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

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

March 15, 2024

Abstract

Using 150,000 actual trades, we examine the competitiveness of the U.S. equity broker-wholesaler marketplace. We find that retail investors' trading costs have substantial and persistent dispersion across wholesalers within brokers—and yet, most brokers hardly change their routing and continue sending orders to more expensive wholesalers. Our simulated routing based on past performance lowers trading costs by 34%. We also find that after a new wholesaler enters, existing wholesalers reduce trading costs by 14%. Finally, we present a stylized model that explains our results. Overall, our findings suggest that competition among wholesalers is far from perfect.

JEL Classifications: G12, G14, G50

Keywords: retail trading, execution quality, order routing, competition, market microstructure, regulation, broker-dealers, wholesalers

[*]Christopher Schwarz (corresponding author, phone: 949-824-0936, 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 wish to thank Alyssa Moncrief and Sanath Nair for research assistance. This paper benefited from helpful comments by conference participants at the 2023 ESSFM Asset Pricing Conference and 2024 MFA Conference, as well as seminars at Baruch College, Columbia University, Cornell University, Iowa State University, Northeastern University, and Notre Dame University.

[†]All the brokerage accounts referenced were funded directly by the authors with personal money. Jeongmin Lee did not contribute any funds or input to the design and execution of trading. No outside compensation was received from any broker or wholesaler for this study. This paper required no outside approval. All errors are our own.

1. Introduction

Retail investors access to equity markets has vastly improved over the last few decades, spurred by technological advances. In the U.S., the adoption of Regulation NMS (SEC (2005)) led most retail brokers to route customers' orders to specialized OTC market makers, known as "wholesalers." By separating retail orders from institutional investors' orders on stock exchanges, wholesalers can provide greater price improvement, reducing retail investors' trading costs. Furthermore, in late 2019, most retail brokers abolished commissions, reducing trading costs further. Still, an open question is whether the market for executing retail equity trades is sufficiently competitive.

On the one hand, brokers are expected to enforce competition among wholesalers. Under their legal duty of "best execution," brokers must 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)). On the other hand, today's market structure has come under scrutiny from policy-makers. In 2022, the SEC issued four proposals that would fundamentally revise the regulations governing the structure of U.S. equity markets.[1] One major concern is that the wholesaler market is too concentrated (Hu and Murphy (2022)), with only four large players (Citadel, Virtu, G1X, and Jane Street) that compete in the market for retail trade execution services. Another concern is the practice of payment for order flow (PFOF) from wholesalers to brokers, which could also negatively affect price improvement.

[1]The proposed changes include: "Disclosure of Order Execution Information" (SEC (2022a)), which was recently adopted, "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)).

In this paper, we examine the competitiveness of the retail broker-wholesaler marketplace and find that it deviates substantially from perfect competition. We observe that each broker's trading costs have sizable and persistent dispersion across wholesalers. Yet, most brokers hardly adjust their routing strategies, and even the only broker whose adjustments are statistically significant does so modestly. Further, two thirds of the brokers consistently send a larger fraction of their order flows to wholesalers who charge higher trading costs. Our simulated dynamic routing strategies based on past performance would lower trading costs by 34% on average. We also observe that after a new wholesaler enters, the existing wholesalers lower their trading costs by 14% on average. Finally, we present a theoretical model to illustrate the mechanism through which the brokers' limited adjustments allow wholesalers to exercise market power, and to collectively explain our results.

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 experiment. In this experiment, we placed our own retail trades across a randomly selected set of stocks at randomized times through multiple brokerage accounts.[2] 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 of wholesalers.[3] Second, our approach allows us to directly conduct *within-broker* comparison of trading costs—measured by the effective

[2]We generated a total of approximately 172,000 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 (stock), size (dollars and shares traded), direction (buy or sell), and submission time. 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.

[3]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 with randomized timing on a representative sample could provide a more unbiased analysis.

over quoted spread (E/Q)—charged by different wholesalers. Focusing on dispersion *within broker* is important since there are large variations of trading costs *across brokers* (Battalio et al. (2001) and Schwarz et al. (2023)).[4]

To examine the competitiveness of retail trade execution, we test implications of the perfect Bertrand competition model. If the market were perfectly competitive with brokers frictionlessly switching across wholesalers based on their execution, the order flow should only go to the wholesaler(s) with the best execution quality. Thus, we would expect that for a given broker there is no dispersion of execution costs across wholesalers. Instead, we find a substantial dispersion, ranging from 42% to 151% of brokers' average E/Q. Moreover, this dispersion is persistent since execution quality is predictable over time, both at the aggregate and stock levels. For example, when examining persistence at the aggregate level, regression slopes of current trading costs on past trading costs are often close to one, with R-squares over 50%.

Next, we study how brokers respond to such dispersion in trading costs. Despite the predictability of trading costs, a majority of our brokers seem to hardly change their routing for our trades based on past execution. In fact, only one does so at a statistically significant level. Interestingly, the brokers that are the most responsive to past trading costs are also the brokers that have the lowest E/Q dispersion. We therefore examine if there are differences in broker routing patterns and find two different approaches.

Approximately two-thirds of our brokers route stocks to wholesalers using what we call a "proportional" method. These brokers simply take a "slice" of their aggregate order flow and

[4]As shown by Schwarz et al. (2023), brokers experience substantial variations in average execution costs. Such across-broker variations in execution could obfuscate comparisons across wholesalers when examined through the lens of aggregate-level data provided in public disclosures.

send it to each wholesaler; the only variation across wholesalers is the dollar size of their slice. The remaining one-third of brokers use what we call a "selective" routing method ("smart-routing" in industry parlance), where the routing for each stock is selected from wholesaler execution and stock characteristics. We find that proportional brokers have higher E/Q dispersion and are less responsive than selective brokers.

One possible explanation for the limited response of brokers to past execution quality could be that routing has stabilized at a steady state. If this were the case, we would expect brokers to direct orders in a manner that minimizes costs by sending more orders to wholesalers that provide lower execution costs. Instead, our findings show that the majority of brokers, the proportional-routing ones, actually route more orders to wholesalers associated with higher average execution costs for our trades. To illustrate the impact of not changing market share to predictably lower-cost wholesalers, we show that simulated dynamic routing decisions based on our prior price execution could lower trading costs by 34% on average, hypothetically.

We continue to explore competition by examining the entry of Jane Street as a new wholesaler for a particular broker in early 2022. This broker is the most responsive one from our empirical findings, and also has data starting in 2021. Immediately after this entry, we find that, even with their reduced market share, incumbent wholesalers significantly improved their price execution at that broker. Overall, trading costs are reduced by 14%. This suggests that retail investors at that broker benefited from this increased wholesaler competition.

To shed light on the underlying economic intuition, we develop a stylized model of order routing. In this model, a broker can route orders to two wholesalers but faces switching costs when changing the market share. Switching costs capture potential frictions that limit the ability or willingness of brokers to respond to dispersion in execution quality. For example, these costs could represent the time and cost of monitoring execution and implementing technology for active routing. Alternatively, brokers could be setting maximum or minimum limits on market shares across wholesalers, reflecting for instance capacity constraints at the wholesaler level, broker diversification requirements, and so on.

Equilibrium outcomes depend on the size of switching costs. When switching costs are absent or very small, the equilibrium resembles that of Bertrand competition. There is no dispersion among execution costs since brokers can just change to the lowest cost wholesaler. Therefore, there is no relation between trading costs and market share. Next, when switching costs are high, wholesalers with the largest share have an incentive to raise trading costs because the resulting increase in their profit margin more than offsets any reduction in their market share.[5] This leads to a positive relation between market share and trading costs, consistent with what we observe for proportional brokers. Finally, when switching costs are moderate, a large wholesaler with lower marginal production costs has an incentive to decrease trading costs as the drop in profit margin is more than offset by an increased market share. This leads to a negative relation between market share and trading costs, which is similar to what we see with selective brokers.

As noted, the SEC has proposed several rules to increase competition in the broker-wholesaler marketplace. These proposals attempt to accomplish this goal through changes in best execution requirements, tick sizes, disclosure, and how orders are executed in the

[5]This is because, from the viewpoint of wholesalers, higher trading costs translate into greater revenues for them.

marketplace. The direct implication of our results is that more disclosure would be beneficial for retail investors. Under the recently adopted disclosure requirements (SEC (2022a)), the 605 forms (disclosures of order execution information by market centers) will be required not only for market centers, but also for retail brokers. This would help investors pick brokers that provide lower execution costs. However, the proposal still does not require disclosing execution costs by broker-wholesaler pair. As we show, disaggregated data at broker-wholesaler level would be more informative about wholesaler performance because it controls for broker order-flow characteristics. Such public disclosures would create additional pressure on brokers for better monitoring and execution, creating more competition in this market.

Finally, it is important to note the limitations of our study. Our experiment was based solely on placing small "market" orders for equities during the day; we do not evaluate other types of orders or options trading. Most of our orders are odd-lots, i.e., less than 100 shares, although we also experimented with round lots, with similar results.[6] Even so, it should be noted that odd lots are becoming increasingly important, now accounting for 60% of orders and close to 20% of trading volume.[7] Further, we only examine execution quality in terms of price improvement, while other aspects may be important as well. Lastly, we do not observe the entire order flow of our brokers.

[6]For example, if some brokers only set allocations at the overall order flow level either due to technological limitations or due to prioritizing one segment of orders (such as trade size) or one set of symbols (tickers), then any segment of their flow that is not prioritized will exhibit sluggish adjustment since execution within that segment does not impact routing.

[7]See https://www.sec.gov/marketstructure/datavis/ma_overview.html. Currently, the 605 forms do not cover odd lots. So, it is hard to know whether execution quality systematically differs from round lots. Bartlett (2021) uses a regression discontinuity approach, using 2020 data, to show a drop in PI from 45% to 41% for odd lots, implying a slightly worse execution. On the other hand, Fidelity, one of the largest retail brokers, reports that, during 2022Q4, its odd-lot bucket has actually better price improvement, by about 1 ppt. See https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/FIF-FBS-retail-execution-quality-stats.pdf.

Our paper is most closely related to Dyhrberg et al. (2023). The authors use public 605 forms by market centers to investigate the competitiveness of the wholesaler marketplace.[8] Their general conclusion is that the wholesaler market is competitive. They find that lower cost wholesalers have higher market share and that Jane Street's entry into the market did not cause competitors to change their pricing. While we are able to replicate their main empirical finding, we find more mixed results based on more granular data. By observing specific routing information and comparing execution across wholesalers *within broker*, we find that competition among wholesalers is far from perfect. Other previous work on execution quality has mostly focused on different market centers and types of trades. Also, the market environment has drastically changed since the advent of zero commissions in 2019.[9]

Our paper is also related to Ernst and Spatt (2023) 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 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.

The rest of the paper is organized as follows. Section 2 reviews the institutional background, in particular interactions between brokers and wholesalers. Section 3 then describes

[8]For details on Form 605, see SEC (2005) ("Disclosure of Order Execution Information"). One benefit of the 605 reports is that the filers observed the direction of trades, so that there is no need to use an algorithm to do so. Indeed, the standard Trade and Quote (TAQ) database has no information on wholesalers, brokers, direction of trade, and so forth. So, the typical approach in empirical research based on TAQ uses the Lee and Ready (1991) algorithm, or the more recent Boehmer et al. (2021) method. These assign buy (sell) signals from trades executed above (below) the midpoint or based on the amount of subpenny price improvement, respectively. The problem, however, is that these algorithms misclassify many entries at or beyond the midpoint, which systematically understates the extent of price improvement.

[9]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.

our self-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. Section 6 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, who as broker-dealers execute trades, with generally better pricing. 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.[10]

Interestingly, the relation between brokers and wholesalers is rather loose, reflecting the nature of their private arrangements.[11] 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, and wholesalers do not commit to any set price improvement. Notably, brokers receive no

[10] 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 and Spatt (2023), 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, as was the practice of retail broker-dealers internalizing trades, although to a much lesser extent than nowadays.

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

indicative quotes, pre-trade, from wholesalers.

From the viewpoint of wholesalers, they only see the order flow from each broker in aggregate, i.e., without knowing the identity of the clients placing orders. Wholesalers can execute the order flow they receive, either internally, i.e., into or from their own inventory, or externally, i.e., pass them along to another venue for execution, with the majority — close to 90% — internalized (SEC (2022b)).

At this point, it is useful to detail the best execution requirements, which apply to both the originating broker and the wholesaler, who acts as "executing" broker. 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. They cannot see execution quality for the same trades executed at the same venues for other brokers, which motivates the proposal to expand the 605 disclosure of execution information to brokers (SEC (2022a)). 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; 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.

Admittedly, 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. Indeed, different 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.[12] 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 important. For instance, it may be beneficial to keep small allocations to some venues to enable broader and continuous comparisons of execution

[12]Brokers provide anonymized rankings across their wholesalers. Reportedly, this is generally done across trade size segment, e.g., odd-lots, then in "buckets" of 100-499, 500-1999, 2000-4999, and above 5,000 shares.

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. Brokers may even be hesitant to route a majority of their orders to one wholesaler in order to diversify against operational issues such as outages. Relatedly, some wholesalers may not have the technical capabilities to handle multiples of their current trading volumes. More generally, the quality and breadth of services offered by the wholesaler also certainly matters.

3. Data and Descriptive Evidence

3.1. Data

Our main source of data is self-generated. In summary, we place simultaneous identical trades (i.e., trades in the same stock of the same order size at the same time) across multiple brokerage accounts. These include E*Trade (ET), Fidelity (FD), Interactive Brokers (IBKR), Robinhood (RH), Schwab (SC), and TD Ameritrade (TD). For IBKR, we opened first an account with commissions (IBKR Pro), and then without commissions (IBKR Lite). Otherwise, all other brokers charge no commission. After (before) the opening (closing) auction, we place round-trip "market" orders for 138 stocks. We use market orders because they are the most widely used trade type by retail investors.[13] Our order target size is $100, with a minimum order size of a full share. An early version of this dataset is used in both Schwarz et al. (2023) and Barber et al. (2023). Further details about the experiment can be found in those manuscripts.

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

In total, we placed 171,634 trades equivalent to $22.4 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. TAQ also provides a broad classification of trade locations, but most of our trades are off-exchanges and only coded as "D." Next, we use return and volume data from CRSP. Finally, we were able to identify the actual venues to which our trades were sent, by relying 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. This information is crucial to evaluate routing decisions and wholesaler execution.

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 that brokerage account, whereas the column totals give the total number of trades sent to that wholesaler across all of our accounts. Panel A shows the total number of trades across brokers and wholesalers, whereas Panel B shows percentages by brokers.

[**Insert Table 1 about here**]

Most of our commission-free trades are sent to four wholesalers -- Citadel, Virtu, Jane Street, and G1X. The rest is sent mostly to Two Sigma (mainly by Robinhood), to UBS, which shrunk its market share over this period, and other venues. Routing patterns in the IBKR Pro account are very different, however, with most of the trades sent to IBKR's own Alternative Trading System (ATS) and to exchanges.[14] The routing patterns are also shown

[14]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

in Figure 1.

[Insert Figure 1 about here]

We use effective over quoted (E/Q) spread to measure transaction 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 $1 - 2 \times PI$. Therefore, lower E/Q is equivalent to greater PI and indicates reduced trading costs.

Our data reveal substantial variation in order routing as well as effective spreads, both over time and in the cross-section. To provide a visual representation of the details in our dataset, Figure 2 plots 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 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]

Our data have two main advantages over publicly available order routing information. Firstly, 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. Specifically, brokers are required to file SEC Rule 606 reports that disclose the fraction of orders routed to the venues used. However, there is no execution information.

SEC and FINRA oversight. They generally charge no execution fees or fees that are lower than exchanges.

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. In contrast, we can precisely trace the routing and execution for each one of our trades and therefore compare wholesalers at the same broker.[15] Thus, we have the full joint distributions instead of only the marginal ones. Secondly, our trades are largely odd lots, which represent a majority of retail trades but are not reported in 605 filings.

That being said, since most of our orders are odd lot orders and we only examine market orders for equities, it is important to acknowledge that our results may not generalize to larger orders or other order types.[16]

3.2. Stock Routing Patterns

Brokers can follow different approaches to routing. The first is to simply route a percentage of their entire order flow to a wholesaler, which we refer to as "proportional" routing. Each wholesaler then receives a slice of the broker's order flow, with the same relative weights (or composition) across stocks, perhaps only differing in the share of total volume sent. One benefit of this approach is that, since each wholesaler receives the same composition for the order flow, the broker can compare execution directly at the aggregate level across wholesalers. In addition, discussions with the industry suggest that wholesalers prefer this proportional approach because the order flow is more diversified as well as more predictable over time, making it easier to manage inventory, thus possibly leading to better execution.

[15] We compare our trades routing to the 606 data for all of our brokers, except IBKR who does not separate their LITE and PRO accounts. The fit was excellent, with an R-square around 90%. As expected, it was slightly worse for Robinhood, which actively changes routing across stocks and over time. So, our routing sample is representative of the brokers' routing.

[16] Schwarz et al. (2023), however, did place larger trades and trades over 100 shares. They find similar execution compared to their $100 trades. Thus, we would expect orders in larger sizes to be handled similarly.

The second approach is for brokers to route orders on a stock-by-stock basis, which we call "selective" routing, usually described as "smart" routing by the industry. Under this scenario, brokers evaluate execution for each stock individually and then normally increase the fraction of each stock routed to wholesalers that have provided the best execution for that stock. With selective routing, however, different wholesalers would receive different order flow composition, making overall execution quality not directly comparable across wholesalers, making performance evaluation much more complex.

To illustrate, for each broker, we first compute the percentage of our orders for each stock with more than 100 trades that are routed to Citadel. We then sort each stock from the lowest to the highest percentage. We display results in Figure 3 for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, and TD Ameritrade, in Panels A to F, respectively. In Panel G, we plot results for IBKR Pro, except that we calculate the percentage sent to its own ATS instead of Citadel. The horizontal axis corresponds to each of our stocks. Each entry describes the average percentage routed to the selected venue with a circle in the middle of whiskers that represent 95% confidence bands. Red lines indicate that the percentage is significantly different from the overall average for that broker.

[**Insert Figure 3 about here**]

If brokers were using a "proportional" method to route orders, we should observe that all of our stocks have essentially the same percentage of their orders routed to Citadel, or that the average line should be flat. On the other hand, 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 use a method close to proportional routing: E*Trade, Fidelity, Schwab, and TD Ameritrade. The graphs for IBKR Lite and Pro and Robinhood, on the other hand, indicate that these two brokers use selective order routing. Indeed, these two brokers advertise this feature.[17]

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 can be directly observed by the broker. Following the idea that more trades should be routed to venues that provide lower execution cost, the first variable is the prior calendar month's E/Q ratio for that stock at that venue minus the average E/Q for that same stock across venues, all at the same broker (*Venue Excess E/Q (t-1)*). This variable measures the cost of price execution for this stock traded at this wholesaler relative to other venues. We would expect a negative coefficient with selective routing. The second variable is the percentage of trades that were routed to that venue during the prior calendar month (*Venue % (t-1)*), which controls for persistence in routing decisions.

We also include a number of other control variables. Stock characteristics include the log of the stock price, log-volume, and both the day's return and the absolute value of the return. We also include the spread at the time of the trade, the trade direction (1/0 indicates a buy/sell), and a S&P500 index membership indicator (1 if so, or 0). The last variable indicates whether our last trade went to the same venue. In all models, we include

[17]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/.

day fixed effects.[18] We report results in Table 2 for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, and TD Ameritrade in Panels A to F, respectively.[19]

[Insert Table 2 about here]

The logistic models lead to the same conclusions as Figure 3. Four brokers have nearly proportional routing. The coefficients on prior execution costs are mostly not significant. Thus, these brokers clearly do not allocate order flow on the basis of past stock-level execution.

On the other hand, many variables are significant for Robinhood and IBKR Lite. In both cases, a wholesaler is more likely to get an order if that stock's prior month execution was better than other wholesalers. Order flow is also persistent over time. In many cases, the selective-routing systems for these two brokers agree. For example, both are more likely to route trades for stocks with high volumes to Citadel. We find similar results for IBKR Pro. That brokerage account sends orders to different venues (its own ATS, exchanges, and wholesalers) persistently based on prior execution as well.

Interestingly, based on the findings of Schwarz et al. (2023), the three accounts with the lowest overall price improvement are those using selective routing. Other factors, however, may be the primary drivers of the observed differences in execution, such as the toxicity of order flow.[20] Likewise, perhaps brokers with more toxic order flows have to work especially

[18] Because the number of days that we have traded exceeds the number of stocks, we cannot cluster by stock and include day fixed effects in the same model. We find similar conclusions if we remove the day fixed effects and cluster by stock, or if we include month fixed effects and cluster by stock.

[19] 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.

[20] Also, brokers may face conflicts of interests when operating their own ATS that may affect execution quality. Anand et al. (2021) examine institutional brokers that operate their own ATS. They argue that this setup can create potential conflicts of interest. For example, such brokers would avoid paying exchange fees

hard at improving their execution quality. In any event, the key issue is the quality of actual, not past, trade execution.

4. Competitiveness of Retail Trade Execution

In the prior section, we described brokers' routing patterns. In this section, we examine competition of the wholesaler marketplace where our retail trades are sent.

4.1. Price Execution Dispersion and Persistence

We begin with 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 these numbers for each broker-wholesaler pair. We show the average E/Q for each broker, as well as the excess E/Q that represents the deviation of a given wholesaler's execution from the broker's average for each stock. These averages are computed over all trades within a month for each pair. We also report time-series statistics 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]

Across all brokers, we observe statistically and economically large differences across wholesalers, even within the largest ones. For E*Trade, Fidelity, and Schwab, the spread

that they typically absorb by using affiliated venues. Also, they benefit from higher volumes on affiliated venues because other participants typically pay fees to their ATS. As a result, such brokers may prefer affiliated ATSs over other venues, even if not optimal for the client. Indeed, these authors report that such trading in affiliated ATSs is associated with lower execution quality.

between the best and worst execution cost E/Q is greater than 0.20. TD has the smallest spread, at 0.10, but also the lowest average execution cost.

Next, we find that the dispersion in price execution is persistent over time, both at the aggregate and stock levels. To evaluate persistence, we regress this excess E/Q against its average over the last one- and three-month periods. 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 in Table 4.

[Insert Table 4 about here]

Panels A and B show persistence at the overall level. We find that wholesaler comparative performance is extremely persistent at the aggregate level. Across all brokers, the prior one-month (three-month) coefficient is 0.74 (0.83), which is very high and statistically significant, and the R-square is very high, above 50%. The slope coefficient is lower than one, which suggest slow reversion to the broker mean, or mild amelioration of performance over time. These results hold up in our two subgroups of proportional and selective-routing brokers, with similar slope coefficients. We also find high persistence at each broker. Figure 4 illustrates this very strong relationship.

[Insert Figure 4 about here]

Next, we perform the same analysis at the individual stock level. At each broker, excess E/Q is now calculated as the wholesaler E/Q for a stock at a broker minus the E/Q for that stock across all wholesalers for that broker, both averaged over the month. Even though we found minor effects of stock characteristics on routing in Table 2, this allows us to control for possible tilts in the stocks routed across wholesalers, which could create artificial persistence in excess E/Q. Additionally, this 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. On the other hand, using data at the stock level is surely noisier than at the aggregate level, leading to estimated coefficients that can be biased downward if the right-hand-side variables are affected by greater errors in the variables.[21] Results are shown in Panels C and D of Table 4.

The results are consistent with those at the aggregate level. Across all brokers, subgroups, and individual brokers, we find that price execution is persistent based on the prior one- and three-month price execution. Across all brokers, the slope coefficients are 0.21 and 0.48, respectively, and also highly statistically significant. The values of the coefficients, however, are systematically lower than those at the aggregate level in Panels A and B. Also, the R-squares are on the order of four to eight percent instead of 50 percent. As indicated, the lower coefficients surely reflect the greater degree of noise when evaluating PI at the stock level.

Overall, these findings suggest that, for a given broker, there is a substantial and persistent dispersion in price executions across wholesalers.

4.2. Broker Response to Prior Price Improvement

Since wholesalers' price execution is predictable, brokers can use prior data to make changes to their routing practices to obtain greater price improvement for their customers.

[21] This issue is akin to tests of asset pricing models where the right-hand-side variable consists of stock-level historical betas, which are affected by estimation error, reflecting the usual sampling variability. Traditional methodology then groups stocks into portfolios to decrease this error, and thus lower the bias in slope coefficients. See Kim (1995), for example. In this case, individual stock inventory differences, which likely drive execution, would be diversified at the aggregate level.

Indeed, brokers are required, under their best execution obligations, to evaluate the execution quality of venues over time and to act upon this evaluation. Doing so should be an essential practice to maintain competition in the wholesaler market. As previously mentioned, brokers reportedly comply with these obligations "by establishing routing allocations based on this historical performance." In this section, we examine if the evidence supports this assertion.

4.2.1. Routing Changes and Prior Price Execution

To determine how brokers react to prior execution, we first compute the monthly change in the percentage of orders routed from each broker to each wholesaler, using all trades across all stocks. Next, we regress these changes against the past excess E/Q for that wholesaler at that broker, 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 relationship, reflecting less routing to wholesalers with higher execution costs. For our brokers that use proportional routing, changes should be observed at the overall routing level. For selective-routing brokers, changes should occur at the stock level, perhaps obscuring changes at the overall level. Our models include appropriate brokers dummies and cluster standard errors by month. We report results at the overall levels in Panels A and B of Table 5.

[Insert Table 5 about here]

Across all brokers, we find some evidence that brokers change their routing toward wholesalers that provided better execution the prior month. The slopes are barely statistically significant, however, and the economic magnitude, for example of the -0.029 coefficient, 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 $-2 \times -0.029 = 6.2\%$ 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.[22]

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 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 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.

[22] 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.

4.2.2. Market Share and Prior Price Execution

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, they focus on market share levels. 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 96% 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 -0.25 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 0.3% 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.

For the "Top 4" subsample (Panel B), for all brokers, however, results now 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.[23]

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.

[**Insert Figure 6 about here**]

4.2.3. Counterfactual

Next, we run a counterfactual analysis to evaluate the potential improvement in our execution from routing orders based on past execution. Each month for each broker, we simulate routing 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

[23] 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.

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 aggregate 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 aggregate execution, the effective spread decreases from an average across brokers of 0.313 to 0.233, which is a major reduction in trading costs of 26% across brokers. While statistically significant for all brokers, the 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 lead 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 modulating the change would result in large E/Q improvements. So, the open question is why brokers are not rerouting trades more actively. Regardless of the reason, our results suggest that the wholesaler market is not perfectly competitive and therefore has room for improvement.

4.3. Impact of a New Wholesaler on Price Execution

One potential reason why brokers are not altering their routing is a lack of 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.[24] 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 22, 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 23, 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]

[24]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.

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 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 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.

4.4. Discussions and Implications for 605 Forms

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.

More generally, the current 605 forms obfuscate within-broker execution quality. This can be demonstrated as follows. First, note that Robinhood's average execution is worse than that of other brokers, perhaps reflecting its order flow toxicity. Next, we noted that, following its entry in 2022, Jane Street gained market share at Robinhood while Virtu significantly lost market share. This certainly represents an improvement for investors. However, looking at the 605 form for the smallest reported trades, from 100 to 499 shares, which best match Robinhood's retail clients, Jane Street actually shows a worsening of execution quality from January to April 2022. This worsening is an artifact of the higher fraction of trades executed for Robinhood. On the flip side, Virtu's 605 form shows an improvement in E/Q. However, we do not observe changes in the within-broker E/Q for Jane Street and Virtu. So, the changes in 605 data almost certainly reflect the addition and subtraction of Robinhood trades, respectively, as opposed to true changes in execution quality.

The conclusion is that changes in the clientele served by wholesalers could create misleading changes in aggregate execution numbers shown in the 605 forms. This demonstrates the superiority of within-broker analysis relative to the aggregate reporting in the 605 forms, and the need for expanding 605 reports to the broker-wholesaler pairs.

5. Model of Order Routing with Switching Costs

Overall, our empirical findings indicate that the wholesaler marketplace is imperfectly competitive. While many of our results are intuitive, such as the increased competition from Jane Street's entry improving the execution quality of other wholesalers, others are more puzzling. Specially, why would proportional brokers route larger shares of their orders to wholesalers with lower price execution quality? If they were randomly allocating trades, we would expect the relation to be random, not significantly positive. More generally, if price execution is predictable, why do brokers not respond to persistent dispersion across different wholesalers? To provide economic insights into these questions, this section develops a stylized model of order routing including brokers' switching costs.

5.1. Setup and Equilibrium

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. 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.

A key assumption in the model is that the broker incurs quadratic switching costs when adjusting the market share between the wholesalers.[25] Switching costs serve as a simple modeling device to capture potential frictions that limit the ability or willingness of brokers to respond to dispersion in execution quality, consistent with what we document in Section 4. For example, switching costs could reflect the time and cost it takes for brokers to monitor wholesalers' performance, managerial/organizational inertia ("nobody gets fired for buying IBM"), the desire/requirement to supply stable order flows to wholesalers, or the lack of technology to implement complex routing, etc.

Specifically, we assume that switching costs are quadratic in the adjustment of market share, given by $\frac{s}{2}\Delta^2$, where $s > 0$ and $\Delta \in [-\sigma, 1-\sigma]$ is the additional market share allocated to wholesaler X. Then the broker optimally chooses an adjustment of the market share Δ to minimize the sum of the execution costs and the 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 to process and make markets for the broker's (customers') orders. These costs may be heterogeneous, denoted by f_X and f_Y for wholesalers X and Y, respectively, where $f_X \leq f_Y$ without loss of generality. Wholesaler X optimally chooses the execution cost p_X to maximize its profits:

$$\max_{p_X} (p_X - f_X)(\sigma + \Delta). \tag{2}$$

[25]Switching costs have long been used and studied in the economics literature (see, e.g., Klemperer (1987)).

Similarly, wholesaler Y chooses the execution cost p_Y to maximize its profits:

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

We focus on a pure-strategy Nash equilibrium, which is found when (1) the broker optimally chooses its adjustment in the market share, Δ, as a function of the execution costs charged by the wholesalers; (2) the wholesalers optimally chooses execution costs, p_X and p_Y, given the broker's strategy and each other's pricing; and (3) the broker's and the two wholesalers' strategies, Δ, p_X, and p_Y, are all consistent. The proposition below fully characterizes equilibrium. The proof is in Appendix A.

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*

$$p_X = \frac{2f_X + f_Y}{3} + \frac{s(1+\sigma)}{3} \qquad \textit{and} \qquad p_Y = \frac{f_X + 2f_Y}{3} + \frac{s(2-\sigma)}{3}, \tag{4}$$

and the broker adjusts its market share by

$$\Delta = \frac{f_Y - f_X}{3s} + \frac{1 - 2\sigma}{3}. \tag{5}$$

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

$$p_X = f_Y - s(1-\sigma), \qquad p_Y = f_Y, \qquad \textit{and} \qquad \Delta = 1 - \sigma, \tag{6}$$

in other words, the lower-cost wholesaler X captures the entire market.

In the following subsections, we provide the intuition and study properties of the equilibrium characterized above to connect the theoretical implications with our empirical findings. To clarify economic intuition, we analyze three cases separately: (1) the benchmark case absent switching costs, (2) the case with switching costs and the same marginal costs for the wholesalers, and (3) the case with switching costs and different marginal costs.

5.2. Equilibrium Absent Switching Costs

As a benchmark, first consider the case when there is no switching cost (i.e., $s = 0$). Then we have a classic Bertrand (or price) competition between the two wholesalers.

When the wholesalers have the same marginal cost, there is perfect competition even with just two competitors, with their execution prices being equal to the marginal cost ($p_X = p_Y = f_X = f_Y$); this case is also well known as the "Bertrand paradox." Any dispersion in prices makes the broker shift its entire order flow to the lower-cost wholesaler. Such responsive routing in turn makes the wholesalers compete vigorously with one another, eliminating price dispersion and driving down prices to the marginal cost.

When the wholesalers have different marginal costs ($f_X < f_Y$), however, the market is imperfectly competitive. Competition between the two wholesalers drives down prices only to the higher marginal cost, allowing the low-cost wholesaler to remain profitable. Thus, responsive routing and Bertrand competition alone do not guarantee perfect competition.

Regardless of whether wholesalers have the same marginal cost or not, Bertrand competition implies that there is no dispersion among execution costs charged by wholesalers. Even if there were a dispersion, we should not observe this in the data because the lowest-cost wholesaler should capture a 100% market share. Further, the broker's responsive routing is inconsistent with our finding that most brokers do not respond to prior execution costs.

5.3. Equilibrium When the Wholesalers Have the Same Marginal Cost

Now, suppose that the broker incurs switching costs (i.e., $s > 0$), and assume, for now, that the two wholesalers have the same marginal cost (i.e., $f_X = f_Y$).

The broker finds the optimal adjustment of market share, Δ, by solving Equation (1). From the first order condition, we have[26]

$$\Delta = \frac{p_Y - p_X}{s}. \tag{7}$$

Thus, the broker always moves towards the lower-cost wholesaler (i.e., $\Delta > 0$ if and only if $p_X < p_Y$). But the extent to which it does so depends on and decreases in the switching costs. Higher switching costs make it difficult for the broker to adjust market share drastically even when there is wide price dispersion.

In equilibrium (Proposition 1), the wholesalers charge:

$$p_X = f_X + \frac{s(1+\sigma)}{3} \qquad \text{and} \qquad p_Y = f_X + \frac{s(2-\sigma)}{3}. \tag{8}$$

As switching costs increase, both wholesalers charge more and earn higher margins relative to the marginal cost. The broker's limited ability to adjust market share provides the wholesalers with a scope to exercise their market power.

Further, the extent to which wholesalers extract rents from the broker (or more precisely,

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

its customers) depends on and increases in their market share. The wholesaler with larger pre-existing market share (i.e., X if $\sigma > 1/2$ and Y if $\sigma < 1/2$) charges higher costs than the other wholesaler. While the broker does move away from the higher-cost, large wholesaler, that wholesaler remains large after the adjustment.[27] For example, if wholesaler X was initially large with $\sigma = 2/3$, then it remains large after the adjustment with $\Delta = -1/9$ and $\sigma + \Delta = 5/9$, even though p_X is higher, at $p_X = p_Y + s/9$.

Thus, there is a positive relation between market share and execution costs, consistent with what we find for proportional brokers in Figure 6. See also Panel A in Figure 7. While the positive relation might initially appear counter-intuitive, the intuition now becomes clear. It is not that brokers route more order to some wholesalers because they charge higher costs, but rather that some wholesalers can charge higher costs because brokers route more orders to them. Note that this result does not require that brokers have substantial switching costs. Any switching costs imply that the broker routes more order flows to the more expensive wholesaler.[28]

5.4. Equilibrium When the Wholesalers Have the Different Marginal Costs

In the prior cases, we would find either no or a positive relation between execution costs and market share. In this context, the negative relation that we document for selective brokers requires further analysis. While selective brokers somewhat respond to execution, their routing behaviors still reflect the presence of strictly positive switching costs.

To understand the negative relation between market share and execution costs, we next allow wholesalers to have different marginal costs (i.e., $f_X < f_Y$ since we assume $f_X \leq f_Y$

[27] In equilibrium, $\sigma + \Delta = (1 + \sigma)/3$ such that $\sigma + \Delta > 1/2$ if and only if $\sigma > 1/2$.

[28] This is because when $f_X = f_Y$, the second case in Proposition (1) does not apply.

without loss of generality), while the broker still incurs switching costs (i.e., $s > 0$).

There are three cases depending on the size of switching costs; see Panel C of figure 7. When switching cost are low (i.e., $s \leq (f_Y - f_X)/(2-\sigma)$), there is no relation between market share and execution costs since the equilibrium resembles Bertrand competition with no switching costs (Section 5.2). The lower-cost wholesaler drives the other wholesaler to zero profits and obtains a 100% market share. Interestingly, switching costs reduce the prices the broker pays because the broker would not shift the entire order flow to the low-cost wholesaler unless it is compensated for the switching costs.

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, as in Panel B of Figure 7. From Equation (4), we obtain that the lower-cost wholesaler (X) charges lower costs and maintains a larger market share, if it starts as the larger wholesaler ($\sigma > 1/2$). The larger wholesaler here has two competing incentives. On the one hand it can charge more to take advantage of its market share and extract larger rents from the broker. On the other hand, it can charge less to take advantage of its lower costs and obtain an even larger market share. When the broker's switching costs are not too high, the incentive to reduce costs outweighs the incentive to raise costs, resulting in the larger wholesaler charging lower costs than the smaller wholesaler in equilibrium.

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, market share and execution costs have positive relationship, as in the case with the same marginal costs in Section 5.3. Given the high switching costs, the broker cannot easily change the market share, and the wholesaler with a larger market share charges higher costs and maintains a larger market share.

Notice, these results are the most consistent with our empirical findings that selective brokers, who tend to respond to execution albeit modestly and thus exhibit smaller switching costs than proportional brokers, also exhibit a negative relation between market share and execution costs, while proportional brokers generate a positive relation between market share and execution costs.

We mention in passing that the difference in switching costs between proportional and selective brokers may be explained by regulatory pressure. Prior literature has shown that large variations in execution costs across brokers likely reflect different characteristics (or toxicity) of order flows (e.g., Schwarz et al. (2023)). Despite the varying degrees of toxicity in order flows, brokers' executions are often benchmarked against the same regulatory standard, such as NBBO. Thus, brokers with more toxic order flows have stronger incentives to improve their execution quality than those with less toxic order flows. If brokers can make costly investments to reduce their switching costs, those with more toxic order flows are likely to have lower switching costs than their counterparts with less toxic order flows. In fact, selective brokers, whose behaviors are consistent with having lower switching costs than proportional brokers, are also the ones with the two highest execution costs, indicating a higher prevalence of more toxic order flows (Table 3).

6. Conclusions

Retail trading has reached record volumes in the last several years, spurred by technological advances as well as the advent of commission-free trading. Even so, the current market structure has attracted the attention of regulators. Many worry about potential conflicts of interest, such as payment for order flow. Others worry about the competitiveness of the wholesaler marketplace given that it has only four large players. These concerns ultimately led the SEC to introduce four proposals aimed at increasing competition in order to improve trading execution. The counterargument is that competition across wholesalers is already enforced by retail brokers, as required by best execution standards.

Using a self-created dataset of over 150,000 trades that allows us to look at within-broker interactions between brokers and wholesalers, we find significant evidence that the retail trade wholesaler marketplace is not perfectly competitive. First, we find a large dispersion in price execution across wholesalers for the same broker, which is inconsistent with Bertrand competition. Second, while wholesaler's price execution is persistent, brokers largely do not change their routing. Third, surprisingly a majority of our brokers actually route more of their trades to wholesalers with worse execution for our trades. Finally, Jane Street's entry into Robinhood's wholesaler marketplace forced other wholesalers to provide better pricing. We also provide a stylized model that provides economic insights into our results through the lens of switching costs.

Competition could be increased potentially through several methods. For example, greater disclosure would put more pressure on brokers to manage trading costs actively. While the proposed extension of 605 forms to brokers does represent an improvement, further extension of disclosures to each broker-wholesaler pair would help apply such pressure. Even without more disclosure, brokers could allocate their trades more actively, yielding better execution for their retail investors and enforcing pricing discipline on existing wholesalers. Finally, new entrants may also create more competition, forcing existing wholesalers to improve their execution for the ultimate benefit of retail investors.

References

Anand, A., M. Samadi, J. Sokobin, and K. Venkataraman (2021). Institutional order handling and broker-affiliated trading venues. *Review of Financial Studies 34*, 3364–3402.

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. (2021). Modernizing odd lot trading. *Columbia Business Law Review*, 520–568.

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). Absolute and relative wholesaler execution quality in May 2022. 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.

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

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

Ernst, T. and C. Spatt (2023). Payment for order flow and asset choice. https://ssrn.com/abstract=4056512. Working Paper.

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

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.

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

Kim, D. (1995). The errors in the variables problem in the cross-section of expected stock returns. *Journal of Finance 50*, 1605–1634.

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

Lee, C. and M. Ready (1991). Inferring investor behavior from intraday data. *Journal of Finance 46*, 733–746.

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

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.

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.

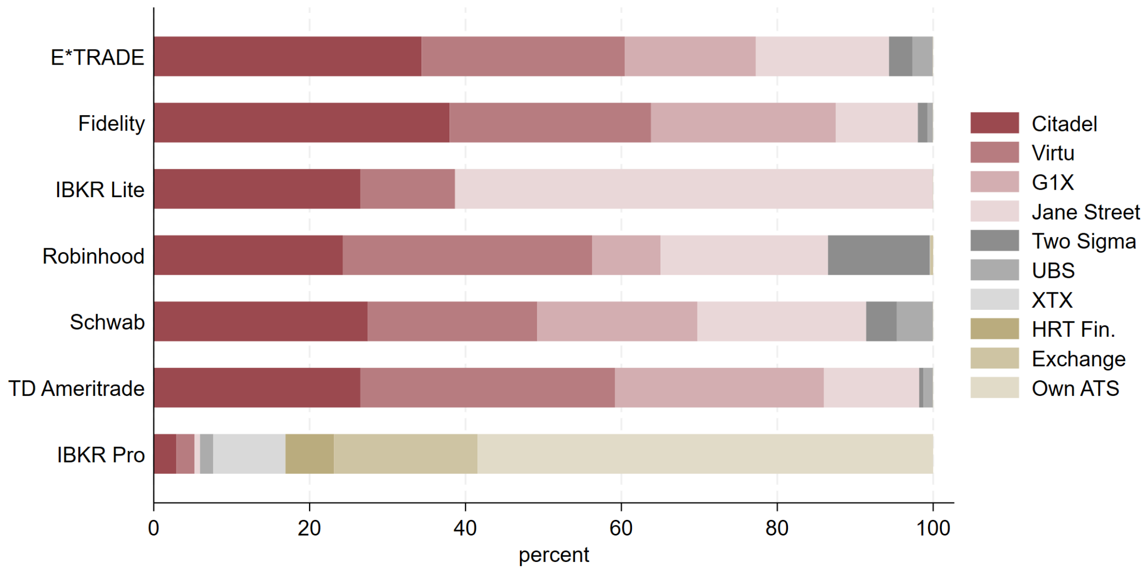
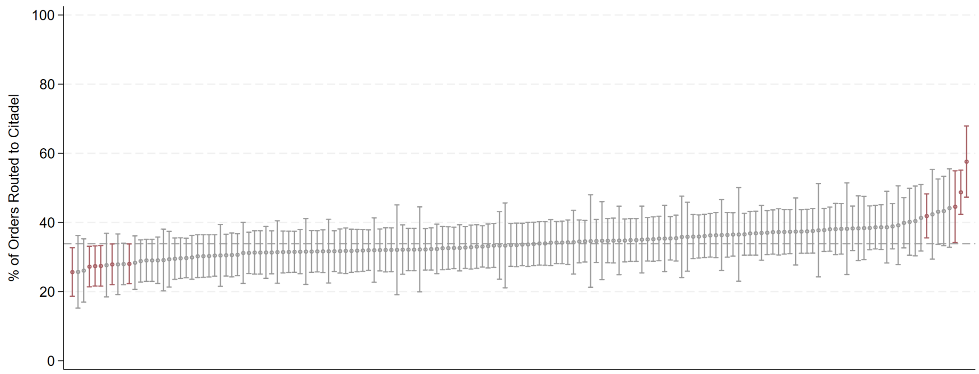


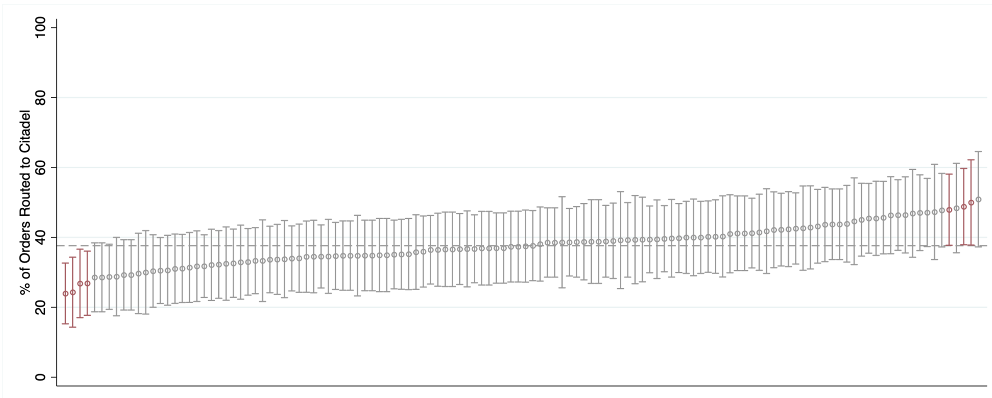
Figure 1: Wholesaler Share by Broker
This figure shows the percent of our orders that went to each wholesaler for each brokerage account. The raw data are in Table 1, Panel B.

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.

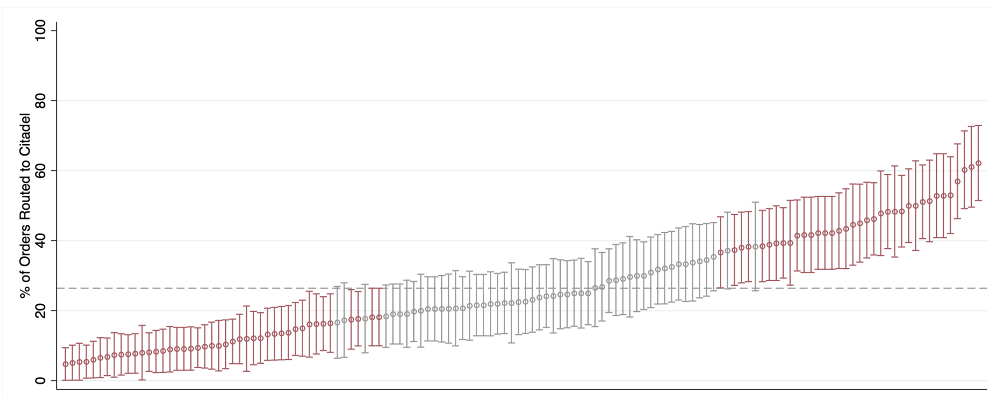
Panel A: E*Trade to Citadel



Panel B: Fidelity to Citadel



Panel C: IBKR Lite to Citadel



Panel D: Robinhood to Citadel

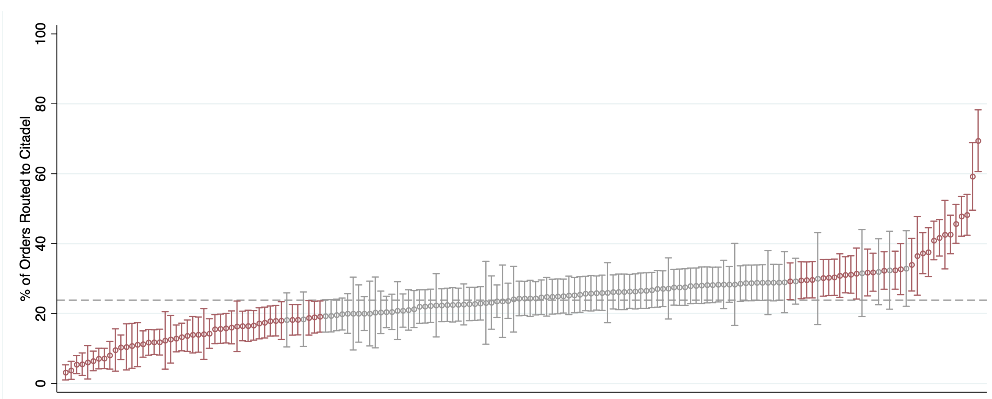
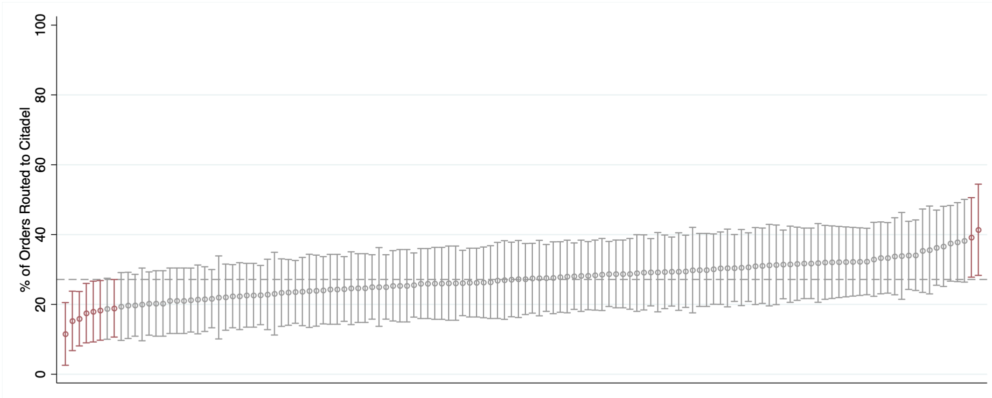
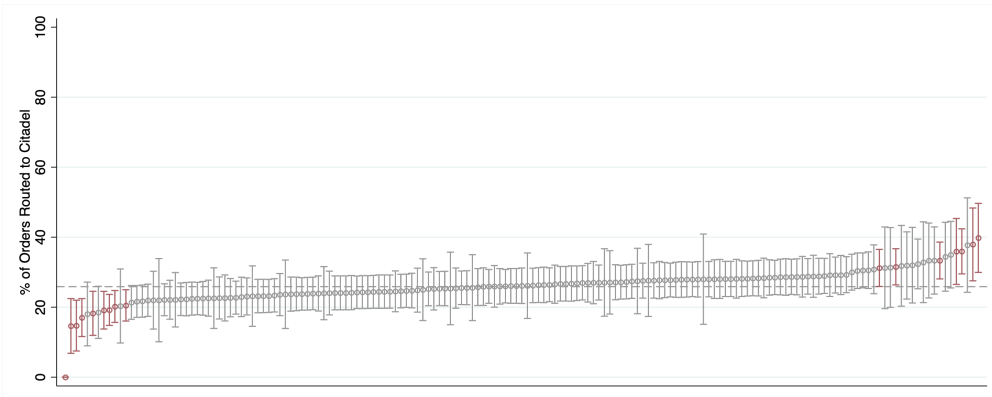


Figure 3: (Continued on the following page.)

Panel E: Schwab to Citadel



Panel F: TD Ameritrade to Citadel



Panel G: IBKR Pro to Own ATS

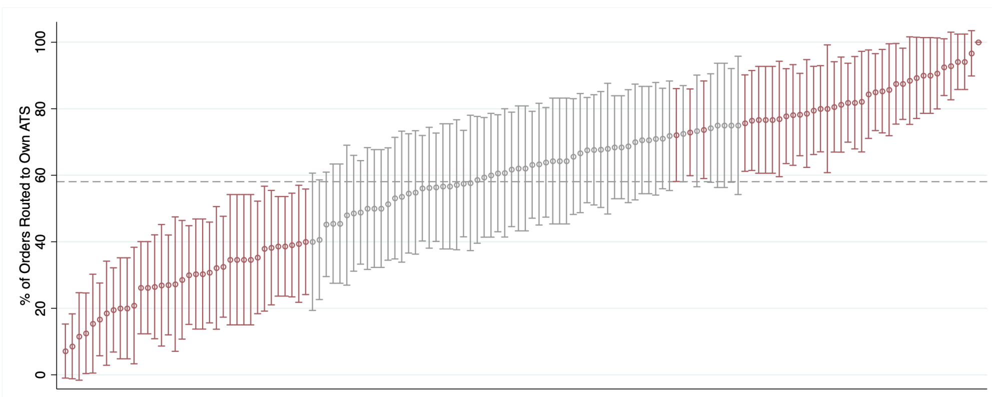


Figure 3: Order Routing Patterns across Individual Stocks
This figure shows the percentage of our orders for each stock that are routed to a specific venue. Panels A to F report order routing for E*Trade, Fidelity, IBKR Lite, Robinhood, Schwab, and TD Ameritrade to Citadel, while Panel G reports order 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.

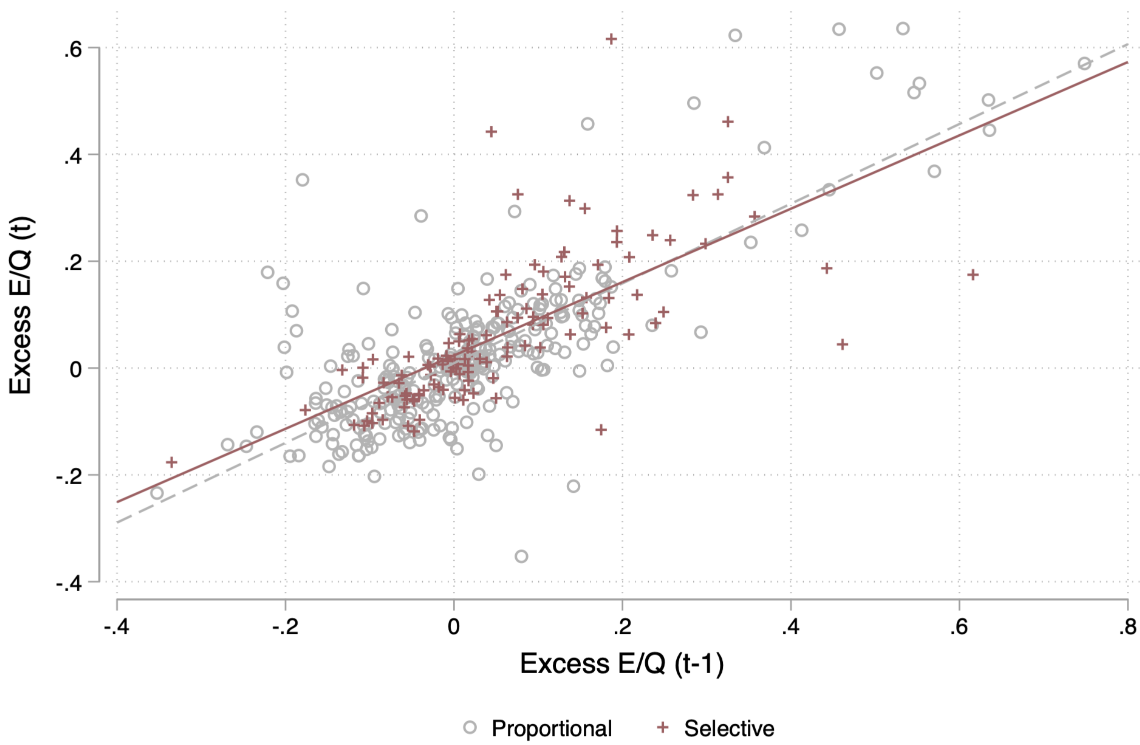


Figure 4: Relation between Current and Prior Month Effective Spreads
This plots, for each broker-wholesaler pair, the excess effective over quoted trade spread (E/Q) for the current month on the vertical axis against that for the prior month. Excess E/Q is computed as the average E/Q for each wholesaler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

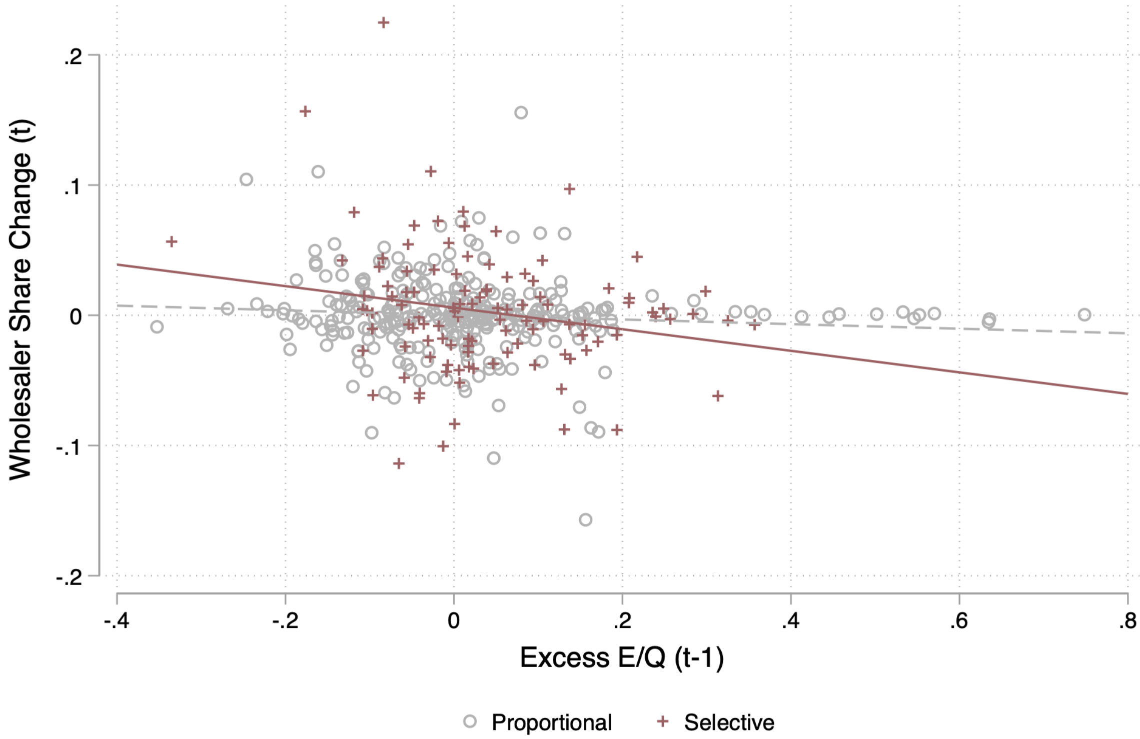


Figure 5: Relation between Market Share Changes and Prior Month Effective Spreads
This plots, for each broker-wholesaler pair, the change in wholesaler market share for the current month on the vertical axis against the prior month excess effective over quoted trade spread (E/Q). Excess E/Q is computed 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.

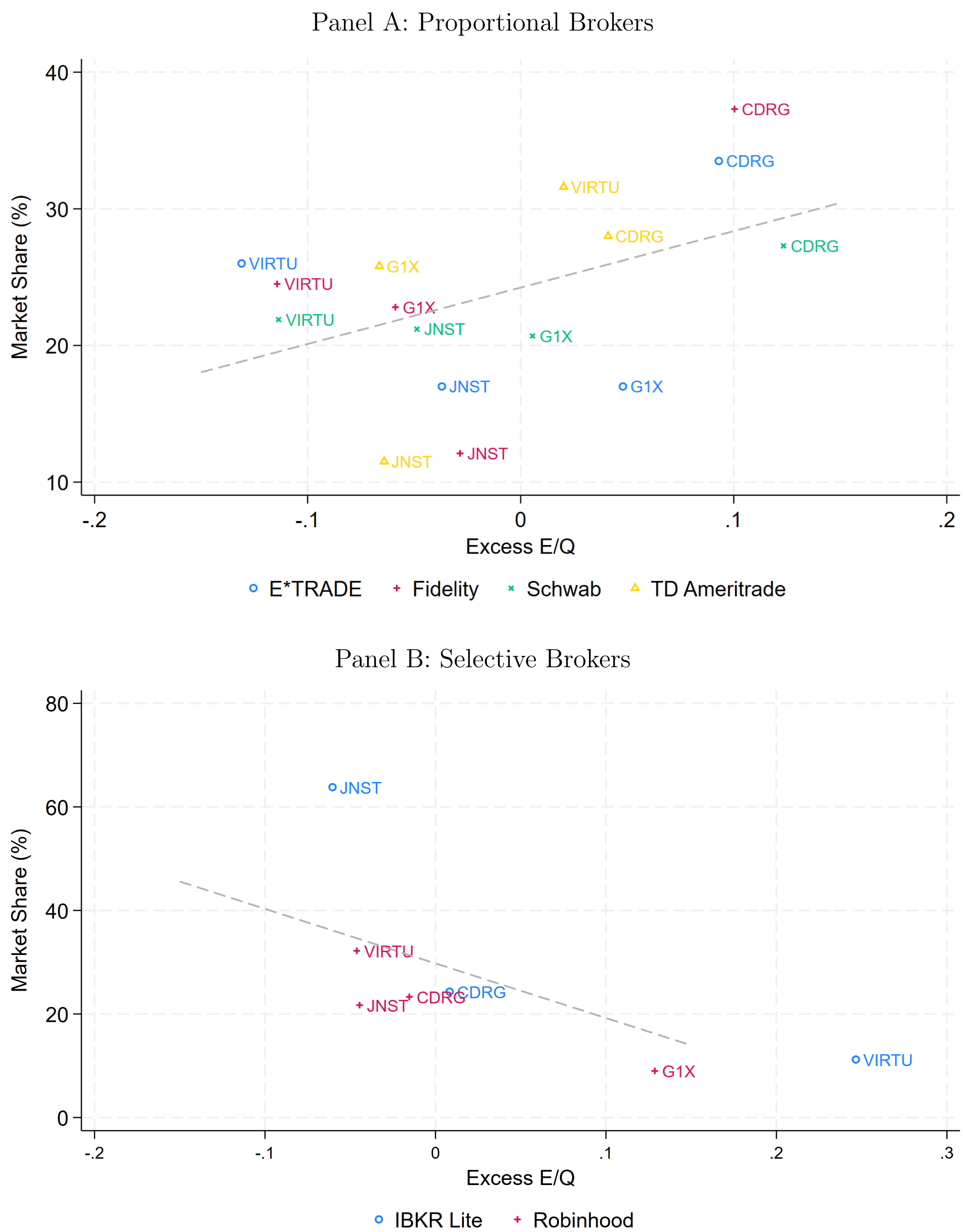


Figure 6: Wholesaler Price Improvement and Market Share
This figure describes the relation, for each broker, between wholesaler market share and its price improvement as measured by its excess effective over quoted spread. The wholesaler sample consists of the "Top 4," including Citadel (CDRG), Virtu,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

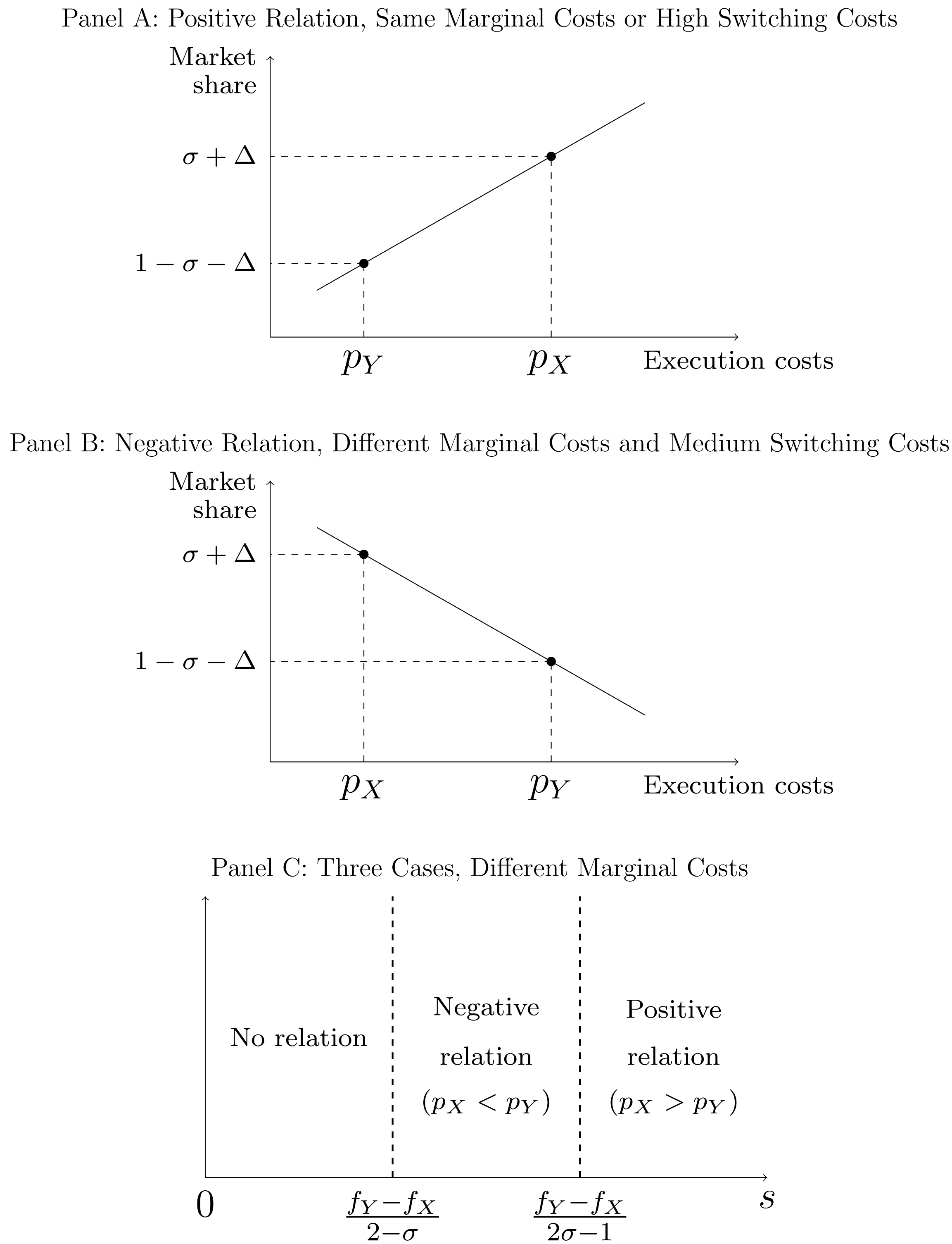


Figure 7: Model
Panel A plots the positive relation between execution and share, which arises when the wholesalers have the same marginal cost or when switching costs are high. Panel B plots the negative relation, which arises when switching costs are medium. Panel C shows how the size of switching costs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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Broker-Wholesaler Routing Count										
	Citadel	Virtu	Jane Street	G1X	Two Sigma	UBS	Exchange	Other	Own ATS	Total
E*Trade	10,651	8,085	5,286	5,202	919	811	-	-	-	30,954
Fidelity	4,561	3,102	1,262	2,847	148	83	-	226	-	12,229
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	-	-	-	11,084
TD Ameritrade	11,088	13,660	5,161	11,160	224	497	-	520	-	42,310
IBKR Pro	126	96	31	-	-	71	1,464	-	2,477	4,265
Total	41,976	41,407	29,251	24,995	6,939	1,974	1,614	880	2,477	151,513

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%		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%				100%
TD Ameritrade	26%	32%	12%	26%	1%	1%		1%		100%
Average	29%	25%	24%	16%	4%	2%	0%	1%		
IBKR Pro	3%	2%	1%			2%	34%		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. Each month, we compute the execution cost for each wholesaler within each broker. "Excess" execution cost is then measured as the difference between the wholesaler average effective over quoted spread (E/Q) at that broker and the overall E/Q for that 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's average effective over quoted spread for reference purposes in the first column. t-values are in parentheses. **,* represents significance at the 1%, 5% levels respectively.

	Broker-level E/Q	Broker-adjusted E/Q (Deviation from broker-level E/Q)				Dispersion (Max-Min) Normalized by Broker-level E/Q
		Citadel	Virtu	Jane Street	G1X	
Fidelity	0.142	0.100**	-0.114**	-0.028	-0.059**	150.70%
		(6.3)	(-9.0)	(-1.4)	(-4.4)	
TD Ameritrade	0.093	0.041**	0.020	-0.064**	-0.066**	115.05%
		(4.1)	(1.7)	(-3.5)	(-9.4)	
Schwab	0.229	0.123**	-0.114**	-0.049**	0.005	103.49%
		(11.9)	(-5.7)	(-3.0)	(0.9)	
E*Trade	0.322	0.093**	-0.131**	-0.037**	0.048**	69.57%
		(6.1)	(-17.1)	(-3.0)	(3.9)	
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)	

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 Overall Price Improvement (Prior Month)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1)	0.735**	0.746**	0.488**	0.627*	0.505**	0.857**	0.687**	0.916**	0.599**
	(12.31)	(13.35)	(3.75)	(3.04)	(6.00)	(13.08)	(5.83)	(11.09)	(4.19)
R-square	0.569	0.590	0.245	0.344	0.373	0.765	0.489	0.811	0.384
Panel B: Wholesaler Overall Price Improvement (Prior 3-month Average)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1, t-3)	0.834**	0.832**	0.654**	0.924**	0.826**	0.868**	0.819**	0.980**	0.741**
	(13.04)	(12.60)	(3.85)	(5.36)	(16.48)	(11.52)	(7.21)	(10.21)	(4.65)
R-square	0.550	0.550	0.272	0.336	0.614	0.667	0.520	0.866	0.400
Panel C: Wholesaler Stock-level Price Improvement (Prior Month)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1)	0.210**	0.196**	0.187**	0.169**	0.188**	0.219**	0.246**	0.507**	0.190**
	(11.00)	(11.07)	(6.98)	(5.62)	(7.08)	(6.95)	(7.33)	(7.95)	(5.01)
R-square	0.043	0.037	0.037	0.024	0.033	0.048	0.058	0.246	0.035
Panel D: Wholesaler Stock-level Price Improvement (Prior 3-month Average)									
Dep Var:	Broker-adjusted E/Q (t)								
Broker-adjusted E/Q (t-1, t-3)	0.478**	0.480**	0.464**	0.426**	0.495**	0.508**	0.467**	0.712**	0.407**
	(20.73)	(21.15)	(15.27)	(7.62)	(7.73)	(13.61)	(6.18)	(7.33)	(4.30)
R-square	0.080	0.084	0.093	0.051	0.082	0.095	0.068	0.237	0.047

Table 5: Changes in Routing in Response to Price Improvement

This table examines how brokers alter their routing to wholesalers based on their prior overall (Panels A and B) or stock-level (Panels C and D) price improvement. Each period, we compute the broker-adjusted overall (stock-level) price improvement, measured as E/Q, of each wholesaler for each broker by subtracting the broker overall (stock-level) average from the wholesaler overall (stock-level) average for that broker. We then regress the percent change in orders routed next period to that wholesaler against the wholesalers' excess price improvement 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 Overall Price Improvement (Prior Month)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.029*	-0.018	-0.025	-0.043	-0.064	-0.005	-0.083*	-0.067	-0.104*
	(-2.30)	(-1.74)	(-0.67)	(-0.84)	(-0.90)	(-1.23)	(-2.13)	(-0.87)	(-2.36)
R-square	0.013	0.008	0.009	0.022	0.042	0.002	0.041	0.022	0.076
Panel B: Wholesaler Overall Price Improvement (Prior 3-month Average)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1, t-3)	-0.015	-0.014	-0.024	-0.046	0.007	-0.011	-0.026	-0.092	0.017
	(-0.84)	(-0.89)	(-0.34)	(-0.49)	(0.25)	(-1.98)	(-0.54)	(-1.11)	(0.29)
R-square	0.003	0.004	0.005	0.015	0.001	0.005	0.004	0.037	0.002
Panel C: Wholesaler Stock-level Price Improvement (Prior Month)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1)	-0.007	0.007	-0.003	0.014	0.015	0.007	-0.046**	-0.071	-0.039**
	(-1.24)	(1.45)	(-0.44)	(0.76)	(1.33)	(0.75)	(-3.51)	(-1.30)	(-3.45)
R-square	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.002
Panel D: Wholesaler Stock-level Price Improvement (Prior 3-month Average)									
Dep Var:	Percent Change in Routed Orders (t)								
Broker-adjusted E/Q (t-1, t-3)	-0.011	0.003	-0.016	-0.009	0.039	0.010	-0.049	-0.047	-0.049
	(-0.91)	(0.24)	(-0.74)	(-0.18)	(2.95)	(0.45)	(-1.93)	(-0.67)	(-1.73)
R-square	0.000	0.000	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)	-0.254**	-0.153**	-0.044	0.056	0.250	-0.293**	-0.765**	-1.064**	-0.670**
	(-8.46)	(-4.52)	(-0.40)	(0.60)	(2.17)	(-6.81)	(-6.33)	(-10.47)	(-4.16)
R-square	0.071	0.035	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)	-0.144**	0.306**	0.095	0.463**	0.137	0.661**	-0.721**	-1.064**	-0.653**
	(-3.14)	(6.14)	(1.50)	(3.27)	(1.91)	(6.82)	(-5.43)	(-10.47)	(-3.00)
R-square	0.017	0.121	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		-25.6%

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)

Appendix

A. Proofs

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:

$$\min_{\Delta \in [-\sigma, 1-\sigma]} (\sigma + \Delta)p_X + (1 - \sigma - \Delta)p_Y + \frac{s}{2}\Delta^2. \tag{9}$$

The F.O.C. implies:

$$\Delta = \begin{cases} \frac{p_Y - p_X}{s} & \text{if } p_Y - p_X \in [-s\sigma, s(1-\sigma)], \\ 1 - \sigma & \text{if } p_Y - p_X > s(1-\sigma). \\ -\sigma & \text{otherwise.} \end{cases} \tag{10}$$

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

$$\max_{p_X} (p_X - f_X)(\sigma + \Delta) \tag{11}$$

Using Equation (10), the F.O.C. implies

$$p_X = \begin{cases} \frac{s\sigma + p_Y + f_X}{2} & \text{if } p_Y \in (f_X - s\sigma, f_X + s(2-\sigma)), \\ f_X & \text{if } p_Y \leq f_X - s\sigma, \\ p_Y - s(1-\sigma) & \text{otherwise.} \end{cases} \tag{12}$$

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

$$p_Y = \begin{cases} \frac{s(1-\sigma) + p_X + f_Y}{2} & \text{if } p_X \in (f_Y - s(1-\sigma), f_Y + s(1+\sigma)), \\ f_Y & \text{if } p_X \leq f_Y - s(1-\sigma), \\ p_X - s\sigma & \text{otherwise.} \end{cases} \tag{13}$$

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 (12) and (13), we have

$$p_X = \frac{2f_X + f_Y}{3} + \frac{s(1+\sigma)}{3} \qquad \text{and} \qquad p_Y = \frac{f_X + 2f_Y}{3} + \frac{s(2-\sigma)}{3}. \tag{14}$$

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

$$s > \frac{f_Y - f_X}{2 - \sigma}. \tag{15}$$

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 (12) and (13), we have

$$p_X = f_Y - s(1-\sigma) \qquad \text{and} \qquad p_Y = f_Y. \tag{16}$$

Ensuring that these solutions satisfy the conditions, we have

$$s \leq \frac{f_Y - f_X}{2-\sigma}. \tag{17}$$

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 (12) and (13), we have

$$p_X = \frac{f_X + f_Y + s\sigma}{2} \qquad \text{and} \qquad p_Y = f_Y. \tag{18}$$

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

$$f_Y - f_X \in (f_X - s\sigma, f_X + s(2-\sigma)), \tag{19}$$

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

$$f_Y - f_X \geq s(2-\sigma). \tag{20}$$

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

$$\Delta = \begin{cases} 1 - \sigma & \text{if } p_X \leq p_Y \\ -\sigma & \text{otherwise.} \end{cases} \tag{21}$$

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. ■