What Does Best Execution Look Like?

Thomas Ernst[*], Andrey Malenko[†], Chester Spatt[‡], and Jian Sun[§]

November 30, 2023

Abstract

U.S. retail brokers route order flow to wholesalers based on their past performance. Brokers face a strategic choice over how often to reallocate order flow, how aggressively to reward or punish performance, and what history, across time or securities, to consider. This paper analyzes how broker choices for allocating order flow shape competition among wholesalers. Our empirical results are consistent with the theory that prospects for future order flow provide wholesalers with strong incentives to offer price improvement and allow brokers to discipline the provision of liquidity.

[*]University of Maryland, Robert H. Smith School of Business: ternst@umd.edu
[†]Boston College, Carroll School of Management: malenkoa@bc.edu
[‡]Carnegie Mellon University, Tepper School of Business: cspatt@andrew.cmu.edu
[§]Singapore Management University, Lee Kong Chian School of Business: jiansun@smu.edu.sg

I. Introduction

Almost all retail brokerages in the U.S. offer commission-free trading. While trading costs are lower than ever before, regulators continue to vigilantly review the prices retail investors obtain, as well as broader industry practices. The Securities and Exchange Commission has released several proposals to restructure or refine the industry, including more disclosure of execution quality, requirements for order-by-order auctions, reforms to trading fees and tick size, and a legal best-execution obligation for brokers, which would overlap with the existing FINRA rule for best execution. Motivated in part by these proposals, we seek to develop an understanding of a fundamental question: How do retail brokers obtain best execution for their clients?

While the goal of best execution—obtaining the best price possible—is straightforward, achieving it presents a formidable strategic challenge. This paper investigates the economics of achieving best execution, including the decisions brokers make, and how those decisions shape both immediate outcomes and the market structure of retail internalization. Best execution is a classic principal-agent problem: brokers (the principals) must design a system which induces wholesalers (the agents) with private information about their willingness to trade to offer competitive prices. As an example, suppose that a broker uses two market centers, and ranks them based on past performance. For future order flow, a broker should route more orders to the better market center, but exactly how much order flow to route in this manner is a difficult question. Sending 100% of the order flow to the better market center eliminates any incentives for further improvement from the better market center as well as the ability to evaluate the weaker market center, while sending a proportionate 50% of the order flow offers no reward to the market center, nor does it maximize the benefit of superior prices offered by the market center. Moreover, the frequency of evaluation, as well as whether to evaluate trades in different symbols separately or collectively, form important dimensions to the strategic choice.

In practice, retail brokers work with several wholesalers, who are market makers who specialize in executing retail order flow. These wholesalers offer substantially better prices than open market centers (such as exchanges)[1], in part because wholesalers can offer retail investor order flow better

[1] Battalio and Jennings (2022).

prices due to the order flow's characteristic lower adverse selection[2] or lower correlation.[3] Retail brokers do not communicate with wholesalers on each individual trade, but instead evaluate wholesalers on aggregate past performance. This practice reflects the fact that substantial liquidity, both from wholesalers and alternative trading systems, is non-displayed and even discretionary in nature. Rather than focusing on a trade-by-trade view, brokers entice competition for the aggregate order flow. Brokers make strategic choices, informed by past history, to route dynamically more orders to wholesalers offering superior past prices.

To gain insight into these strategic choices of brokers, we obtain proprietary data from three large retail brokers (henceforth Broker A, Broker B, and Broker C), which collectively comprise over 50% of the U.S. equity market's retail brokerage industry. This data grants us considerable insight into the specific practices of these brokers, including best-execution statistics on each of the wholesalers they route to. While each of the brokers focuses on obtaining the best price possible for customers, they use different lengths of past history and consider different sets of symbols when routing.

We begin our analysis by documenting that wholesalers compete vigorously for order flow. Wholesalers who offer better prices obtain more order flow, while wholesalers offering inferior prices receive less order flow. Effective-over-quoted (EFQ) spreads provide a standard measure of price improvement, reflecting the ratio of prices paid by retail investors to the publicly available prices; lower EFQ means larger savings by retail investors. For every one percentage point increase in EFQ, we find that wholesalers obtain between 0.8 and 1.2% less order flow allocation. If we consider the ordinal ranking of wholesalers, a worse ordinal ranking leads to 6 to 10% less order flow allocation. We also show that these results are specific to the history considered by the broker; if we evaluate wholesaler performance based on longer or shorter windows, or for differently sized orders, the explanatory power of history, as measured by R^2 of the regression of order allocation based on past history, drops considerably. In other words, brokers route based on a specific history, such that it is crucial to understand order allocation decisions made by the broker.

While wholesalers compete on aggregate order flow rather than individual orders, market conditions which allow greater price improvement apply to their competitors as well as themselves.

[2] Easley, Kiefer, and O'Hara (1996).
[3] Baldauf, Mollner, and Yueshen (2022).

Using daily execution data, we show that on volatile market days, EFQ ratios decrease, consistent with wholesalers offering greater price improvement to retail investors. Effective spreads charged to retail investors are less volatility-sensitive than public quoted spreads, indicating that when quoted spreads increase, the opportunity for wholesalers to provide price improvement to retail investors is greater, and competitive pressures lead them to pass on these improvements to retail investors.

A large market maker, henceforth wholesaler A5, entered the wholesaling business in late 2021, which allows us to conduct an event study around the entry of a competitor. We use two measures of competition, $First\_To\_Second$, the arithmetic difference in EFQ offered by the first-best vs. second-best wholesaler, and $First\_To\_Average$, the difference between the first-best and volume-weighted average wholesaler. Wholesaler A5 works with Broker A, who routes each symbol on an individual basis. We find that wholesaler A5 gains more market share in securities with a larger $First\_To\_Second$ difference, and a larger EFQ. Following the entry of Wholesaler A5, we find that $First\_To\_Second$ declines when measured among the incumbent wholesalers, but does not decline when including wholesaler A5, consistent with a displacement story, where wholesaler A5 offers superior improvement to gain market share, but displaced competitors respond by reducing the price improvement they give. We find some evidence consistent with an increase in competition, however, as $First\_To\_Average$ and Effective-Over-Quoted spreads decrease in both the incumbent wholesalers and the wholesalers including wholesaler A5.

During the sample we study, Broker B makes a focus change in their routing decisions, facilitating an event study on how broker priorities shape wholesaler behavior. Broker B considers all aspects of past performance, but expanded their primary focus to include particularly small (odd lot) orders. Wholesalers respond rapidly, with EFQs for odd lot orders declining from an average over 80% to an average below 30%. We find a contemporaneous rise in EFQ of large orders, however, suggesting that that the market for retail wholesaling is highly competitive. Brokers are not missing out on Pareto improvement: any change in priorities will bring improvement along one dimension and dis-improvement along other dimensions. The change in prices obtained when broker routing behavior changes is critical in understanding the effect of proposed changes in market structure. In the empirical analysis of the SEC's order competition proposal, for example, the SEC measures substantial mid-quote liquidity that retail investors do not always access. Such attempts to access this liquidity, however, particularly in a way that gives rise to a winner's curse or information leakage, may lead to the liquidity changing in availability. What may seem like a tempting 'obvious' improvement to routing algorithms fails to take into account that the price improvement is offered conditional on the current routing algorithm. Any change to the algorithm potentially would change the conditional liquidity offered.

Wholesalers typically have detailed information on how their performance compares relative to their competitors. Two brokers in our sample evaluate and change routing allocations at a monthly frequency, typically at the beginning of the month. This creates an opportunity to observe how wholesalers behave when changes to their order allocation are imminent. We find that they offer greater price improvement when they are further behind the closest better competitor, but offer less improvement when they are further ahead of their closest worse competitor.

Decisions by brokers in how they divide their order flow change the competitive landscape for wholesalers. One broker in our sample routes all orders according to all past history, while another broker routes each individual stock according to the history in that specific stock. We show that there is greater difference in wholesaler EFQ in large stocks than small stocks, consistent with greater economies of scale in large stocks. We note a similar set of differences between small and large orders, with greater differences in EFQ between wholesalers in orders in the largest stocks. Routing each symbol in individual size and symbol bins may create more competition in smaller symbols and orders, where wholesalers are more equal in ability, but less competition in large stocks or large orders, where wholesalers may have greater differences in their economies of scale.

Finally, we analyze non-marketable limit orders. In contrast to market orders, limit orders must be displayed on the exchange. We show that wholesaler performance in limit order fill rates is far less well explained by wholesaler-specific prior performance, consistent with there being smaller differentiation among wholesalers in their relative ability to handle non-marketable limit orders. Nonetheless, there are considerable differences in the exchange rebates wholesalers would collect for posting these limit orders, as exchanges tie rebates to volume tiers that market makers conduct. On October 2023, the SEC released a proposal for Rule 6b-1, which would restrict such tiering as well as require additional data on tiering. Our results highlight a connection between this rule and wholesaler competition for retail orders, as lower fees or higher rebates allow wholesalers to offer more price improvement to retail orders, while differences in exchange volume tiers may make it difficult for smaller wholesalers to offer price improvement comparable to that offered larger wholesalers.

In sum, our results suggest that broker's routing is intensely competitive. The SEC has proposed an update to Rule 605 with individual broker-wholesaler performance data; our proprietary data resembles this, and highlights the valuable insights such data offer as well as the importance of quoted spreads in making sense of a measure of effective-over-quoted spreads. Contemporaneously, the SEC has also proposed a best-execution rule, though for equities we show that the existing FINRA Best-Execution rule has created an intensely competitive market system. We do not find Pareto shortcomings in the current system, but rather that brokers have developed policies to enforce wholesaler competition. Price improvement offered by wholesalers is conditional on these policies, and any changes to broker's routing criteria change the conditional price improvement obtained.

II. Prior Literature and Contribution

Our work complements closely related work by Dyhrberg, Shkilko, and Werner (2022) which uses public SEC Rule 605 and 606 data to evaluate broker routing to wholesalers. Their paper shows that the market is competitive, that price improvement is much larger than payment for order flow, and that the scale of price improvement wholesalers give to retail investors closely matches the level of price improvement institutions obtain by timing their marketable orders to trade when spreads are narrower. While their focus is on market-wide competition, our data allows us to focus on the *individual* broker-wholesaler relationships. We characterize the specific choices individual brokers make in obtaining best execution, and how these specific choices influence the prices obtained, as well as the broader competitive landscape.

Our work is also closely related to Huang, Jorion, and Schwarz (2023), who use self-generated trading data to evaluate broker-wholesaler competition. On the basis of their self-generated data set of small trades, they argue that "brokers should be able to meaningfully improve order execution by altering routing based on prior execution." Our conclusions differ from theirs on a number of dimensions. First, we show that the specific history matters: if a broker routes orders of a specific size based on 30 days of past history of orders of that size, their routing decisions will not be well explained by different window lengths (like 15 or 45 days) nor by different sizes (smaller or larger orders). Second, Huang et al. (2023) assume that changing routing could increase access to better prices without changing said prices. We analyze a focus change by one of our brokers, where they changed the weight placed on specific past history. This leads to an improvement along one dimension, but dis-improvement along other dimensions, suggesting there is not a Pareto-improvement available to brokers.

Several papers analyze the performance of wholesalers and market segmentation. Easley et al. (1996), Battalio and Holden (2001), and Hu and Murphy (2022) theoretical model cream-skimming by wholesalers, while van Kervel and Yueshen (2023) model anti-competitive reference pricing. Battalio and Jennings (2022) empirically analyze the execution quality given one or more large wholesalers, showing that wholesalers consistently give price improvement relative to the exchange. Our focus is on the competitive broker-wholesaler relationship, with brokers monitoring wholesaler performance and allocating order flow accordingly.

III. Empirical Analysis

A. Background Information on Wholesaler Competition

We obtain proprietary data from three large retail brokers, henceforth Broker A, Broker B, and Broker C. Collectively, these firms comprise over 50% of the U.S. equity market's retail brokerage industry. From each broker, we obtain proprietary data on their routing practices, including the best-execution statistics from each of the wholesalers they route to. While brokers target all specific features of best execution highlighted under FINRA Rule 5310, they do place emphasis on distinct features of wholesaler performance; we highlight some of these focus-points of their best-execution practices in Table I.

All U.S. brokers have a best-execution duty under FINRA Rule 5310, which requires brokers to "use reasonable diligence" to obtain a price "as favorable as possible under prevailing market conditions." For retail orders, brokers are able to obtain far better prices than the prevailing NBBO, as wholesalers are willing to offer superior prices to retail investors. This price improvement, i.e. prices better than the public exchange bid or ask, is non-displayed liquidity; consequently, brokers do not simply route to the best quote, but must define a routing practice which maximizes the extent of non-displayed price improvement their orders obtain. In practice, brokers do not negotiate on individual trades, but send future order flow to wholesalers based on past performance. Brokers regularly evaluate the execution quality they receive, and adjust routing accordingly. Execution quality can cover many aspects of trading, with FINRA Rule 5310 suggesting a holistic process in which brokers consider price improvement, execution likelihood, speed, and size of order fills.

The brokers in our sample work with wholesalers who will always match or improve upon the public national best bid or offer (NBBO). Consequently, brokers route orders to wholesalers based on the past history of how much price improvement, relative to the NBBO, that wholesalers have offered. This quantity is typically measured as an effective-over-quoted spread. An effective spread measures the difference between trade price and mid-quote, while a quoted spread is half the difference between the bid and ask. As an example, suppose the best bid is \$10.00 while the best ask is \$10.10, the mid-quote is \$10.05. If a retail investor uses a market order and buys at \$10.07, they received 3 cents of price improvement relative to the best ask price, and are charged an effective-over-quoted spread of $\frac{10.07-10.05}{1/2(10.10-10.00)} = \frac{2}{5} = 40\%$. In this example, the retail investor's EFQ of 40% means that they paid 40% of what they would have paid in spread if they had traded at the NBBO. Lower EFQ ratios imply larger savings for retail investors.

The first dimension along which our brokers differ is the amount of history they focus upon when making routing decisions. When routing an order, brokers will consider prior EFQ charged by each wholesaler. All brokers holistically evaluate past performance, and consider the entire history of wholesaler performance, but place a primary focus on different periods. Broker A, for example, places a primary focus on 30 days of history, while Brokers B and C both place a primary focus on 90 days of history. Conversations with brokers highlighted a trade-off between focusing on shorter history, which provides a rapid reward to improvement, and a longer history, which provides a reward to persistence in performance.

When using past history to inform routing decisions, brokers differ not just in the length, but in what history they use to route orders. Broker C measures wholesaler performance in all securities, and routes all orders based on the past performance in all securities. In contrast, Broker A evaluates wholesaler performance for past orders in an individual symbol, further divided into five size categories. Within each symbol, routing in each order-size category is determined based on past performance in that order-size category. Broker B takes an intermediate approach, evaluating history in four specific security bins divided into three size categories. When considering past history to determine a routing decision, there is a trade-off between the quantity and relevance of the information. The choice of routing history also impacts market structure, as the specific history a broker considers defines distinct categories in which wholesalers compete.

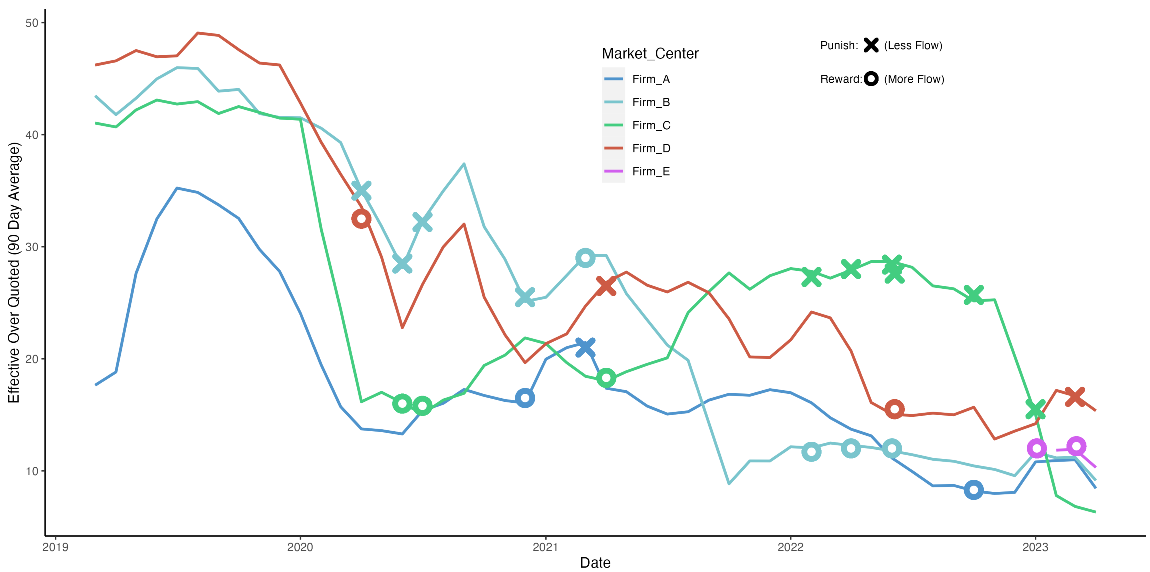
When brokers adjust future order flow, they face a further strategic choice over how to allocate flow across wholesalers. The wholesalers offering the best performance naturally obtain the most order flow, while wholesalers offering the worst performance obtain the least. The reward to the best wholesaler, however, must strike a balance between rewarding past performance while also maintaining the ability to increase the reward (order allocation) should the wholesaler offer even greater price improvement.

Table I: Best Execution Focus Points. All brokers use a holistic process and consider all aspects of wholesaler performance, but have slightly different primary focuses in measuring wholesaler performance. We summarize some of the basic differences here. Category refers to the bin that a broker uses, with all orders in the bin being routed in the same fashion. History refers to the past data which forms the focal point for future routing decisions. Observations refer to the level of detail we have in the data that we obtain from each broker.

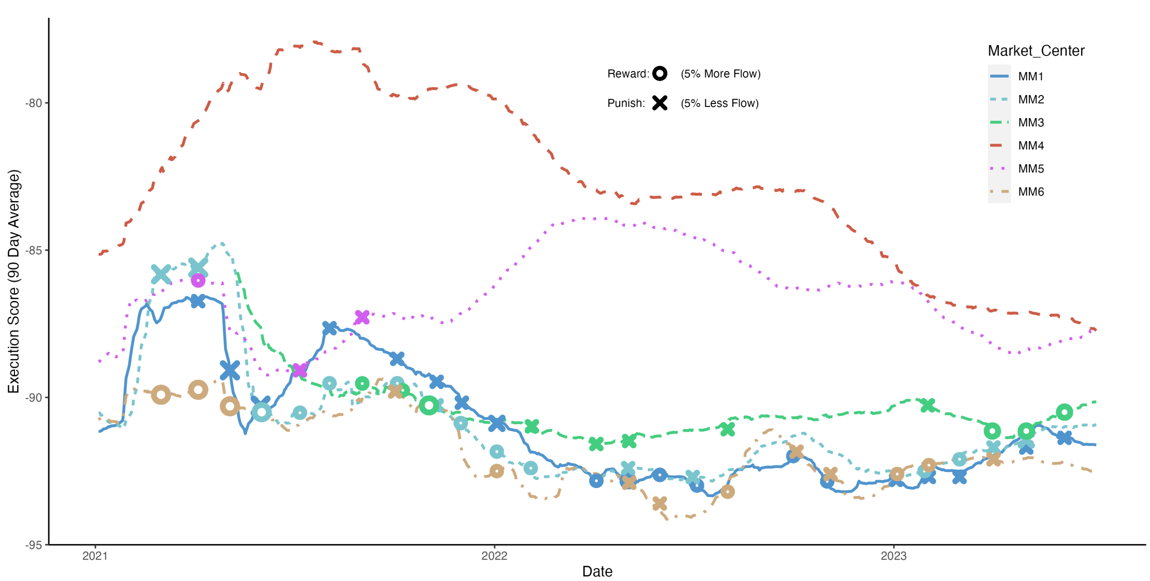
	Broker A	Broker B	Broker C
Category:	Each Symbol 5 Size Categories	Four Security Bins 3 Size Categories	One Bin
History:	30 Days	90 Days	90 Days
Decisions:	Daily (Rolling)	Monthly	Monthly
Observations:	Daily	Monthly + Daily Subsample	Monthly + Daily Subsample

Figure 1. Broker Execution and Routing. Brokers route based on past performance. Wholesalers with a lower effective-over-quoted spread (or higher execution score) obtain more order flow, while wholesalers with a higher effective-over-quoted spread obtain less order flow. For Broker C, we plot the negative of their proprietary score in Panel B, where wholesalers with a more negative score obtain more order flow.

Panel A: Broker B Data



Panel B: Broker C Data



B. Routing-Performance Relationship

Brokers route orders based on past performance, routing more orders to wholesalers who have offered better prices. Routing creates competition among wholesalers to offer the best prices: wholesalers offering superior prices are rewarded with more order flow in the future, while wholesalers offering inferior prices are punished with reduced order flow in the future. To evaluate the routing-performance relationship, we estimate the following regression:

REGRESSION 1: *For each broker i, wholesaler j, time period t and security bin k, we estimate:*

$$OrderShare_{ijkt} = \alpha_0 + \alpha_1 Prior\_EFQ_{ijkt} + X_{ijk} + \epsilon_{ijkt}$$

We use two measurements $Prior\_EFQ$. The first measure of $Prior\_EFQ$ is the prior periods's effective-over-quoted ratio, which is the average effective spread paid divided by the average quoted spread, measured as a percentage. The second measure of $Prior\_EFQ$ is the wholesaler's ordinal rank (with the wholesaler having the best EFQ score in that period receiving a rank of one for that period, the wholesaler with the second-best EFQ score in that period receiving a score of two for that period, and so on). $OrderShare$ measures the share of orders routed to a wholesaler, measured as a percentage of the broker's total order flow.

Our unit of observation depends on the broker, as there are differences in how brokers categorize past wholesaler performance. For Broker C, we use monthly data, with the ninety-day average EFQ of all stocks. For Broker B, we use monthly data with the ninety-day average EFQ of each of the four categories of stocks. For Broker A, we use daily data of the individual stocks. Across all specifications, we fit a fixed effect for each wholesaler and cluster standard errors by wholesaler. We also fit category fixed effects for the data from Brokers A and B.

Results of Regression 1 are presented in Table II. Across all three brokers, there is a strong relationship between performance and order share, with wholesalers with worse EFQ obtaining smaller order share allocations in future periods. For every 1% increase in EFQ (i.e. charging a higher price, as a percentage of the quoted spread), we estimate that wholesalers receive 0.8% less order share from Broker A and 1.2% less order share from Broker B. For each lower rank, wholesalers receive 9.7% less order share from Broker A and 6.0% less order share from Broker B.

With the proprietary Execution Score from Broker C (of which EFQ is an important component), brokers similarly receive 2.9% (cardinal score measure) to 6.5% (ordinal score rank) less order flow from Broker C.

When routing, all brokers must make a choice in determining how much history to consider, and any choice will involve a trade-off. Considering more history gives more information and encourages consistency in performance, while a shorter history allows more rapid adjustments to reflect more recent market conditions. We investigate the importance of window history in explaining broker broker routing history by re-estimating Regression 1 with Broker A's data on four different historical windows: 5 days, 10 days, 30 days, and 45 days. For each window, we examine the explanatory power of prior EFQ on OrderShare. Results are presented in Table III. The R^2 of Regression 1 peaks at 9.3% for the 30-day-window history, which matches the true length of history on which Broker A focuses. We conduct a similar exercise using performance across order size bins. We compare routing of orders in size category 3 (trades between 500 and 1,999 shares), and how it relates to prior EFQ performance of trades in Size 1 ($<$ 100 shares), Size 3, or Size 5 (5,000 to 9,999). The R^2 is 12.5% for the predictability of Size 3 Order Shares based on Size 3 EFQ, and less than 2.5% for Size 1 or Size 5 EFQ, consistent with Broker A's practice of routing orders for each symbol of order size 3 based on past EFQ of size 3 trades in that symbol. Our results highlight the danger of viewing broker routing history with incomplete information: a decision to route based on 30 days history may look illogical if one considers 5 or 45 days of history; similarly, a decision to route in one size category may look illogical if one only sees EFQ in other size categories.

Table II: Order Share and Execution Quality. We estimate Regression 1 which regresses the Order Share allocated to each wholesaler on prior market performance. Order Share is the fraction of orders going to each wholesaler. Prior EFQ refers to effective-over-quoted spread of the wholesaler in the prior period. Prior EFQ Rank is the ordinal rank of each wholesaler in the prior period, with 1 being the best-ranked wholesaler, 2 the second-ranked wholesaler, and so on. Score and Score Rank refer to a proprietary score used by Broker C, of which effective-over-quoted spread is a key component. Observations are at the stock-symbol-wholesaler level for Broker A with a fixed effect for each trade date and standard errors clustered by wholesaler, the stock-category-wholesaler level for Broker B with a fixed effect for each trade date and standard errors clustered by wholesaler, and the wholesaler level for Broker C, with standard errors clustered by wholesaler.

	Dependent variable: OrderShare					
	Broker A Data		Broker B Data		Broker C Data	
	(1)	(2)	(3)	(4)	(5)	(6)
Prior EFQ	-0.807^{***}		-1.183^{***}			
	(0.073)					
Prior EFQ Rank		-9.682^{***}		-6.041^{***}		
		(1.022)				
Prior Score					-2.947^{***}	
Prior Score Rank						-6.549^{***}
Observations	129,526	129,526	742	842	170	176
R^2	0.216	0.302	0.299	0.575	0.603	0.6912
Residual Std. Error	21.233	19.760	10.722	10.986	10.129	9.870
Note:						*p<0.1; **p<0.05; ***p<0.01

Table III: History Windows. When brokers route orders, they focus on a specific amount of past history. We estimate Regression 1 with Broker A's data with a variety of windows. Our interest is in how the R^2 changes with historical window measured, so we fit a constant term but do not fit any fixed effects. In Panel A, we estimate using 4 different windows: 5 days, 10 days, 30 days, and 45 days; broker A uses a 30-day window. In Panel B, we estimate order share for Size 3 (between 500 and 1,999 shares) trades using prior EFQ on Size 1 (< 100 shares), Size 3, or Size 5 (5,000 - 9,999 shares); broker A routes Size 3 orders based on Size 3 performance.

	Dependent variable:			
	OrderShare			
	(1)	(2)	(3)	(4)
Prior 5 Days EFQ	-0.469^{***}			
	(0.006)			
Prior 10 Days EFQ		-0.635^{***}		
		(0.006)		
Prior 30 Days EFQ			-0.835^{***}	
			(0.007)	
Prior 45 Days EFQ				-0.469^{***}
				(0.006)
Observations	129,526	129,526	129,526	129,526
R^2	0.051	0.072	0.093	0.051

	Dependent variable:		
	OrderShare For Trades Size 3		
	(1)	(2)	(3)
Prior EFQ - Size 1	-0.518^{***}		
	(0.031)		
Prior EFQ - Size 3		-0.975^{***}	
		(0.024)	
Prior EFQ - Size 5			0.003
			(0.002)
Observations	11,420	11,420	11,420
R^2	0.024	0.125	0.0002
Note:	*p<0.1; **p<0.05; ***p<0.01		

C. Routing and Market Volatility

Brokers route orders based on past performance, meaning that each order a wholesaler receives will factor into future evaluation, and subsequent routing, by the broker sending the order. While wholesalers also face competition from the public NBBO, as they must match or improve upon the NBBO for any orders they internalize, their primary source of competition is the degree to which they improve upon the NBBO in aggregate relative to competing wholesalers. This creates a pressure for day-by-day competition, as any market conditions in which wholesalers can offer more improvement means their competitors may offer more improvement.

Retail trading flow has lower adverse selection than non-retail flow. Days with higher market volatility will have greater rewards to timing trades; as a result, wholesalers should offer more price improvement. To evaluate the relationship between wholesaler improvement and market conditions, we estimate the following regression:

REGRESSION 2: *For each broker i, wholesaler j, time period t and security bin k, we estimate:*

$$\begin{aligned} Spread_{ijkt} = \quad & \alpha_0 + \alpha_1 Volatility_{ijkt} + \alpha_2 Variance_{ijkt} + Depth_{ijkt} \\ & + Return_{ijkt} + X_{ijk} + \epsilon_{ijkt} \end{aligned}$$

We estimate three specifications with different values for *Spread*: the effective-over-quoted spread, the effective spread paid by retail customers, or the public quoted spread. Results of Regression 2 are presented in Table IV (for Broker C) and Table V (for Broker A). Higher variance, volatility, and returns are all associated with lower effective-over-quoted spreads. Wholesalers give more price improvement in times of volatile markets, consistent with all wholesalers facing competitive pressure to increase the price improvement they give whenever they are able to. We note that EFQ is typically measured not as a weighted average of the individual EFQs, but as the total effective spread charged divided by the total quoted spread, which avoids any incentive by wholesalers to offer improvement only when spreads are narrow.

For wholesalers, competition over the entire window of history means that price improvement on volatile days has the same weight in evaluation as price improvement on calm days. For retail investors, the effect is a more consistent trading experience, with greater price improvement on more volatile days. Comparing Columns (2) and (3) of Table IV and V, effective spreads paid by retail investors are considerably less volatility-sensitive than public quoted spreads.

These results highlight two important aspects of EFQ. The first is that total spreads, rather than averaged spreads, avoid incentives by wholesalers to only offer price improvement when spreads are narrow. Brokers evaluate the total effective spread their customers paid bench-marked against the total quoted spread prevailing when their customers trade. Any action to narrow the spread by wholesalers is rewarded equally.[4] Second, the prevailing quoted spreads at the time retail traders place their orders are very important in evaluating and benchmarking performance. With the proposed updates to 605, the interpretation of the reports would be far more meaningful with bench-marking information on each broker's total quoted spreads, as this offers crucial context to interpreting an EFQ measure.

[4]In contrast, if EFQ is averaged across trades, rather than calculated as total effective divided by total quoted, market makers are incentivized to concentrate their improvement when quoted spreads are narrow, or to concentrate improvement in stocks with narrow quoted spreads. Suppose, for example, that a retail trader places two buy orders, each for 100 shares, and that the half-bid-ask spread is 2 cents at the time of the first order, and 4-cents at the time of the second order. If a market maker fills these orders with 2 cents improvement on the first order (EFQ of 0%), and none on the second (EFQ of 100%), the simple average is an EFQ of 50%. If instead, a wholesaler gives 0 cents improvement on the first order (EFQ of 100%) and 2 cents improvement on the second order (EFQ of 50%), the simple average is an EFQ of 75%. Under a simple weighted average, wholesalers have incentives to avoid giving price improvement when bid-asks spreads are wide. In both cases, however, the wholesaler gave 2 cents of improvement, and the total effective spread over total quoted spread is $\frac{4}{6} = 66\%$, meaning the total effective over quoted spread metric avoids this perverse incentive structure. Battalio and Jennings (2022) provide further discussion of EFQ statistics and the related issue of size improvement.

Table IV: Market Volatility and Retail Outcomes - Broker C. This table presents results of Regression 2. Column (1) uses EFQ, the effective spread divided by quoted spread. Column (2) uses the effective spread paid by retail customers, in basis points. Column (3) uses the public quoted spreads, in basis points.

	Dependent variable:		
	EFQ (%)	Effective Spread (BPS)	Public Quoted Spread (BPS)
	(1)	(2)	(3)
VWAP	0.018	0.105***	0.147***
	(0.028)	(0.030)	(0.028)
Trade Volume	0.013	−0.136***	0.219***
	(0.031)	(0.034)	(0.032)
Variance Ratio 15 Second	−7.950***	10.249***	6.003**
	(2.600)	(2.778)	(2.630)
Variance Ratio 1 Minute	−3.725**	11.386***	29.245***
	(1.841)	(1.967)	(1.862)
Intraday Vol	−0.156	8.962***	20.292***
	(0.471)	(0.503)	(0.477)
Depth (Millions $)	0.236	−0.557	−2.615
	(2.003)	(2.140)	(2.020)
Log Return	−10.075***	−15.001***	25.501***
	(3.767)	(4.025)	(3.810)
Observations	176,266	176,266	176,275
R^2	0.075	0.328	0.736
Residual Std. Error	69.405	74.157	70.210
Note:			*p<0.1; **p<0.05; ***p<0.01

Table V: Market Volatility and Retail Outcomes - Broker A. This table presents results of Regression 2. Column (1) uses EFQ, the effective spread divided by quoted spread. Column (2) uses the effective spread paid by retail customers, in basis points. Column (3) uses the public quoted spreads, in basis points.

	Dependent variable:		
	EFQ (%)	Effective Spread (BPS)	Public Quoted Spread (BPS)
	(1)	(2)	(3)
VWAP	0.001	0.164***	0.071***
	(0.005)	(0.020)	(0.027)
Trade Volume	1.449**	−3.092	9.581**
	(0.722)	(2.840)	(3.937)
Variance Ratio 15 Seconds	0.270	39.559***	35.773***
	(0.906)	(3.560)	(4.937)
Variance Ratio 1 Minute	−4.001***	52.127***	68.921***
	(0.673)	(2.646)	(3.669)
Intraday Vol	−0.581***	10.364***	40.469***
	(0.152)	(0.599)	(0.831)
Depth	−0.820**	0.728	2.457
	(0.325)	(1.279)	(1.774)
Log Return	−0.882	12.709**	46.892***
	(1.435)	(5.639)	(7.820)
Observations	64,906	64,906	64,906
R^2	0.126	0.604	0.557
Note:			*p<0.1; **p<0.05; ***p<0.01

D. *Competition and Wholesaler Entry*

Broker A began working with a new wholesaler, referred to as A5, on December 15, 2021. We examine the impact the entry of wholesaler A5 has on competition for Broker A's order flow. We take as exogenous the exact date, December 15, 2021, on which wholesaler A5 begins receiving order flow from Broker A, and compare market competition before and after this date. The extent to which wholesaler A5 enters the market is endogenous, particularly as Broker A routes each security and order size bin separately. We investigate the extent to which the existing competition impacts wholesaler A5's entry with the following regression:

REGRESSION 3: *For each asset j in order size category k in time period t:*

$$A5\_Post\_Share_{jk(t+1)} = \alpha_0 + \alpha_1 Competition_t + \epsilon_{ijkt}$$

We consider two measures of competition: $First\_To\_Second$, the difference in effective-over-quoted-spread between the first- and second-ranked wholesalers, and $First\_To\_Average$, the difference in effective-over-quoted-spread between the first wholesaler and the volume-weighted average effective-over-quoted spread. We calculate each competition measure in the period from November 15 to December 15, 2021. A5_Post_Share is the order share of wholesaler A5 in the period from March 3 to April 3, 2022.

Results of Regression 3 are presented in Table VI. Wholesaler A5 enters categories with a larger gap between the first and second-ranked wholesalers; for every 1% increase in First_to_Second (the arithmetic difference in effective-over-quoted spread between the first-ranked to second-ranked wholesaler), wholesaler A5 obtains an additional 0.07% of order share in the post-entry period.[5] Similarly, when the top wholesaler has a 1% larger order share, wholesaler A5 has a 0.15% larger order share in the post-entry period. We find no significant relationship between First_To_Average and A5's order share in the post-entry share, however, suggesting the competitiveness of the average wholesaler is less important than that of the highest-ranked wholesaler.

We also consider how competition changes one wholesaler A5 enters with the following regression:

[5]In relative terms, a one standard deviation increase in First_to_Second is associated with a 5% increase in the order share of wholesaler A5 in the post-entry period.

REGRESSION 4: *For each asset j in order size category k in time period t:*

$$Competition_{ijkt} = \alpha_0 + \alpha_1 Post_t + \epsilon_{ijkt}$$

Competition is either the First-To-Second or First-To-Average competition measure, calculated excluding wholesaler A5 as our baseline specification. Post is an indicator for which takes the value 1 for the periods after the entry of wholesaler A5. We estimate using data from November 15 to December 15, 2021 (prior to wholesaler A5's entry), and two time periods following wholesaler A5's entry: March 3 to April 3, 2022 and May 15 to June 15, 2022.

Results are presented in Table VII. Following entry by wholesaler A5, the gap between the first and second-ranked wholesaler narrows by 1.4%, while the first-to-average gap narrows by 2%. Effective-over-quoted spreads decrease by 6%, consistent with an increase in competition, though we note that we cannot control for time trends due to the nature of the data.

We find some evidence consistent with displacement, whereby entry by wholesaler A5 aligns with exit by another wholesaler. From Panel A of Table VII, the HHI index increases for the firms excluding A5, and from Panel B of Table VII, there is no change to the gap between the First-to-Second EFQ offered by wholesalers, while the top wholesaler market share increases by 9.6% in the post-entry period. Thus in categories where wholesaler A5 gains market share, they often gain considerable market share from competitors. Nonetheless, effective-over-quoted spreads and the gap between the First-To-Average spread decrease.

We also estimate Regression 4 with the level or order share (as opposed to a *Post* indicator) obtained by wholesaler A5 in the March 3 to April 3, 2022 time period. While category-specific order share obtained by wholesaler A5 is endogenous (depending on the effective-over-quoted spread provided by A5 and by competing wholesalers), in categories where wholesaler A5 obtains a larger proportion of the order flow, there are larger decreases in the First-To-Average difference in effective-over-quoted spreads, the HHI, and the effective-over-quoted spread itself.

Table VI: Wholesaler Entry Decision. We estimate Regression 3, which measures the impact of a new wholesaler entering the market. Our outcome variable is A5_PostShare measures the share of wholesaler A5 obtained in each category in the March 3 to April 3, 2022 data. Categories are the unique symbol and order size bins for each stock. Within each category, we measure First-To-Second (the difference between the effective-over-quoted spread of the first vs. second wholesaler), First-To-Average (the difference between the effective-over-quoted spread of the first wholesaler vs. volume-weighted average), and First Firm Order Share (the share of orders obtained by the top firm in that category). These variables are measured between November 15 to December 15, 2021 and March 3 to April 3, 2022), and we fit a fixed effect for each stock, order size bin, and each top wholesaler.

	Dependent variable: A5_PostShare					
	(1)	(2)	(3)	(4)	(5)	(6)
First-To-Second	0.073^{**}					0.001
	(0.033)					(0.057)
First-To-Avg		0.054				0.060
		(0.045)				(0.073)
First Firm Order Share			0.147^{***}			-0.095
			(0.030)			(0.120)
HHI				17.102^{***}		9.990
				(2.809)		(13.297)
Effective-Over-Quoted Spread					0.053	0.148^{**}
					(0.036)	(0.062)
Observations	1,461	1,461	1,586	1,586	1,586	1,461
R^2	0.444	0.442	0.396	0.403	0.384	0.447
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table VII: Wholesaler Entry and Competition. We estimate Regression 4, which measures the impact of a new wholesaler entering the market. We use data from November 15 to December 15, 2021 (before the wholesaler enters) and contrast it with data from March 3 to April 3, 2022. Post takes the value 1 for the periods after wholesaler A5 enters, while A5_Share is the order share of wholesaler A5.

Panel A: Excluding Wholesaler A5

	Dependent variable:				
	First-To-Second	First-To-Avg	First Firm Order Share	HHI	Effective Over Quoted Spread
	(1)	(2)	(3)	(4)	(5)
Post	-1.367^{**}	-2.021^{***}	0.159	0.112^{***}	-6.020^{***}
	(0.589)	(0.405)	(0.599)	(0.013)	(0.499)
Observations	3,133	3,133	3,467	2,100	3,467
R^2	0.262	0.233	0.384	0.470	0.555

Panel B: Including Wholesaler A5

	Dependent variable:				
	First-To-Second	First-To-Avg	First Firm Order Share	HHI	Effective Over Quoted Spread
	(1)	(2)	(3)	(4)	(5)
Post	0.674	-2.977^{***}	9.617^{***}	-0.070^{***}	-7.109^{***}
	(0.586)	(0.432)	(1.188)	(0.006)	(0.478)
Observations	3,157	3,157	2,106	3,467	3,467
R^2	0.293	0.205	0.441	0.421	0.560

Panel C: Wholesaler A Entry As Independent Variable

	Dependent variable:				
	First-To-Second	First-To-Avg	First Firm Order Share	HHI	Effective Over Quoted Spread
	(1)	(2)	(3)	(4)	(5)
A5_Share	-0.027	-0.068^{***}	149.499	-0.004^{***}	-0.198^{***}
	(0.020)	(0.015)	(302.094)	(0.0002)	(0.017)
Observations	3,157	3,157	2,106	3,467	3,467
R^2	0.294	0.197	0.418	0.458	0.548
Note:					*p<0.1; **p<0.05; ***p<0.01

E. Broker Focus Change and Wholesaler Responses

While brokers evaluate wholesalers based on specific criteria, wholesalers in turn are aware of, and extremely responsive to, these criteria. Any changes brokers make in how they allocate orders will impact the responses of wholesalers, including the average price improvement they are able to provide. We examine a focus change by Broker B, whereby Broker B changed the weighting of different historical information, and evaluate the strategic responses of wholesalers to this focus change.

We gain critical insight into how a change to existing evaluation criteria influences wholesaler behavior, as well as a deeper view of the competitiveness of wholesalers. Contemporary work by Huang et al. (2023) argues that changes in broker routing criteria could deliver superior execution, though the authors "assume that such rerouting would not alter our trade execution nor the competitive dynamics of the wholesaler market." We offer a direct empirical test of that assumption: when a broker changes routing criteria, how do prices change? We reach the opposite conclusion: changes in routing practices change the prices obtained, and we note that there is no free Pareto improvement. While there are improvements along one dimension, they are accompanied by dis-improvements along another dimension, consistent with an intensely competitive market between wholesalers.

Broker B implemented its focus change on January 1, 2020. While Broker B uses a holistic process which considers all aspects of order execution both before and after this date, beginning January 1, Broker B include odd-lot orders (orders for less than 100 shares) in their primary focus.[6] Prior to the focus change, all orders executed within the NBBO, but odd lot orders frequently executed at a price only slightly better than the public quoted spread (an EFQ close to 100%). Following the change, the EFQ spread charged on odd-lot orders rapidly diminishes, as illustrated in Figure 2, with the average EFQ on odd-lot orders declining to below 50% over the first six months, and below 30% over the next six months.

At the same time that EFQ declines dramatically for odd-lot orders, however, EFQ rises for large orders (orders for more 2,000 or more shares). Figure 2 Panel B illustrates the gradual rise in EFQ for large orders, with most wholesalers charging an average effective spread approximately

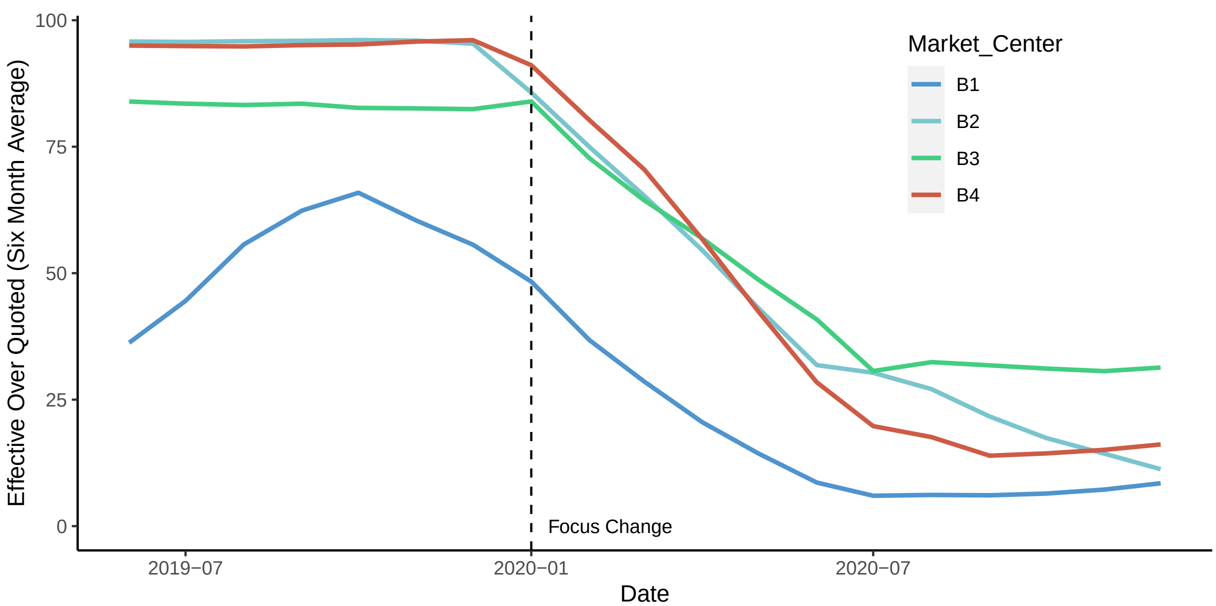
[6] Broker B monitors executions for all order sizes but placed particular emphasis on orders between 100 and 1,999 shares. Following the focus change, executions between 1 and 1,999 shares became a primary focus.

equal to the quoted spread on January 1, 2020, and charge an effective spread closer to 150 to 200% of the average quoted spread six months after the focus change. Note that our data from Broker B has the EFQ measured against the NBBO and not the depth-weighted NBBO; thus any marketable order which "walks the book" will be measured as paying an EFQ larger than 100%. Nonetheless, these orders may still receive price improvement relative to the depth-weighted NBBO: for example, if the best bid is \$9.92 while the best ask consists of 100 shares offered at \$10.00 and 100 shares offered at \$10.10, a market order to buy 200 shares filled by a wholesaler at an average price of \$10.04 would have received price improvement of 1 cent. Compared to the half-bid-ask spread of 4 cents, however, the order would be measured as paying an effective-over-quoted spread of $\frac{10.04-9.96}{1/2(10.00-9.92)} = 200\%$ despite receiving price improvement. Separately, we obtain exception reports from Broker B, which track any time an order of any size trades at a price worse than the NBBO. Most months have zero such orders, with an occasional month have a low single digit of such orders; both before and after the focus change, we detect no change in the number of order exceptions, indicating that any orders with an EFQ greater than 100% are trading at worse prices than the top of the NBBO, but equal or superior prices to the appropriate depth-weighted NBBO which would be obtained by "walking the book."

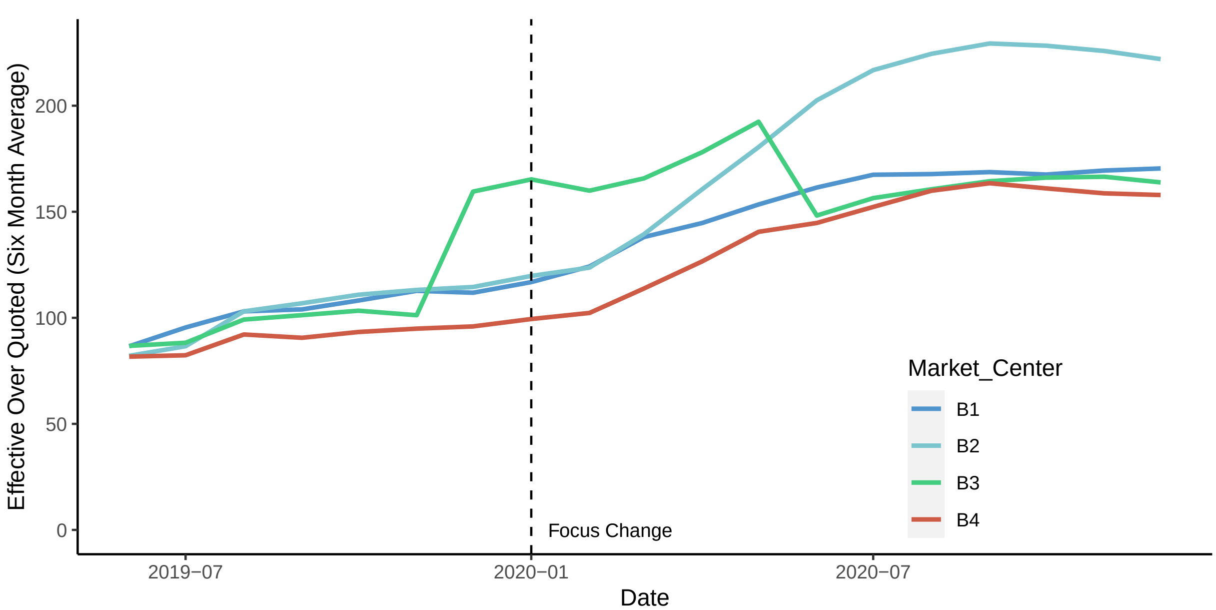
Any change to market structure will change the behavior of participants. Ernst, Spatt, and Sun (2022) highlight how order-by-order auctions expose market makers to a winner's curse, and consequently the price improvement they provide in auctions can sometimes be worse than that obtained in the current system of broker's routing. Within the SEC's proposal for order competition, for example, the SEC highlights the existence of widespread mid-quote liquidity, but cannot measure what the availability of such liquidity would be under a different market structure. Our results here provide a unique opportunity to directly observe how prices offered change under different broker routing priorities, and offer further empirical support for the competitiveness of broker's routing. Prioritizing one dimension of improvement leads to price improvement in that dimension, but that priority must come at the expense of a reduced priority along other dimensions, and potential price dis-improvement. Contrary to the idea that brokers could achieve obvious improvements by changing their routing practices, we show that they are currently Pareto optimal.

Figure 2. Focus Change by Broker B. Broker B holistically evaluates all aspects of order execution, but began placing more emphasis on odd lot orders on January 1, 2020. While odd lots always execute within the NBBO (i.e. customers paid smaller effective spreads than the publicly quoted spreads, paying an EFQ ratio less than 100%), EFQs narrow further once the focus change goes into effect. Conversely, the EFQ ratio widens for large orders. Note that for large orders, the order size is often larger than the available depth at the NBBO; an EFQ of 150% therefore represents a worse price than the NBBO but can still represent a superior price than if the order were to "walk the book" on a public exchange.

Panel A: Odd Lot Orders



Panel B: Large Orders



F. Strategic Wholesaler Behavior

On every order, wholesalers face a trade-off. Charging a higher bid-ask spread increases their revenue on an individual trade, but harms their average EFQ which may lead to reductions in their order flow allocation in future months. Brokers frequently share information with wholesalers about their performance, including where it falls relative to their competitors, and each wholesaler will also have access to public SEC 605 and 606 reports. Consequently, wholesalers have abundant information about their position relative to competitors.

Figure 3 plots the difference between the first-ranked and second-ranked wholesalers, across months. We plot the difference in overall score, the proprietary measure of Broker C which includes EFQ as a component, between the first-ranked and second-ranked wholesaler in Panel A, and the difference between the second-ranked and third-ranked wholesaler in Panel B. The month-end for each period is plotted via a gray vertical line. Order allocation decisions are made in the beginning of each month, thus any improvement at the end of the month can immediately lead to a larger allocation share, while improvement in the middle of the month will not pay off with a larger allocation share until the beginning of the next month. We note that there are some rapid shifts in the ninety-day-average execution scores around the month-ends. To formally analyze how wholesalers use information about their position to change their execution quality, particularly around the month-end, we estimate the following regression.

REGRESSION 5: *For each wholesaler j, time period t and security bin k, we estimate:*

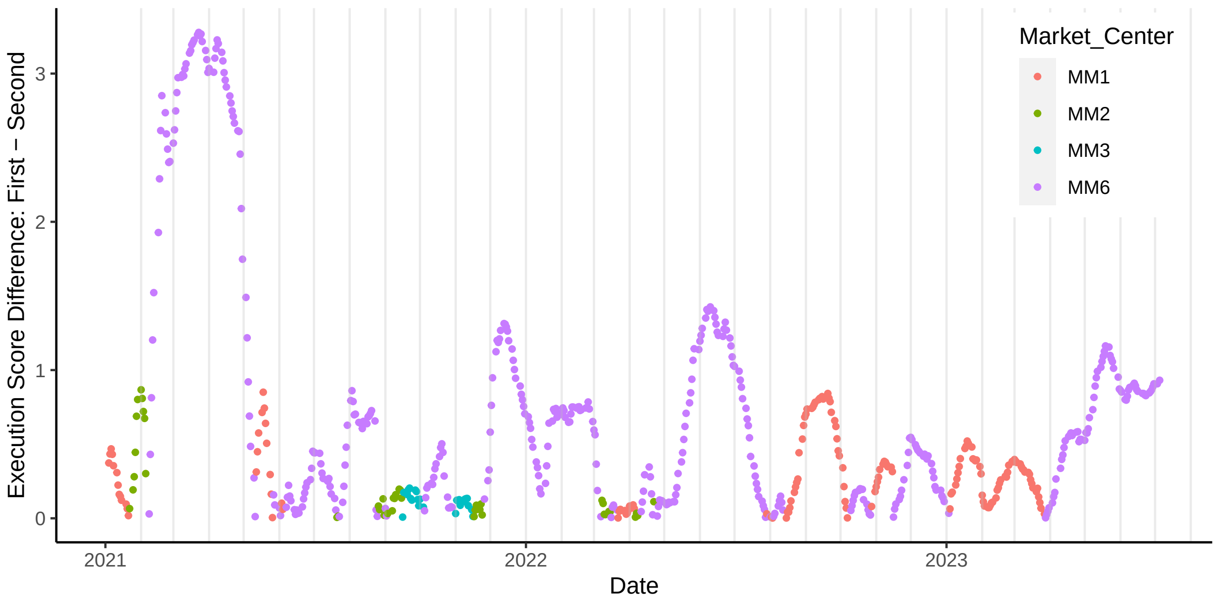
$$EFQ_{jkt} = \alpha_0 + \alpha_1 Distance\_Below\_Better_{jkt} + \alpha_2 Distance\_Above\_Worse_{jkt} + X_{jk} + \epsilon_{jkt}$$

For each wholesaler j, we define $Distance\_Below\_Better_j$ as the difference between wholesaler j and the closest better-ranked wholesaler (or zero in the case that wholesaler j is the best ranked). We define $Distance\_Above\_Worse_j$ as the difference between wholesaler j and the closest worse-ranked wholesaler (or zero in the case that wholesaler j is the worst ranked). We also include fixed effects for each market center, each rank, and each trade date, and cluster standard errors by market center. For Broker B, which has four separate routing tables based on security bins, we also include a fixed effect for each security bin.

Results of Regression 5 are presented in Table VIII. For Broker B, wholesalers who are behind a better-ranked wholesaler by a larger amount offer more price improvement on individual days. When their closest better competitor offers an extra 1% better EFQ, wholesalers offer an extra 0.5% EFQ. Results are stronger for the last 10 days of the month, with an estimated coefficient of 0.535%, compared to when there are more than 10 days to the month-end, with an estimated coefficient of 0.405%, consistent with market makers becoming more responsive to EFQ differences when the re-ranking is more imminent. Wholesalers who are ahead of their closest competitor by a larger amount offer less price improvement, though the result is only statistically significant in the sub-sample of days which are less than 10 days from the month end. For Broker C, wholesalers with a larger gap to a better-ranked wholesaler offer substantially more improvement, with a 1% larger gap in overall score associated with 0.6% increase in their own proprietary score. Larger leads over the next-best wholesaler are associated with less improvement, but the result is no longer statistically significant once standard errors are clustered by wholesaler.

Figure 3. Wholesaler Score Differences. Broker C uses a proprietary execution score to rank wholesalers, of which EFQ is an important component. In Panel A, we plot the execution score difference between the first-ranked and second-ranked wholesaler, while in Panel B we plot the execution score difference between the second-ranked and third-ranked wholesaler. We highlight each month-end with a gray vertical line, and note that order allocation decisions are made near the beginning of each month.

Panel A: Difference between First and Second-Best Wholesaler



Panel B: Difference between Second and Third-Best Wholesaler

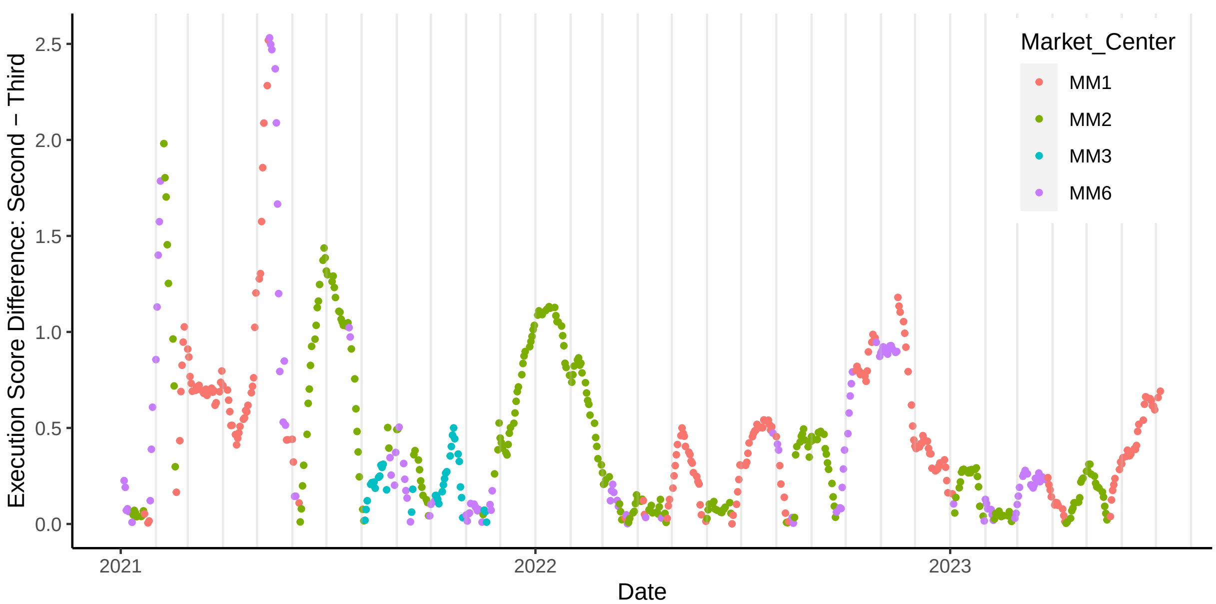


Table VIII: Wholesaler Strategic Improvement. *Distance_Below_Better* refers to the EFQ difference between a wholesaler and the closest better-ranked wholesaler (and takes the value 0 if the wholesaler is the best-ranked wholesaler). For Broker C, we use the proprietary score rather than the EFQ. *Distance_Above_Worse* refers to the EFQ difference between a wholesaler and the closest worse-ranked wholesaler (and takes the value 0 if the wholesaler is the worst-ranked). We estimate using three groups: the full sample in column (1), the sub-sample of dates at least 10 days from the month end in column (2), and the sub-sample of dates at most 10 days from the month end in column (3). We fit a fixed effect for each wholesaler, trade date, wholesaler rank, and, for Broker B's data, each category. Standard errors are clustered by wholesaler.

Panel A: Broker B Data

	Dependent variable: EFQ		
	Full	> 10 Days	< 10 Days
	(1)	(2)	(3)
Distance_Below_Better	0.448**	0.405*	0.535**
	(0.129)	(0.131)	(0.131)
Distance_Above_Worse	−0.137*	−0.148	−0.098**
	(0.052)	(0.063)	(0.018)
Observations	8,889	5,968	2,631
R^2	0.263	0.251	0.284

Panel B: Broker C Data

	Dependent variable: Proprietary Score		
	Full	> 10 Days	< 10 Days
	(1)	(2)	(3)
Distance_Below_Better	0.623***	0.682***	0.532***
	(0.064)	(0.066)	(0.093)
Distance_Above_Worse	−0.076	−0.045	−0.083
	(0.084)	(0.084)	(0.114)
Observations	3,679	2,477	1,080
R^2	0.763	0.797	0.730
Note:			*p<0.1; **p<0.05; ***p<0.01

G. *Symbol History Bundling and Wholesaler Economics*

When brokers route an order, they may consider historical wholesaler performance in past orders in the same symbol, or they may consider historical wholesaler performance in past orders in a set of related symbols, or historical wholesaler performance in all symbols. Each choice changes the competitive landscape for wholesalers.

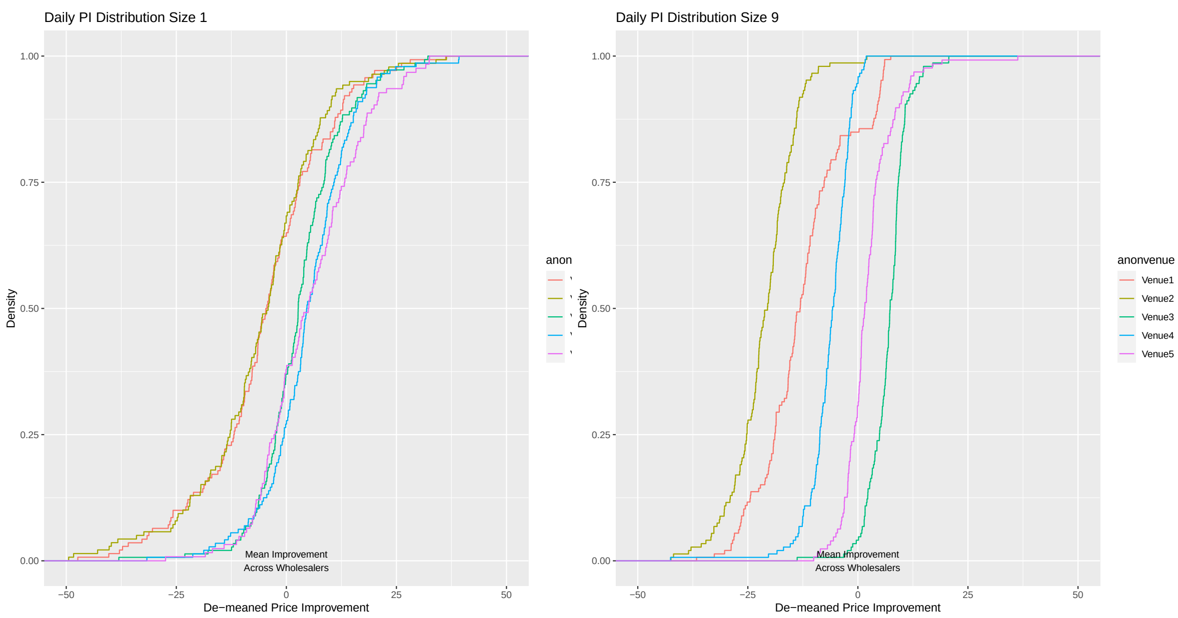
Broker A routes each symbol separately, with each symbol further divided into separate size bins, which allows us to gain insight into the nature of wholesaler competition across symbols. For each stock-day, we calculate the average level of price improvement given by each wholesaler. We then plot the cumulative distribution function of price improvement in Figure 4. Panel A plots the distributions for least active stocks, and has a tightly clustered pattern with each wholesaler offering a very similar level of price improvement. Panel B plots the distributions for the most active stocks, with much more variation in outcome: Venues 3 and 5 give considerably better price improvement than venues 1 and 2, for example. Taken together, these results suggest that there is more competitive variation in most active stocks than less active stocks stocks, with greater relative outperformance in the highest volume stocks.

Small stocks have larger bid-ask spreads, but may also have larger inventory holding costs. Dyhrberg et al. (2022) highlight this trade-off between spreads and inventory holding costs, and separately postulate that fixed costs are high in the wholesaling industry, with economies of scale rewarding the largest wholesalers. We note that if fixed costs are important, they are particularly important for narrow-stock spreads rather than wide-stock spreads. Consistent with this observation, we observe that the differences in price improvement given by wholesalers seem to be smallest in the smallest-stocks, consistent with economies of scale mattering less for these stocks. Conversely, differences in price improvement given by wholesalers are larger in the largest stocks, consistent with important economies of scale (relative to the narrow bid-ask spread).

A similar pattern occurs between the smallest and largest orders, as illustrated in Figure 5. For the smallest orders, differences in price improvement between wholesalers are small, while for the largest orders, differences in price improvement are sometimes large. We also note that for the largest orders, price improvement is measured relative to best bid or ask, and not the depth-weighted NBBO; consequently, an order which would "walk the book" would obtain negative price

Figure 4. : Price Improvement by Size Decile. We plot the distribution of de-meaned price improvement, with 0 corresponding to the mean price improvement across all wholesalers. We divide our data into ten deciles by trading volume. Panel A presents the distribution for the least active stocks, while Panel B presents the distribution for the most active stocks. Note that the mean is value-weighted, and wholesalers with superior performance will obtain a larger portion of order flow.

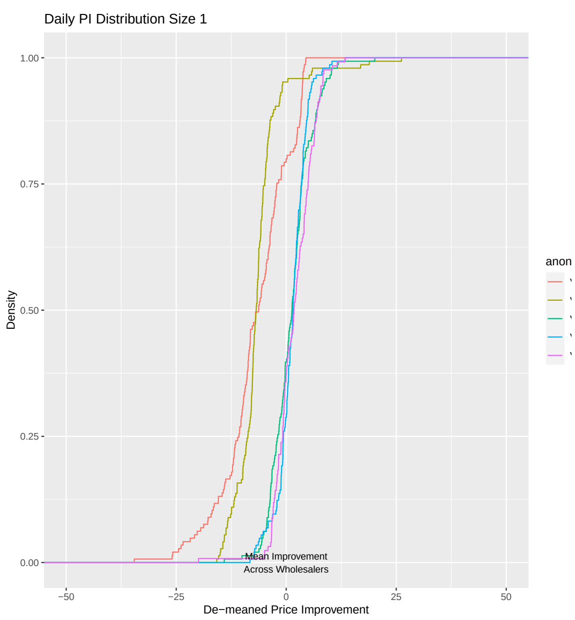
Panel A: Lowest Volume Stocks **Panel B: Highest Volume Stocks**



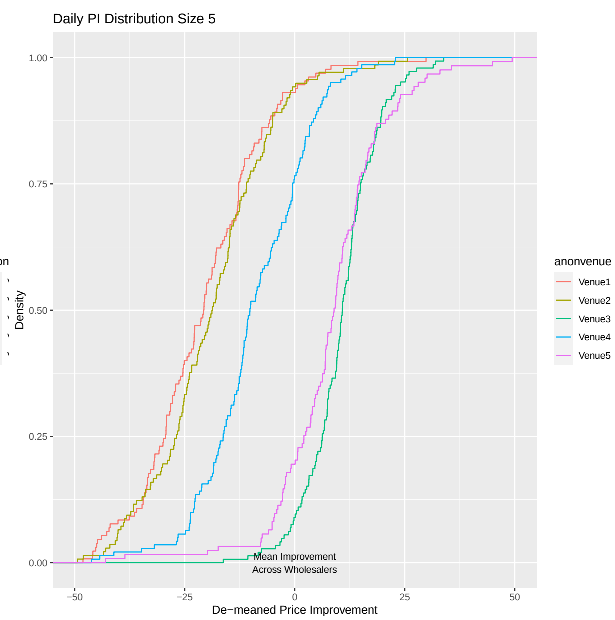
improvement relative to the best bid or ask, even if it executes at a price superior to that of the depth-weighted NBBO.

Figure 5. : Price Improvement by Order Size. We plot the distribution of de-meaned price improvement, with 0 corresponding to the mean price improvement across all wholesalers. Panel A presents the distribution for the smallest category of order sizes (odd lots), while Panel B presents the distribution for the largest category of order sizes (greater than 5,000 shares). Note that the mean is value-weighted, and wholesalers with superior performance will obtain a larger portion of order flow.

Panel A: Smallest Orders



Panel B: Largest Orders



H. Limit Orders and Fill Rates

While our primary focus is on the routing of market orders, we also examine the routing of limit orders. In addition to best-execution obligations of brokers, limit orders are also regulated by the Order Display Rule.[7] Any non-marketable limit orders which a wholesaler receives and does not immediately fill must be displayed on an exchange. Effectively, the wholesalers serve largely as a pass-through, receiving limit orders from brokers and posting them to exchanges. Spatt (2019) highlights an important reason for this pass-through behavior: exchange pricing is often volume-tiered. Limit orders are eligible for a market-making rebate, and the more a market participant trades, the larger the rebate they earn. Market makers who are active in the market may qualify for the best rebate tier, while a small or even medium broker may not qualify for the best rebate tier. We compare the persistence of performance between market and limit orders with the following regression:

REGRESSION 6: *For each wholesaler j, time period t and security bin k, we estimate:*

$$RelativePerformance_{jkt} = \alpha_0 + \alpha_1 Prior\_RelativePerformance_{jkt} + X_{jk} + \epsilon_{jkt}$$

Relative Performance measures the arithmetic difference between wholesaler j and the average wholesaler performance. For market orders, performance is measured via effective-over-quoted spreads, as a percentage, while for limit orders, performance is measured as the average monthly fill rate, as a percentage. Controls include a fixed effect for each wholesaler, and for each asset category (we have asset category information only for market orders, not limit orders). Results of Regression 6 are presented in Table IX. Of primary interest is the share of explained variation of performance, which we measure with R^2. For limit orders, our simple regression explains just 1% of future performance variation, while for market orders, our simple regression explains 22% of future performance variation.

Our results are consistent with wholesaler variation in limit order performance being minimal, which is natural given the order display rule applies equally to all wholesalers. Whatever orders they do not immediately fill are simply displayed on exchange, with no wholesaler having a

[7] There are, in fact, two distinct rules covering the display of limit orders. SEC Rule 11AC1-4 and FINRA Rule 6460: Display of Customer Limit Orders.

Table IX: Limit Order Persistence. We estimate Regression 6, which measures the persistence of wholesaler performance, using data from Broker B. For limit orders, performance is defined as the arithmetic difference between the fill rate for wholesaler i compared to the average fill rate across all wholesalers. For market orders, performance is defined as the arithmetic difference between the EFQ for wholesaler i compared to the average EFQ across all wholesalers. We include wholesaler fixed effects, an asset fixed effect for market orders (where the data we obtain from the broker has multiple asset categories), and cluster standard errors by wholesaler.

	Dependent variable: Performance	
	Limit Orders	Market Orders
	(1)	(2)
LM_Performance	−0.001	−0.311
	(0.040)	(0.214)
Wholesaler FE	X	X
Asset FE		X
Observations	618	1,612
R^2	0.013	0.227
Adjusted R^2	0.006	0.223
Note:		*p<0.1; **p<0.05; ***p<0.01

better or worse display technology. We note, however, that to the extent that there is variation in exchange volume-based pricing between wholesalers, different wholesalers will obtain different financial rewards from posting limit orders to the exchange. Even if the performance, as measured by fill rates obtained by customers, is the same, wholesaler profits will not be.

Our results highlight an important facet of exchange volume tiering. In October 2023, the SEC released a proposal for Rule 6b-1, which introduce restrictions on exchange volume tiering as well as requiring additional disclosure of volume tiers. While these results have obvious immediate impact on exchange customers, wholesalers are a key demographic of exchange customers, and reductions in the exchange price differentials which wholesalers face have the potential to change the nature of competition among wholesalers for retail orders. The largest wholesalers may shoulder a larger portion of the costs of exchange trading, potentially reducing their ability to offer price improvement to retail traders, while smaller wholesalers may see an increased ability to compete for retail order flow if they obtain the same fee and rebate levels.

IV. Conclusion

U.S. retail brokers need to repeatedly decide on how to divide order flow among wholesalers. This decision is non-trivial and has both information and incentive components. From the information point of view, past performance provides information to a broker about quality of different wholesalers. From the incentive point of view, allocation of future order flow provides incentives for wholesalers to offer price improvements today. There is a clear conflict between these two roles. While the information role of order flow allocation favors routing almost all order flow to the wholesaler the broker perceives to be the best, such allocation is likely to be detrimental for incentives, since the best wholesaler may have weak incentives to offer price improvement today if it knows that the beliefs of the broker do not change much. In this paper, we provide empirical analysis of this question using proprietary data from three large retail brokers. We find that broker routing is intensely competitive, brokers allocate order flow to wholesalers based on their past performance, and wholesalers are aware of it and offer price improvements accordingly.

The proprietary data we analyze is similar to the updated 605 that the SEC has proposed, and our work highlights the rich information contained in this data. As the SEC considers changing exchange volume tiering, our work highlights how this also can impact wholesaler competition for retail orders. Our results also indicate that the current system is highly competitive, which runs contrary to the narrative behind the SEC proposals for order-by-order auctions and a separate best-execution rule which would overlap with the existing efficacious FINRA rule.

In ongoing work, we are developing a dynamic principal-agent model to study competition among wholesalers for order flow in light of different order flow reallocation rules. Our goal is to capture the key empirical patterns and use the model to shed light on pros and cons of various reallocation rules.

REFERENCES

Baldauf, Markus, Joshua Mollner, and Bart Z. Yueshen, 2022, Siphoned Apart: A Portfolio Perspective on Order Flow Fragmentation, *Available at SSRN 4173362* .

Battalio, Robert, and Craig W Holden, 2001, A Simple Model of Payment for Order Flow, Internalization, and Total Trading Cost, *Journal of Financial Markets* 4, 33–71.

Battalio, Robert, and Robert Jennings, 2022, Why do Brokers Who do Not Charge Payment for Order Flow Route Marketable Orders to Wholesalers? .

Dyhrberg, Anne Haubo, Andriy Shkilko, and Ingrid M Werner, 2022, The Retail Execution Quality Landscape, *Fisher College of Business Working Paper* 014.

Easley, David, Nicholas M Kiefer, and Maureen O'Hara, 1996, Cream-skimming or Profit-Sharing? The Curious Role of Purchased Order Flow, *Journal of Finance* 51, 811–833.

Ernst, Thomas, Chester S Spatt, and Jian Sun, 2022, Would Order-By-Order Auctions Be Competitive?, *Available at SSRN: 4300505* .

Hu, Edwin, and Dermot Murphy, 2022, Competition for Retail Order Flow and Market Quality, *Available at SSRN 4070056* .

Huang, Xing, Philippe Jorion, and Christopher Schwarz, 2023, Who Is Minding the Store? Monitoring and Competition in Retail Trade Execution, *Working Paper. Available at SSRN: 4609895* .

Spatt, Chester S, 2019, Is Equity Market Exchange Structure Anti-Competitive?

van Kervel, Vincent, and Bart Zhou Yueshen, 2023, Anticompetitive Price Referencing, *Available at SSRN: 4545730* .