because the broker did not send any trades to that venue, or because the number of observations is too small, e.g., for UBS. We chose a cutoff point of at least 100 trades to include venues in our analysis.

potential routing determinants are mostly insignificant. In contrast, many variables are significant for Robinhood, IBKR Lite, and IBKR Pro.

[**Insert Table 2 about here**]

4. Competitiveness of Retail Trade Execution

In this section, we examine competition within the wholesaler marketplace where our retail trades are directed. We conduct three analyses to assess the implications for perfect price (Bertrand) competition. We begin by analyzing the dispersion in execution costs across wholesalers within each broker. While the degree of dispersion varies among the brokers in our sample, we generally find substantial and persistent dispersion. Given these predictable execution costs, we then investigate whether brokers adjust their routing based on past execution, a behavior typically assumed to be a prerequisite for a competitive market. Finally, we further explore the degree of competitiveness in the marketplace by studying the impact of a new wholesaler entry on the incumbents.

4.1. Execution Costs Dispersion and Persistence

We begin by examining the implications of the perfect Bertrand competition hypothesis. Under this hypothesis, where brokers can frictionlessly switch among wholesalers based on price execution, only those wholesalers offering the best price execution would attract order flows. This would result in no dispersion in execution costs across wholesalers for a given broker. In Table 3, we present the average E/Q for each broker in Column (1). We note that selective brokers, IBKR Lite and Robinhood, have higher average execution costs, reflecting the greater toxicity of their order flows.

To measure the dispersion, we calculate broker-adjusted E/Q, representing the deviation of a given wholesaler's execution cost from the broker's average, which is shown in Columns (2) through (5). These deviations are averaged across all trades within a month and subsequently averaged across all months in our sample period. The standard error is based on the Fama and MacBeth (1973) approach using Newey and West (1987) with one lag to control for autocorrelation.

[Insert Table 3 about here]

For all brokers, we observe statistically and economically large dispersion in execution costs across wholesalers. When measured relative to the broker-level average E/Q, the gap between the maximum and minimum execution costs varies from 42% to 151%. That dispersion is more pronounced among proportional brokers (70% to 151% of the broker's average) compared to selective brokers (42% to 58% of the broker's average), reflecting the latter's higher execution costs.

Next, we find that the dispersion in execution costs is persistent over time, both at the aggregate and stock levels. To evaluate persistence, we regress the monthly average broker-adjusted E/Q against lagged one-month or three-month average broker-adjusted E/Q. The regression is estimated across all brokers, then separately for proportional and selective-routing brokers, and for each broker individually, with standard errors clustered by month. Results are shown in Table 4.

[Insert Table 4 about here]

Panels A and B show that wholesaler's broker-adjusted performance is highly persistent at the aggregate level. Across all brokers, the prior one-month (three-month) coefficient is 0.72 (0.83), which is economically large and statistically significant, with the R-square exceeding 50%. The slope coefficient close to one suggests a slow reversion to the broker mean, or a gradual adjustment of performance over time. This strong persistence in execution costs is present in both subgroups—proportional and selective-routing brokers—with similar slope coefficients. We also find high persistence within each broker. Figure 4 provides graphical evidence to illustrate this strong relation.

[**Insert Figure 4 about here**]

Next, we perform the same analysis at the individual stock level. For each broker, monthly broker-adjusted E/Q is now calculated at the individual stock level, as the deviation of a given wholesaler's E/Q for a stock from the broker-level average E/Q for that stock, both averaged over a month. Such stock-level analysis is representative of selective routing, where brokers are relying on persistence in price execution at the individual stock level to make routing adjustments. Results are shown in Panels C and D of Table 4.

The results are consistent with those at the aggregate level. We find that execution cost is persistent across all brokers, in both subgroups, and within each broker. For all brokers, the slope coefficients are 0.21 and 0.48, based on the prior one- and three-month execution costs, respectively, and statistically significant. However, both the values of the coefficients and R-squares are systematically lower than those observed at the aggregate level in the previous panels.[27]

Overall, these findings suggest that, for a given broker, there is substantial and persistent

[27]The smaller magnitude is because execution costs at the stock level may be noisier than at the aggregate level, which leads to downward-biased coefficient estimates due to errors in the right-hand-side variables.

dispersion in price executions across wholesalers, which is a prerequisite for active routing decisions.

4.2. Broker Response to Prior Execution Costs

From an economic perspective, since wholesalers' price execution is predictable, brokers can use prior data to adjust their routing practices to reduce execution costs for their customers. From a regulatory standpoint, under best execution obligations, brokers are required to evaluate the execution quality of venues over time and act upon this evaluation. This practice is essential for maintaining competition in the wholesaler market. Brokers reportedly comply with these obligations "by establishing routing allocations based on this historical performance."[28] To explore this empirically, in this section, we examine the extent to which brokers adjust their routing practices in response to wholesalers' past execution performance.

4.2.1. Routing Changes and Prior Execution Costs

To determine how brokers respond to prior execution costs, we first measure a broker's adjustment in routing by computing the monthly change in the percentage of orders routed from the broker to each wholesaler. Next, we regress these routing changes from each broker to each wholesaler against the past broker-adjusted E/Q of that wholesaler, measured over the prior one- and three-month periods. The regression is estimated across all brokers, then separately for proportional and selective-routing brokers, and finally for each broker individually.

We expect a negative relation, reflecting less routing to wholesalers with higher execution costs. Given different routing styles, we conduct analysis both at aggregate level and at

[28] Schwab (2023), p.14.

individual stock level. Our models include brokers dummies when appropriate and cluster standard errors by month.

[Insert Table 5 about here]

We report results at the aggregate level in Panels A and B of Table 5. Across all brokers, we find some evidence that brokers change their routing toward wholesalers that provided lower execution costs the prior month. The slopes are barely statistically significant, however, and the economic magnitude, -0.032, is quite small. Recall that E/Q=+1 is the worst possible pricing and that E/Q=-1 is the best, with the latter meaning buying at the bid and selling at the ask. So, if E/Q were to go down from $+1$ to -1, which is an extreme move, the wholesaler would have its share increased by $-2 \times -0.032 = 6.4\%$ only. Furthermore, these results are almost driven completely by Robinhood, a selective broker. Also, note that only the prior month execution has any statistically significant impact on routing changes.[29]

To further illustrate these routing patterns, we plot changes in market share against prior month excess E/Q in Figure 5. We observe that the slope for proportional brokers is flat while slightly negative for selective brokers.

[Insert Figure 5 about here]

Next, we consider stock-by-stock routing as a function of prior stock execution, with results in Panels C and D. As mentioned, brokers with selective routing should be expected to send relatively more trades for individual stocks to wholesalers with better execution for that stock. Indeed, we find that selective-routing brokers do make changes to their stock

[29]In untabulated results, we also examine weekly changes on prior week excess E/Q and monthly changes on prior two-month execution. The results are similar to those in Table 5.

routing patterns based on prior execution of individual stocks. Again, we find no response for proportional brokers.

Overall, our results are not consistent with perfect competition. Most brokers either do not or cannot make changes to their routing patterns that are likely to improve price execution for our types of trades. Either these brokers are unable or unwilling to actively change routing across wholesalers, or the wholesaler market is not sufficiently competitive. Alternatively, brokers could be prevented from altering their routing due to frictions such as the practical considerations mentioned in Section 2.

4.2.2. Market Share and PriorExecution Costs

In contrast, Dyhrberg et al. (2023) conclude that the wholesaler market is competitive using Form 605 data. They find that wholesalers with better price execution tend to have a higher percentage of trades routed to them. While we analyze changes in market shares, they focus on levels instead. Hence, for comparison purposes, we also run our analysis using levels instead of changes. Each month, we regress the percentage of our trades routed to each wholesaler against its prior month's excess E/Q for each broker. We present results for the full wholesaler sample as well as a "Top 4" subsample (Citadel, Virtu, Jane Street, and G1X), which receives 94% of our orders. Results are shown in Table 6.

[Insert Table 6 about here]

For the full set of wholesalers and brokers, which is the most comparable to these authors' analysis, we indeed match their results. The negative slope of -0.29 indicates that better price execution, or lower cost, is associated with greater market share for wholesalers. For every 0.01 lower excess E/Q, a wholesaler receives 0.29% more share. However, the table also shows that this result is much stronger for our selective-routing brokers than our proportional-routing brokers. Notably, only three of our six brokers show significant coefficients with the expected negative signs.

The previous regressions, however, put equal weights on all wholesalers, including those with tiny market shares. In contrast, for the "Top 4" subsample (Panel B), results differ sharply across the two broker groups. While selective-routing brokers still have a very significant negative relation, our proportional brokers now display a positive, significant relation.

To illustrate this point, consider Fidelity. As shown in Panel D of Figure 2, Citadel is, on average, the laggard among the "Top 4" wholesalers for our trade execution at that broker. Virtu has the best execution, with Jane Street and G1X in between. So, one would expect a similar ranking of market shares. However, Panel C shows that Citadel receives the most orders, with a share around 40%, which has been relatively stable over our 13 months of trading.[30]

To illustrate the evidence across brokers, Figure 6 plots the overall relation between market share and execution costs. Panels A and B break down the sample into proportional and selective brokers, respectively. Selective brokers display the expected negative relation between higher cost and lower shares. In contrast, this relation is positive for proportional brokers. This positive sign seems puzzling but has an economic interpretation, as we shall

[30]As an aside, it is interesting to note that execution costs for Fidelity have sharply decreased over this period, from an average E/Q of 0.30 to around 0.10, which is a remarkable improvement. As discussed previously, best execution certainly has many dimensions, across types of orders, trade sizes, and execution metrics. Our sample focuses on E/Q for our odd-lot market orders and our statistics involve equally weighted averages across all orders. It is certainly possible that Fidelity receives better execution than average from Citadel for non-odd lot orders. Even so, it should be feasible to establish different routing patterns across trade sizes so that small retail investors using odd lots could enjoy better execution.

see in the next section.

[Insert Figure 6 about here]

Finally, we investigate whether these results extend to our sample of approximately 100 round-trip trades in round lots, which are compared to simultaneous trades in odd lots for two brokers, Robinhood and TD. As described in the Appendix, we continue to find a negative relation between market share and execution costs for our selective broker, in contrast with that for our proportional broker. So, there is no evidence that the relation we found for our odd lots would differ for round lots, albeit with a limited sample.

4.2.3. Counterfactual

Next, we run a counterfactual analysis to evaluate the potential improvement in our execution from actively routing orders based on past execution. Each month for each broker, we simulate rerouting all of our orders to the wholesaler who had the best execution during the prior month. All of our trades are then assigned the average execution for each stock which that broker received from that wholesaler in the current month. We also run an analysis where we perform stock-by-stock routing based on prior month stock execution. Note that in both cases our hypothetical experiment must assume that such rerouting would not alter our trade execution nor the competitive dynamics of the wholesaler market -- which is reasonable given the small size of trades. We report our results in Table 7 with overall and stock-by-stock execution in Panels A and B, respectively.

[Insert Table 7 about here]

The columns show the original E/Q, the hypothetical E/Q, as well as the absolute and relative difference, for each broker. When using overall, aggregate execution, the effective spread decreases from an average across brokers of 0.313 to 0.233. Averaging relative changes across brokers gives a reduction in trading costs of 34%, which is a major improvement. While statistically significant for all brokers, the absolute changes for the proportional-routing brokers are much larger than for the selective-routing brokers. Using stock-by-stock routing creates some improvement as well, but the changes are more muted. This is likely due to the lower stock-by-stock persistence we documented earlier, but also less extreme allocations across wholesalers.

Overall, these results show that the lack of routing changes documented in the prior section leads to significantly worse price execution than is theoretically achievable. In practice, we concede that this method would create extreme swings in routing that are not realistic. Still, even partial adjustments are likely to result in large E/Q improvements.

4.3. Impact of a New Wholesaler on Price Execution

One potential reason why brokers are not altering their routing is a lack of perfect competition in the wholesaler market. In this section, we examine how the entry of a new firm impacts competition. Specifically, Jane Street entered the retail wholesaler market in 2020, gaining significant market share progressively across brokers.[31] Unfortunately, most of these entrances predate our trading experiment, which starts in early 2022. However, Jane Street did not become a wholesaler for Robinhood until the first quarter of 2022, which is in our sample.

When we began trading, none of our orders were routed to Jane Street. By February 24,

[31]Based on Form 606 filings, Jane Street became a market center for Fidelity in the second quarter of 2020, for E*Trade in the second quarter of 2021, and for TD Ameritrade in the fourth quarter of 2021.

2022, almost a quarter of our trades were routed to Jane Street, as shown in Panel A in Figure 2. During the initial period, Jane Street provided very low trading costs, even negative (Panel B). This amount was not economically sustainable and, once Robinhood started allocating more trades to Jane Street, its trading cost went back to a level comparable to the best other wholesalers.

To formally evaluate the impact of this new entrant, we examine changes in two wholesaler characteristics, i.e., market share and E/Q for Robinhood's wholesalers, before and after February 24, 2022. If Jane Street increased competition, we should see a lower allocation to other venues and a decrease in execution costs. Table 8 shows changes in venue routing and execution costs in Panels A and B, respectively.

[**Insert Table 8 about here**]

The entry of Jane Street significantly impacted the wholesaler market for Robinhood. Its allocation to Jane Street went from 3% to 23%, leading to large drops in shares for Virtu, Citadel, and G1X. We also see that, while not always statistically significant, execution costs decreased for all wholesalers after the new entry. Citadel decreased its E/Q cost sharply, from 0.54 to 0.40. For Robinhood overall, the average execution cost decreased from 0.55 to 0.47, which is economically significant. Overall, these results suggest that the wholesaler market benefited from this additional competition.

There are two potential explanations for the observed increase in overall execution quality at Robinhood. The first is that existing wholesalers raise their execution quality across all trades in response to increased competition. The second is that, as part of its selective-routing system, Robinhood systematically reroutes individual stock trades with the worst execution from its existing wholesalers to Jane Street. To investigate this latter explanation, we regress the change in the wholesaler's share in that stock across periods against its initial excess execution cost. If the change was driven by selective-routing decisions, we should see negative, significant coefficients, meaning that higher E/Q should lead to lower share allocations. Table 9 shows the regression results, which include some stock-level controls.

[**Insert Table 9 about here**]

The table shows insignificant coefficients in the first row. Thus, routing changes were not driven by the wholesaler's individual stock execution quality relative to its peers in the weeks leading up to Jane Street's addition. This suggests that Robinhood used Jane Street's entrance to benefit from better execution across wholesalers.

Our results are in contrast to Dyhrberg et al. (2023) who conclude that the entry of Jane Street did not impact execution between the second and fourth quarters of 2021. Several factors could drive the different conclusions. First, our data is more detailed, providing exact execution data for specific stocks routed by a specific broker on a daily basis; in contrast, the SEC Rule 605 reports provide stock-level execution statistics for each market center aggregated across all their clients on a monthly basis. Second, we examine one broker only, whereas 605 reports provide averages across all brokers. This allows us to focus on an actual date of entry for Jane Street, which should be more precise, given that its addition to each broker's list of venues was spread over different periods. Third, almost all of our trades are odd lots, which are not reported in 605 reports. So, at least for odd-lot trades at Robinhood, Jane Street's entry significantly altered order routing and improved price execution.

5. Theoretical Framework

We now present a theoretical framework based on price (Bertrand) competition with frictions. Our goal is to provide economic insights into how brokers' limited response allows wholesalers to exercise market power, leading to the puzzling positive relation between market share and execution costs observed for proportional brokers, as well as the negative relation observed for selective brokers (in Section 4.2.2).

5.1. *Setup*

Consider a generic broker, which can route the orders from its retail customers to two wholesalers X and Y. The broker and the two wholesalers are risk neutral. The size of all orders is normalized to one.[32] The initial market share of wholesaler X is given by $\sigma \in [0, 1]$ with the remaining $1 - \sigma$ routed to wholesaler Y. Denote by p_X and p_Y the execution costs (i.e., E/Q) charged to the broker's customers by the wholesalers X and Y, respectively.

To represent brokers' limited response to execution quality we document in Section 4.2.1, we assume that brokers face frictions when adjusting the market shares between the wholesalers. Specifically, allocating an additional market share $\Delta \in [-\sigma, 1 - \sigma]$ to wholesaler X incurs "switching" costs of $\frac{s}{2}\Delta^2$, where $s > 0$.

These switching costs could reflect various real-world frictions, such as the time and cost it takes for brokers to monitor wholesalers' performance, the lack of technology to implement complex routing allocations, and managerial or organizational inertia. Alternatively, large switching costs may result from brokers prioritizing other order segments, as discussed later

[32] In Section 6.1, we consider a model extension with multiple order segments.

in Section 6.1.

The broker optimally chooses an adjustment of the market share Δ, taking as given the unit execution costs charged by the two wholesalers p_X and p_Y, to minimize total costs, including both execution and switching costs:

$$\min_{\Delta \in [-\sigma, 1-\sigma]} (\sigma + \Delta)p_X + (1 - \sigma - \Delta)p_Y + \frac{s}{2}\Delta^2. \tag{1}$$

Wholesalers incur constant marginal costs f to process and make markets for the broker's (customers') orders.[33] Each wholesaler optimally chooses its unit execution cost to maximize its profits, taking as given the other wholesaler's execution cost and the broker's response to both execution costs:

$$\max_{p_X} (p_X - f)(\sigma + \Delta); \tag{2}$$

$$\max_{p_Y} (p_Y - f)(1 - \sigma - \Delta). \tag{3}$$

5.2. Equilibrium

From the broker's objective function in Equation (1), the first-order condition implies[34]

$$\Delta = \frac{p_Y - p_X}{s}. \tag{4}$$

The broker always moves towards the lower-cost wholesaler (i.e., $\Delta > 0$ if and only if $p_X < p_Y$), but the adjustment is limited by the switching costs.

[33] More generally, marginal costs may differ across wholesalers. We address this in Section 5.3.

[34] The solution requires that the difference in execution costs are sufficiently small, which is satisfied in equilibrium; see the proof in the Appendix for details.

Taking into account the broker's strategy in Equation (4), jointly solving the first-order conditions from the wholesalers' objective functions in Equations (2) and (3) yields

$$p_X = f + \frac{s(1+\sigma)}{3} \qquad \text{and} \qquad p_Y = f + \frac{s(2-\sigma)}{3}. \tag{5}$$

The broker's limited adjustments allow the wholesalers to exercise their market power, with the markup increasing in the switching cost parameter s.

Further, wholesalers extract more rents from the broker (or more precisely, its customers) as their market share increases. From Equation (5), we have

$$p_X > p_Y \qquad \text{if and only if} \qquad \sigma > 1/2. \tag{6}$$

Hence, the wholesaler with a larger pre-existing market share (i.e., X if $\sigma > 1/2$ and Y if $\sigma < 1/2$) charges more than the other wholesaler.

The broker does move away from the higher-cost, larger wholesaler. Substituting Equation (5) into Equation (4), we have the broker's equilibrium adjustment:[35]

$$\Delta = \frac{1-2\sigma}{3}. \tag{7}$$

Notice that the equilibrium execution costs ensure that the larger wholesaler remains

[35] Since the broker moves away from the larger wholesaler, the market share between the two wholesalers becomes more balanced after the adjustment. Theoretically, if this adjustment continues over multiple periods, market shares could eventually be evenly split. However, this assumes a perfectly static environment. In reality, many factors can change. For example, the broker's order flow may change as the customer base and preferences evolve. This can affect wholesalers' marginal costs differently, preventing market shares from converging to an even split. In addition, some realistic restrictions on maximum and minimum values for the *levels* of weights, instead of changes, may prevent market shares from being evenly split, even after a very long period of time.

large even after the adjustment (i.e., $\sigma + \Delta = (1+\sigma)/3 > 1/2$ if and only if $\sigma > 1/2$), even though its unit execution cost is higher.

Thus, at the equilibrium, this model predicts a positive relation between market share and execution costs, consistent with what we find for proportional brokers in Figure 6. This relation is also illustrated in Panel A of Figure 7. While this positive relation may seem counter-intuitive at first, our theoretical framework provides economic insight. Brokers do not route more orders to some wholesalers because they charge higher costs; instead, some wholesalers can charge higher costs because brokers route more orders to them. Brokers' limited adjustments allow wholesalers to exercise market power, and the optimal exercise of this power generates a positive relation between market share and execution costs.

[**Insert Figure 7 about here**]

5.3. Extension

Next, we consider an extension where the two wholesalers have different marginal costs ($f_X < f_Y$). As we shall see, this extension generates both the positive relation between market share and execution costs, as discussed above, and the negative relation we observe for selective brokers (Panel B in Figure 6). The full characterization of equilibrium is in the Appendix (Proposition 1).

Holding the difference between the two wholesalers constant, there are three cases depending on the size of switching costs (see Panel C of Figure 7). When switching costs are low (i.e., $s \leq (f_Y - f_X)/(2-\sigma)$), there is no relation between market share and execution costs. The lower-cost wholesaler drives the other wholesaler to zero profits and obtains a

100% market share, as in textbook Bertrand competition.

When switching costs are medium (i.e., $s \in \left(\frac{f_Y - f_X}{2-\sigma}, \frac{f_Y - f_X}{2\sigma - 1}\right)$), the relation between market share and execution costs can be negative, consistent with what we find for selective brokers in Figure 6. This is described in Panel B of Figure 7. If the lower-cost wholesaler (X) initially has a larger market share ($\sigma > 1/2$), it continues to charge lower unit execution costs and maintains a larger market share. The larger wholesaler has two competing incentives: charging more to exploit its market share or charging less to leverage its lower costs for a bigger market share. When the broker's switching costs are not too high, the incentive to lower costs prevails because this maximizes total profits, leading the larger wholesaler to charge lower costs than the smaller wholesaler in equilibrium. This case predicts a negative relation between market share and execution costs.

Finally, when switching costs are high (i.e., $s \geq (f_Y - f_X)/(2\sigma - 1)$), or when switching costs are medium but the lower-cost wholesaler initially has a smaller market share, there is a positive relation between market share and execution costs, as discussed in Section 5.2. The optimal exercise of market power implies that the larger wholesaler charges more but keeps the difference moderate to maintain its large market share.

In sum, the relation between market share and execution costs can be positive with high switching costs, and negative with low switching costs. This is what we observe for proportional and selective brokers, respectively. Thus, our empirical findings are consistent with proportional brokers facing higher switching costs relative to selective brokers.

6. Discussion

In this section, we discuss three main points. First, we show how broker priorities across different segments can be incorporated into our theoretical framework. Second, we explore how varying characteristics of broker order flows, such as order toxicity, may explain the heterogeneity across brokers. Third, we examine how broker heterogeneity can complicate the interpretation of aggregate public disclosures and discuss the implications for disclosure policies.

6.1. Broker Priorities and Switching Costs

In the model, switching costs are used as a simple device to represent brokers' limited response to execution, consistent with our empirical evidence, helping us understand how wholesalers exercise market power and how this explains the puzzling patterns between market share and execution costs. While switching costs may derive from various real-world frictions brokers face, as discussed previously, switching costs may also be alternatively interpreted to reflect broker priorities across different segments of their order flows.

We document that proportional brokers hardly respond to execution quality (Section 4.2). Since we do not observe the entire order flow and our data is based on small orders, it is possible that proportional brokers may still respond to other order segments, such as larger, round-lot orders. Although fully investigating this alternative hypothesis would require a different dataset,[36] we provide an extension of our model that accommodates this possibility.

[36] As indicated in Section 4.2.2, however, we find that execution costs are highly correlated across order sizes based on a limited sample of round-lot trades that we placed.

Order Routing with Two Segments. Consider a proportional broker with two segments of order flows: segment 1 with a fraction w and segment 2 with a fraction $1-w$, where the weights are defined for example in terms of dollar ratios. Assume that the broker prioritizes segment 2 in its scorecard, such that the broker's assigned weight on segment 1, denoted by w', is less than w. Then the broker's optimal routing problem in Equation (1) becomes

$$\min_{\Delta\in[-\sigma,1-\sigma]} (\sigma+\Delta)\left(w'\,p_{1X}+(1-w')p_{2X}\right)+(1-\sigma-\Delta)\left(w'\,p_{1Y}+(1-w')p_{2Y}\right)+\frac{s}{2}\Delta^2, \quad (8)$$

where p_{1j} and p_{2j} denote the execution costs charged by wholesaler $j=X,Y$ to the first and second segments, respectively.

For simplicity, assume that the two wholesalers have the same marginal costs, which are equal to zero. Then the wholesalers' profit maximization problems are

$$\max_{p_{1X},\,p_{2X}} (w\,p_{1X}+(1-w)\,p_{2X})(\sigma+\Delta); \quad (9)$$

$$\max_{p_{1Y},\,p_{2Y}} (w\,p_{1Y}+(1-w)\,p_{2Y})(1-\sigma-\Delta). \quad (10)$$

Intuitively, for any level of p_{1X} and $p_{2X}>0$, wholesaler X can always lower the execution costs of segment 2 ($p_{2X}-\epsilon$) and raise the execution costs of segment 1 ($p_{1X}+(1-w)/w\cdot\epsilon$). This keeps the same total markup but makes the wholesaler more attractive to the broker since the broker prioritizes the second segment ($w'<w$). The same logic applies to wholesaler Y. Thus, in equilibrium, both wholesalers optimally lower the execution costs for segment 2 to zero.

With $p_{2X} = p_{2Y} = 0$, the broker's routing problem in Equation (8) now becomes

$$\min_{\Delta \in [-\sigma, 1-\sigma]} (\sigma + \Delta)(w' \, p_{1X}) + (1 - \sigma - \Delta)(w' \, p_{1Y}) + \frac{s}{2}\Delta^2, \tag{11}$$

which is equivalent to the original optimization problem (1) without order segmentation, except with the new, higher switching cost parameter of $s' := s/w' \geq s$. The wholesalers' problems are identical to Equations (2) and (3). Thus, the equilibrium is identical to that in Section 5.2, where the effective switching costs are higher as brokers prioritize other segments more.

Here, wholesalers understand the broker's priorities and optimally exercise market power by increasing the markup on one segment while decreasing it on another segment. This cross-subsidization is consistent with Ernst et al. (2023), who find that wholesalers respond to changes in broker priorities by increasing execution costs for the non-prioritized segment.

6.2. *Broker Heterogeneity and Order Toxicity*

Our empirical analysis reveals substantial heterogeneity in broker routing patterns, which can be interpreted through the lens of heterogeneous switching costs. Selective brokers implement more active routing choices, which correspond to lower switching costs. However, these selective brokers tend to have higher average execution costs, as shown in Table 3. This relation is somewhat puzzling because, all else being equal, active routing to cheaper venues by selective brokers should lead to lower, not higher, execution costs.

One possible explanation is that switching costs can be endogenously driven by broker characteristics, which are also associated with execution costs.[37] This creates an endogenous

[37]Another potential explanation in the case when Robinhood chose to invest significantly in technology

relation between switching costs and overall execution costs, leading to seemingly puzzling results. Appendix A.2 explores this explanation in detail by endogenizing brokers' switching costs. The main intuition is that brokers with order flows facing higher marginal costs are more prone to violating the NBBO threshold, which creates a stronger incentive for these brokers to select lower execution costs. However, since the reduction in switching costs may not fully offset the higher marginal costs, these brokers still end up with higher overall execution costs in equilibrium.

Among various order flow characteristics, order flow toxicity has received considerable attention in recent literature. "Toxicity" generally refers to continued directionality, or momentum, in order flows. This could arise when informed traders split large orders into consecutive trades to mitigate market impact (e.g., Interactive Brokers investors). Alternatively, directionality could also arise when retail investors coordinate on online platforms, or herd on the same market sentiment (e.g., Robinhood investors). Such directional patterns represent major inventory risks for market makers because prices are pushed systematically against them. In response, to protect themselves against such toxic order flow, wholesalers increase effective spreads, resulting in worse price execution.

Although toxicity is an elusive concept, and difficult to measure,[38] there are indications that order flows characteristics systematically differ across brokers. Eaton et al. (2022) study brokerage platform outages to examine the impact of retail investors on markets. They find that outage-induced reductions in Robinhood retail activity are associated with

to lower switching costs is the $65 million settlement reached by Robinhood with the SEC in 2020. The agreement required Robinhood to pay particular attention to the execution quality of customer orders. See https://www.sec.gov/litigation/admin/2020/33-10906.pdf

[38]Such patterns cannot be detected using conventional U.S. databases because they do not disclose information about the type of order, the customer, the broker, and the execution venue.

lower spreads; in contrast, imbalances deteriorate during outages at other brokers (E*Trade, TD Ameritrade, and Schwab). This is consistent with the hypothesis that Robinhood order flows are more momentum-oriented, hence justifying greater effective spreads than for other brokers.[39]

6.3. Simpson's Paradox: Implications for 605 Reports

Heterogeneity across brokers also has implications for the interpretation of 605 reports. As illustrated above, brokers with more toxic order flows have an incentive to be responsive and route a larger fraction of their orders to relatively cheaper wholesalers. From the viewpoint of wholesalers, however, such rerouting increases the fraction of their executed trades coming from more toxic flows that require wider effective spreads, thereby worsening their overall execution quality when aggregated at the top level across all clients.

This is akin to Simpson's paradox, a well-known phenomenon in probability and statistics, in which a trend appears in several groups of data but disappears or reverses when the groups are combined. For instance, more highly skilled doctors may not produce higher survival rates because they often treat more difficult and sicker patients. This clientele effect distorts comparisons of survival rates.

Likewise, the current 605 forms obfuscate comparisons of execution quality across market centers. This issue might also explain why our findings seem to differ from Dyhrberg et al. (2023) based on these 605 forms. Overall, this discussion demonstrates the advantage of within-broker analysis relative to the aggregate reporting in the 605 forms, and the need for

[39]In addition, Schwarz et al. (2023) examine the effect of order imbalances (OIB) on effective spreads for their trades and find that OIB can generate variations in price execution consistent with the observed economic magnitude of execution differences across brokers.

expanding 605 reports to the broker-wholesaler pairs for more meaningful comparisons.

7. Conclusions

The U.S. equity market is one of the most liquid markets globally. Recently, retail trading has reached record volumes, spurred by technological advances and commission-free trading, which have led to major reductions in trading costs. Even so, there might be room for further improvement given the specific features of this market. Retail brokers who handle their customers' orders lack a direct economic stake in execution quality since all execution costs are borne by their clients. Customers also lack the necessary information to evaluate brokers along this dimension. Only four major wholesalers execute most retail orders. Finally, while brokers have a legal duty of best execution, this term is broadly defined and may not benefit every category of retail investors.

Using over 150,000 independently-generated trades that allows us to look at within-broker interactions between brokers and wholesalers, we find three main results inconsistent with perfect competition in this market. First, we find a large and persistent dispersion in execution costs across wholesalers for the same broker. Second, we find substantial heterogeneity in brokers' routing practices, with several brokers continuing to route more orders to wholesalers with worse execution. Finally, we document a case where the entry of a new wholesaler immediately reduces existing wholesalers' execution costs. We also provide a stylized model that provides economic insights into our results through the lens of switching costs.

Overall, our results suggest that even in this highly liquid market, retail customers could benefit from more active broker routing, increased wholesaler competition, and better public disclosures.

References

Barber, B., X. Huang, P. Jorion, T. Odean, and C. Schwarz (2023). A (sub)penny for your thoughts: Tracking retail investor activity in TAQ. *Journal of Finance*, forthcoming.

Bartlett, R., J. McCrary, and M. O'Hara (2023). The market inside the market: Odd-lot quotes. *Review of Financial Studies 36*, forthcoming.

Battalio, R. (2018). What has changed in four years? Are retail broker routing decisions in 4Q2016 consistent with the pursuit of best execution? In Walter Mattli, ed. *Global Algorithmic Capital Markets*, 147–164. Oxford University Press.

Battalio, R., S. Corwin, and R. Jennings (2016). Can brokers have it all? On the relation between make-take fees and limit order execution quality. *Journal of Finance 71*, 2193–2238.

Battalio, R. and R. Jennings (2023). Wholesaler execution quality. https://papers.ssrn.com/abstract=4304124. Working Paper.

Battalio, R., R. Jennings, and J. Selway (2001). The potential for clientele pricing when making markets in financial securities. *Journal of Financial Markets 4*, 85–112.

Bresnahan, T. F. (1987). Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics*, 457–482.

Daughety, A. (1984). Regulation and industrial organization. *Journal of Political Economy 92*, 932–953.

Dyhrberg, A., A. Shkilko, and I. Werner (2023). The retail execution quality landscape. https://papers.ssrn.com/abstract_id=4313095. Working Paper.

Eaton, G. W., T. C. Green, B. S. Roseman, and Y. Wu (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics 146*(2), 502–528.

Einav, L. and J. Levin (2010). Empirical industrial organization: A progress report. *Journal of Economic Perspectives 24*(2), 145–162.

Ernst, T., A. Malenko, C. Spatt, and J. Sun (2023). What does best execution look like? . Working Paper.

Ernst, T., C. Spatt, and J. Sun (2024). Would order-by-order auctions be competitive? https://ssrn.com/abstract=4300505. Working Paper.

Fama, E. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

Farrell, J. and P. Klemperer (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of Industrial Organization 3*, 1967–2072.

FINRA (2014). Rule 5310: Best execution and interpositioning. https://www.finra.org/rulesguidance/rulebooks/finra-rules/5310. Financial Industry Regulatory Authority.

FINRA (2015). Notice 15-46: Best execution. https://www.finra.org/sites/default/

files/notice_doc_file_ref/Notice_Regulatory_15-46.pdf. Financial Industry Regulatory Authority.

Goolsbee, A. and C. Syverson (2008). How do incumbents respond to the threat of entry? evidence from the major airlines. *Quarterly Journal of Economics 123*, 1611–1633.

Hu, E. and D. Murphy (2022). Competition for retail order flow and market quality. https://ssrn.com/abstract=4070056. Working Paper.

Klemperer, P. (1987). Markets with consumer switching costs. *The Quarterly Journal of Economics 102*(2), 375–394.

Levy, B. (2022). Price improvement and payment for order flow: Evidence from a randomized controlled trial. https://ssrn.com/abstract=4189658. Working Paper.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

O'Hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *Journal of Finance 69*, 2199–2236.

Porter, R. H. (1983). A study of cartel stability: the joint executive committee, 1880-1886. *The Bell Journal of Economics 14*(2), 301–314.

Schwab (2022). U.S. equity market structure: Order routing practices, considerations, and opportunities. https://content.schwab.com/web/retail/public/about-schwab/Schwab-2022-order-routing-whitepaper.pdf. White Paper.

Schwab (2023). SEC market structure proposals. https://www.sec.gov/comments/s7-32-22/s73222-20162957-332913.pdf. Comment on SEC Proposals.

Schwarz, C., B. Barber, X. Huang, P. Jorion, and T. Odean (2023). The 'actual retail price' of equity trades. https://ssrn.com/abstract=4189239. Working Paper.

SEC (2005). Regulation NMS. https://www.sec.gov/rules/final/34-51808.pdf. Securities and Exchange Commission Release No. 34-51808.

SEC (2022a). Disclosure of order execution information. https://www.sec.gov/rules/proposed/2022/34-96493.pdf. Securities and Exchange Commission Release No. 34-96493.

SEC (2022b). Order competition rule. https://www.sec.gov/rules/proposed/2022/34-96495.pdf. Securities and Exchange Commission Release No. 34-96495.

SEC (2022c). Regulation best execution. https://www.sec.gov/rules/proposed/2022/34-96496.pdf. Securities and Exchange Commission Release No. 34-96496.

SEC (2022d). Regulation nms: Minimum pricing increments, access fees, and transparency of better priced orders. https://www.sec.gov/rules/proposed/2022/34-96494.pdf. Securities and Exchange Commission Release No. 34-96494.

SEC (2024). Final rule: Disclosure of order execution information. https://www.sec.gov/

rules/final/2024/34-99679.pdf. Securities and Exchange Commission Release No. 34-99679.

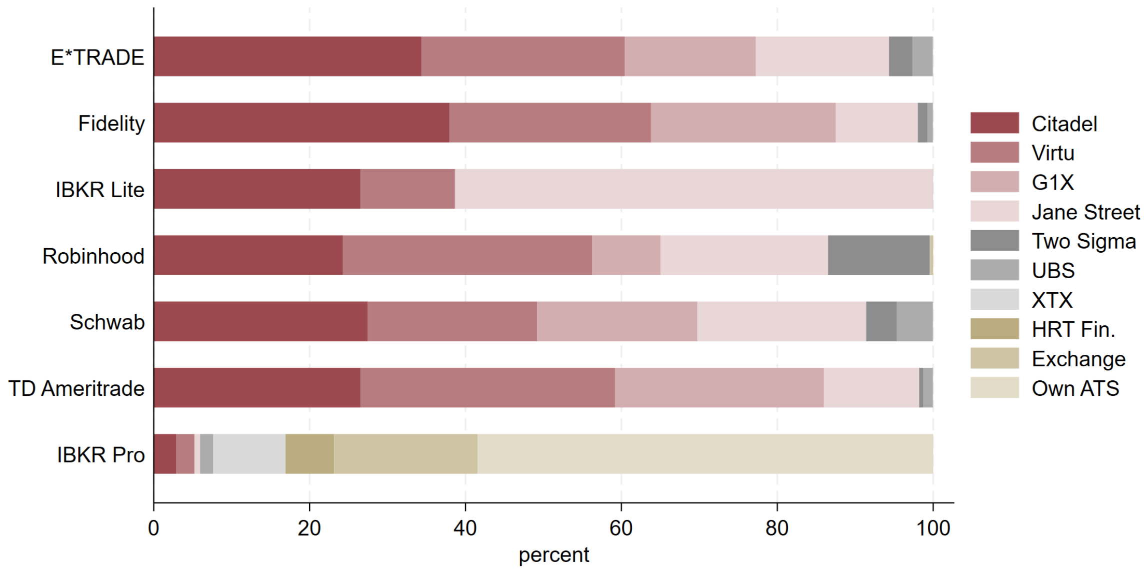


Figure 1: Wholesaler Share by Broker
This figure shows the percent of our orders that went to each wholesaler for each brokerage account. The raw data are in Table 1, Panel B.

Figure 2: Wholesaler Data for Robinhood and Fidelity
This figure graphs the time series of the fraction of our Robinhood and Fidelity orders that go to each wholesaler (Panels A and C) as well as the execution quality measured by the effective over quoted spread (E/Q) from each wholesaler for these trades (Panels B and D). In both cases, we use a rolling average over the last five trading days.

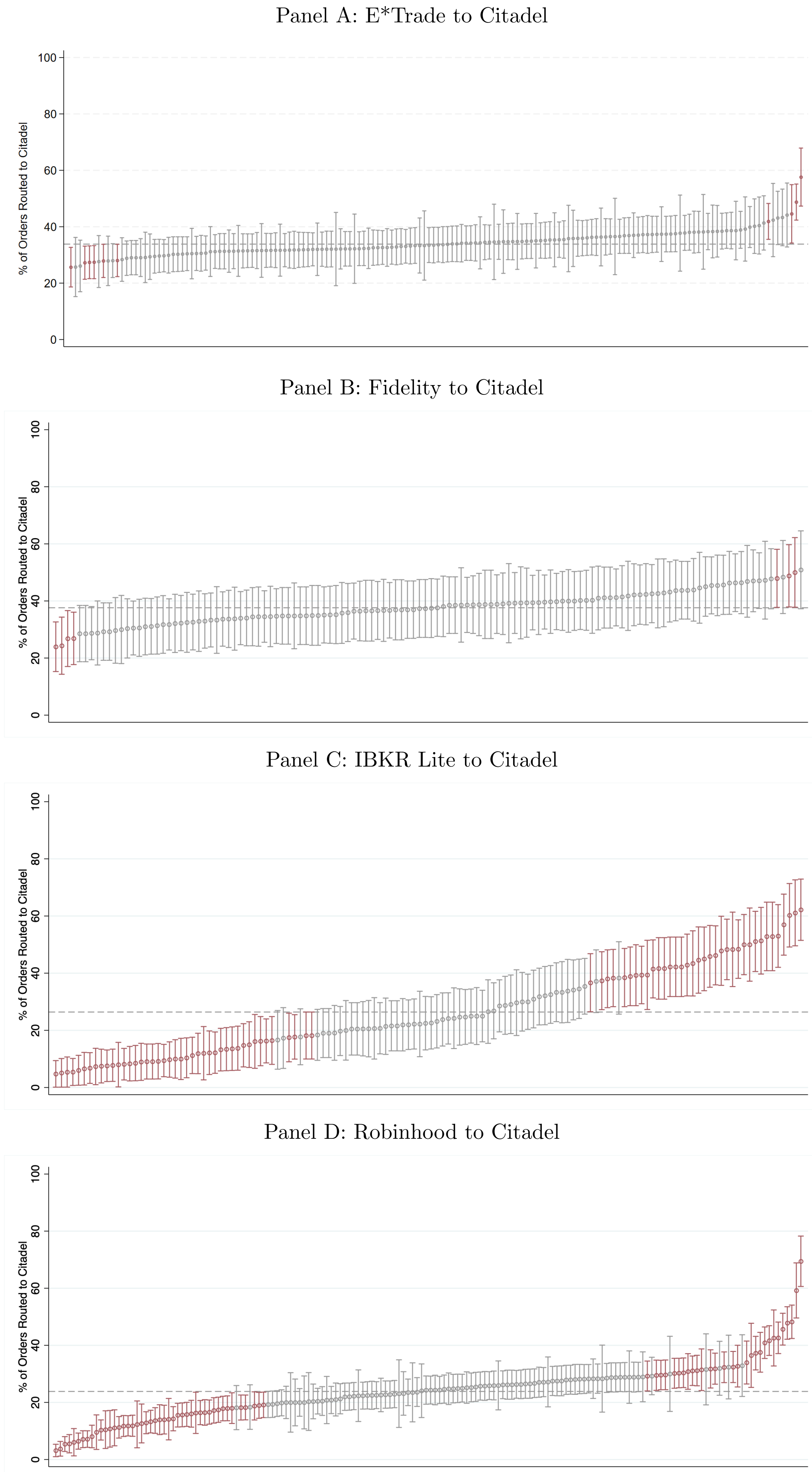
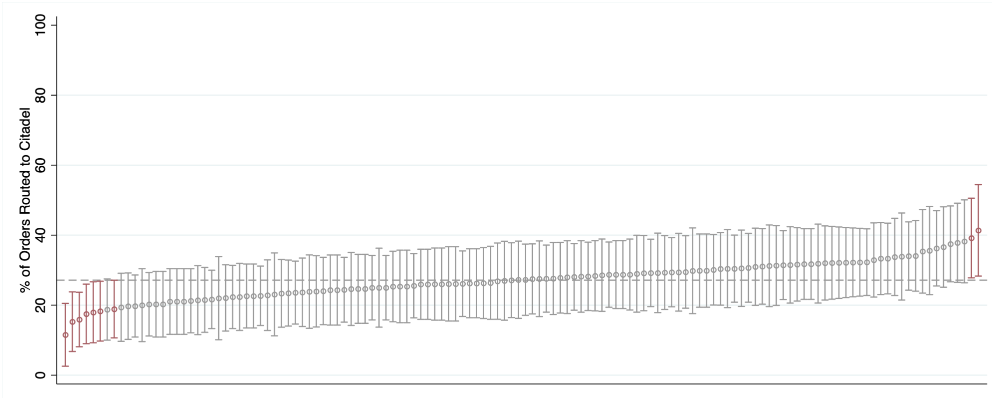
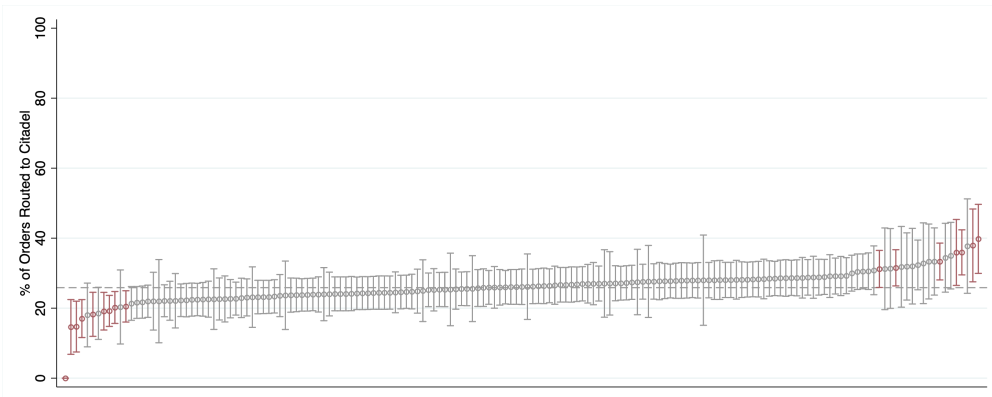


Figure 3: (Continued on the following page.)

Panel E: Schwab to Citadel



Panel F: TD Ameritrade to Citadel



Panel G: IBKR Pro to Own ATS

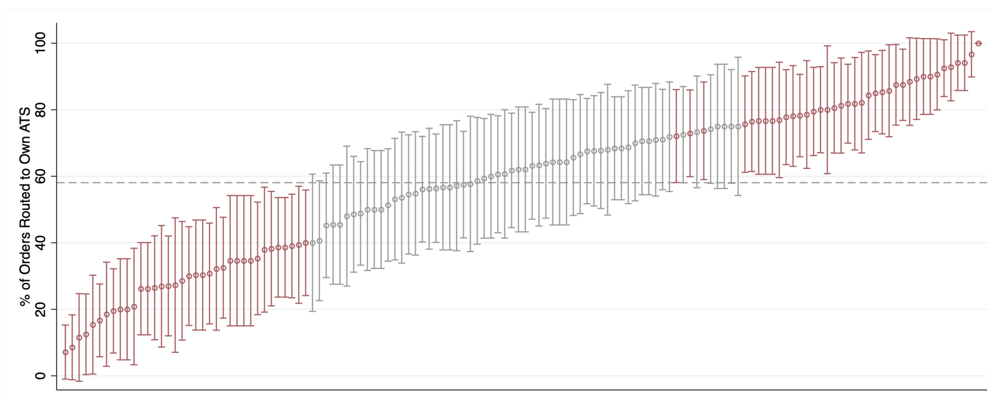


Figure 3: Order Routing Patterns across Individual Stocks
This figure shows the percentage of our orders for each stock that are routed to a specific venue. Panels A to F report order routing for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, and TD Ameritrade to Citadel, while Panel G reports order routing for IBKR Pro account to IBKR's own ATS. Each vertical bar represents one stock, with whiskers showing 95% confidence intervals. A stock requires at least 100 trades to be included. If a stock percentage is significantly different from the average at the 5% level, lines are shown in red; otherwise, lines are in black.

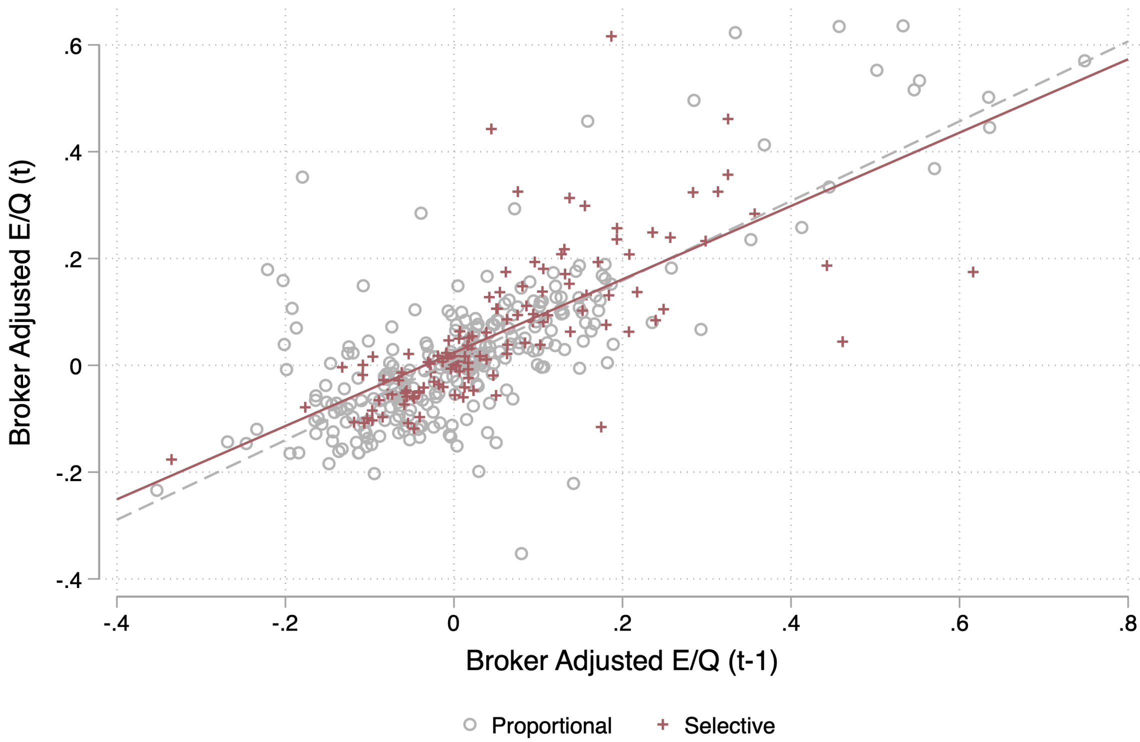


Figure 4: Relation between Current and Prior Month Effective Spreads
This plots, for each broker-wholesaler pair, the excess effective over quoted trade spread (E/Q) for the current month on the vertical axis against that for the prior month. Excess E/Q is computed as the average E/Q for each wholesaler at that broker minus the average for all wholesalers for that broker. Circles represent proportional brokers; crosses represent selective brokers.

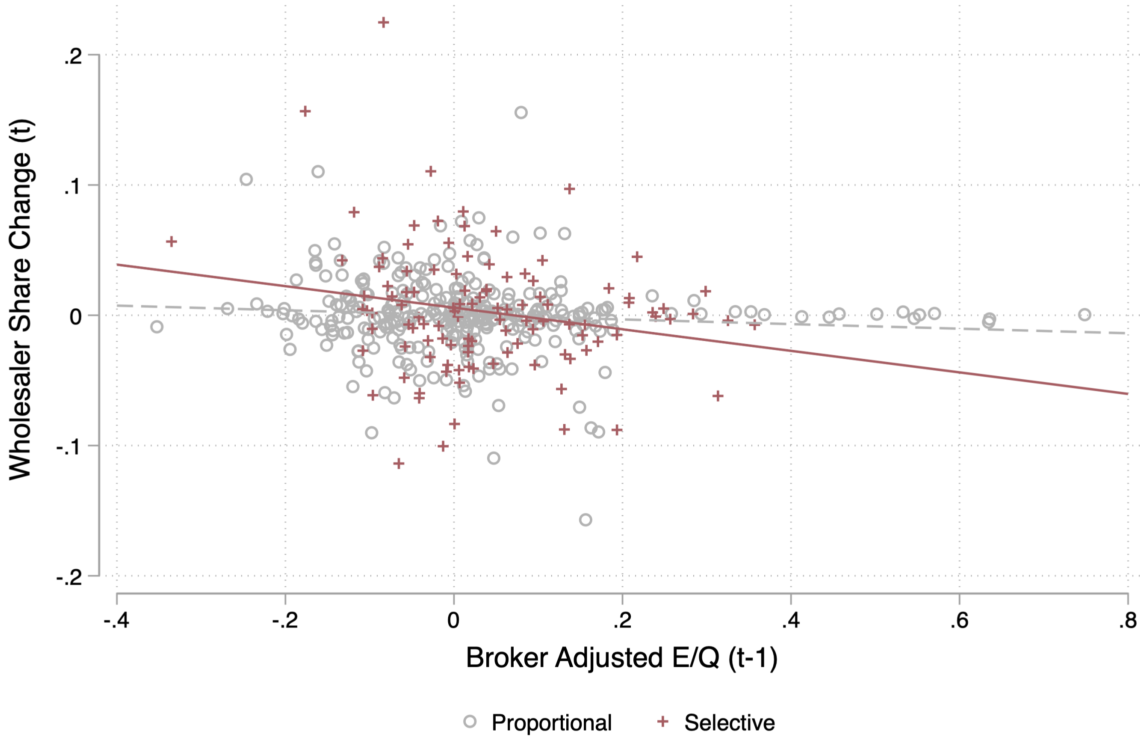
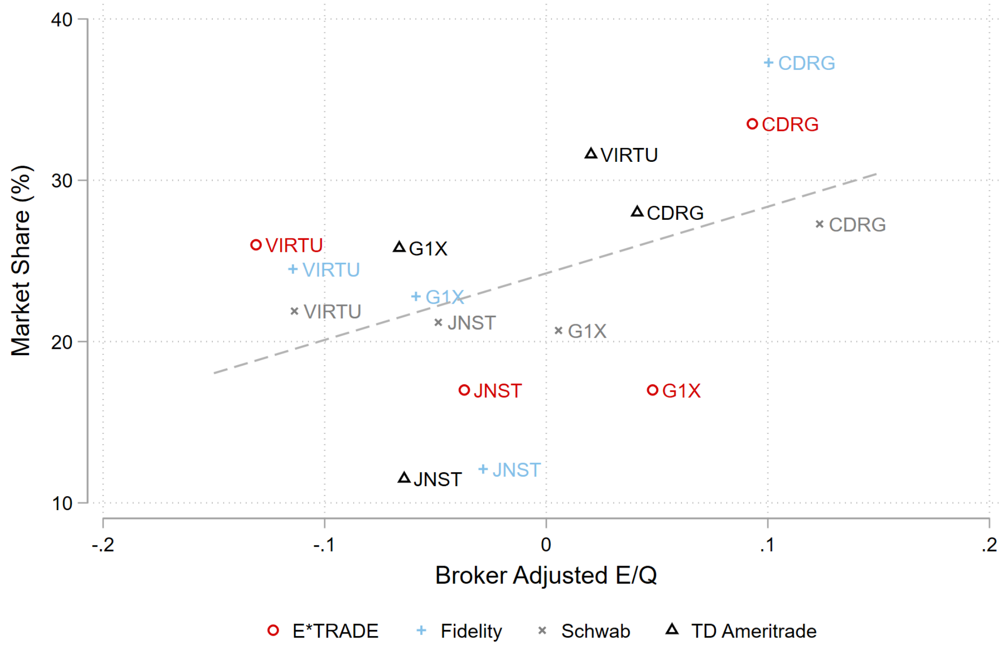


Figure 5: Relation between Market Share Changes and Prior Month Effective Spreads
This plots, for each broker-wholesaler pair, the change in wholesaler market share for the current month on the vertical axis against the prior month excess effective over quoted trade spread (E/Q). Excess E/Q is computed as the average E/Q for each wholesaler at that broker minus the average for all wholesalers for that broker. Circles represent proportional brokers; crosses represent selective brokers.

Panel A: Proportional Brokers



Panel B: Selective Brokers

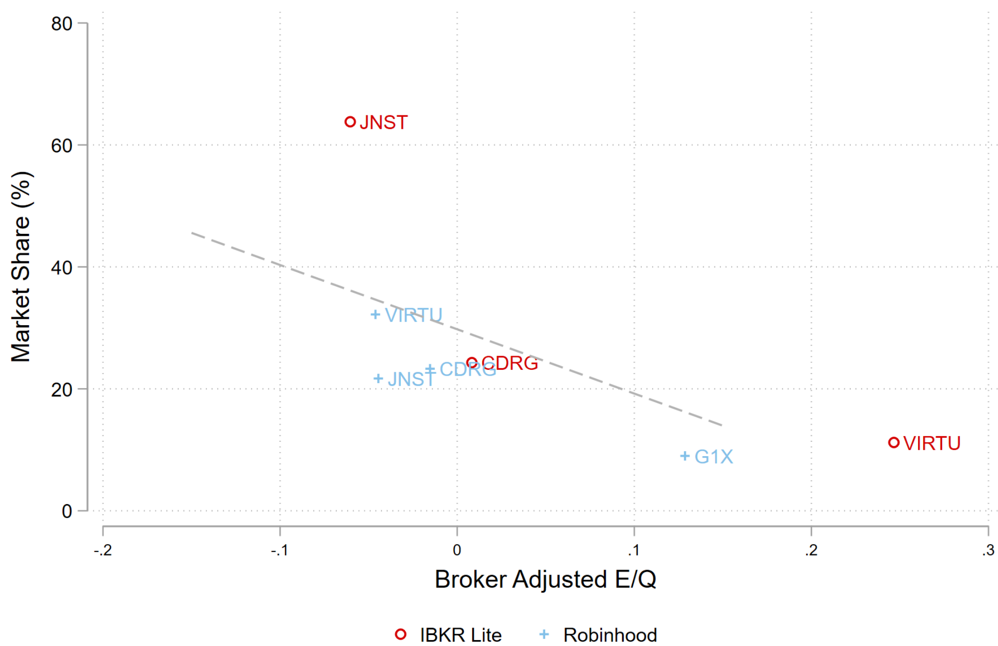


Figure 6: Wholesaler Price Improvement and Market Share
This figure describes the relation, for each broker, between wholesaler market share and its price improvement as measured by its excess effective over quoted spread (E/Q). The wholesaler sample consists of the "Top 4," including Citadel (CDRG), Virtu, Jane Street (JNST) and G1X. Panel A plots the relation for proportional brokers while Panel B plots the relation for selective brokers.

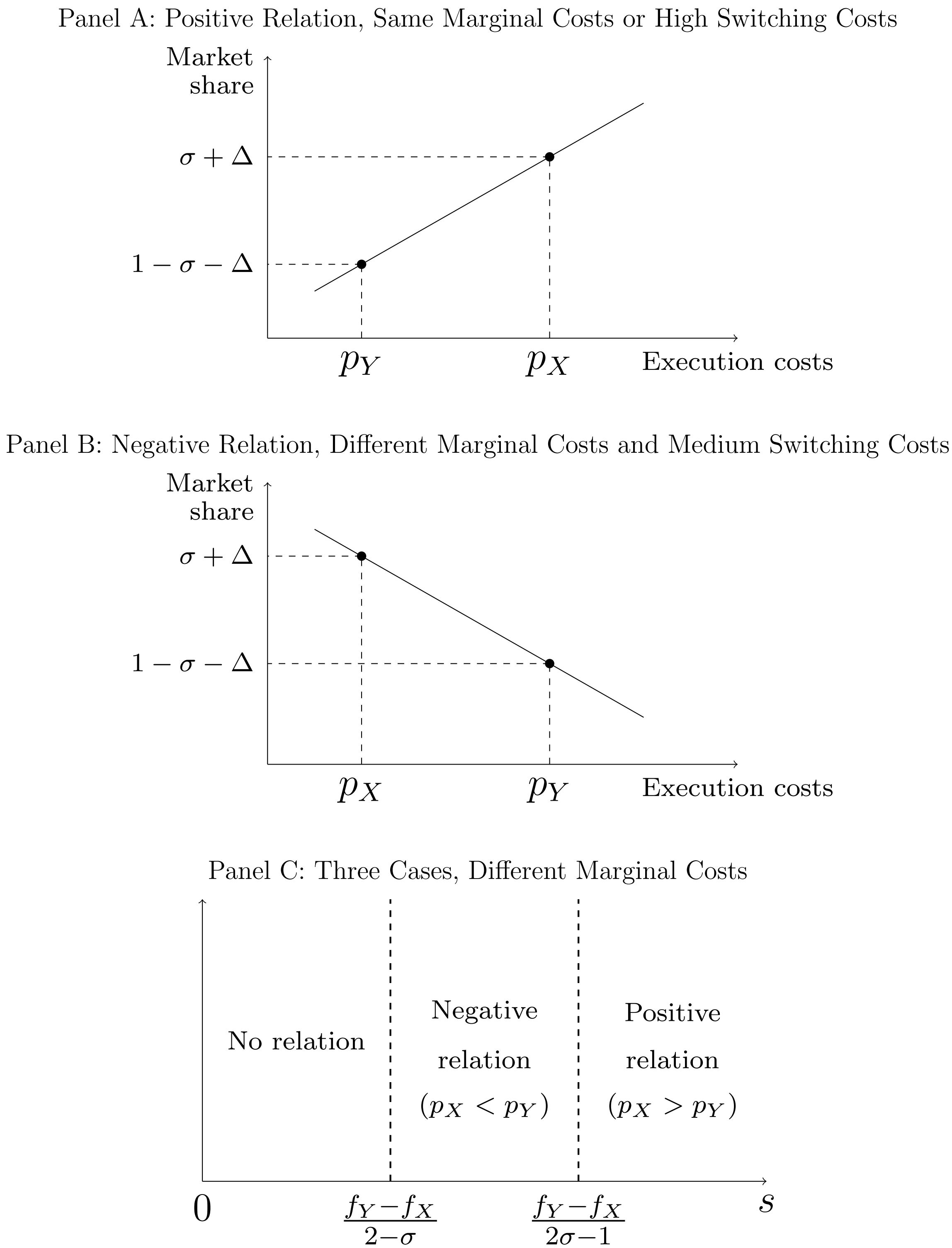


Figure 7: Model
Panel A plots the positive relation between execution cost and market share, which arises when the wholesalers have the same marginal cost or when switching costs s are high. Panel B plots the negative relation, which arises when switching costs are medium. Panel C shows how the size of switching costs affects the sign of the relation when the marginal costs of the wholesalers (f_X, f_Y) differ.

Table 1: Summary Statistics on Order Routing

This table presents summary statistics on order routing for our trades. We placed parallel trades at six brokers from December 2021 through May 2023. We requested and obtained routing information from the brokers through SEC rule 606(b)(1). The table reports the number of trades at each broker that go to each wholesaler in Panel A, as well as the percent of orders for each broker in Panel B. Averages in Panel B exclude IBKR Pro.

Panel A: Broker-Wholesaler Routing Count

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Citadel	Virtu	Jane Street	G1X	Two Sigma	UBS	Exchange	Other	Own ATS	Total
E*Trade	10,699	8,206	5,327	5,267	938	818	10	-	-	31,265
Fidelity	4,561	3,103	1,262	2,847	148	83	1	228	-	12,233
IBKR Lite	2,834	1,293	6,543	-	-	-	-	-	42	10,712
Robinhood	9,669	12,761	8,568	3,507	5,212	-	150	92	-	39,959
Schwab	3,047	2,410	2,400	2,279	436	512	-	1	-	11,085
TD Ameritrade	11,088	13,660	5,161	11,160	224	497	-	520	-	42,310
Total	41,898	41,433	29,261	25,060	6,958	1,910	161	841	42	147,564
IBKR Pro	126	96	31	-	-	71	776	688	2,477	4,265

Panel B: Routing Percentages by Broker

	Citadel	Virtu	Jane Street	G1X	Two Sigma	UBS	Exchange	Other	Own ATS	Total
E*Trade	34%	26%	17%	17%	3%	3%	0%	-	-	100%
Fidelity	37%	25%	10%	23%	1%	1%	0%	2%	-	100%
IBKR Lite	26%	12%	61%	-	-	-	-	-	0%	100%
Robinhood	24%	32%	21%	9%	13%	-	0%	0%	-	100%
Schwab	27%	22%	22%	21%	4%	5%	-	0%	-	100%
TD Ameritrade	26%	32%	12%	26%	1%	1%	-	1%	-	100%
Average	29%	25%	24%	16%	4%	2%	0%	1%	0%	100%
IBKR Pro	3%	2%	1%			2%	18%	16%	58%	100%

Table 2: Drivers of Routing Decisions
This table examines how brokers route orders to wholesalers, or venues. For each broker, we run a logistic regression where the dependent variable is one if the trade is routed to that wholesaler and zero otherwise. For regressors, we include E/Q (effective over quoted spread) for that stock at that venue in excess of the average for that stock across venues (*Venue Excess E/Q (t-1)*), the percent of orders routed to that wholesaler the previous month (*Venue % (t-1)*), as well as a dummy variable set at one if our last order was routed to that venue (*Prior Same Venue*). We also include a number of stock characteristics including the log of the stock price, the trade date's log volume, return, absolute return, the spread at the time of the trade, and a dummy variable for stocks in the S&P 500 index. Finally, we include a dummy variable reflecting whether the trade was a buy or a sell (*Buy (1/0)*). Models include day fixed effects. **,* represents significance at the 1%, 5% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: E*Trade						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	-0.024	-0.016	0.011	0.100*	0.245	-0.027
Venue % (t-1)	0.121	-0.139	0.204	-0.083	2.521**	-9.734**
Log(Price)	0.008	0.026	0.010	0.033*	-0.013	0.025
Log(Volume)	0.013	-0.005	0.007	0.008	0.017	-0.003
Return	0.058	0.042	-0.045	0.159	-0.209	0.078
Abs(Return)	0.057	0.043	-0.046	0.157	-0.219	0.079
Spread	-0.073	0.038	-0.036	0.065	-0.287	-0.052
Buy (1/0)	-0.011	-0.031*	-0.013	0.033	0.065	0.047
SP500	-0.050	-0.080	0.006	-0.083	0.112	0.299
Prior Same Venue	-0.070*	-0.030	0.011	-0.014	-0.440	-0.025
Panel B: Fidelity						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	0.084	-0.019	-0.141	0.114		
Venue % (t-1)	-0.031	0.058	0.277	0.400		
Log(Price)	-0.010	0.057	0.002	0.029		
Log(Volume)	0.003	0.011	0.013	-0.065**		
Return	-0.069	0.053	-0.117	0.174		
Abs(Return)	-0.069	0.054	-0.118	0.164		
Spread	-0.003	-0.054	0.065	-0.223		
Buy (1/0)	-0.018	-0.009	-0.002	0.032		
SP500	0.053	-0.109	0.100	0.032		
Prior Same Venue	0.121*	0.023	0.017	0.050		

	(1)	(2)	(3)	(4)	(5)	(6)
Panel C: IBKR Lite						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	-0.822**	-0.724**		-0.523**		
Venue % (t-1)	1.039**	2.224**		1.490**		
Log(Price)	-0.047	-0.294**		0.170**		
Log(Volume)	0.110**	0.078**		-0.148**		
Return	-0.138	-0.063		0.155		
Abs(Return)	-0.142	-0.062		0.155		
Spread	-0.388*	-0.284		-0.002		
Buy (1/0)	-0.130**	0.504**		-0.092**		
SP500	0.192*	0.132		-0.185*		
Prior Same Venue	-0.131*	-0.109		-0.130**		
Panel D: Robinhood						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	-0.410**	-0.341**	-0.755**	-0.260**	-0.271**	
Venue % (t-1)	2.423**	1.680**	3.638**	1.853**	2.355**	
Log(Price)	-0.002	0.026*	0.070**	-0.070**	0.017	
Log(Volume)	0.031**	-0.054**	0.047**	-0.009	-0.041**	
Return	-0.063	0.220**	0.001	-0.275**	0.187*	
Abs(Return)	-0.062	0.221**	0.004	-0.279**	0.189*	
Spread	-0.099*	-0.065	-0.012	0.180**	-0.116	
Buy (1/0)	-0.038**	-0.029*	0.017	0.013	0.014	
SP500	-0.105*	0.070	0.147*	-0.093	-0.228**	
Prior Same Venue	0.072*	0.054*	0.029	0.085*	0.141**	
Panel E: Schwab						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	0.050	0.117	0.036	-0.041	-0.024	-0.241
Venue % (t-1)	-0.318	0.023	0.075	-0.281	1.131	-0.474
Log(Price)	-0.015	0.053*	-0.017	0.027*	-0.017	-0.064
Log(Volume)	0.012	-0.012	0.028	-0.019	-0.010	-0.047
Return	0.127	-0.065	0.068	-0.094	0.595	0.199
Abs(Return)	0.130	-0.064	0.063	-0.093	-0.751	0.049
Spread	0.007	-0.244*	0.105	0.113	-0.412	-0.320
Buy (1/0)	0.007	0.005	0.020	-0.044**	-0.024	-0.067
SP500	-0.086	-0.014	-0.045	0.012	0.065	0.522
Prior Same Venue	0.108	0.115	-0.005	0.089	0.763**	1.336**

	(1)	(2)	(3)	(4)	(5)	(6)
Panel F: TD Ameritrade						
Dep Var: The trade is routed to	Citadel	Virtu	G1X	Jane Street	Two Sigma	UBS
Venue Excess E/Q (t-1)	-0.009	0.117*	0.004	0.041		
Venue % (t-1)	-0.040	0.182	0.054	0.845**		
Log(Price)	-0.005	0.007	0.002	-0.018		
Log(Volume)	-0.008	0.013*	0.030**	0.023**		
Return	-0.047	-0.058	0.067	-0.027		
Abs(Return)	-0.047	-0.060	0.067	-0.027		
Spread	0.011	0.018	0.031	0.100*		
Buy (1/0)	0.008	0.006	-0.019	-0.007		
SP500	0.079	-0.039	-0.097*	-0.102		
Prior Same Venue	-0.012	-0.096*	-0.127**	-0.284**		

Panel G: IBKR Pro			
Dep Var: The trade is routed to	IBKR ATS	Exchange	Wholesaler
Venue Excess E/Q (t-1)	-0.485**	-0.088	-0.800**
Venue % (t-1)	2.245**	0.889**	1.510**
Log(Price)	0.218**	-0.292**	-0.094*
Log(Volume)	0.014	-0.023	-0.038
Return	-0.024	0.202	0.118
Abs(Return)	-0.024	0.194	0.128
Spread	1.118**	-1.402	-1.016*
Buy (1/0)	0.023	-0.123*	0.058
SP500	-0.112	-0.020	0.202
Prior Same Venue	-0.249**	0.384**	-0.322**

Table 3: Average Excess Execution Cost by Wholesaler

This table examines wholesaler execution cost within each broker, using the effective over quoted spread (E/Q). Each month, we compute the execution cost for each wholesaler within each broker. We then compute the average across our sample period using Fama and MacBeth (1973) while standard errors are computed using Newey and West (1987) with one lag. We also report the broker-level average E/Q for reference purposes in the first column. t-values are in parentheses. **,* represents significance at the 1%, 5% levels respectively.

	Broker-level	Broker-adjusted E/Q (Deviation from broker-level E/Q)				Dispersion (Max-Min) Normalized by
	E/Q	Citadel	Virtu	Jane Street	G1X	Broker-level E/Q
E*Trade	0.322	0.093**	-0.131**	-0.037**	0.048**	69.57%
		(6.1)	(-17.1)	(-3.0)	(3.9)	
Fidelity	0.142	0.100**	-0.114**	-0.028	-0.059**	150.70%
		(6.3)	(-9.0)	(-1.4)	(-4.4)	
IBKR Lite	0.527	0.008	0.247**	-0.060**		58.25%
		(0.4)	(9.6)	(-4.7)		
Robinhood	0.421	-0.015	-0.046**	-0.045	0.129**	41.57%
		(-1.0)	(-5.5)	(-1.7)	(6.6)	
Schwab	0.229	0.123**	-0.114**	-0.049**	0.005	103.49%
		(11.9)	(-5.7)	(-3.0)	(0.9)	
TD Ameritrade	0.093	0.041**	0.020	-0.064**	-0.066**	115.05%
		(4.1)	(1.7)	(-3.5)	(-9.4)	

Table 4: Persistence of Wholesaler Price Improvement

This table examines the persistence of price improvement by wholesaler, measured as E/Q in excess of the broker averages, at the overall level (Panels A and B) or stock-level level (Panels C and D). We regress the broker-adjusted price improvements against prior period values, measured over the last one- and three-month averages. t-values are in parentheses (based on standard errors clustered by month.) **,* represents significance at the 1%, 5% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Proportional Brokers					Selective Brokers		
	Brokers	All	E*Trade	Fidelity	Schwab	TD	All	IBKR Lite	Robinhood
Panel A: Wholesaler Aggregate Execution Cost (Prior Month)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1)	0.719**	0.732**	0.488**	0.627*	0.505**	0.857**	0.686**	0.916**	0.599**
	(12.01)	(13.20)	(3.75)	(3.04)	(6.00)	(13.08)	(5.84)	(11.09)	(4.19)
R-square	0.578	0.577	0.245	0.344	0.373	0.765	0.491	0.811	0.384
Panel B: Wholesaler Aggregate Execution Cost (Prior 3-month Average)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1, t-3)	0.831**	0.835**	0.654**	0.924**	0.826**	0.868**	0.818**	0.980**	0.741**
	(12.24)	(12.01)	(3.85)	(5.36)	(16.48)	(11.52)	(7.14)	(10.21)	(4.65)
R-square	0.551	0.550	0.272	0.336	0.614	0.667	0.520	0.866	0.400
Panel C: Wholesaler Stock-level Execution Cost (Prior Month)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1)	0.208**	0.196**	0.187**	0.169**	0.188**	0.219**	0.245**	0.507**	0.190**
	(11.22)	(11.10)	(6.98)	(5.62)	(7.08)	(6.95)	(7.31)	(7.95)	(5.01)
R-square	0.046	0.038	0.037	0.024	0.033	0.048	0.058	0.246	0.035
Panel D: Wholesaler Stock-level Execution Cost (Prior 3-month Average)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1, t-3)	0.477**	0.480**	0.464**	0.426**	0.495**	0.508**	0.467**	0.712**	0.407**
	(21.37)	(21.09)	(15.27)	(7.62)	(7.73)	(13.61)	(6.18)	(7.33)	(4.30)
R-square	0.082	0.085	0.093	0.051	0.082	0.095	0.069	0.237	0.047

Table 5: Changes in Routing in Response to Prior Execution Cost

This table examines how brokers adjust their routing to wholesalers based on their prior aggregate (Panels A and B) and stock-level (Panels C and D) execution cost. Each month, for each broker, we compute the wholesaler broker-adjusted execution cost (E/Q) as the deviation of the wholesaler's execution cost from the broker-level average. This measure is calculated at both the aggregate level and the individual stock level. We then regress the percentage point change in orders routed to a given wholesaler against the wholesalers' lagged execution cost, based on the prior one- and three-month periods, respectively. t-values are in parentheses (based on standard errors clustered by month.) **,* represents significance at the 1%, 5% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Proportional Brokers					Selective Brokers		
	Brokers	All	E*Trade	Fidelity	Schwab	TD	All	IBKR Lite	Robinhood
Panel A: Wholesaler Aggregate Execution Cost (Prior Month)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.032*	-0.019	-0.025	-0.043	-0.064	-0.005	-0.088*	-0.067	-0.104*
	(-2.27)	(-1.65)	(-0.67)	(-0.84)	(-0.90)	(-1.23)	(-2.14)	(-0.87)	(-2.36)
R-square	0.020	0.008	0.009	0.022	0.042	0.002	0.053	0.022	0.076
Panel B: Wholesaler Aggregate Execution Cost (Prior 3-month Average)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1, t-3)	-0.018	-0.014	-0.024	-0.046	0.007	-0.011	-0.033	-0.092	0.017
	(-0.95)	(-0.93)	(-0.34)	(-0.49)	(0.25)	(-1.98)	(-0.62)	(-1.11)	(0.29)
R-square	0.013	0.004	0.005	0.015	0.001	0.005	0.020	0.037	0.002
Panel C: Wholesaler Stock-level Execution Cost (Prior Month)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.006	0.006	-0.003	0.014	0.015	0.007	-0.045**	-0.071	-0.039**
	(-1.39)	(1.30)	(-0.44)	(0.76)	(1.33)	(0.75)	(-3.42)	(-1.30)	(-3.45)
R-square	0.004	0.005	0.000	0.000	0.000	0.000	0.006	0.003	0.002
Panel D: Wholesaler Stock-level Execution Cost (Prior 3-month Average)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1, t-3)	-0.010	0.002	-0.016	-0.009	0.039	0.010	-0.049	-0.047	-0.049
	(-0.80)	(0.20)	(-0.74)	(-0.18)	(2.95)	(0.45)	(-1.92)	(-0.67)	(-1.73)
R-square	0.001	0.001	0.000	0.000	0.002	0.000	0.001	0.001	0.001

Table 6: Levels of Venue Routing based on Prior Price Execution

This table examines how wholesalers' market shares are related to their prior price improvement. Each month, we compute the price improvement, measured as the effective over quoted spread (E/Q), of all our trades for each wholesaler by broker. We also compute the percentage of our orders routed to each wholesaler. We then regress the percent of orders routed to the wholesaler this month against the price improvement the prior month. Panel A examines the relation using all wholesalers. The second model only examines the "Top 4" wholesalers (Citadel, Virtu, Jane Street, and G1X), which filled 96% of our trades. Regressions are run with broker dummy variables where appropriate. t-values are in parentheses (based on standard errors clustered by month.) **,* represents significance at the 1%, 5% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Proportional Brokers					Selective Brokers		
	Brokers	All	E*Trade	Fidelity	Schwab	TD	All	IBKR Lite	Robinhood
Panel A: All Wholesalers									
Dep Var:	Percentage of Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.294**	-0.160**	-0.044	0.056	0.250	-0.293**	-0.837**	-1.064**	-0.670**
	(-8.61)	(-4.27)	(-0.40)	(0.60)	(2.17)	(-6.81)	(-6.87)	(-10.47)	(-4.16)
R-square	0.197	0.037	0.002	0.002	0.085	0.184	0.331	0.474	0.513
Panel B: "Top 4" Wholesalers									
Dep Var:	Percentage of Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.213**	0.311**	0.095	0.463**	0.137	0.661**	-0.866**	-1.064**	-0.653**
	(-3.69)	(6.09)	(1.50)	(3.27)	(1.91)	(6.82)	(-5.97)	(-10.47)	(-3.00)
R-square	0.118	0.129	0.014	0.241	0.204	0.270	0.281	0.474	0.426

Table 7: Hypothetical Price Improvement from Rerouting
This table presents results on hypothetical price improvement if our trades were rerouted based on prior execution. For each broker, we compute the average execution cost, measured as the effective over quoted spread, received on each stock from each wholesaler. In Panel A, we compute the overall cost we received from each wholesaler across all stocks in the prior month. We then reroute all of this month's trades to the best wholesaler for the prior month. In Panel B, we compute the average cost we received for each stock from each wholesaler the prior month. We then route all of this month's trades for that stock to the wholesaler that had the best cost the prior month. In each panel, we report the average original cost, the hypothetical cost, as well as the difference and the change relative to the original price improvement, across all the months in our sample. t-statistics are computed using Fama and MacBeth (1973). **,* represents significance at the 1%, 5% levels respectively.

Panel A: Execution Cost from Rerouting using Overall Execution

Broker	Original	Hypothetical	Change	t-value	PI Change %
Proportional:					
E*Trade	0.352	0.223	-0.129	-12.61**	-36.6%
Fidelity	0.192	0.093	-0.099	-4.87**	-51.6%
Schwab	0.240	0.160	-0.079	-4.80**	-33.3%
TD Ameritrade	0.118	0.047	-0.071	-8.15**	-60.2%
Selective:					
IBKR Free	0.575	0.530	-0.046	-3.77**	-7.8%
Robinhood	0.402	0.346	-0.056	-5.23**	-13.9%
Average	0.313	0.233	-0.080		-33.9%

Panel B: Execution Cost from Rerouting using Stock Level Execution

Broker	Original	Hypothetical	Change	t-value	PI Change %
Proportional:					
E*Trade	0.349	0.311	-0.038	-3.41**	-10.9%
Fidelity	0.197	0.149	-0.047	-3.08**	-24.4%
Schwab	0.232	0.190	-0.042	-3.09**	-18.1%
TD Ameritrade	0.108	0.021	-0.086	-7.64**	-80.6%
Selective:					
IBKR Free	0.571	0.590	0.020	1.37	3.3%
Robinhood	0.397	0.454	0.057	3.77**	14.4%
Average	0.309	0.286	-0.023		-7.4%

Table 8: Changes for Robinhood Wholesalers after Jane Street Addition
This table shows changes in Robinhood's wholesaler market after Jane Street became an additional venue for that broker. Panel A reports the percentage of our trades routed to each wholesaler before and after the start date of February 24, 2022. Panel B reports the (E/Q) execution cost again before and after. In both panels, t values use standard errors clustered by stock. The prior period covers the six weeks before the start date. The posterior period runs from February 24 to April 15, 2022. In both cases, we take the average across all trades each day and then average across days, computing t-values using Fama and MacBeth (1973)). **, * represents significance at the 1%, 5% levels respectively.

Panel A: Wholesaler Shares

	Pre-Jane Street	Post-Jane Street	Difference	t-value	% Change
Virtu	39.0%	28.6%	-10.4%	-7.58**	-26.7%
Citadel	26.6%	21.5%	-5.1%	-4.13**	-19.2%
Two-Sigma	18.1%	18.7%	0.6%	0.57	3.2%
G1X	13.0%	8.1%	-4.9%	-4.43**	-37.7%
Jane Street	2.7%	22.5%	19.8%	19.57**	733.3%

Panel B: Execution Cost (E/Q)

	Pre-Jane Street	Post-Jane Street	Difference	t-value	% Change
Overall	0.548	0.470	-0.078	-5.66**	-14.3%
Virtu	0.483	0.448	-0.035	-1.15	-7.2%
Citadel	0.536	0.398	-0.138	-5.91**	-25.7%
Two-Sigma	0.612	0.597	-0.015	-0.42	-2.4%
G1X	0.703	0.643	-0.061	-2.53*	-8.6%
Jane Street	0.238	0.391	0.154	2.00	64.7%

Table 9: Drivers of Changes in Stock Allocation after Jane Street Addition
This table analyzes the drivers behind Robinhood's routing of stocks after Jane Street's addition to its routing venues. We separate the sample into the six weeks prior to February 24, 2022, and a post period from February 24 to April 15, 2022. For each stock-wholesaler observation, we compute the execution cost, measured as the effective over quoted spread, on that stock relative to the overall venue average (Excess E/Q) over the prior period. We then regress the change in the wholesaler's share in that stock from the pre- to post- period against Excess E/Q. We also include stock-level controls, i.e., the average spread, log of the stock price, log of the stock volume, and the average daily return. t-values are in parentheses. **,* represents significance at the 1%, 5% levels respectively.

	Citadel	Virtu	G1X	Two Sigma
Excess E/Q	0.092	0.027	-0.009	-0.064
	(1.12)	(0.24)	(-0.10)	(-0.99)
Avg. Spread	-0.023	-0.054	0.026	0.027
	(-0.65)	(1.27)	(0.74)	(0.79)
Log(Price)	0.007	0.036**	-0.024*	-0.016
	(0.73)	(3.08)	(-2.37)	(-1.78)
Log(Volume)	-0.012	-0.022*	-0.012	0.020**
	(-1.58)	(-2.56)	(-1.59)	(2.74)
Avg. Daily Return	-0.015	-0.003	0.038**	-0.035**
	(-1.19)	(-0.16)	(2.99)	(-2.97)

A.1. Proofs of Main Model

Proposition 1 *There are two cases. If either the wholesalers have the same marginal cost (i.e., $f_X = f_Y$), or the switching costs are sufficiently high (i.e., $s > \frac{f_Y - f_X}{2-\sigma}$, where $f_Y > f_X$), then in equilibrium the wholesalers charge*

$$p_X = \frac{2f_X + f_Y}{3} + \frac{s(1+\sigma)}{3} \qquad \textit{and} \qquad p_Y = \frac{f_X + 2f_Y}{3} + \frac{s(2-\sigma)}{3}, \tag{12}$$

and the broker adjusts its market share by

$$\Delta = \frac{f_Y - f_X}{3s} + \frac{1-2\sigma}{3}. \tag{13}$$

Otherwise, meaning that the wholesalers have different marginal costs and the switching costs are not sufficiently high (i.e., $0 < s \leq (f_Y - f_X)/(2-\sigma)$), in equilibrium

$$p_X = f_Y - s(1-\sigma), \qquad p_Y = f_Y, \qquad \textit{and} \qquad \Delta = 1 - \sigma, \tag{14}$$

assuming that when the broker is indifferent to routing, it sends all orders to wholesaler X.

Proof of Proposition 1. First, suppose that the broker incurs switching costs ($s > 0$). The wholesalers' marginal costs may or may not be the same ($f_X \leq f_Y$). The broker optimally chooses Δ that solves:

$$\min_{\Delta \in [-\sigma, 1-\sigma]} (\sigma + \Delta)p_X + (1 - \sigma - \Delta)p_Y + \frac{s}{2}\Delta^2. \tag{15}$$

The F.O.C. implies:

$$\Delta = \begin{cases} \frac{p_Y - p_X}{s} & \text{if } p_Y - p_X \in [-s\sigma, s(1-\sigma)], \\ 1-\sigma & \text{if } p_Y - p_X > s(1-\sigma). \\ -\sigma & \text{otherwise.} \end{cases} \tag{16}$$

Each wholesaler optimally chooses prices to maximize profits. Wholesaler X solves

$$\max_{p_X} (p_X - f_X)(\sigma + \Delta) \tag{17}$$

Using Equation (16), the F.O.C. implies

$$p_X = \begin{cases} \frac{s\sigma + p_Y + f_X}{2} & \text{if } p_Y \in (f_X - s\sigma, f_X + s(2-\sigma)), \\ f_X & \text{if } p_Y \leq f_X - s\sigma, \\ p_Y - s(1-\sigma) & \text{otherwise.} \end{cases} \tag{18}$$

Similarly, wholesaler Y's F.O.C. implies

$$p_Y = \begin{cases} \frac{s(1-\sigma) + p_X + f_Y}{2} & \text{if } p_X \in (f_Y - s(1-\sigma), f_Y + s(1+\sigma)), \\ f_Y & \text{if } p_X \leq f_Y - s(1-\sigma), \\ p_X - s\sigma & \text{otherwise.} \end{cases} \tag{19}$$

Theoretically, there is a total of nine cases. However, given that $f_X \leq f_Y$, p_Y cannot be less than f_X. It is also not optimal for wholesaler X to charge $p_X > f_Y + s(1+\sigma)$ such that wholesaler Y receives the entire order flow. Thus, a total of four cases remains.

Case 1: $p_X \in (f_Y - s(1-\sigma), f_Y + s(1+\sigma))$ **and** $p_Y \in (f_X - s\sigma, f_X + s(2-\sigma))$**.** From Equations (18) and (19), we have

$$p_X = \frac{2f_X + f_Y}{3} + \frac{s(1+\sigma)}{3} \qquad \text{and} \qquad p_Y = \frac{f_X + 2f_Y}{3} + \frac{s(2-\sigma)}{3}. \tag{20}$$

Ensuring that these solutions satisfy the conditions for the prices' ranges, we have

$$s > \frac{f_Y - f_X}{2-\sigma}. \tag{21}$$

Case 2: $p_X \in (f_Y - s(1-\sigma), f_Y + s(1+\sigma))$ **and** $p_Y \geq f_X + s(2-\sigma)$**.** Again from Equations (18) and (19), we have

$$p_X = f_Y - s(1-\sigma) \qquad \text{and} \qquad p_Y = f_Y. \tag{22}$$

Ensuring that these solutions satisfy the conditions, we have

$$s \leq \frac{f_Y - f_X}{2-\sigma}. \tag{23}$$

Case 3: $p_X \leq f_Y - s(1-\sigma)$ **and** $p_Y \in (f_X - s\sigma, f_X + s(2-\sigma))$**.** From Equations (18) and (19), we have

$$p_X = \frac{f_X + f_Y + s\sigma}{2} \qquad \text{and} \qquad p_Y = f_Y. \tag{24}$$

Substituting these solutions to the conditions yields contradictions. $p_Y = f_Y \in (f_X - s\sigma, f_X + s(2-\sigma))$ implies:

$$f_Y - f_X \in (f_X - s\sigma, f_X + s(2-\sigma)), \tag{25}$$

while $p_X = \frac{f_X + f_Y + s\sigma}{2} \leq f_Y - s(1-\sigma)$ implies:

$$f_Y - f_X \geq s(2-\sigma). \tag{26}$$

Both conditions cannot be satisfied simultaneously.

Case 4: $p_X \leq f_Y - s(1-\sigma)$ and $p_Y \geq f_X + s(2-\sigma)$. In this case, $p_Y = f_Y$ and wholesaler X has no incentive to reduce prices strictly below $f_Y - s(1-\sigma)$ since it already receives the entire order flow. Thus, the result is identical to Case 2.

Finally, if the broker does not incur switching costs (i.e., $s = 0$), the broker's optimal strategy is

$$\Delta = \begin{cases} 1-\sigma & \text{if } p_X \leq p_Y \\ -\sigma & \text{otherwise.} \end{cases} \tag{27}$$

Given this, wholesalers X and Y's optimal strategies are $p_X = p_Y = f_Y$. Wholesaler Y has no incentive to raise or reduce prices, since any other prices imply zero or negative profits. Wholesaler X also has no incentive to raise or reduce prices. Raising prices imply zero market share. Reducing prices only lower profits since it is already receiving the entire order flow. ■

Proposition 2 *Holding the difference in marginal costs ($f_Y - f_X$) constant, optimal switching costs decrease in marginal costs, while equilibrium execution costs increase in marginal* *costs.*

Proof of Proposition 2. The overall execution costs $\bar{p}$ is defined by

$$\bar{p} := (\sigma + \Delta)p_X + (1 - \sigma - \Delta)p_Y. \tag{28}$$

Substituting Δ, p_X and p_Y in Proposition 1 into above yields

$$\bar{p} = f_X + \left(\frac{5 - 2\sigma + 2\sigma^2}{9}\right) s - (f_Y - f_X)\left(\frac{f_Y - f_X}{9s} + \frac{4 + \sigma}{9}\right). \tag{29}$$

Notice, $\bar{p}$ is always strictly increasing in switching costs s, and also increasing in the marginal cost f_X, holding $\Delta_F := f_Y - f_X$ constant.

Thus, the optimal switching costs are given by

$$s = \min\{s_{\max}, s^*\}, \tag{30}$$

where s^* is a unique solution to

$$0 = G(f_X, s) := f_X + \left(\frac{5 - 2\sigma + 2\sigma^2}{9}\right) s - \Delta_F\left(\frac{\Delta_F}{9s} + \frac{4 + \sigma}{9}\right) - p_{\max}. \tag{31}$$

Since $G(\cdot)$ is increasing in f_X and s, we have

$$\frac{ds^*}{df_X} < 0 \qquad \text{and} \qquad \frac{ds}{df_X} \leq 0. \tag{32}$$

Then the equilibrium execution costs are given by

$$\bar{p} = p_{\max} + \min\{0, G(f_X, s_{\max})\}. \tag{33}$$

Since $G(\cdot)$ is increasing in f_X, we have

$$\frac{d\bar{p}}{df_X} \geq 0 \qquad \text{and} \qquad \frac{d\bar{p}}{df_X} > 0 \text{ if } s = s_{\max}. \tag{34}$$

Since Δ_F is held constant, the derivatives with respect to f_Y remain the same. ■

A.2. Endogenous Switching Costs

Our empirical analysis documents that overall execution costs are higher for selective brokers than for proportional brokers (Table 3). This is somewhat puzzling because, all else being equal, the smaller switching costs of selective brokers, leading to more active routing decisions, should imply that their execution costs are lower (rather than higher) than those of proportional brokers. We now endogenize brokers' switching costs, which sheds light on this result.

Timeline. There are three points in time: $t = 0, 1, 2$. At $t = 0$, brokers choose optimal switching costs, as described below. At $t = 1$, brokers optimally route their order flows to wholesalers X and Y with given switching costs as described in Section 5.1. At $t = 2$, investors pay the execution costs based on the broker's routing decisions. Full equilibrium is solved by backward induction.

Broker's Decision. At $t = 0$, each broker solves the following optimization problem to choose its switching cost:

$$\min_{s \in [0, s^{\max}]} g(s^{\max} - s) \qquad \text{subject to} \qquad \bar{p} \leq p_{\max}. \tag{35}$$

Here, $g(\cdot)$ is the investment required to reduce the switching cost. When the investment is zero, the switching cost is at the maximum $s^{\max}$. The function g is strictly increasing in the gap $(s^{\max} - s) > 0$. Next, $\bar{p}$ stands for the overall execution costs that the broker's customers pay wholesalers, determined in equilibrium at $t = 1$. Lastly, $p_{\max}$ indicates the maximum execution costs allowed by regulators. The mandate that retail investors receive prices at or better than the prevailing best quotes (NBBO) implies that $p^{\max}$ (E/Q) is set at one.[40] In summary, the broker opts for minimal investments, which translate into high switching costs, while complying with regulatory requirements.

It is important to note that customers entirely cover execution costs, in which brokers do not have a direct economic stake. Although brokers' revenues rely on the size of order flow through PFOF, this size is considered exogenous (and thus fixed) here because customers are unlikely to respond to it. Indeed, retail customers typically lack information on their broker's execution quality and may consider other factors, such as convenience and popularity, when selecting brokers.

Full Equilibrium. Now, we solve for full equilibrium. For the sake of clarity, we assume that the marginal costs are uniform across wholesalers for each broker. (For further details and the general case, refer to Proposition 2 in the Appendix,)

The overall execution cost $\bar{p}$ is the equilibrium-weighted average of execution costs from the two wholesalers:

$$\bar{p} := (\sigma + \Delta)p_X + (1 - \sigma - \Delta)p_Y. \tag{36}$$

At $t = 1$, $\bar{p}$ is determined in partial equilibrium characterized in Section 5.1. Substituting Δ, p_X and p_Y in Proposition 1 into above, we have

$$\bar{p} = f + \left(\frac{5 - 2\sigma + 2\sigma^2}{9}\right) s. \tag{37}$$

[40] In practice, brokers are likely to maintain execution costs strictly below the NBBO spreads to allow for a buffer and avoid the risk of violating any regulations.

Taking this into account, brokers find the optimal switching costs that solve the optimization problem (35) at $t = 0$, which yields:

$$s = \begin{cases} \frac{9}{5-2\sigma+2\sigma^2}\left(p^{\max} - f\right), & \text{if } f > \bar{f} \\ \\ s^{\max}, & \text{otherwise,} \end{cases} \tag{38}$$

where $\bar{f} := p^{\max} - \left(\frac{5-2\sigma+2\sigma^2}{9}\right) s^{\max}$. Note that, without loss of generality, f can be assumed to not exceed $p^{\max}$, as wholesalers can always send the orders to the exchange if marginal (execution) costs were to exceed 1, i.e., effective spreads were greater than quoted spreads.

Comparative Statics. The optimal switching costs in (38) generally decrease in the marginal costs. The relation is strict when the marginal costs are sufficiently high but becomes weak when the marginal costs are low such that switching costs reach the maximum.

Intuitively, brokers with order flows facing higher marginal costs (i.e., more "toxic" order flows) face greater regulatory pressure than brokers with order flows facing lower marginal costs (i.e., order flows are less "toxic"), as the regulations enforce the same standards (NBBO) across brokers. Thus, brokers with more toxic order flows are likely to have smaller switching costs than those with less toxic order flows because they are more prone to violating the threshold.

The negative relation between switching costs and marginal costs offers valuable insights into understanding the empirical finding that selective brokers exhibit larger execution costs compared to proportional brokers, despite also exhibiting smaller switching costs. If smaller switching costs associated with selective brokers are indicative of and driven by their more toxic order flows, it is not surprising that selective brokers experience larger execution costs than proportional brokers. This is because the reduction in switching costs may not fully compensate the increase in marginal costs, as shown in Figure 8.

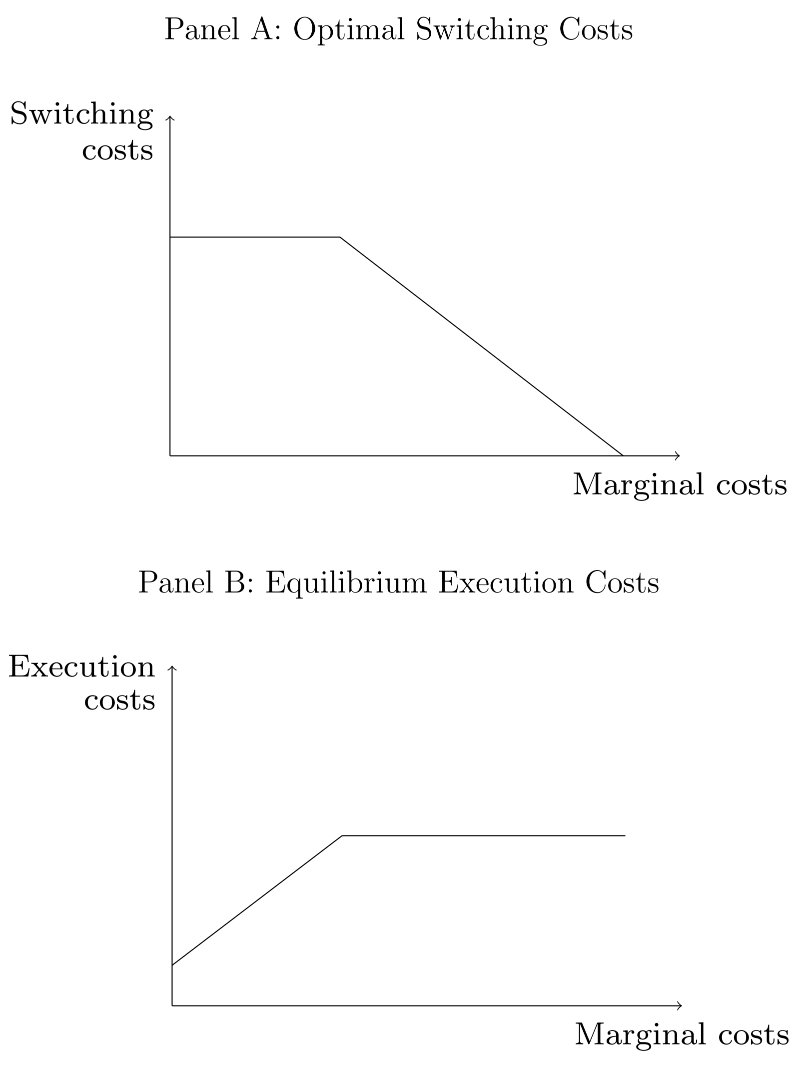


Figure 8: Endogenous Switching Costs
Panel A plots the negative relation between optimal switching costs and marginal costs. Panel B plots the positive relation between equilibrium execution costs and marginal costs.

A.3. Odd vs. Round Lots

Our experiment primarily focused on odd-lot trades, worth around \$100. This Appendix provides limited evidence about round lots of 100 or more shares. Odd lots have a special status under Reg NMS.[41] So perhaps they are not the main focus of execution quality. A reasonable concern is that our odd lot results may not be representative of the broader universe of trades. Odd lots could conceivably have systematically worse execution than round lots. The question is whether the relation between routing and trade execution could systematically differ for round lots.

Our trading experiment provides some guidance to investigate this issue. At the beginning of our trading experiment, we extended our standard trading size of \$100 to additional simultaneous \$1,000 trades for Robinhood and TD Ameritrade. We confirmed that the execution quality was essentially identical across these two trade sizes, and then closed the extended experiment after three months. Because these data also contain round lots, they can be used to generate two categories of simultaneous round-lot and odd-lot trades, each with about 100 round-trip observations, for these two brokers.

First, we examine the correlation between execution costs for odd lots and round lots across the four major wholesalers. This is very high, at 0.91 for TD and 0.82 for Robinhood, with both numbers highly significant. So, there is no indication that odd lots are systematically treated differently.

Next, we can check whether the relation between market share and execution quality differs across odd lot and round lots. For consistency and greater precision, market shares

[41] NBBO only uses round lots. The current 605 forms do not report execution statistics for odd lot trades. TAQ did not even include odd lot trading data before October 2013.

are taken from those in Figure 6. For TD, a proportional broker, the analysis is simpler. Both the odd-lot and round-lot trades have the same routing, so our odd lots provide a perfect reflection of market shares for the total order flow. Hence, the only way our odd-lot results could be overturned would be if the wholesalers with larger shares had lower execution costs for round lots, vs. the higher costs we observe for odd lots. This is certainly possible if TD were to focus solely on execution quality for round lots, in which case wholesalers could strategically game the objective function, by providing better execution for round lots, while taking advantage of ignored odd lot trades.

Panels A and B in Figure 9 compare the relation between market share and execution cost for our odd lots and round lots at TD. The figure shows a similar and flattish relation, thus broadly confirming our previous results for odd lots. Likewise, Panels C and D display the comparison for our Robinhood's trades. In both panels, we observe a strongly negative relation, which is similar for odd lots and round lots.

[**Insert Figure 9 about here**]

Admittedly, this evidence is based on a limited sample of trades. Even so, there is no evidence that the relation we found for our odd lots would differ for round lots.

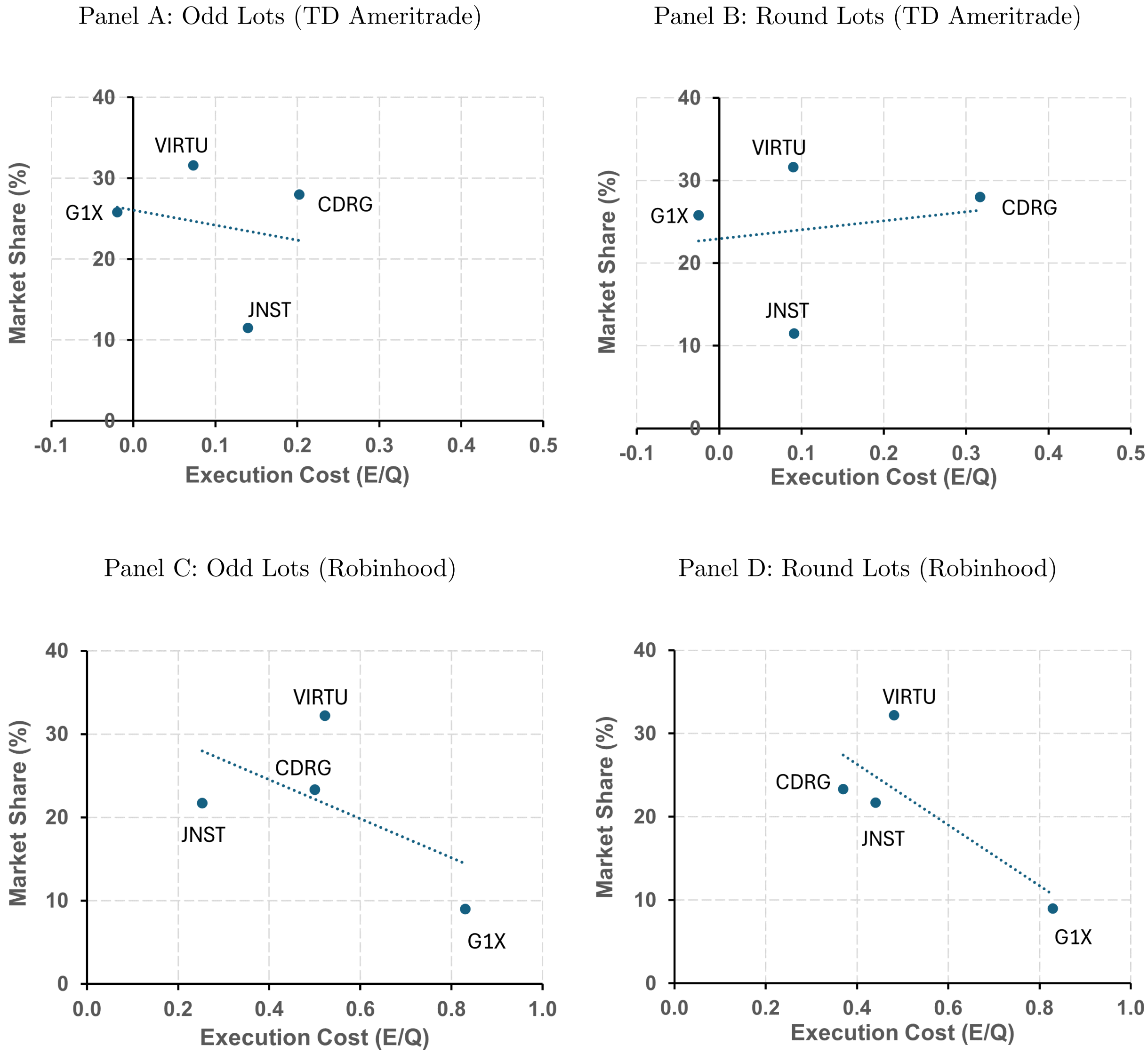


Figure 9: Market Share and Execution Costs for our Odd and Round Lots
These figures plot the relation between market share and execution costs for two brokers for which we have simultaneous trades of odd-lot and round-lot sizes, about \$100 and \$1000, respectively. Execution costs are derived from our approximately 100 round-trip trades for each of the two categories executed at TD America (Panels A and B) and Robinhood (Panels C and D). Market shares are the same as in Figure 6.