The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXX, NO. 4 • AUGUST 2025



# Would Order-By-Order Auctions Be Competitive?

THOMAS ERNST, CHESTER SPATT, and JIAN SUN*

## ABSTRACT

We model two methods of executing segregated retail orders: brokers' routing, whereby brokers allocate orders using the market maker's overall performance, and order-by-order auctions, where market makers bid on individual orders, a recent U.S. Securities and Exchange Commission proposal. Order-by-order auctions improve allocative efficiency, but face a winner's curse reducing retail investor welfare, particularly when liquidity is limited. Additional market participants competing for retail orders fail to improve total efficiency and investor welfare when entrants possess information superior to incumbent wholesalers. Our results hold when new entrants are less informed or the information structure differs. We also examine the cross-subsidization of brokers' routing.

RETAIL ORDER FLOW IN U.S. equities is segregated, with retail brokers routing almost all of their retail customer orders directly to market makers.

*Thomas Ernst is at the Robert H. Smith School of Business, University of Maryland. Chester Spatt is at the Tepper School of Business, Carnegie Mellon University. Jian Sun is at the Lee Kong Chian School of Business, Singapore Management University. We are grateful for the comments of two reviewers, the Associate Editor, and Urban Jermann (the Editor). We are thankful to Kerry Back, Yashar Barardehi, Robert Battalio, Svetlana Bryzgalova, Eric Budish, Qi Fan, Larry Glosten, Terrence Hendershott, Pete Kyle, Dmitry Livdan, Mark Loewenstein, Katya Malinova, Josh Mollner, Cecilia Parlatore, Andriy Shkilko, Alberto Teguia, Xin Wang, Bart Zhou Yueshen, and seminar and conference participants at the 2023 Asian Meeting of the Econometric Society, Arizona State University, Baruch College, Carnegie Mellon–Penn State–Pittsburgh joint conference, Central Bank Conference on the Microstructure of Financial Markets (Federal Reserve Board of Governors), China International Conference in Finance (CICF), Eastern Finance Association, HEC (Paris), Imperial College, Johns Hopkins, London Business School, University of Maryland, the Microstructure Exchange, New York University Stern Market Microstructure Conference, Northern Finance Association, Office of Financial Research (U.S. Treasury), Securities and Exchange Commission Economics Conference, Singapore Management University, Society for Financial Studies Cavalcade, Utah Winter Finance Conference, Western Finance Association, as well as numerous anonymous industry participants. We have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose. Jian Sun acknowledges that this study was financed in part by Singapore Management University through research grant 22-LKCSB-SMU-010 from the Ministry of Education Academic Research Fund Tier 1.

Correspondence: Chester Spatt, Carnegie Mellon University, Tepper School of Business, 5000 Forbes Ave, Pittsburgh, PA 15213; e-mail: cspatt@andrew.cmu.edu



DOI: 10.1111/jofi.13449

These market makers assume a best-execution obligation once they receive the order, whether they privately internalize trades off-exchange, or they fill the retail orders using liquidity obtained from other sources, like exchanges or alternative trading systems (ATS) (such as "dark pools"). Retail trades are attractive to market makers, due to either lower adverse selection (Easley, Kiefer, and O'Hara (1996)) or their trades being less correlated (Baldauf, Mollner, and Yueshen (2024)). In both cases, market makers provide retail investors with better prices than the exchanges. Recently, the U.S. Securities and Exchange Commission (SEC) has proposed a change in market structure in an attempt to increase competition for retail orders.[1] While previous academic work has explored *whether* retail flow should be segregated and its value, if segregated, the question of *how* segregated retail flow should be executed has received less attention.

We model and evaluate two distinct methods of executing segregated retail trades: brokers' routing and order-by-order auctions. Our brokers' routing model closely resembles the current market structure, with retail brokers determining where to route each order to maximize execution quality. While retail brokers use recent competing market-maker performance to inform routing decisions, they do not communicate with any market maker prior to routing each individual order. Our order-by-order auction models the SEC's proposed Rule 615, which would mandate auctions for retail trades. These auctions would only be available to retail market orders, where any market participant could bid on each individual order.

We evaluate both models with a focus on inventory cost and competition. In our model, a broker receives an order from a retail investor and chooses one market maker to execute the order. Executing the order incurs inventory costs for market makers. We assume that each market maker $i$ receives a cost shock $y_i$ just before the retail broker routes the order and this shock is private information. The inventory cost is affected by both the market maker's cost shock and the average cost shock of all market makers. Intuitively, each cost shock can be thought of as a shock to the inventory position of market maker $i$, with a market maker's willingness to trade depending on both his own inventory and the aggregate liquidity of all market makers. To obtain the order, market participants submit their spreads simultaneously to the broker, and the one with the lowest bid obtains the order. The key difference between brokers' routing and order-by-order auctions is market participants' information set when they submit their spreads. Under an order-by-order auction, each market maker $i$ submits the spread after observing the realization of $y_i$, while under brokers' routing, the spread is submitted when each market maker $i$ only observes a noisy signal about $y_i$. We solve the market equilibrium under

[1] In remarks before the Securities Industry and Financial Markets Association's (SIFMA's) 2022 Annual Meeting, SEC Chair Gensler stated that "I've also sought recommendations around how to instill greater competition for retail market orders on an order-by-order basis, through auctions. With greater competition, more market participants would have access to these retail market orders."

both trading mechanisms and then identify differences in welfare distribution, inventory management, and order allocation efficiency that arise under the two market structures.

In the brokers' routing setting, each market maker $i$ observes only a noisy signal of its cost shock $y_i$ when submitting the spread. This is motivated by our observation that, under brokers' routing, when market makers compete for order flows they are effectively competing for future orders and thus do not have precise information about the cost of executing the orders. This closely mirrors the current system of order routing in equities, where market makers agree to accept order flow from brokers but there is no pretrade communication about individual orders—brokers route to a market maker, and the market maker must accept the order. In practice, evaluation of trades is done on a periodic (e.g., daily, weekly, or monthly) basis, and market makers strive to secure future order flows by improving current execution quality (Ernst et al. (2024)). However, when competing on improving current execution quality, market makers do not observe the actual inventory costs associated with the orders that they will obtain in the future. We obtain a symmetric equilibrium for brokers' routing, where the bidding (spread) strategy of this equilibrium is monotone in the noisy signal. A broker routes the order to the market maker who submits the lowest spread. This delivers a highly competitive outcome with relatively low expected market-maker profits, because market makers are less informationally heterogeneous and their bidding strategies are less dependent on their observed noisy signals. While the brokers' routing setting delivers strong competition among market makers, the lack of communication on any specific trade means that a trade may be routed to a market maker that has high ex-post inventory cost, leading to inefficient order allocation and inventory management.

In the order-by-order auction model, in contrast, each market maker $i$ bids after observing its cost shock $y_i$. This is motivated by the SEC proposal that all retail orders be auctioned order-by-order, and if a market maker wins the auction, it executes the order immediately. When market makers compete for individual retail orders, they have more accurate information about the marginal inventory cost of executing the order. In the auction, market makers' symmetric equilibrium bid (spread) is increasing in their cost shock $y_i$, and thus, in equilibrium, the participant with the lowest realized inventory cost will always win the auction with the most aggressive bid, leading to first-best allocative efficiency. The common-value nature of the auction, however, creates a winner's curse problem: the participant who wins the auction learns that all other participants had higher-cost signals. Consequently, market participants bid conservatively in the auction, and thus, they earn more profit from the auction due to the strategic concern.

The welfare implications of switching from brokers' routing to order-by-order auctions are thus determined by the trade-off between more efficient inventory management and the higher rent earned by market makers due to concerns about the winner's curse. As previously noted, our inventory cost has two components: the common-value component, which depends on the average cost shock across all market makers, and the independent private-value (IPV) component, which depends only on the cost shock of the individual market maker. The private-value component of inventory cost underlies the considerations of allocative efficiency: order-by-order auctions ensure that the lowest-cost market maker always obtains the order, which leads to a welfare gain shared by investors and market makers. The common-value component, which is the same for all market makers, does not influence allocative efficiency, but it does lead to a more severe winner's curse problem under order-by-order auctions. As a result, all market makers will bid more conservatively, leading to larger rents for market makers and worse welfare for retail investors. In summary, while order-by-order auctions offer higher allocative efficiency than brokers' routing, they produce less competition than brokers' routing.

Based on the above intuition, both total welfare and market makers' welfare increase when switching from brokers' routing to order-by-order auctions. The change in investor welfare depends on the trade-off between the gain from more efficient order allocation and more rent earned by market makers due to the winner's curse. When the common-value component is a larger share of the inventory cost, for instance, for small stocks or during times with low liquidity, order-by-order auctions deliver worse welfare for retail investors than the current system of brokers' routing. For these stocks, the potential allocative efficiency gains of the auction are smaller than the welfare loss to retail investors from the increase in the winner's curse. The welfare results also depend on the number of participating bidders: when there are many bidders, the welfare gain from allocative efficiency dominates, and thus investor welfare is higher under order-by-order auctions. While forecasts of the number of bidders in the proposed auctions are inherently speculative, existing data offer reasonable suggestions. Using SEC 605 reports, Dyhrberg, Shkilko, and Werner (2022) show that retail investor trading volume greatly exceeds institutional volume for less liquid stocks; if institutional traders were to hypothetically switch to executing all of their trading volume through auctions, there would be less than one institutional bidder on average per retail order in these less-liquid stocks.

The SEC proposal for discrete order-by-order auctions would replace existing retail liquidity programs (RLPs), which implement an order-by-order system of competition by means of a continuous double auction. Both systems offer auction-format competition for segmented retail order flow, and RLPs offer important lessons, particularly about the number of market participants interested in bidding for retail order flow. Ernst, Spatt, and Sun (2024) examine RLPs and show that while there is ample bidding for liquid stocks in calm times, there is limited voluntary liquidity provision in times of market volatility or in less liquid stocks. During times of stress, market participants place almost zero bids in RLPs. These results suggest that the number of bidders interested in segmented retail order flow may be highly dependent on market conditions. Our model further suggests that when there is a limited number of bidders, ex-ante competition produces superior retail investor welfare relative to ex-post competition.

We next extend our baseline model to include institutional traders, as a key objective of the SEC proposal is to enable institutional traders to trade directly with retail investors in auctions. While the entry of institutional traders can potentially increase the number of bidders in an auction, we consider the fact that they also have superior information about the fundamental value of the asset. Incumbent wholesalers, who have no information about the fundamental value, respond by bidding more cautiously in the auction, which leads to less competition. As a result, total welfare can be lower in order-by-order auctions, which contrasts with the results of the baseline model. This welfare loss is due to more-informed institutional traders crowding out the participation of incumbent wholesalers. The overall welfare of retail investors can thus decline further in the switch to order-by-order auctions. Moreover, we find a market segmentation result due to asymmetric information. When information asymmetries between institutional traders and incumbent wholesalers are sufficiently large, only institutional traders will compete for high-quality (low-cost) orders, while all market participants compete for low-quality (high-cost) orders. This leads to a heterogeneous impact of switching to order-by-order auctions on orders with different qualities. When we endogenize the participation decisions of both wholesalers and institutional traders, our analysis shows that in scenarios with severe information asymmetry, if the opportunity cost of institutional traders exceeds that of incumbent wholesalers (but remains relatively low to guarantee their participation), then allowing institutional traders to participate in order-by-order auctions always leads to lower investor welfare. This finding highlights the potential negative effects of institutional traders' entry on investor welfare.

On the theoretical side, our extension to institutional traders features a first-price auction with asymmetric bidders. Although auctions with asymmetric bidders are generally complex (Hubbard and Kirkegaard (2015)), our model offers a tractable framework that admits linear strategies and closed-form solutions for welfare analysis. While we have applied this framework to the context of retail order routing, we believe that the economic model can be applied to broader settings in economics. In addition, our model suggests that ex-post competition (when bidders have more information) may guarantee higher profits for bidders if their new information is independent, as opposed to the case in which their new information is common (Milgrom and Weber (1982)).

Finally, we consider several extensions of our model. First, we consider the case in which newly entered institutional traders are less informed than incumbent wholesalers. Under this extension, the fundamental trade-off and primary results of our main model continue to hold. The key difference between this scenario and that in which institutional traders are more informed is that total welfare is always higher under order-by-order auctions, because the entry of institutional traders in this context does not crowd out the participation of incumbent wholesalers. We next consider a different information structure from our baseline model. In our baseline, under brokers' routing, wholesalers observe their cost shock $y_i$ with some positive probability. In this extension, in contrast, we assume that wholesalers access fundamentally different information under these two trading mechanisms. For instance, they might base their decisions more on macroeconomic information under brokers' routing, while focusing on position information in order-by-order auctions. We confirm that the main results of our baseline model are robust to this extension. Lastly, we consider the impact of order-by-order auctions on cross-sectional liquidity of stocks. Under our model of brokers' routing, wholesalers compete (or commit to a spread) before observing the characteristics of each order. Brokers can evaluate wholesalers based on their performance across all orders, including stocks with different sizes and liquidity levels, allowing for cross-subsidization. In equilibrium, wholesalers may incur losses when trading small, high-cost stocks, which are compensated by profits from trading large, low-cost stocks. This results in smoothed trading costs across stocks. However, switching to order-by-order auctions makes wholesalers earn a profit on each individual order, which leads to more variation in spreads among stocks.

*Related Literature*. Several papers consider whether retail segmentation is optimal as a market design. For instance, Easley, Kiefer, and O'Hara (1996) suggest that Payment for Order Flow (PFOF) is related to skimming trades with lower adverse selection, Battalio and Holden (2001) argue that retail investors benefit from segmentation due to their lower adverse selection, and Baldauf, Mollner, and Yueshen (2024) argue that retail investors are less correlated. Segregating retail flow allows retail traders to obtain better prices,[2] but different mechanisms for segregating retail flow are not explored. Motivated by the recent SEC call for order-by-order competition, our paper provides a theoretical analysis of two possible methods of executing segmented retail trades: the current system of brokers' routing and a hypothetical order-by-order system. We show that the proposed order-by-order system would potentially increase allocative efficiency, but would also decrease retail investor welfare in less-liquid stocks due to lower competition.

Previous studies (Bernhardt and Hughson (1997) and Biais, Martimort, and Rochet (2000)) show that market makers can earn positive profits when competing for orders. Bernhardt and Hughson (1997) emphasize the importance of order splitting in the duopoly case, while Biais, Martimort, and Rochet (2000) study common-value auctions in which multiple market makers compete for an informed order. In both papers, the key friction is the asymmetric information from the liquidity demand side, which corresponds to informed traders. Our study also predicts that market makers will earn positive profits in both the brokers' routing and order-by-order auction settings. However, in contrast to the previous studies, there is no asymmetric information from the liquidity demand side in our model since retail orders are typically uninformed. Milgrom and Weber (1982) show that bidders' profit (wholesaler profit in our model) is lower if bidders all receive the same new information. Our model suggests that this is not necessarily the case if bidders receive independent new information.

[2] This "best execution" guarantee practice may be problematic, as discussed by van Kervel and Yueshen (2023).

Our extension to institutional traders is related to the auction literature with asymmetric bidders (Lebrun (1999), Maskin and Riley (2000, 2003), Kim (2008), Kaplan and Zamir (2012)). In finance settings, Gorbenko and Malenko (2024) examine the endogenous initiation of auctions, and show that bidder asymmetry emerges endogenously. Other implications include equity auctions (Liu (2016)), takeover contests (Povel and Singh (2006)), mergers and acquisitions (Chen and Wang (2023)), and banking (He, Huang, and Parlatore (2024)). Our model is based on a classical model (Menezes and Monteiro (2004), Klemperer (2018)), which we extend to a setting with asymmetrically informed bidders. We obtain a tractable linear equilibrium. In addition, we find an interesting market segmentation result that does not have a counterpart in the aforementioned papers: less-informed bidders tend to bid only on high-cost orders, whereas more-informed bidders bid across all types of orders.

Several studies, including Eaton et al. (2022) and Parlour and Rajan (2003), examine how retail participants themselves impact market liquidity. Our analysis focuses solely on the execution quality obtained by marketable retail orders, as these are the orders that would be subject to the SEC order-by-order proposal. Hu and Murphy (2022), Jain et al. (2024), and Schwarz et al. (2023) explore variation in execution quality among brokers. Market makers can offer two possible forms of superior prices: PFOF (payments from market makers to brokers) and price improvement (payments from market makers directly to retail customers). Brokers may or may not pass on the total amount of PFOF revenue back to their customers in the form of lower commissions, as documented by Battalio, Jennings, and Selway (2001), while Schwarz et al. (2023) and Battalio and Jennings (2022) show that brokers prioritize execution quality, even along dimensions not reflected in SEC 605 reports. Our welfare analysis focuses on the extent to which wholesalers compete for order flow, rather than the extent to which brokers rebate any PFOF received back to retail traders.

Liquidity varies considerably in the cross-section of stocks. Corwin and Coughenour (2008) argue that specialists allocate attention to more-liquid stocks during times of market stress, while Foley et al. (2022) show how tying designated market maker (DMM) assignments in large and small stocks can lead to substantial increases in liquidity for small stocks with little to no observed harm for large stocks. In an extension to our model, we show that brokers' routing can lead to a similar cross-subsidization that is not possible under order-by-order auctions.

At the core of our theoretical analysis is the comparison between ex-ante and ex-post competition. As we demonstrate, ex-post competition need not be stronger than ex-ante competition. In our setting, greater heterogeneity ex-post actually reduces the degree of competition compared to ex-ante competition. The mechanism design literature offers a related perspective by highlighting the value of ex-ante pre-commitment for designing mechanisms (e.g., Myerson (1979), Harris and Townsend (1981)).

One possible analog to order-by-order trading exists in option markets, where a considerable share of volume executes in auctions. Bryzgalova, Pavlova, and Sikorskaya (2023) show that these auctions are correlated with retail trading measures, while Ernst and Spatt (2022) present empirical analysis of specific rules, such as a price-match guarantee and out-sized allocation, which prevent competition in option auctions. Hendershott, Khan, and Riordan (2022) present a model and empirical evidence that auctions in option markets are imperfectly competitive. Another empirical analogue is in bond markets, where Hendershott, Livdan, and Schürhoff (2021) highlight a winner's curse in bond auctions, which is consistent with our model extension to opportunistic traders, which shows that additional bidders do not always improve outcomes due to the winner's curse. In equities, RLPs provide a continuous double auction for retail orders, Ernst, Spatt, and Sun (2024) show that RLPs have limited bidding interest for less-liquid securities and during times of market stress. Proposed auctions may at times attract no bidders at all, and Battalio and Jennings (2023) look at the potential for the National Best Bid and Offer (NBBO) to move during the auction under the SEC's current proposed auction time of 100 to 300 milliseconds.

The remainder of the paper is organized as follows. Section I introduces our baseline model. Section II examines the entry of institutional traders. Section III explores several extensions. Section IV discusses the empirical implications of our model. Section V concludes.

## I. Baseline Model

The model consists of two dates, time 0 and time 1, and there is no discounting. There are three types of market players: a (retail) investor, a broker, and $N \geq 2$ ex-ante identical wholesalers indexed by $i \in \{1, 2, \ldots N\}$. All players in our model are risk neutral. Our broker's objective is to minimize the bid-ask spread paid by the investor.[3]

At time 0, the broker receives a one-unit sell order from the investor and then sends this sell order to a wholesaler to execute the order by the end of time 0.[4] We assume that the retail investor is trading only for liquidity reasons, with the direction of the order independent of the asset value. If wholesaler $i$ executes the order, it must hold the position until time 1, which incurs (marginal) inventory cost $\zeta_i$. The structure of $\zeta_i$ is specified later in this section. Let $s_i$ be the half bid-ask spread offered by wholesaler $i$. The profit that wholesaler $i$ receives at time 1 is then

$$s_i - \zeta_i.$$

[3] Some brokers take PFOF, which accrues to the broker rather than the retail investor. Financial Industry Regulatory Authority (FINRA) best-execution rules prevent considering PFOF in routing decisions. In practice, brokers set one PFOF rate across all wholesalers, so that conditional on setting this PFOF rate, they route to minimize spread. While we do not capture any trade-offs in setting a high PFOF rate (which may impact spreads), our broker's incentives are consistent with a broader goal of extracting welfare from wholesalers, via combined PFOF and spread.

[4] For simplicity, we assume that the costs considered in this model are nondirectional, so the direction of the order does not change our results.

We consider a tractable framework of common-value auctions with linear equilibrium (Klemperer (2018), Menezes and Monteiro (2004)). At time 0, each wholesaler $i$ is hit by an independent and identically distributed (i.i.d.) inventory cost shock $y_i$. For simplicity, we assume that $y_i$ is drawn from a uniform distribution $U[-\frac{1}{2}, \frac{1}{2}]$. The inventory cost $\zeta_i$ has the structure

$$\zeta_i = c_0 + c_1 \frac{1}{N} \sum_{j=1}^{N} y_j + c_2 y_i, \tag{1}$$

where $c_0$, $c_1$, and $c_2$ are positive constants. The cost function consists of three components. The first term, $c_0$, is the unconditional expected inventory cost of executing the order, which is the same for all wholesalers. The second term,

$$c_1 \frac{1}{N} \sum_{j=1}^{N} y_j,$$

represents $\zeta_i$'s exposure to the aggregate cost shock $\frac{1}{N}\sum_{j=1}^{N} y_j$. The third term,

$$c_2 y_i,$$

represents $\zeta_i$'s exposure to the individual cost shock $y_i$. In our model, the coefficients $c_0$, $c_1$, and $c_2$ are exogenous and are determined by stock characteristics.[5] Note that when $N$ is fixed, $\frac{c_1}{c_2}$ measures the correlation of wholesalers' inventory cost, $\mathrm{corr}(\zeta_i, \zeta_j)$, for any $i \neq j$. When $\frac{c_1}{c_2}$ is higher, wholesalers' inventory cost is more correlated.

We consider a unified trading mechanism that integrates both brokers' routing and order-by-order auctions within the framework. In this framework, we assume that each wholesaler $i$ submits the spread $s_i$ after privately observing a noisy signal $w_i$ about the cost shock $y_i$. We further assume that the signal $w_i$ has the following structure.

ASSUMPTION 1: *For any $i$, with probability $p$, $w_i = y_i$, while with probability $1-p$, $w_i$ is drawn from uniform distribution $U\left[-\frac{1}{2}, \frac{1}{2}\right]$, which is independent of all other random variables in the model.*

For any $i$, the noisy signal $w_i$ is either equal to the cost shock $y_i$ with probability $p$ or is completely uninformative with probability $(1-p)$. The broker observes spreads offered by all wholesalers and sends the order to the winner with the lowest spread at the end of time 0.[6] At time 1, all random variables are realized and all players collect their payoffs.

[5] For example, a stock that is about to announce earnings may have a very high $c_1$, with wholesalers very concerned about aggregate inventory imbalances. In contrast, a tick-constrained stock with low informational asymmetries may have a very low $c_1$ value, with wholesalers not very concerned about aggregate inventories.

[6] If more than one wholesaler submits the lowest spread, the winner is chosen randomly among those who submit the lowest spread.

The key difference between brokers' routing and order-by-order auctions is their respective information sets when wholesalers compete for order flow. Under brokers' routing, $p < 1$, which implies that wholesalers do not receive accurate signals about inventory cost when they compete. In practice, brokers and wholesalers establish long-term relationships. Wholesalers' performance is evaluated in the aggregate, not order-by-order, and wholesalers cannot choose when to accept order flow from the broker (see Ernst et al. (2024)); when brokers send an order, wholesalers must fulfill it either by internalizing the order, or by filling the order at another ATS or exchange (and pay any associated trading fees, such as take fees). Under order-by-order auctions, since all wholesalers compete for each order in real time, they have accurate information about their cost shocks when submitting their spreads, which corresponds to the case $p = 1$ in this framework.

In summary, in this framework, the key difference between brokers' routing and order-by-order auctions is the parameter $p$, which measures the precision of wholesalers' information about their cost shocks. Under brokers' routing, $p < 1$, and under order-by-order auctions, $p = 1$.

We focus on symmetric equilibria of the trading mechanism with a general $p \in [0, 1]$. We conjecture (and verify later) that there exists a symmetric equilibrium in which all wholesalers choose the same strategy $s_i(w; p) = s(w; p)$, where

$$s(w; p) = K_0(p) + K_1(p)w,$$

and both $K_0(p)$ and $K_1(p)$ are solved in equilibrium.

PROPOSITION 1: *For any $p \in [0, 1]$, there exists a linear symmetric equilibrium in which the spread offered by wholesaler $i \in \{1, 2, \ldots N\}$ is*

$$s(w_i; p) = K_0(p) + K_1(p)w_i,$$

*where*

$$K_0(p) = c_0 + \frac{p}{2N}\left[c_1\left(\frac{1}{N} + \frac{N}{2} - \frac{1}{2}\right) + c_2\right]$$

*and*

$$K_1(p) = \frac{N-1}{N}p\left[c_1\left(\frac{1}{2} + \frac{1}{N}\right) + c_2\right].$$

In our model, the order is always executed, but inventory cost and equilibrium spreads differ between order-by-order auctions (when $p = 1$) and brokers' routing (when $p < 1$). Let $W_W(p)$ be the expected total profit of all wholesalers, which is given as the expected equilibrium spread minus the incurred inventory cost,

$$W_W(p) = \mathbb{E}\left\{\mathbb{E}\left[K_0(p) + K_1(p)r - c_0 - c_1\frac{1}{N}\sum_{j=1}^{N} y_j - c_2 y_i | w_i = \min_j w_j = r\right]\right\},$$

let $W_I(p)$ be the investor's expected profit, which equals the negative expected equilibrium spread,

$$W_I(p) = -\mathbb{E}\left\{\mathbb{E}\left[K_0(p) + K_1(p)r|\min_i w_i = r\right]\right\},$$

and let $W_{total}(p)$ be the expected total profit, which satisfies $W_{total}(p) = W_W(p) + W_I(p)$. We then obtain the following results.

LEMMA 1: *For any $p \in [0, 1]$, the welfare outcomes of the equilibrium characterized in Proposition 1 are*

$$W_W(p) = \frac{p(c_1 + Nc_2)}{N(1+N)},$$

$$W_I(p) = -\left[c_0 + p\frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right],$$

$$W_{total}(p) = W_W(p) + W_I(p) = -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right).$$

Based on the above results, we consider the welfare comparison between order-by-order auctions and brokers' routing.

PROPOSITION 2: *For any $p \in [0, 1)$, $W_{total}(p) < W_{total}(1)$, $W_W(p) < W_W(1)$, and $W_I(p) < W_I(1)$ if and only if $N(N-3) > 2\frac{c_1}{c_2}$.*

Proposition 2 is a direct result of Lemma 1. The difference between total welfare under order-by-order auctions and brokers' routing is

$$W_{total}(1) - W_{total}(p) = (1-p)\frac{N-1}{N+1}\frac{c_2}{2},$$

which is positive for any $p < 1$. This is because the only welfare loss in our model is from inefficient inventory management. Order-by-order auctions implement the first-best allocation, as the order is always executed by the wholesaler with the lowest inventory cost, while under brokers' routing, it is possible that a wholesaler that obtains the order has ex-post high inventory cost, leading to inefficiency.

The difference between wholesaler welfare under order-by-order auctions and brokers' routing is

$$W_W(1) - W_W(p) = \frac{(1-p)(c_1 + Nc_2)}{N(1+N)},$$

which is positive for any $p < 1$. Intuitively, since wholesalers obtain more precise signals about their cost shocks under order-by-order auctions, their strategies are more responsive to their signals under order-by-order auctions. This makes their offered spreads more heterogeneous, leaving more rent for wholesalers under order-by-order auctions.

The comparison of investor welfare under order-by-order auctions and brokers' routing is unclear, reflecting the trade-off between more efficient inventory management and higher rent earned by wholesalers driven by their heterogeneity in information and concerns about the winner' s curse. First, when the number of wholesalers $N$ is sufficiently large, that is, when $N(N-3) > 2\frac{c_1}{c_2}$, investor welfare will always be higher under order-by-order auctions. Second, the winner' s curse concern arises from the common-value component of the inventory cost, which is related to the correlation of wholesalers' true inventory cost, represented by $\frac{c_1}{c_2}$. Wholesalers tend to bid more conservatively because of the winner's curse, and this effect is more pronounced when $\frac{c_1}{c_2}$ is higher. The effect of the winner's curse is to increase the (expected) equilibrium spread, which hurts the investor. We can also derive intuition for this result from standard auction theory. The revenue equivalence theorem for IPV auctions suggests that investor welfare in first-price auctions is determined by the second-highest honest bid. With more precise information available in order-by-order auctions, the second-highest honest bid tends to be lower, increasing investor welfare. When $N$ is large, the law of large numbers brings the inventory cost structure closer to an IPV setting. Similarly, a lower $\frac{c_1}{c_2}$ reduces the relative importance of the common-value component, which also makes the setting closer to that of an IPV setting.

## II. The Entry of Institutional Traders

Under the SEC's proposal, the entry of institutional traders has been highlighted as a key feature of order-by-order auctions.[7] The SEC hopes that, relative to the current brokers' routing system, order-by-order auctions will allow institutional traders to increase competition for retail trades.[8] This hope, however, ignores the fact that institutional traders typically have superior information about asset quality compared to wholesalers,[9] and their entry decisions are endogenous. The benefits of additional traders must therefore be weighed against the potential increases in informational asymmetry among bidders. In this section, we first analyze the entry of institutional traders: more institutional traders increases the allocative efficiency of the market, but the superior

[7] As SEC chairman Gary Gensler noted, "…individual investors don't necessarily get the best prices that they could get if institutional traders, like pension funds, could systematically and directly compete for their orders" Gensler (2021).

[8] "…a wholesaler is often chosen by a formula that depends on past execution quality of the wholesaler, its relationship with the broker-dealer, and other factors. In addition, the bilateral nature of the wholesaler business model not only restricts contemporaneous competition among wholesalers, it also restricts opportunities for other market participants" Securities and Exchange Commission (2022).

[9] Hendershott, Livdan, and Schürhoff (2015) demonstrate that institutional trading volume predicts several types of important news announcements. In Yang and Zhu (2020), market makers attempt to profit by trading on the information of institutional traders, but they obtain only partial knowledge of institutional traders' information based on past observed trades, leaving them less informed than institutional traders.

information of institutional traders negatively impacts bidding by incumbent wholesalers. Which effect dominates depends on the level of information asymmetry in the market. We next endogenize the entry decisions of institutional traders and show that when the opportunity cost of institutional traders exceeds that of wholesalers (but is not too high), retail investors will surely be made worse by the equilibrium level of entry chosen by institutional traders in order-by-order auctions.

To extend our model to include institutional traders, we make two minimal changes to the baseline model. First, in addition to the $N$ wholesalers who always provide market-making services, there are $N_0 \geq 2$ institutional traders who can also provide liquidity, but only in order-by-order auctions. This setup is consistent with the market design suggested in the SEC proposal, under which institutional traders are absent in the current brokers' routing system but can be active and provide more competition in order-by-order auctions. We assume that institutional traders $i \in \{1, 2, \ldots N_0\}$ also receive i.i.d. cost shocks $y_i$ at time 0, that follows uniform distribution $U\left[-\frac{1}{2}, \frac{1}{2}\right]$. The cost shock $y_i$ plays a role similar to that for wholesalers, as discussed later in the model.

Second, we consider the following (new) inventory cost structure,

$$\widetilde{\zeta}_i = \widetilde{c}_0 + c_1 \frac{1}{\widetilde{N}} \sum_{j=1}^{\widetilde{N}} y_j + c_2 y_i, \tag{2}$$

where $\widetilde{c}_0$ is a random variable that can be $c_0 - \delta_c$ or $c_0 + \delta_c$[10] with equal probabilities and $\widetilde{N}$ is the number of active market makers for this order.[11] In brokers' routing, only wholesalers compete for retail orders, so our equilibrium features $\widetilde{N} = N$; and in order-by-order auctions, both wholesalers and institutional traders can compete for retail orders, so in this case, $\widetilde{N}$ depends on the number of active market makers. Note that $\mathbb{E}\left(\widetilde{c}_0\right) = c_0$, that is, the unconditional expectation of inventory cost remains the same in this extension.

Institutional traders have an information advantage over wholesalers. Specifically, all institutional traders can observe the realization of $\widetilde{c}_0$ at time 0, while wholesalers only know the distribution of $\widetilde{c}_0$. This implies that when competing for retail orders, institutional traders can condition their bids on the realization of $\widetilde{c}_0$, while wholesalers can only use distributional information about $\widetilde{c}_0$.[12] This information asymmetry captures the notion that institutional traders are more informed about the characteristics of traded assets, market conditions, or future price movement.

We first consider the market equilibrium in the brokers' routing system. Since institutional traders are absent in brokers' routing, the only difference between this extension and our baseline model is the structure of inventory

[10] Without loss of generality, we assume $\delta_c \geq 0$.

[11] "Active market makers" refer to those who have a nonzero probability of obtaining the order in equilibrium.

[12] We could also interpret $\pm\delta_c$ as private information about asset quality and keep the inventory cost structure unchanged. This would not change the outcomes of our model.

cost. The additional randomness in expression (2) has no impact on market equilibrium, because all wholesalers are risk neutral and thus care only about the expectation of $\widetilde{c}_0$. Recall that $\mathbb{E}\left(\widetilde{c}_0\right) = c_0$, which is the same as in the baseline model.

PROPOSITION 3: *With inventory cost structure (2), under brokers' routing, the equilibrium bidding strategies and welfare outcomes are the same as characterized by Proposition 1 and Lemma 1.*

The market equilibrium is unchanged under brokers' routing, and thus, we view it as a suitable benchmark for the (new) model with institutional traders. Under order-by-order auctions, however, the equilibrium does change due to the entry of institutional traders.

First, consider the case in which $\delta_c = 0$, and thus institutional traders have no informational advantage compared to wholesalers. In this case, the only effect is increased competition in order-by-order auctions, which is a direct result of the increased number of bidders competing for the retail order. Relative to Proposition 1 of our baseline model, to obtain the new market equilibrium of order-by-order auctions, we simply take $p = 1$ and replace the number of bidders $N$ in Proposition 1 with $(N + N_0)$, because institutional traders are ex ante identical to wholesalers in this case. The following proposition characterizes this equilibrium. Since all wholesalers and institutional traders can observe their own cost shock $y_i$ with probability one, we write their equilibrium strategies explicitly as functions of $y$ in the proposition.

PROPOSITION 4 (No Information Advantage): *When $\delta_c = 0$, under order-by-order auctions, there exists a linear symmetric equilibrium in which all wholesalers and institutional traders choose the same strategy, and the spread offered by any wholesaler or institutional trader i is*

$$\widetilde{s}_i(y_i) = \widetilde{k}_0 + \widetilde{k}_1 y_i,$$

*where*

$$\widetilde{k}_0 = c_0 + \frac{p}{2(N+N_0)}\left[c_1\left(\frac{1}{N+N_0} + \frac{N+N_0}{2} - \frac{1}{2}\right) + c_2\right]$$

*and*

$$\widetilde{k}_1 = \frac{N+N_0-1}{N+N_0}p\left[c_1\left(\frac{1}{2} + \frac{1}{N+N_0}\right) + c_2\right].$$

Denote by $\widetilde{W}_i^j$ total welfare ($i = total$), investor welfare ($i = I$), wholesaler welfare ($i = W$), and institutional trader welfare ($i = IT$) under order-by-order auctions ($j = OBO$) and brokers' routing ($j = BR$). To begin, we consider how participation of institutional traders changes welfare outcomes under order-by-order auctions.

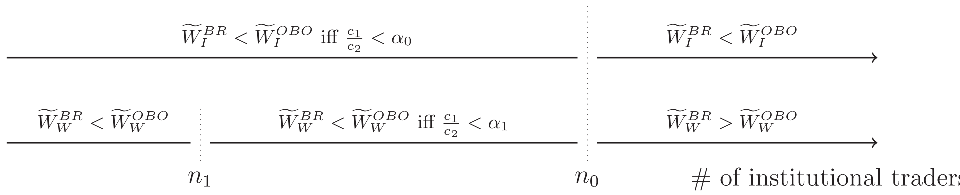


**Figure 1. Investor and wholesaler welfare comparison**: $\widetilde{W}_I^{BR}$ **versus** $\widetilde{W}_I^{OBO}$ **and** $\widetilde{W}_W^{BR}$ **versus** $\widetilde{W}_W^{OBO}$**, when** $\delta_c = 0$.

PROPOSITION 5: *When $\delta_c = 0$, both $\widetilde{W}_{total}^{OBO}$ and $\widetilde{W}_I^{OBO}$ are increasing functions of the number of institutional traders $N_0$, and $\widetilde{W}_W^{OBO}$ is a decreasing function of $N_0$.*

Proposition 5 shows that regulators focusing on total or investor welfare will have different preferences regarding the participation of institutional traders compared to wholesalers. This finding aligns with our discussions with practitioners. Institutional traders enhance order allocation efficiency in order-by-order auctions, ultimately benefiting retail investors, but this leads to increased competition, negatively impacting wholesalers. It is important to note that our discussion of Proposition 5 is within the context of order-by-order auctions. We also consider the welfare implications of transitioning from the current brokers' routing framework to order-by-order auctions with institutional traders. These two trading mechanisms differ not only in the participation of institutional traders, but also in the information environment, as discussed in the baseline model. Our next proposition summarizes these results.

PROPOSITION 6: *When $\delta_c = 0$, we always have $\widetilde{W}_{total}^{BR} < \widetilde{W}_{total}^{OBO}$. In addition, there exist $n_0 > n_1 > 0$ such that:*

1. *If $N_0 \geq n_0$, we always have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$; if $N_0 < n_0$, $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$ if and only if $\frac{c_1}{c_2} < \alpha_0$, where $\alpha_0 > 0$ is solved in (B5).*
2. *If $N_0 \leq n_1$, we always have $\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO}$; if $N_0 \geq n_0$, we always have $\widetilde{W}_W^{BR} > \widetilde{W}_W^{OBO}$; if $n_1 < N_0 < n_0$, $\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO}$ if and only if $\frac{c_1}{c_2} < \alpha_1$, where $\alpha_1 > 0$ is solved in (B7).*

Figure 1 illustrates the welfare results in Proposition 6. Proposition 6 shows that in the absence of adverse selection ($\delta_c = 0$), total welfare is higher under order-by-order auctions. This outcome is driven by two factors. First, as explained in our baseline model, order-by-order auctions implement a first-best allocation, which increases total welfare. Second, introducing institutional traders increases the number of bidders in the auction. Since each bidder draws an independent cost shock $y_i$, increasing the number of bidders decreases the expected lowest private component of inventory cost, $\min_i y_i$.

The impact of institutional trader participation on investor and wholesaler welfare is less straightforward. According to Proposition 5, investor welfare increases with $N_0$, while wholesaler welfare decreases. Consequently, for sufficiently large $N_0$ ($N_0 \geq n_0$), investor welfare is higher in order-by-order auctions, while wholesaler welfare is lower in order-by-order auctions. For sufficiently small $N_0$ ($N_0 \leq n_1$), wholesaler welfare is always higher in order-by-order auctions.

Proposition 6 additionally highlights the importance of the correlation structure in inventory costs, represented by $\frac{c_1}{c_2}$. In cases in which $N_0$ is less than $n_0$, both the investor and wholesalers can (potentially) benefit from switching to order-by-order auctions, provided $\frac{c_1}{c_2}$ is sufficiently low. Lower correlations in inventory costs increase the likelihood that both the investor and wholesalers benefit more under order-by-order auctions. This is attributed to improved total efficiency in order execution when switching from brokers' routing to order-by-order auctions, particularly when inventory costs between participants are less correlated. The resulting surplus is shared by the retail investor and wholesalers (and institutional traders).

We next consider the case in which institutional traders have at least some informational advantage, that is, $\delta_c > 0$. We still focus on symmetric linear equilibria whereby all institutional traders choose the same linear strategy and all wholesalers choose the same linear strategy. When $\delta_c > 0$, institutional traders can condition their offered spreads on the realization of $\widetilde{c}_0$. Intuitively, when observing $\widetilde{c}_0 = c_0 - \delta_c$, institutional traders submit lower spreads, and when $\widetilde{c}_0 = c_0 + \delta_c$, they submit higher spreads. In contrast, wholesalers cannot condition their spreads on the realizations of $\widetilde{c}_0$, but only on the distributional information $c_0$. Wholesalers are more likely to win auctions when $\widetilde{c}_0 = c_0 + \delta_c$ than when $\widetilde{c}_0 = c_0 - \delta_c$, as institutional traders bid more aggressively in the latter case. This leads to adverse selection for wholesalers, as they are more likely to win auctions when $\widetilde{c}_0 > c_0$. A winner's curse argument implies that wholesalers will submit more conservative bids in equilibrium. When $\delta_c$ is sufficiently large, the winner's curse concern becomes so severe that all wholesalers will be completely out of contention for high-quality (low-cost) stocks, only obtaining the retail order when $\widetilde{c}_0 = c_0 + \delta_c$. Consequently, when $\widetilde{c}_0 = c_0 - \delta_c$, institutional traders face no competition from wholesalers, which can reduce retail investor welfare. The following proposition characterizes this equilibrium.

PROPOSITION 7: *Let $\widetilde{s}^-(y;\delta_c)$ and $\widetilde{s}^+(y;\delta_c)$ be two bidding strategies, where*

$$\widetilde{s}^-(y;\delta_c) = \widetilde{k}_0^-(\delta_c) + \widetilde{k}_1^-(\delta_c)y$$

*with*

$$\widetilde{k}_0^-(\delta_c) = c_0 - \delta_c + \frac{p}{2N_0}\left[c_1\left(\frac{1}{N_0} + \frac{N_0}{2} - \frac{1}{2}\right) + c_2\right]$$

$$\widetilde{k}_1^-(\delta_c) = \frac{N_0 - 1}{N_0} p\left[c_1\left(\frac{1}{2} + \frac{1}{N_0}\right) + c_2\right],$$

*and*

$$\widetilde{s}^+(y; \delta_c) = \widetilde{k}_0^+(\delta_c) + \widetilde{k}_1^+(\delta_c) y$$

*with*

$$\widetilde{k}_0^+(\delta_c) = c_0 + \delta_c + \frac{p}{2(N+N_0)}\left[c_1\left(\frac{1}{N+N_0} + \frac{N+N_0}{2} - \frac{1}{2}\right) + c_2\right]$$
$$\widetilde{k}_1^+(\delta_c) = \frac{N+N_0-1}{N+N_0} p\left[c_1\left(\frac{1}{2} + \frac{1}{N+N_0}\right) + c_2\right].$$

*There exists a threshold $\underline{\delta} > 0$ that satisfies the condition*

$$\widetilde{k}_0^-(\underline{\delta}) + \widetilde{k}_1^-(\underline{\delta})\frac{1}{2} < \widetilde{k}_0^+(\underline{\delta}) - \widetilde{k}_1^+(\underline{\delta})\frac{1}{2},$$

*such that when $\delta_c > \underline{\delta}$, there exists an equilibrium of order-by-order auctions in which*

1. *wholesalers choose bidding strategy $\widetilde{s}^+(y; \delta_c)$, and*
2. *institutional traders choose bidding strategy $\widetilde{s}^+(y; \delta_c)$ when observing $c_0 + \delta_c$ and $\widetilde{s}^-(y; \delta_c)$ when observing $c_0 - \delta_c$.*

In this equilibrium, when the true state is $\widetilde{c}_0 = c_0 - \delta_c$, the highest possible spread offered by institutional traders is still lower than the lowest possible spread offered by wholesalers, and thus wholesalers will never obtain the order, irrespective of their signal realizations. When the true state is $\widetilde{c}_0 = c_0 + \delta_c$, wholesalers and institutional traders will choose the same bidding strategy $\widetilde{s}^+(y; \delta_c)$, and thus, all players will obtain the order with the same probabilities.

If we interpret the random variable $\widetilde{c}_0$ as the heterogeneous quality of stocks, then in equilibrium, only institutional traders compete for retail orders of high-quality stocks, while all participants compete for orders of low-quality stocks. The market for low-quality stocks becomes more competitive due to an increase in the number of bidders, while the market for high-quality stocks may become less competitive as institutional traders are the only effective bidders. The presence of adverse selection can weaken competition and potentially harm total welfare, as illustrated by our following proposition.

PROPOSITION 8: *In the equilibrium characterized by Proposition 7, there exist $t^*$, $t_1 \geq t_0$ and $t_3 > t_2$, such that*

1. *$\widetilde{W}_{total}^{OBO} > \widetilde{W}_{total}^{BR}$ if and only if $N_0 > t^*$,*

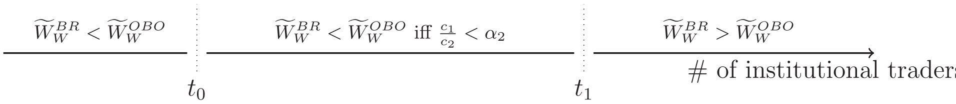


**Figure 2. Wholesaler welfare comparison $\widetilde{W}_W^{BR}$ versus $\widetilde{W}_W^{OBO}$, when $\delta_c > \underline{\delta}$.**

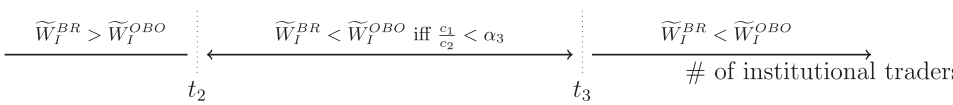


**Figure 3. Investor welfare comparison $\widetilde{W}_I^{BR}$ versus $\widetilde{W}_I^{OBO}$, when $\delta_c > \underline{\delta}$.**

2. *if $N_0 \le t_0$, we always have $\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO}$; if $N_0 \ge t_1$, we always have $\widetilde{W}_W^{BR} > \widetilde{W}_W^{OBO}$; if $t_0 < N_0 < t_1$, $\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO}$ if and only if $\frac{c_1}{c_2} < \alpha_2$, where $\alpha_2 > 0$ is defined in (B13), and*
3. *if $N_0 \ge t_3$, we always have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$; if $N_0 \le t_2$, we have $\widetilde{W}_I^{BR} > \widetilde{W}_I^{OBO}$; if $t_2 < N_0 < t_3$, we have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$ if and only if $\frac{c_1}{c_2} < \alpha_3$, where $\alpha_3 > 0$ is defined in (B15).*

Figures 2 and 3 illustrate the welfare results in Proposition 8. Comparison of total welfare between brokers' routing and order-by-order auctions is ambiguous because of the adverse selection. The transition from brokers' routing to order-by-order auctions improves total welfare only when a sufficient number of institutional traders provide liquidity. As we show above, in equilibrium, only $N_0$ institutional traders effectively compete for high-quality (low-cost) orders, and all institutional traders and wholesalers ($N_0 + N$ in total) compete for low-quality (high-cost) orders. When the number of institutional traders $N_0$ is small, the loss of welfare in high-cost stocks is greater than the gain in welfare in low-cost stocks.

Proposition 8 also highlights a divergence in preferences between wholesalers and retail investors with respect to the shift from brokers' routing to order-by-order auctions. Specifically, when the number of institutional traders, $N_0$, is sufficiently large, the investor strictly prefers order-by-order auctions, while wholesalers prefer brokers' routing. Conversely, with a smaller $N_0$, the investor prefers brokers' routing and wholesalers prefer order-by-order auctions. These results also hold in our earlier analysis under no adverse selection ($\delta_c = 0$). However, when $\delta_c = 0$, investors will always realize greater welfare under order-by-order auctions when the correlation of wholesalers' inventory cost is low enough, that is, when $\frac{c_1}{c_2}$ is low enough, but this result does not hold when adverse selection is severe. In Proposition 8, when $N_0 \le t_2$, the investor is always worse off in order-by-order auctions, regardless of the value of $\frac{c_1}{c_2}$. This result implies that the presence of adverse selection further hurts investor welfare.

Despite this misalignment, it remains possible to identify scenarios in which both wholesalers and the investor favor order-by-order auctions, as confirmed by the following lemma.

LEMMA 2: *First, $t_3 > t_0$. Second, when $N \geq 3$, there exists a threshold $p^*$ such that $t_1 > t_2$ if and only if $p < p^*$.*

When $t_1 > t_2$, a specific range emerges ($\max\{t_2, t_0\} < N_0 < \min\{t_1, t_3\}$) such that both the investor and wholesalers realize greater welfare under order-by-order auctions when the $\frac{c_1}{c_2}$ ratio is sufficiently low. This follows directly from Proposition 8. Such a scenario is feasible only if the signal precision $p$ under brokers' routing is low. The rationale behind this result is as follows. For both wholesalers and the investor to benefit from order-by-order auctions, a significant overall welfare improvement is necessary when transitioning from brokers' routing. With a lower $p$, total welfare under brokers' routing is lower, increasing the welfare improvements offered by order-by-order auctions. Under these conditions, both the investor and wholesalers are more likely to benefit from the switch to order-by-order auctions.

### *A. The Endogenous Entry of Institutional Traders*

The entry and exit decisions of wholesalers and institutional traders are endogenous. As we discussed earlier, the entry of institutional traders will reduce wholesalers' profit, and thus, wholesalers may exit the market. This force tends to reduce liquidity provision from wholesalers and thus investor welfare. In this section, we formalize the endogenous entry and exit of liquidity providers—wholesalers and institutional traders. To isolate the effect of endogenous entry, we focus primarily on the order-by-order auction framework. In other words, we consider and compare two cases: order-by-order auctions when institutional traders are excluded, and order-by-order auctions when their entry is allowed.[13] We extend the model in Section II and introduce new assumptions about wholesalers and institutional traders' opportunity cost (outside option). For each wholesaler, we assume that the opportunity cost of participating in order-by-order auctions is $\pi_W$. This cost can be viewed as the maximum profit wholesalers could achieve using their current resources, should they opt out of the market-making business. There exists a pool of identical institutional traders who can potentially participate in order-by-order auctions. The opportunity cost of entering the business is $\pi_I$ for each institutional trader, which may relate to regulatory compliance, hiring needs,

[13] We do not consider the case of allowing endogenous entry of institutional traders in the current brokers' routing system. This is because no such mechanism exists in the current system, and there is no indication that it will be introduced in the near future, and thus, to keep our analysis relevant, we have excluded this scenario. In practice, institutional traders are not market makers aiming to both buy and sell; rather, they are interested in taking or exiting positions. The directional nature of their trading poses a significant challenge for any potential entry into the current brokers' routing system, where brokers expect all marketable orders that they route—both buys and sells—to be immediately filled by wholesalers.

or infrastructure setup.[14] Our goal is to solve for the equilibrium numbers of wholesalers, $\widetilde{N}$, and institutional traders, $\widetilde{N}_0$, when the entry of institutional traders is allowed, and the equilibrium number of wholesalers, $\hat{N}$, when the entry of institutional traders is not allowed. We focus on equilibria such that, if institutional traders are active, at least two institutional traders participate. This allows us to apply the analysis from the previous sections consistently.

In our previous analysis, we consider two cases: one with no adverse selection ($\delta_c = 0$), and one with severe adverse selection ($\delta_c > \underline{\delta}$). The second case is more empirically relevant and provides new economic insights. Thus, in the rest of this section, we assume that adverse selection is sufficiently pronounced, and hence focus on the equilibria under this condition. For completeness, we provide the analysis for the case $\delta_c = 0$ in Appendix A. First, to focus on the equilibria when both wholesalers and institutional traders are possibly active, we impose the following assumption on $\pi_W$.

ASSUMPTION 2: $\pi_W \leq \frac{1}{24}\left(\frac{c_1}{3} + c_2\right)$.

Assumption 2 guarantees that the opportunity cost $\pi_W$ is not too high, such that when two institutional traders are operating in equilibrium, at least one wholesaler would still like to enter. When the entry of institutional traders is disallowed, the expected profit of each of the $\hat{N}$ wholesalers is

$$\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}} + c_2\right),$$

and thus, the equilibrium condition is

$$\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}} + c_2\right) \geq \pi_W > \frac{1}{\hat{N}+1}\frac{1}{\hat{N}+2}\left(\frac{c_1}{\hat{N}+1} + c_2\right). \tag{3}$$

When the entry of institutional traders is allowed, suppose that there are $\widetilde{N} \geq 1$ wholesalers and $\widetilde{N}_0 \geq 2$ institutional traders in equilibrium. Then, based on our previous results, the expected profit of each wholesaler is

$$\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0} + c_2\right)$$

and the equilibrium condition for wholesalers is

$$\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0} + c_2\right)}{\widetilde{N}+\widetilde{N}_0+1} \geq \pi_W > \frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0+1} + c_2\right)}{\widetilde{N}+\widetilde{N}_0+2}. \tag{4}$$

Conditions (3) and (4) jointly imply that

$$\hat{N} \geq \widetilde{N} + \widetilde{N}_0.$$

[14] We assume that both wholesalers and institutional traders choose to enter the market if they are indifferent.

The above analysis implies that when the entry of institutional traders is allowed, if both wholesalers and institutional traders operate in equilibrium, then the total number of liquidity providers in this case, $\widetilde{N}+\widetilde{N}_0$, is weakly smaller than the number of liquidity providers when the entry of institutional traders is disallowed, $\hat{N}$, suggesting that the welfare of the retail investors is lower if the entry of institutional traders is allowed. The following proposition characterizes the full equilibrium.

PROPOSITION 9: *When adverse selection is severe, that is, when $\delta_c$ is sufficiently high, there exist $\pi_3 > \pi_2 > \pi_1 > 0$ such that*

1. *When $\pi_I > \pi_3$, no equilibrium exists with $\widetilde{N}_0 \geq 2$.*
2. *When $\pi_I \in (\pi_1, \pi_3]$, there exists an equilibrium such that $\widetilde{N} \geq 1$ and $\widetilde{N}_0 \geq 2$. In all such equilibria, investor welfare must be strictly lower when the entry of institutional traders is allowed.*
3. *When $\pi_I \in (0, \pi_2]$, there exists an equilibrium such that $\widetilde{N} = 0$ and $\widetilde{N}_0 \geq 2$. In all such equilibria, if $\pi_I \geq \pi_W$ ($\pi_I \leq \pi_W$), investor welfare must be weakly lower (higher) when the entry of institutional traders is allowed.*

*Since $\pi_2 > \pi_1$, multiple equilibria may exist when $\pi_I \in (\pi_1, \pi_2]$.*

Proposition 9 consists of two nontrivial cases. The first is probably the most empirically relevant case, in which both institutional traders and wholesalers coexist in the market. In this equilibrium, we find that the total number of liquidity providers (both wholesalers and institutional traders) is weakly lower than the total number of liquidity providers (solely wholesalers) in the equilibrium in which institutional traders are disallowed. This leads to lower welfare for the investor. The intuition is as follows. Institutional traders have an informational advantage over wholesalers. The reason wholesalers can still exist in equilibrium is that institutional traders have a relatively higher opportunity cost compared to wholesalers. In this case, institutional traders demand higher expected profits to break even, which lowers the competitiveness of the market. As a result, there will be fewer liquidity providers in equilibrium, which ultimately hurts investor welfare. We highlight this result in the following remark.

REMARK 1: If both wholesalers and at least two institutional traders coexist in equilibrium, then investor welfare must be strictly lower compared to the case in which institutional traders are disallowed.

When the opportunity cost of institutional traders is low (when $\pi_I < \pi_2$), it is possible that no wholesaler will exist in equilibrium. Note that the threshold $\pi_2$ satisfies $\pi_2 > \pi_W$. This means that even if institutional traders have higher opportunity cost compared to wholesalers, it is still possible that only institutional traders operate in equilibrium. This is because they have a significant informational advantage compared to wholesalers, and thus are able to capture higher profit. However, if $\pi_I \in [\pi_W, \pi_2)$, then in this equilibrium, investor welfare must be weakly lower, because in this case, there will be weakly fewer liquidity providers in equilibrium compared to the case in which institutional traders are disallowed. Similarly, we show that when $\pi_I \leq \pi_W$, investor welfare will be (weakly) higher if the entry of institutional traders is allowed. Whether the condition $\pi_I \leq \pi_W$ holds is inherently an empirical question. Our prior is that this condition is not likely to hold in practice, as wholesalers focus on its market-making business while institutional traders typically focus on investing rather than liquidity provision, which suggests that they have a higher opportunity cost $\pi_I$.

## III. Extensions

### *A. More Informed Wholesalers*

In our previous analysis with institutional traders, we assume that institutional traders are more informed than wholesalers, and we show that this additional layer of asymmetric information may have an additional negative impact on welfare. In this extension, we consider a setting in which wholesalers are more informed than institutional traders, which may arise if institutional traders' information is not correlated with short-term price movements. To begin, we consider the same cost structure (2) and assume that all wholesalers observe the realization of $\widetilde{c}_0$, while all institutional traders only know the distributional information about $\widetilde{c}_0$. For simplicity, we use the same notation in this extension (there are $N \geq 2$ wholesalers in brokers' routing, and $N \geq 2$ wholesalers and $N_0 \geq 0$ institutional traders in order-by-order auctions). To focus on the most interesting case, we consider the case in which the adverse selection is severe, which corresponds to the case in which $\delta_c$ is sufficiently high.

First, under brokers' routing, since wholesalers observe the realization of $\widetilde{c}_0$, which can be $c_0 - \delta_c$ or $c_0 + \delta_c$, their bidding strategies will depend on the realization of $\widetilde{c}_0$. Note that under any realization $\widetilde{c}_0$, the market is (almost) the same as that discussed in Section I, with the only difference being the value of $c_0$. We then obtain the following equilibrium and welfare results.

PROPOSITION 10: *When there are N wholesalers in brokers' routing, there exists a linear symmetric equilibrium in which the spread submitted by each wholesaler i is*

$$\widetilde{t}(w_i, \widetilde{c}_0) = \widetilde{K}_0 + \widetilde{K}_1 w_i,$$

*where*

$$\widetilde{K}_0 = \widetilde{c}_0 + \frac{p}{2N}\left[c_1\left(\frac{1}{N} + \frac{N}{2} - \frac{1}{2}\right) + c_2\right]$$

*and*

$$\widetilde{K}_1 = \frac{N-1}{N}p\left[c_1\left(\frac{1}{2} + \frac{1}{N}\right) + c_2\right].$$

*Under this equilibrium, investor's welfare is*

$$\widetilde{W}_I^{BR} = -\left[c_0 + p\frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right].$$

The proof replicates the proof of Proposition 1. We next consider the market equilibrium under order-by-order auctions. In this extension, we have $N$ wholesalers who observe $\widetilde{c}_0$ and $N_0$ institutional traders who do not know $\widetilde{c}_0$. Note that in Section II, we have $N_0$ institutional traders who observe $\widetilde{c}_0$ and $N$ wholesalers who do not. Due to the symmetry and based on our results in Proposition 7, we obtain the following result.

PROPOSITION 11: *Let $\widetilde{t}^-(y;\delta_c)$ and $\widetilde{t}^+(y;\delta_c)$ be two bidding strategies, where*

$$\widetilde{t}^-(y;\delta_c) = \widetilde{K}_0^-(\delta_c) + \widetilde{K}_1^-(\delta_c)y$$

*with*

$$\widetilde{K}_0^-(\delta_c) = c_0 - \delta_c + \frac{c_1}{4N}\left(N - 1 + \frac{2}{N}\right) + \frac{c_2}{2N}$$
$$\widetilde{K}_1^-(\delta_c) = \frac{N-1}{N}\left(\frac{c_1}{2}\frac{N+2}{N} + c_2\right),$$

*and*

$$\widetilde{t}^+(y;\delta_c) = \widetilde{K}_0^+(\delta_c) + \widetilde{K}_1^+(\delta_c)y$$

*with*

$$\widetilde{K}_0^+(\delta_c) = c_0 + \delta_c + \frac{c_1}{4(N+N_0)}\left(N + N_0 - 1 + \frac{2}{N+N_0}\right) + \frac{c_2}{2(N+N_0)}$$
$$\widetilde{K}_1^+(\delta_c) = \frac{N+N_0-1}{N+N_0}\left(\frac{c_1}{2}\frac{N+N_0+2}{N+N_0} + c_2\right).$$

*Then there exists a threshold $\underset{\sim}{\delta}$ such that when $\delta_c > \underset{\sim}{\delta}$, there exists an equilibrium in which*

1. *institutional traders choose bidding strategy $\widetilde{t}^+(y;\delta_c)$, and*
2. *wholesalers choose bidding strategy $\widetilde{t}^+(y;\delta_c)$ if observing $c_0 + \delta_c$ and $\widetilde{t}^-(y;\delta_c)$ if observing $c_0 - \delta_c$.*

*Under this equilibrium, investor welfare is*

$$\widetilde{W}_I^{OBO} = -c_0 - \frac{1}{2}\left[\frac{c_1}{(N+N_0)(N+N_0+1)} - \frac{(N+N_0-3)c_2}{2(N+N_0+1)} + \frac{c_1}{N(N+1)} - \frac{(N-3)c_2}{2(N+1)}\right].$$

The proof replicates the proof of Proposition 7. In this extension, we obtain an equilibrium similar to that characterized in Proposition 7. With sufficiently pronounced asymmetric information between wholesalers and institutional traders, uninformed players (institutional traders in this extension) completely opt out of providing liquidity for high-quality (low-cost) stocks. This implies that in order-by-order auctions, introducing uninformed institutional traders does not change the welfare of retail investors trading low-cost stocks. We obtain the following result regarding investor welfare.

PROPOSITION 12: *When $\delta_c > \underline{\delta}$, we have the following results on welfare comparison:*

1. $\widetilde{W}^{OBO}_{total} > \widetilde{W}^{BR}_{total}$
2. $\widetilde{W}^{OBO}_{I} > \widetilde{W}^{BR}_{I}$ *if and only if*

$$\left[\begin{array}{l}\frac{1}{(N+N_0)(N+N_0+1)}\\ +\frac{1}{N(N+1)}(1-2p)\end{array}\right]\frac{c_1}{c_2} < \frac{1}{2}\left(\frac{N+N_0-3}{N+N_0+1}-\frac{N-3}{N+1}\right)+\frac{N-3}{N+1}(1-p). \quad (5)$$

In Proposition 12, we consider the welfare outcomes under brokers' routing versus order-by-order auctions. Total welfare is always higher under order-by-order auctions, even with severe informational asymmetry. This stems from the fact that, in this extension, wholesalers are better informed and thus always active in providing liquidity, independent of the order type. This results in increased participation of institutional traders without a decrease in wholesaler participation, a notable contrast to our findings in Proposition 8, where total welfare could decline under order-by-order auctions due to the exit of wholesalers from low-cost orders.

For investor welfare, the comparison remains ambiguous and depends on condition (5). Notably, when $N_0 = 0$, this condition is reduced to that in Proposition 2. The left-hand side (LHS) of condition (5) decreases in $N_0$, while its right-hand side (RHS) increases in $N_0$. Consequently, introducing more institutional traders into order-by-order auctions makes this condition more easily satisfied. This implies that if investor welfare is higher under order-by-order auctions in the baseline model with no information asymmetry about $\widetilde{c}_0$, then wholesalers being more informed does not alter this outcome.

However, wholesalers being more informed weakens the potential advantages of introducing institutional traders. To see this, consider the case with a very high number of institutional traders ($N_0 \to \infty$). Then condition (5) becomes

$$\frac{1-2p}{N}\frac{c_1}{c_2} < 2 + (N-3)(1-p),$$

which may or may not hold depending on specific parameters. However, if institutional traders are more informed, as per Proposition 8, then retail investor welfare is always higher under order-by-order auctions when $N_0 \to \infty$. This raises questions about the efficacy of introducing more institutional traders to order-by-order auctions. Our findings suggest that even with sufficiently many institutional traders, retail investors may not benefit from the shift from brokers' routing to order-by-order auctions if wholesalers have more information than the newly entered institutional traders.

### *B. Alternative Information Structure*

In our baseline model, wholesalers receive the same information with different probabilities under both brokers' routing and order-by-order auctions. We now consider the possibility that the nature of the information itself differs between these two trading mechanisms. For instance, in order-by-order auctions, their information might primarily reflect a transient effect influenced by wholesalers' current position, which can be fast-moving, while under brokers' routing, their information could be more closely related to their financing costs, which can be slow-moving and influenced by factors like current interest rates and financial conditions. Despite these possible differences, in this extension, we show that the core insights and trade-offs identified in our baseline model continue to hold.

To explore this possibility, in this extension, we shift our focus from the probability of receiving the same information (as in the baseline model) to the type of information obtained under each trading mechanism. We still consider the same inventory cost structure (1) for both trading mechanisms, and assume that for each $i$,

$$y_i = w_i + l_i, \tag{6}$$

where $\{w_i, l_i\}_i$ are independent of each other and satisfy the following assumption.

ASSUMPTION 3: *$A$ is a nonnegative integer. All $w_i$ are i.i.d. and follow uniform distribution $U\left[-\frac{1}{2}\frac{1}{2A+1}, \frac{1}{2}\frac{1}{2A+1}\right]$. All $l_i$ are i.i.d. and follow the distribution*

$$Prob(l_i = l) = \frac{1}{2A+1}$$

*for all $l \in \frac{1}{2A+1}\left\{-A, -(A-1)\ldots 1, 0, 1, \ldots (A-1), A\right\}$.*

The following lemma shows that the distribution of $y_i$ remains the same as in our baseline model under this new specification.

LEMMA 3: *Under Assumption 3, $\{y_i\}_i$ are i.i.d. and follow uniform distribution $U\left[-\frac{1}{2}, \frac{1}{2}\right]$.*

In this extension, we assume that any wholesaler $i$ observes only $w_i$ under brokers' routing, while they observe both $w_i$ and $l_i$ (and thus $y_i$) under order-by-order auctions. The first component $w_i$ in (6) represents factors like current interest rates and financial conditions, which partly influence actual inventory costs and are observable under brokers' routing (due to their slow-moving nature). The second component, $l_i$, represents additional transient information including liquidity position that is rapidly changing and can only be observed under order-by-order auctions. Note that $\mathbb{E}(l_i) = 0$, which means that the effects of the second term average out under brokers' routing and will not impact wholesalers' decisions under brokers' routing. When $A$ is larger, the first component $w_i$ is less important for true inventory costs. When $A = 0$, the brokers' routing and order-by-order auctions become equivalent.

Under the above specifications, we obtain the equilibrium outcomes and welfare results for both brokers' routing and order-by-order auctions.

PROPOSITION 13: *Under Assumption 3, there exists a linear symmetric equilibrium of brokers' routing in which the spread offered by wholesaler $i \in \{1, 2 \ldots N\}$ is*

$$s(w_i; A) = A_0 + A_1 w_i,$$

*where*

$$A_0 = c_0 + c_1 \frac{1}{2N(2A+1)}\left(\frac{1}{N} + \frac{N}{2} - \frac{1}{2}\right) + c_2 \frac{1}{2N(2A+1)}$$

*and*

$$A_1 = \frac{N-1}{N}\left(\frac{c_1}{2}\frac{N+2}{N} + c_2\right).$$

Similar to our baseline model, we use $W_W(A)$, $W_I(A)$, and $W_{total}(A)$ to represent wholesaler welfare, investor welfare, and total welfare, respectively, under any $A$. As we discuss earlier, the welfare outcomes under order-by-order auctions (when both $w_i$ and $l_i$ are observable) are equivalent to those when $A = 0$. We then obtain the following welfare results.

PROPOSITION 14: *For any strictly positive integer $A$, $W_{total}(A) < W_{total}(0)$, $W_W(A) < W_W(0)$, and $W_I(A) < W_I(0)$ if and only if $N(N-3) > 2\frac{c_1}{c_2}$.*

Our welfare results in Proposition 14 mirror those in Proposition 2 in our baseline model. Total welfare is always higher under order-by-order auctions due to more efficient order allocation. Wholesaler welfare is always higher under order-by-order auctions due to their more heterogeneous and accurate information. Meanwhile, investor welfare may be higher or lower under order-by-order auctions depending on the number of wholesalers and the value of $\frac{c_1}{c_2}$. These results confirm the robustness of our baseline results, which do not rely on specific assumptions about the information structure.

### C. Heterogeneous Stocks and Cross-Subsidization

In our baseline model, we consider one unit order from a single stock, and the equilibrium and welfare outcomes depend on $(c_0, c_1, c_2)$. In this section, we extend our baseline model to heterogeneous stocks with different characteristics $(c_0, c_1, c_2)$. For order-by-order auctions, this extension is straightforward, as the stock characteristics $(c_0, c_1, c_2)$ are publicly observable when market participants compete. The market equilibrium (spread, allocation, and welfare outcomes) can then still be captured by our baseline model. The extension is less straightforward, however, for brokers' routing. This is because brokers' routing features the long-term relationship between brokers and wholesalers, and thus, competition among wholesalers happens before the order arrives and the order characteristics are observed. As a result, market outcomes among heterogeneous stocks under brokers' routing will be less differentiated compared to that under order-by-order auctions. Our model predicts that compared to order-by-order auctions, there is less variation in equilibrium spreads among stocks under brokers' routing. Based on this observation, we can also highlight a cross-subsidization effect: under brokers' routing, the equilibrium spreads of high-cost stocks are relatively low (compared to that under order-by-order auctions), while the equilibrium spreads of low-cost stocks are relatively high. This cross-subsidization effect implies that switching from brokers' routing to order-by-order auctions changes not only the expected welfare of the retail investors, but also the welfare distribution when retail investors have different portfolio holdings.

To capture this idea, we consider a pool of orders characterized by a joint cumulative distribution $G(c_0, c_1, c_2)$, and the realization of $(c_0, c_1, c_2)$ is independent of all other variables in the model. For simplicity, we assume that $G$ has full support on $(0, \infty) \times (0, \infty) \times (0, \infty)$ and is continuously differentiable everywhere. We consider a model with the following timeline:

1. At time $-1$, the cumulative distribution function $G(c_0, c_1, c_2)$ becomes public information, and each wholesaler $i$ observes his private noisy signal $w_i$.
2. At time 0, an order with characteristics $(c_0, c_1, c_2)$ is drawn from distribution $G$, and each wholesaler $i$ observes his private signal $y_i$. The broker then sends the order $(c_0, c_1, c_2)$ to one wholesaler.
3. At time 1, all random variables are realized and all players collect their payoffs.

Under brokers' routing, wholesalers compete and submit their spreads at time $-1$, while under order-by-order competition, they submit their spreads at time 0.[15] We first introduce the following variables for all $j = 0, 1, 2$:

$$\bar{c}_j = \iiint c_j dG(c_0, c_1, c_2) = \mathbb{E}(c_j).$$

In order-by-order auctions, since the order characteristics $(c_0, c_1, c_2)$ are public, the equilibrium and welfare outcomes are the same as those characterized by Proposition 1 (when $p = 1$) in our baseline model. In brokers' routing, since only distributional information $G$ is available when wholesalers compete at time $-1$, the equilibrium strategy will depend on distributional information $G$ but not on the specific values of $(c_0, c_1, c_2)$. The new equilibrium of brokers' routing is characterized by the following proposition.

PROPOSITION 15: *In the extension of heterogeneous stocks, under brokers' routing, there exists an equilibrium in which any wholesaler who observes signal $w$ chooses to submit spread*

$$\bar{T}(w) = \bar{K}_0 + \bar{K}_1 w$$

*where*

$$\bar{K}_0 = \bar{c}_0 + \frac{p}{2N}\left[\bar{c}_1\left(\frac{1}{N} + \frac{N}{2} - \frac{1}{2}\right) + \bar{c}_2\right]$$

*and*

$$\bar{K}_1 = \frac{N-1}{N} p\left[\bar{c}_1\left(\frac{1}{2} + \frac{1}{N}\right) + \bar{c}_2\right].$$

Since all wholesalers are risk neutral and the equilibrium is linear in the baseline model, we still obtain a linear equilibrium in this extension. Consider $K_0(p)$ and $K_1(p)$ in Proposition 1 as functions of $(c_0, c_1, c_2)$. The equilibrium strategy in this extension satisfies

$$\bar{T}(w) = \mathbb{E}(s(w; p)) = \mathbb{E}(K_0 + K_1 w) = \bar{K}_0 + \bar{K}_1 w,$$

[15] The literature on combinatorial auctions (e.g., De Vries and Vohra (2003), Pekeč and Rothkopf (2003)) considers auctions of multiple goods with nonadditive values. In our model, the value derived from executing multiple retail orders is additive for wholesalers. Other papers (e.g., Adams and Yellen (1976), McAfee, McMillan, and Whinston (1989), Jehiel, Meyer-Ter-Vehn, and Moldovanu (2007)) examine models of multi-object auctions with additive values, and show that bundling can sometimes outperform selling multiple goods separately. Given this context, a natural question in this extension is: can the investor benefit from an optimal design of order bundling? Suppose that the broker can select a subset $B \subseteq A$, where $A$ represents the set of all possible values of $(c_0, c_1, c_2)$, and conduct two auctions. One auction includes orders with $(c_0, c_1, c_2) \in A$, while the other includes orders from $(c_0, c_1, c_2) \in A - B$. As we will see below, the investor's profit is linear in (the expected value of) $(c_0, c_1, c_2)$, implying that neither separating nor bundling orders increases the investor's profit in our extension, and thus considering optimal bundling of orders is unnecessary in this extension of brokers' routing.

where $s(w; p)$ is solved in Proposition 1. Then wholesalers choose an "average" bidding strategy in this extension. Note that both $K_0$ and $K_1$ are increasing functions of $c_0$, $c_1$, and $c_2$. This result implies that, compared to our baseline model, the equilibrium spread in this extension is relatively low for stocks with high-cost characteristics and high for stocks with low-cost characteristics.

The welfare impacts when we consider the welfare outcomes for any specific order with characteristics $(c_0, c_1, c_2)$ are also heterogeneous. The following lemma summarizes our results.

LEMMA 4: *In the equilibrium characterized by Proposition 15, investor welfare $\bar{W}^{BR}_{heter,I}$, total welfare $\bar{W}^{BR}_{heter,total}$, and wholesaler welfare $\bar{W}^{BR}_{heter,W}$ are given by*

$$\bar{W}^{BR}_{heter,I} = -\left[\bar{c}_0 + p\frac{2\bar{c}_1 - (N-3)N\bar{c}_2}{2N(1+N)}\right],$$

$$\bar{W}^{BR}_{heter,total} = -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right),$$

$$\begin{aligned}\bar{W}^{BR}_{heter,W} &= \bar{W}^{BR}_{heter,total} - \bar{W}^{BR}_{heter,I}\\ &= (\bar{c}_0 - c_0) + \frac{p}{2(N+1)}[(N-1)c_2 - (N-3)\bar{c}_2] + p\frac{\bar{c}_1}{N(1+N)}.\end{aligned}$$

Total welfare in Lemma 4 is the same as in the baseline model. Note that in our model, total welfare is determined only by inventory cost, but not the equilibrium spread, as the spread is just a transfer between wholesalers and the investor. Since the order is always obtained by the wholesaler with the lowest signal $w_i$, and introducing heterogeneity in stocks does not change allocative efficiency, we conclude that total welfare is the same as in the baseline model for any order $(c_0, c_1, c_2)$ in this extension. The equilibrium spread, however, does change. Specifically, investor welfare now becomes

$$-\left[\bar{c}_0 + p\frac{2\bar{c}_1 - (N-3)N\bar{c}_2}{2N(1+N)}\right],$$

which depends only on the average levels $(\bar{c}_0, \bar{c}_1, \bar{c}_2)$, not on the specific values of $(c_0, c_1, c_2)$. Note that under order-by-order auctions, investor welfare is

$$-\left[c_0 + \frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right],$$

which depends on order characteristics $(c_0, c_1, c_2)$. It follows that investors will be worse off after switching to order-by-order auctions if for their orders $c_0$ is high, $c_1$ is high, or $c_2$ is low (when $N > 3$). This result highlights our cross-subsidization effect under brokers' routing that wholesalers charge relatively low spreads for high-cost stocks and relatively high spreads for low-cost stocks. This cross-subsidization effect also implies that switching from brokers' routing to order-by-order auctions may have unintended effects on retail investors' welfare distribution. For example, investors who mainly trade small, illiquid stocks with high average inventory cost $c_0$ will be worse off after switching to order-by-order auctions, while those who trade large, liquid stocks with low average inventory cost $c_0$ will be better off.

Wholesaler welfare is

$$(\bar{c}_0 - c_0) + \frac{p}{2(N+1)}[(N-1)c_2 - (N-3)\bar{c}_2] + p\frac{\bar{c}_1}{N(1+N)},$$

which depends on the difference between the order characteristics $(c_0, c_1, c_2)$ and the average levels $(\bar{c}_0, \bar{c}_1, \bar{c}_2)$. Under order-by-order auctions, wholesaler welfare

$$\frac{1}{N+1}\left(\frac{c_1}{N} + c_2\right)$$

is always positive. However, under brokers' routing, wholesalers make more profit from stocks with relatively low inventory cost and incur a loss from stocks with high inventory cost, and hence, the welfare (or net profit) from executing a specific order may be negative. This cross-sectional prediction is consistent with the empirical findings of Foley et al. (2022). On average, market makers make positive expected profit from market making. Our result implies that after switching to order-by-order auctions, wholesalers will only submit spreads that are high enough such that they can earn a positive expected profit on every individual order.

## IV. Empirical Implications

Although our paper focuses primarily on the theoretical implications of order-by-order auctions—a proposal that has yet to be adopted or implemented—our model still sheds light on the current design of retail order execution and delivers testable predictions.

First, in our baseline model, we characterize the equilibrium under different signal qualities $p$. In particular, Proposition 2 implies that investor welfare may be an increasing or decreasing function of $p$ depending on the number of wholesalers $N$ and the correlation of wholesalers' inventory costs, represented by $\frac{c_1}{c_2}$. In practice, brokers use different windows to evaluate wholesalers' performance (Ernst et al. (2024)), with the evaluation window varying from 30 to 90 days. In our model, the parameter $p$ measures the accuracy with which wholesalers evaluate their future inventory costs. Intuitively, the longer the window for performance evaluation, the noisier the signal for predicting future inventory costs. As a result, our model suggests that investor welfare might increase or decrease with the length of the performance evaluation window, depending on the competitive landscape and the structure of wholesalers' inventory costs.

Second, our discussion of the endogenous entry of institutional investors shows that the entry of additional participants in market making may improve or reduce investor welfare, depending on how informed the new entrant is. The market-making business has been dominated by a few large wholesalers, but in December 2021, a new wholesaler entered the market (Dyhrberg, Shkilko, and Werner (2022)). Our model suggests that the private information of the new entrant is the key parameter affecting the change in investor welfare.

Finally, although no existing mechanism perfectly replicates the proposed order-by-order auction, several analogous auction formats exist. First, continuous auctions for retail orders operate as RLPs in the current equity market; in our companion paper Ernst, Spatt, and Sun (2024), we show that RLPs' quote interest decreases in times of volatility, in line with a low number of interested bidders in volatile markets. In options markets, Ernst and Spatt (2022) and Hendershott, Khan, and Riordan (2022) show that auctions attract few bidders, consistent with a winner's curse that prevents aggressive bidding by competitors.

## V. Conclusion

Under the current market structure, retail brokers establish relationships with market makers and send individual orders to individual market makers. While market makers are evaluated on the aggregate execution quality they deliver, there is no pretrade communication with respect to individual orders. The SEC concept for order-by-order auctions would require that each individual order be exposed in a bidding process.

Our model shows that a switch to order-by-order auctions comes with trade-offs. Allocative efficiency is improved, as order-by-order auctions ensure that an incoming retail market order is always routed to the market maker who has observed the lowest cost shock. Given the common-value nature of the auction, however, the problem of the winner's curse is more severe under order-by-order auctions, and market makers obtain higher profits in the auction relative to the brokers' routing system. Thus, retail investors can be worse off in the switch to order-by-order auctions, particularly in illiquid stocks or at times when interest in voluntary liquidity provision is low.

We also explore the role of institutional traders in order-by-order auctions and its impact on welfare outcomes and equilibrium structure. The level of information asymmetry between institutional traders and wholesalers plays a key role in determining these outcomes. In the absence of information asymmetry, switching to order-by-order auctions improves total welfare. When informational asymmetry is pronounced, however, this result is reversed, as adverse selection is too severe despite the increase in the number of bidders. Our main results are robust to settings in which institutional traders are less informed or the information structure is different. We also highlight a cross-subsidization effect among heterogeneous stocks, which implies that switching to order-by-order auctions has a heterogeneous impact on retail investors with particular holdings.

Initial submission: June 6, 2023; Accepted: May 15, 2024
Editors: Antoinette Schoar, Urban Jermann, Leonid Kogan, Jonathan Lewellen, and Thomas Philippon

## Appendix A: Extensions

*Endogenous Entry of Institutional Traders—No Adverse Selection*

In this appendix, we present the complete analysis of endogenous entry of institutional traders in the case with no adverse selection. In this case, suppose that there are $\widetilde{N}$ wholesalers and $\widetilde{N}_0$ institutional traders in equilibrium. Then, the profit for each wholesaler or institutional trader is

$$\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0}+c_2\right).$$

Since wholesalers and institutional traders have opportunity cost $\pi_W$ and $\pi_I$, respectively, and there are an unlimited supply of potential wholesalers and institutional traders, we can conclude that if $\pi_W > \pi_I$, there exists an equilibrium such that only institutional traders will operate in the market, with the equilibrium number $\widetilde{N}_0$ solved by

$$\frac{1}{\widetilde{N}_0}\frac{1}{\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}_0}+c_2\right) \geq \pi_I > \frac{1}{\widetilde{N}_0+1}\frac{1}{\widetilde{N}_0+2}\left(\frac{c_1}{\widetilde{N}_0+1}+c_2\right).$$

If $\pi_W \leq \pi_I$, there exists an equilibrium such that only wholesalers will operate in the market, and the equilibrium number $\widetilde{N}$ is solved by

$$\frac{1}{\widetilde{N}}\frac{1}{\widetilde{N}+1}\left(\frac{c_1}{\widetilde{N}}+c_2\right) \geq \pi_W > \frac{1}{\widetilde{N}+1}\frac{1}{\widetilde{N}+2}\left(\frac{c_1}{\widetilde{N}+1}+c_2\right).$$

Note that in the benchmark case of order-by-order auctions without institutional traders, the equilibrium number of wholesalers $\hat{N}$ is solved by

$$\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}}+c_2\right) \geq \pi_W > \frac{1}{\hat{N}+1}\frac{1}{\hat{N}+2}\left(\frac{c_1}{\hat{N}+1}+c_2\right).$$

In this case, the investor will get strictly higher welfare when the entry of institutional traders is allowed only if $\pi_W > \pi_I$. As we discuss in the main text, our prior is that $\pi_W < \pi_I$, and hence, allowing institutional traders to enter the market will not improve the investor's welfare.

## Appendix B: Proofs

*Proof of Proposition 1*

Consider any $i \in \{1, 2 \ldots N\}$ and $(x, w) \in \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$. Let

$$G(x|w) = \text{Prob}\left[\min_{-i} w_{-i} \leq x | w_i = w\right]$$

and $g(x|w) = \frac{dG(x|w)}{dx}$. It is easy to show that $G(x|w) = 1 - \left(\frac{1}{2} - x\right)^{N-1}$ and $g(x|w) = (N-1)\left(\frac{1}{2} - x\right)^{N-2}$. Let

$$\begin{aligned}
v(x,w) =& \mathbb{E}\left[\zeta_i |\min_{-i} w_{-i} = x; w_i = w\right] \\
=& p\mathbb{E}\left[\zeta_i |\min_{-i} w_{-i} = x; w_i = w; \exists j \neq i, w_j = y_j = x\right] \\
&+ (1-p)\mathbb{E}\left[\zeta_i |\min_{-i} w_{-i} = x; w_i = w; \nexists j \neq i, w_j = y_j = x\right] \\
=& p\left[c_0 + c_1\left(\frac{x + (N-2)p\frac{\frac{1}{2}+x}{2} + pw}{N}\right) + c_2 pw\right] \\
&+ (1-p)\left[c_0 + c_1\left(\frac{(N-2)p\frac{\frac{1}{2}+x}{2} + pw}{N}\right) + c_2 pw\right] \\
=& c_0 + c_1\left(\frac{px + (N-2)p\frac{\frac{1}{2}+x}{2} + pw}{N}\right) + c_2 pw.
\end{aligned}$$

We focus on symmetric equilibria. Suppose that all of wholesaler $i$'s opponents use a continuous increasing bidding strategy $B(w) = K_0 + K_1 w$ at time 0. When wholesaler $i$ observes signal $w$ and reports signal $z$, its expected profit is

$$\begin{aligned}
U_i(z,w) &= \mathrm{Prob}\left(z \leq \min_{-i} w_{-i} | w_i = w\right)\left[B(z) - \mathbb{E}\left(\zeta_i | z \leq \min_{-i} w_{-i}, w_i = w\right)\right] \\
&= \left[1 - G(z|w)\right]\left[B(z) - \frac{1}{1 - G(z|w)}\int_z^{\frac{1}{2}} g(x|w)v(x,w)dx\right] \\
&= \left[1 - G(z|w)\right]B(z) - \int_z^{\frac{1}{2}} g(x|w)v(x,w)dx.
\end{aligned}$$

Wholesaler $i$'s marginal incentive is characterized by

$$\begin{aligned}
\frac{\partial U_i(z,w)}{\partial z} &= -g(z|w)B(z) + \left(1 - G(z|w)\right)B'(z) + g(z|w)v(z,w) \\
&= g(z|w)\left[-B(z) + \left(\frac{1 - G(z|w)}{g(z|w)}\right)B'(z) + v(z,w)\right] \\
&= g(z|w)\left[\begin{array}{c} -K_0 - K_1 z + \frac{\frac{1}{2}-z}{N-1}K_1 + c_0 \\ +c_1\left(\frac{pz + (N-2)p\frac{\frac{1}{2}+z}{2} + pw}{N}\right) + c_2 pw \end{array}\right].
\end{aligned}$$

We conjecture that in equilibrium, we have

$$\left.\frac{\partial U_i(z,w)}{\partial z}\right|_{z=w} = 0. \tag{B1}$$

This implies

$$-(K_0 + K_1 w) + \frac{\frac{1}{2} - w}{N-1}K_1 + c_0 + c_1\left(\frac{pw + (N-2)p\frac{\frac{1}{2}+w}{2} + pw}{N}\right) + c_2 pw = 0.$$

Since the above condition holds for all $w$, we have that $K_0, K_1$ are solved by

$$-K_0 + \frac{\frac{1}{2}K_1}{N-1} + c_0 + c_1\frac{N-2}{4N}p = 0,$$

$$-K_1 - \frac{K_1}{N-1} + c_2 p + \frac{2c_1 p}{N} + \frac{c_1(N-2)p}{2N} = 0.$$

We then get

$$K_1 = \frac{N-1}{N}p\left(c_1\left(\frac{1}{2} + \frac{1}{N}\right) + c_2\right), \tag{B2}$$

$$K_0 = c_0 + \frac{p}{2N}\left[c_1\left(\frac{N}{2} - \frac{1}{2} + \frac{1}{N}\right) + c_2\right]. \tag{B3}$$

We also need to verify that condition (B1) is a sufficient condition for optimization. Note that $g(z|w) > 0$ and

$$-K_0 - K_1 z + \frac{\frac{1}{2} - z}{N-1}K_1 + c_0 + c_1\left(\frac{pz + (N-2)p\frac{\frac{1}{2}+z}{2} + pw}{N}\right) + c_2 pw$$

is linear in $z$. It is then clear that with (B2) and (B3), for any $w$, $\frac{\partial U_i(z,w)}{\partial z} < 0 \iff z > w$, confirming that (B1) is a sufficient condition for optimization.

*Proof of Lemma 1*

We introduce the random variable $r = \min_i w_i \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. Its cumulative distribution function (CDF) is $H(r) = 1 - \left(\frac{1}{2} - r\right)^N$, and its probability density function (PDF) is $h(r) = N\left(\frac{1}{2} - r\right)^{N-1}$. Then, the total expected profit of wholesalers is

$$
\begin{aligned}
W_W =& \mathbb{E}\left\{\mathbb{E}\left[s(r;p) - c_0 - c_1\frac{1}{N}\sum_{j=1}^{N} y_j - c_2 y_i | w_i = \min_j w_j = r\right]\right\} \\
=& \mathbb{E}\left\{K_0 + K_1 r - c_0 - c_1\left(\frac{pr + (N-1)p\left(\frac{1}{2}+r\right)\frac{1}{2}}{N}\right) - c_2 pr\right\} \\
=& \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{p}{N}(c_1 + c_2 N)\left(\frac{1}{2} - r\right)^N dr \\
=& \frac{p(c_1 + Nc_2)}{N(1+N)},
\end{aligned}
$$

the expected total profit of investors $W_I$ is

$$
\begin{aligned}
W_I &= -\mathbb{E}\left[\mathbb{E}\left[K_0 + K_1 r | \min_i w_i = r\right]\right] \\
&= -\int_{-\frac{1}{2}}^{\frac{1}{2}} (K_0 + K_1 r) N\left(\frac{1}{2} - r\right)^{N-1} dr \\
&= -\left[c_0 + p\frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right],
\end{aligned}
$$

and total welfare $W_{total}$ is

$$
\begin{aligned}
W_{total} &= \mathbb{E}\left\{\mathbb{E}\left[-c_0 - c_1\frac{1}{N}\sum_{j=1}^{N} y_j - c_2 y_i | w_i = \min_j w_j = r\right]\right\} \\
&= -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right).
\end{aligned}
$$

*Proof of Proposition 2*

Both $W_{total}(p) < W_{total}(1)$ and $W_W(p) < W_W(1)$ are obvious. For the last result,

$$
W_I(p) < W_I(1) \Longleftrightarrow N(N-3)\frac{c_2}{c_1} > 2.
$$

*Proof of Proposition 4*

When $\delta_c = 0$, institutional traders and wholesalers receive i.i.d. signals, and they are symmetric. We first conjecture that all market makers choose the same linear equilibrium strategy

$$
\tilde{\beta}_i(y_i; \delta_c = 0) = \tilde{k}_0(\delta_c = 0) + \tilde{k}_1(\delta_c = 0) y_i.
$$

The total number of wholesalers and institutional traders is $N+N_0$. Following the proof of Proposition 1, we have

$$
\begin{aligned}
\tilde{v}(x,y) &= \mathbb{E}\left[\tilde{\xi}_i \,|\, \min_{-i} y_{-i} = x, y_i = y\right] \\
&= c_0 + c_1 \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N+N_0} y_j \,|\, \min_{-i} y_{-i} = x, y_i = y\right] + c_2 \mathbb{E}\left[y_i \,|\, \min_{-i} y_{-i} = x, y_i = y\right] \\
&= c_0 + \left(\frac{c_1}{N+N_0} + c_2\right) y + \frac{c_1}{N+N_0} x + c_1 \frac{N+N_0-2}{N+N_0}\frac{1}{2}\left(\frac{1}{2} + x\right) \\
&= \left(c_0 + c_1 \frac{N+N_0-2}{4(N+N_0)}\right) + \frac{c_1}{2} x + \left(\frac{c_1}{N+N_0} + c_2\right) y,
\end{aligned}
$$

which is the $v(x,y)$ function in Proposition 1 with $(N+N_0)$ wholesalers and $p=1$. The rest of the proof follows the proof of Proposition 1, and we can show that the equilibrium strategy is equivalent to that in Proposition 1 with the number of wholesalers being $N+N_0$ and $p=1$.

*Proof of Proposition 5*

First, we have $\widetilde{W}^{OBO}_{total} = -c_0 + \frac{N+N_0-1}{N+N_0+1}\frac{c_2}{2}$, so it is clear that $\widetilde{W}^{OBO}_{total}$ is an increasing function of $N_0$. Second, $\widetilde{W}^{OBO}_{W} = \frac{N}{N+N_0}\frac{(c_1+(N+N_0)c_2)}{(N+N_0)(1+N+N_0)}$, and it is clear that $\widetilde{W}^{OBO}_{W}$ is a decreasing function of $N_0$. Finally,

$$
\begin{aligned}
\widetilde{W}^{OBO}_{I} &= -\left[c_0 + \frac{c_1}{(N+N_0)(N+N_0+1)} - \frac{N+N_0-3}{2(N+N_0+1)} c_2\right] \\
&= -c_0 - \frac{c_1}{(N+N_0)(N+N_0+1)} + \frac{N+N_0+1-4}{N+N_0+1}\frac{c_2}{2} \\
&= -c_0 + \frac{c_2}{2} - \frac{c_1}{(N+N_0)(N+N_0+1)} - \frac{4}{N+N_0+1}\frac{c_2}{2},
\end{aligned}
$$

and it is clear that $\widetilde{W}^{OBO}_{I}$ is an increasing function of $N_0$.

*Proof of Proposition 6*

Since the equilibrium of brokers' routing is the same as that in the baseline model, we have $\widetilde{W}^{BR}_{total} = -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right)$, $\widetilde{W}^{BR}_{I} = -\left[c_0 + p\frac{2c_1-(N-3)Nc_2}{2N(1+N)}\right]$, and $\widetilde{W}^{BR}_{W} = \frac{p(c_1+Nc_2)}{N(1+N)}$. For order-by-order auctions, the welfare outcomes are $\widetilde{W}^{OBO}_{total} = -\left(c_0 - \frac{N+N_0-1}{N+N_0+1}\frac{c_2}{2}\right)$, $\widetilde{W}^{OBO}_{I} = -\left[c_0 + \frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2\right]$, and $\widetilde{W}^{OBO}_{W} = \frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{c_1}{N+N_0} + c_2\right)$. First, it is obvious that $\widetilde{W}^{OBO}_{total} > \widetilde{W}^{BR}_{total}$, because $p \in (0,1)$ and $N_0 \geq 2$. Second, regarding the comparison between $\widetilde{W}^{BR}_{I}$

and $\widetilde{W}_I^{OBO}$,

$$
\begin{aligned}
&\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO} \\
\iff & p\frac{2c_1-(N-3)Nc_2}{2N(1+N)} > \frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2 \\
\iff & p\frac{2-(N-3)N\frac{c_2}{c_1}}{2N(1+N)} > \frac{1}{(N+N_0)(N+N_0+1)} - \frac{N+N_0-3}{2(N+N_0+1)}\frac{c_2}{c_1} \\
\iff & \left(\frac{N+N_0-3}{2(N+N_0+1)} - p\frac{(N-3)}{2(1+N)}\right)\frac{c_2}{c_1} > \frac{1}{(N+N_0)(N+N_0+1)} - \frac{p}{N(1+N)}.
\end{aligned} \tag{B4}
$$

We always have $\frac{N+N_0-3}{2(N+N_0+1)} - p\frac{(N-3)}{2(1+N)} > 0$ for any $N \geq 2$, $N_0 \geq 2$, and $p \in (0,1)$. Let $n_0 > 0$ be the solution to $\frac{1}{(N+n_0)(N+n_0+1)} - \frac{p}{N(1+N)} = 0$, which is equivalent to $(N+n_0)(N+n_0+1) = \frac{N(1+N)}{p}$. Since $p < 1$, $n_0$ exists and is unique.

If $N_0 \geq n_0$, condition (B4) always holds, and thus, we always have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$. If $N_0 < n_0$, then condition (B4) holds if and only if

$$
\frac{c_1}{c_2} < \alpha_0 = \frac{\frac{N+N_0-3}{2(N+N_0+1)} - p\frac{(N-3)}{2(1+N)}}{\frac{1}{(N+N_0)(N+N_0+1)} - p\frac{1}{N(1+N)}}. \tag{B5}
$$

Finally, regarding the comparison between $\widetilde{W}_W^{BR}$ and $\widetilde{W}_W^{OBO}$,

$$
\begin{aligned}
&\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO} \\
\iff & \frac{p(c_1+Nc_2)}{N(1+N)} < \frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{c_1}{N+N_0}+c_2\right) \\
\iff & \frac{p\left(1+N\frac{c_2}{c_1}\right)}{N(1+N)} < \frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{1}{N+N_0}+\frac{c_2}{c_1}\right) \\
\iff & \frac{p}{N(1+N)} + \frac{p}{1+N}\frac{c_2}{c_1} < \frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{1}{N+N_0}+\frac{c_2}{c_1}\right) \\
\iff & \left(\frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p}{1+N}\right)\frac{c_2}{c_1} > \\
& -\left(\frac{N}{(N+N_0)^2}\frac{1}{N+N_0+1} - \frac{p}{N(1+N)}\right) \\
\iff & \left(\frac{1}{(N+N_0)(N+N_0+1)} - \frac{p}{N(1+N)}\right)N\frac{c_2}{c_1} > \\
& -\frac{\left(\frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p(N+N_0)}{(1+N)N}\right)}{N+N_0}.
\end{aligned} \tag{B6}
$$

Let $n_1 > 0$ be the solution of

$$\frac{N}{N+n_1}\frac{1}{N+n_1+1} - \frac{p}{(1+N)}\frac{(N+n_1)}{N} = 0,$$

which is equivalent to $(N+n_1)^2(N+n_1+1) = \frac{N^2(1+N)}{p}$. Since $p < 1$, we know $n_1$ exists and is unique. Note that $n_0$ satisfies $\frac{1}{(N+n_0)(N+n_0+1)} - \frac{p}{N(1+N)} = 0$, which is equivalent to $(N+n_0)(N+n_0+1) = \frac{N(1+N)}{p}$. It is clear that $n_0 > n_1$.

If $N_0 \leq n_1$, then

$$\frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p}{1+N} > \frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p}{(1+N)}\frac{(N+N_0)}{N} \geq 0,$$

in which case condition (B6) always holds and we always have $\widetilde{W}_W^{BR} < \widetilde{W}_W^{OBO}$.

If $N_0 \geq n_0$, we know

$$0 \geq \frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p}{1+N} > \frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p}{(1+N)}\frac{(N+N_0)}{N},$$

in which case condition (B6) never holds and we always have $\widetilde{W}_W^{BR} > \widetilde{W}_W^{OBO}$.

If $N_0 \in (n_1, n_0)$, condition (B6) holds if and only if

$$\frac{c_1}{c_2} < \alpha_1 = -\frac{\frac{N}{N+N_0+1} - \frac{p(N+N_0)}{1+N}}{\frac{N}{N+N_0}\frac{1}{N+N_0+1} - \frac{p(N+N_0)}{(1+N)N}}, \tag{B7}$$

where $\alpha_1$ must be positive.

*Proof of Proposition 7*

Let $\delta_{c1}$ be the value that satisfies $\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} = \widetilde{k}_0^+ - \widetilde{k}_1^+ \cdot \frac{1}{2}$, that is,

$$c_0 - \delta_{c1} + \frac{c_1}{4N_0}\left(N_0 - 1 + \frac{2}{N_0}\right) + \frac{c_2}{2N_0} + \frac{N_0-1}{N_0}\left(\frac{c_1}{2}\frac{N_0+2}{N_0} + c_2\right)\frac{1}{2} \tag{B8}$$

$$= c_0 + \delta_{c1} + \frac{c_1\left(N+N_0-1+\frac{2}{N+N_0}\right)}{4(N+N_0)} + \frac{c_2}{2(N+N_0)}$$

$$-\frac{N+N_0-1}{N+N_0}\left(\frac{c_1}{2}\frac{N+N_0+2}{N+N_0} + c_2\right)\frac{1}{2}. \tag{B9}$$

Then, when $\delta_c > \delta_{c1}$, $\left[\widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}, \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}\right] \cup \left[\widetilde{k}_0^+ - \widetilde{k}_1^+ \cdot \frac{1}{2}, \widetilde{k}_0^+ + \widetilde{k}_1^+ \cdot \frac{1}{2}\right] = \emptyset$. This implies that under the equilibrium conjectured, when $\widetilde{c}_0 = c_0 - \delta_c$, only institutional traders can obtain the order no matter what signals market participants observe.

We first verify that it is optimal for any institutional trader to choose $\widetilde{s}^-(y;\delta_c)$ if observing $c_0 - \delta_c$, given other market participants' strategies. When

$\widetilde{c}_0 = c_0 - \delta_c$, then clearly $\widetilde{s}^-(y;\delta_c)$ is an equilibrium if we only have $N_0$ institutional traders in the market, as suggested by Proposition 1. This is essentially the baseline model of order-by-order auctions with $N_0$ bidders and unconditional expected inventory cost being $c_0 - \delta_c$. This means that it is optimal for any institutional trader to choose $\widetilde{s}^-(y;\delta_c)$ if there are only $N_0 - 1$ other institutional traders who also choose $\widetilde{s}^-(y;\delta_c)$ and no wholesalers in the market. Having $N$ wholesalers choose $\widetilde{s}^+(y;\delta_c)$ does not change this optimality, because given other institutional traders' choice $\widetilde{s}^-(y;\delta_c)$, the $N$ wholesalers will never obtain any order in any state when $\widetilde{c}_0 = c_0 - \delta_c$.

We next verify that it is optimal for any institutional trader to choose $\widetilde{s}^+(y;\delta_c)$ if observing $c_0 + \delta_c$, given other market participants' strategies. Following our Proposition 1 in the baseline model of order-by-order auctions ($p = 1$), $\widetilde{s}^+(y;\delta_c)$ is an equilibrium with $N + N_0$ wholesalers and unconditional expected inventory cost $c_0 + \delta_c$. It is therefore optimal for any institutional trader to choose $\widetilde{s}^+(y;\delta_c)$ if there are $N + N_0 - 1$ other wholesalers that also choose $\widetilde{s}^+(y;\delta_c)$.

Finally, we verify that it is optimal for any wholesaler $i$ to choose $\widetilde{s}^+(y;\delta_c)$, given other market participants' strategies. Suppose that wholesaler $i$ observes a cost shock $y_i$. Then, the wholesaler's utility is $U_i = \frac{1}{2}U_1(\widetilde{s}_i) + \frac{1}{2}U_2(\widetilde{s}_i)$, where $U_1$ ($U_2$) is wholesaler $i$'s profit when the state is $c_0 - \delta_c$ ($c_0 + \delta_c$). It is clear that for any $y_i$, we have $\widetilde{s}^+(y_i;\delta_c) = \arg\max_{\widetilde{s}_i \in \left(\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}, \infty\right)} U_i$. This is because when $\widetilde{s}_i \in \left(\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}, \infty\right)$, $\left[\widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}, \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}\right] \cup \left[\widetilde{k}_0^+ - \widetilde{k}_1^+ \cdot \frac{1}{2}, \widetilde{k}_0^+ + \widetilde{k}_1^+ \cdot \frac{1}{2}\right] = \emptyset$, and thus, we always have $U_1(\widetilde{s}_i) = 0$. Further, by definition, $\widetilde{s}^+(y_i;\delta_c) = \arg\max_{\widetilde{s}_i \in \left(\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}, \infty\right)} U_2$. It is also clear that the wholesaler will never choose $\widetilde{s}_i < \widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}$, as any $\widetilde{s}_i < \widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}$ is dominated by $\widetilde{s}_i = \widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}$. Suppose that the wholesaler chooses $\widetilde{s}_i \in \left[\widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}, \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}\right]$. Note that $\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[(c_0 - \delta_c) - \frac{c_1+c_2}{2}\right] > 0$, and the upper bound of the profit in the case $c_0 - \delta_c$ is $\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[(c_0 - \delta_c) - \frac{c_1+c_2}{2}\right]$, because $\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}$ is the highest spread in $\widetilde{s}_i \in \left[\widetilde{k}_0^- - \widetilde{k}_1^- \cdot \frac{1}{2}, \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2}\right]$ and $(c_0 - \delta_c) - \frac{c_1+c_2}{2}$ is the lowest inventory cost. In the case $c_0 + \delta_c$, wholesaler $i$ will obtain the order with probability one, and an upper bound of the profit is $\widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[(c_0 + \delta_c) - \frac{c_1+c_2}{2}\right]$. Then, $U_1 \leq \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[(c_0 - \delta_c) - \frac{c_1+c_2}{2}\right]$ and $U_2 \leq \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[(c_0 + \delta_c) - \frac{c_1+c_2}{2}\right]$. This implies $U_i \leq \frac{1}{2}U_1 + \frac{1}{2}U_2 \leq \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[c_0 - \frac{c_1+c_2}{2}\right]$. Hence,

$$U_i < 0 \Longleftarrow \widetilde{k}_0^- + \widetilde{k}_1^- \cdot \frac{1}{2} - \left[c_0 - \frac{c_1 + c_2}{2}\right] < 0 \Longleftrightarrow \delta_c > \delta_{c2},$$

where

$$\delta_{c2} = \frac{c_1}{4N_0}\left(N_0 - 1 + \frac{2}{N_0}\right) + \frac{c_2}{2N_0} + \widetilde{k}_1^- \cdot \frac{1}{2} + \frac{c_1 + c_2}{2}. \tag{B10}$$

Let $\underline{\delta} = \max\{\delta_{c1}, \delta_{c2}\}$, where $\delta_{c1}$ is defined by (B9) and $\delta_{c2}$ is defined by (B10). Then, when $\delta_c > \underline{\delta}$, we have $\widetilde{s}^+(y_i; \delta_c) = \widetilde{k}_0^+ + \widetilde{k}_1^+ y_i = \arg\max_{\widetilde{s}_i \in (-\infty,\infty)} U_i$ for any $y_i$, which implies that it is optimal for any wholesaler $i$ to choose $\widetilde{s}^+(y; \delta_c)$, given other market participants' strategies.

*Proof of Proposition 8*

Since the equilibrium of brokers' routing is the same as that in the baseline model, we have

$$\widetilde{W}_{total}^{BR} = -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right),$$

$$\widetilde{W}_I^{BR} = -\left[c_0 + p\frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right],$$

and

$$\widetilde{W}_W^{BR} = \frac{p(c_1 + Nc_2)}{N(1+N)}.$$

For order-by-order auctions, the welfare outcomes are

$$\begin{aligned}\widetilde{W}_{total}^{OBO} &= -\frac{1}{2}\left(c_0 + \delta_c - \frac{N+N_0-1}{N+N_0+1}\frac{c_2}{2}\right) - \frac{1}{2}\left(c_0 - \delta_c - \frac{N_0-1}{N_0+1}\frac{c_2}{2}\right)\\ &= -c_0 + \frac{1}{2}\left(\frac{N+N_0-1}{N+N_0+1} + \frac{N_0-1}{N_0+1}\right)\frac{c_2}{2},\end{aligned}$$

$$\begin{aligned}\widetilde{W}_I^{OBO} &= -\frac{1}{2}\left[c_0 + \delta_c + \frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2\right]\\ &\quad -\frac{1}{2}\left[c_0 - \delta_c + \frac{1}{N_0(N_0+1)}c_1 - \frac{N_0-3}{2(N_0+1)}c_2\right]\\ &= -c_0 - \frac{1}{2}\left[\frac{\frac{c_1}{(N+N_0)}}{(N+N_0+1)} - \frac{(N+N_0-3)c_2}{2(N+N_0+1)} + \frac{c_1}{N_0(N_0+1)} - \frac{(N_0-3)c_2}{2(N_0+1)}\right],\end{aligned}$$

and

$$\widetilde{W}_W^{OBO} = \frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{c_1}{N+N_0} + c_2\right).$$

First,

$$\begin{aligned}\widetilde{W}_{total}^{OBO} > \widetilde{W}_{total}^{BR} &\Longleftrightarrow -c_0 + \frac{1}{2}\left(\frac{N+N_0-1}{N+N_0+1} + \frac{N_0-1}{N_0+1}\right)\frac{c_2}{2} > -\left(c_0 - p\frac{N-1}{N+1}\frac{c_2}{2}\right)\\ &\Longleftrightarrow \frac{1}{2}\left(\frac{N+N_0-1}{N+N_0+1} + \frac{N_0-1}{N_0+1}\right)\frac{c_2}{2} > p\frac{N-1}{N+1}\frac{c_2}{2}.\end{aligned}$$

The LHS of the above condition is increasing in $N_0$. Let $t^*$ be the maximum of zero and the solution to

$$\frac{1}{2}\left(\frac{N+t^*-1}{N+t^*+1}+\frac{t^*-1}{t^*+1}\right)=p\frac{N-1}{N+1}. \tag{B11}$$

It follows that the LHS is an increasing function of $t^*$, and hence, there must exist a unique solution of $t^*$. We therefore have $N_0 > t^* \iff \widetilde{W}^{OBO}_{total} > \widetilde{W}^{BR}_{total}$.

Second,

$$\begin{aligned}
&\widetilde{W}^{BR}_W < \widetilde{W}^{OBO}_W\\
&\iff \frac{p(c_1+Nc_2)}{N(1+N)} < \frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{c_1}{N+N_0}+c_2\right)\\
&\iff \frac{p\left(1+N\frac{c_2}{c_1}\right)}{N(1+N)} < \frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{1}{N+N_0}+\frac{c_2}{c_1}\right)\\
&\iff \frac{p}{N(1+N)}+\frac{p}{1+N}\frac{c_2}{c_1} < \frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}\left(\frac{1}{N+N_0}+\frac{c_2}{c_1}\right)\\
&\iff \left(\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{1+N}\right)\frac{c_2}{c_1} >\\
&-\left(\frac{1}{2}\frac{N}{(N+N_0)^2}\frac{1}{N+N_0+1}-\frac{p}{N(1+N)}\right)\\
&\iff \left(\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{1+N}\right)\frac{c_2}{c_1} >\\
&-\frac{\left(\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p(N+N_0)}{(1+N)N}\right)}{N+N_0}.
\end{aligned} \tag{B12}$$

Note that $\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{1+N} > \frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{(1+N)}\frac{(N+N_0)}{N}$. Let $t_0$ be the maximum of zero and the solution to $\frac{1}{2}\frac{N}{N+t_0}\frac{1}{N+t_0+1}-\frac{p}{(1+N)}\frac{(N+t_0)}{N}=0$, and let $t_1$ be the maximum of zero and the solution to $\frac{1}{2}\frac{N}{N+t_1}\frac{1}{N+t_1+1}-\frac{p}{1+N}=0$. It is clear that we must have $t_0 \le t_1$.

If $N_0 \le t_0$ and $N_0 \ge 2$ (if this is not an empty set), we must have $\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{(1+N)}\frac{(N+N_0)}{N} \ge 0$ and $\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{1+N} > 0$, in which case condition (B12) is always satisfied. We must therefore have $\widetilde{W}^{BR}_W < \widetilde{W}^{OBO}_W$.

If $N_0 \ge t_1$, then $\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{1+N} \le 0$ and $\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{(1+N)}\frac{(N+N_0)}{N} < 0$. We must therefore have $\widetilde{W}^{BR}_W > \widetilde{W}^{OBO}_W$.

If $t_0 < N_0 < t_1$, then we know $\widetilde{W}^{BR}_W < \widetilde{W}^{OBO}_W$ if and only if

$$\frac{c_1}{c_2} < \alpha_2 = -\frac{\frac{1}{2}\frac{N}{N+N_0+1}-\frac{p(N+N_0)}{1+N}}{\frac{1}{2}\frac{N}{N+N_0}\frac{1}{N+N_0+1}-\frac{p}{(1+N)}\frac{(N+N_0)}{N}}. \tag{B13}$$

Finally,

$$\begin{aligned}
&\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO} \\
&\Longleftrightarrow -\left[c_0 + p\frac{2c_1 - (N-3)Nc_2}{2N(1+N)}\right] < -c_0 - \frac{1}{2}\left[\begin{array}{c} \frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2 \\ +\frac{1}{N_0(N_0+1)}c_1 - \frac{N_0-3}{2(N_0+1)}c_2 \end{array}\right] \\
&\Longleftrightarrow \left(\begin{array}{c} \frac{1}{(N+N_0)(N+N_0+1)} \\ +\frac{1}{N_0(N_0+1)} - \frac{2p}{N(1+N)} \end{array}\right)\frac{c_1}{c_2} < \frac{1}{2}\left(\frac{N+N_0-3}{N+N_0+1} + \frac{N_0-3}{N_0+1} - 2p\frac{N-3}{N+1}\right). \quad \text{(B14)}
\end{aligned}$$

Let $t_3$ be the unique positive solution of[16]

$$\frac{1}{(N+t_3)(N+t_3+1)} + \frac{1}{t_3(t_3+1)} - \frac{2p}{N(1+N)} = 0,$$

so $N_0 \geq t_3 \Longleftrightarrow \frac{1}{(N+N_0)(N+N_0+1)} + \frac{1}{N_0(N_0+1)} - \frac{2p}{N(1+N)} \leq 0$. Let $t_2$ be the maximum of zero and the solution to

$$\frac{N+t_2-3}{N+t_2+1} + \frac{t_2-3}{t_2+1} - 2p\frac{N-3}{N+1} = 0,$$

so $N_0 \geq t_2 \Longleftrightarrow \frac{N+N_0-3}{N+N_0+1} + \frac{N_0-3}{N_0+1} - 2p\frac{N-3}{N+1} \geq 0$. Let $\alpha_3$ be given by

$$\alpha_3 = \frac{1}{2}\frac{\frac{N+N_0-3}{N+N_0+1} + \frac{N_0-3}{N_0+1} - 2p\frac{N-3}{N+1}}{\frac{1}{(N+N_0)(N+N_0+1)} + \frac{1}{N_0(N_0+1)} - \frac{2p}{N(1+N)}}. \quad \text{(B15)}$$

LEMMA 5: *For any $N \geq 2$ and $p \in [0, 1]$, we must have $t_3 > t_2$.*

PROOF: If $t_2 = 0$, this result must be true as we always have $t_3 > 0$. When $t_2 > 0$, note that $t_2$ and $t_3$ satisfy $\frac{N+t_2-3}{N+t_2+1} + \frac{t_2-3}{t_2+1} - 2p\frac{N-3}{N+1} = 0$ and $\frac{1}{(N+t_3)(N+t_3+1)} + \frac{1}{t_3(t_3+1)} - \frac{2p}{N(1+N)} = 0$. When $N = 2$, $\frac{t_2-1}{t_2+3} + \frac{t_2-3}{t_2+1} + 2p\frac{1}{3} = 0$ and $\frac{1}{(2+t_3)(3+t_3)} + \frac{1}{t_3(t_3+1)} - \frac{2p}{6} = 0$. These two conditions are equivalent to

$$\frac{t_2-1}{t_2+3} + \frac{t_2-3}{t_2+1} + 2p\frac{1}{3} = 0 \Longleftrightarrow 2 + \frac{2}{3}p = \frac{4}{t_2+3} + \frac{4}{t_2+1},$$

and

$$\frac{1}{(2+t_3)(3+t_3)} + \frac{1}{t_3(t_3+1)} - \frac{2p}{6} = 0 \Longleftrightarrow \frac{1}{(t_3+2)(t_3+3)} + \frac{1}{t_3(t_3+1)} = \frac{p}{3}.$$

We can verify that for any $p \in [0, 1]$, we must have $t_3 > t_2$, which can be solved from the two conditions above.

[16] Note that we assume $N \geq 2$.

When $N \geq 3$, note that

$$\begin{aligned}
&\frac{N+t_2-3}{N+t_2+1}+\frac{t_2-3}{t_2+1}-2p\frac{N-3}{N+1}=0\\
\Longleftrightarrow &1-\frac{4}{N+t_2+1}+1-\frac{4}{t_2+1}-2p\left(1-\frac{4}{N+1}\right)=0\\
\Longleftrightarrow &\frac{1}{N+t_2+1}+\frac{1}{t_2+1}-\frac{2}{N+1}=\frac{1-p}{2}\left(1-\frac{4}{N+1}\right)\\
\Longleftrightarrow &\frac{N+N^2-2t_2-2t_2^2}{(N+t_2+1)(t_2+1)(N+1)}=\frac{1-p}{2}\left(1-\frac{4}{N+1}\right),
\end{aligned}$$

and hence, we must have

$$N+N^2-2t_2-2t_2^2 \geq 0.$$

In addition, we have

$$\begin{aligned}
&\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2p}{N(1+N)}\\
=&\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2}{N(1+N)}+\frac{2}{N(1+N)}-\frac{2p}{N(1+N)}\\
=&\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2}{N(1+N)}+\frac{2(1-p)}{N(1+N)}\\
=&\frac{\left(N+N^2-2N_0-2N_0^2\right)\left(N^2+2N_0N+2N_0^2+N+2N_0\right)+2N_0^2\left(1+N_0^2\right)}{(N+N_0)(N+N_0+1)N_0(N_0+1)N(1+N)}\\
&+\frac{2(1-p)}{N(1+N)},
\end{aligned}$$

and hence, we must have

$$\left.\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2p}{N(1+N)}\right|_{N_0=t_2}>0.$$

Note that

$$\left.\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2p}{N(1+N)}\right|_{N_0=t_3}=0.$$

We therefore conclude that $t_3 > t_2$. □

When $N_0 \geq t_3$, $\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2p}{N(1+N)} \leq 0$ and $\frac{N+N_0-3}{N+N_0+1}+\frac{N_0-3}{N_0+1}-2p\frac{N-3}{N+1}>0$, and thus, we must have $\widetilde{W}_I^{BR}<\widetilde{W}_I^{OBO}$; when $N_0 \leq t_2$, $\frac{1}{(N+N_0)(N+N_0+1)}+\frac{1}{N_0(N_0+1)}-\frac{2p}{N(1+N)}>0$ and $\frac{N+N_0-3}{N+N_0+1}+\frac{N_0-3}{N_0+1}-2p\frac{N-3}{N+1} \leq 0$, and thus, we must have $\widetilde{W}_I^{BR} > \widetilde{W}_I^{OBO}$; and when $t_2 < N_0 < t_3$, $\frac{1}{(N+N_0)(N+N_0+1)} + \frac{1}{N_0(N_0+1)} - \frac{2p}{N(1+N)} > 0$ and $\frac{N+N_0-3}{N+N_0+1} + \frac{N_0-3}{N_0+1} - 2p\frac{N-3}{N+1} > 0$, and thus, $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$ if and only if $\frac{c_1}{c_2} < \alpha_3$.

In summary, we have the following results: if $N_0 \geq t_3$, we always have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO}$; if $N_0 \leq t_2$, we have $\widetilde{W}_I^{BR} > \widetilde{W}_I^{OBO}$; and if $t_2 < N_0 < t_3$, we have $\widetilde{W}_I^{BR} < \widetilde{W}_I^{OBO} \Longleftrightarrow \frac{c_1}{c_2} < \alpha_3$.

*Proof of Lemma 2*

If $t_0 = 0$, this result must be true as we always have $t_3 > 0$. Now consider the case when $t_0 > 0$: $t_3$ solves $\frac{N}{(N+t_3)(N+t_3+1)} + \frac{N}{t_3(t_3+1)} = \frac{2p}{(1+N)}$, and $t_0$ solves $\frac{N}{N+t_0}\frac{1}{N+t_0+1} = \frac{2p}{(1+N)}\frac{(N+t_0)}{N}$. Since $t_0 > 0$, then we must have $t_3 > t_0$ because $\frac{1}{t_3(t_3+1)} > 0$ and $\frac{2p}{(1+N)}\frac{(N+t_0)}{N} > \frac{2p}{(1+N)}$.

First, since $t_1$ is the maximum of zero and the solution to $\frac{N}{N+t_1}\frac{1}{N+t_1+1} = \frac{2p}{1+N}$, it is clear that $t_1$ is a weakly decreasing function of $p$. Similarly, since $t_2$ is the maximum of zero and the solution to $\frac{N+t_2-3}{N+t_2+1} + \frac{t_2-3}{t_2+1} - 2p\frac{N-3}{N+1} = 0$, it is clear that $t_2$ is a weakly increasing function of $p$. Notice that when $p \to 0$, we have $t_1 \to \infty$ and $t_2 = 0$. Then, there must exist a threshold $p^*$ such that $t_1 > t_2$ if and only if $p < p^*$.

*Proof of Proposition 9*

Let $M_0$ be the unique integer that satisfies

$$\frac{1}{2}\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0} + c_2\right) \geq \pi_W > \frac{1}{2}\frac{1}{1+M_0}\frac{1}{1+M_0+1}\left(\frac{c_1}{1+M_0} + c_2\right),$$

where $M_0$ is the total number of wholesalers and institutional traders, if there is at least one wholesaler and two institutional traders in equilibrium. Our Assumption 2 implies that $M_0 \geq 3$.

First, consider the case in which institutional traders are banned. Based on our results in the baseline model, conjecture that $\hat{N} \geq 2$. We know that the expected profit for each of the $\hat{N}$ wholesalers is $\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}} + c_2\right)$. Then, in equilibrium, $\hat{N}$ satisfies

$$\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}} + c_2\right) \geq \pi_W > \frac{1}{\hat{N}+1}\frac{1}{\hat{N}+2}\left(\frac{c_1}{\hat{N}+1} + c_2\right). \quad \text{(B16)}$$

Assumption 2 implies that $\hat{N} \geq 3$, which is consistent with our conjecture for this solution.

When the entry of institutional traders is allowed, we focus on equilibria in which at least two institutional traders operate, that is, $\widetilde{N}_0 \geq 2$.

If in equilibrium, $\widetilde{N}=0$ and $\widetilde{N}_0\geq 2$, the equilibrium conditions are

$$\frac{1}{2}\frac{1}{1+\widetilde{N}_0}\frac{1}{1+\widetilde{N}_0+1}\left(\frac{c_1}{1+\widetilde{N}_0}+c_2\right)<\pi_W \tag{B17}$$

and

$$\frac{1}{\widetilde{N}_0}\frac{1}{\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}_0}+c_2\right)\geq\pi_I>\frac{1}{\widetilde{N}_0+1}\frac{1}{\widetilde{N}_0+2}\left(\frac{c_1}{\widetilde{N}_0+1}+c_2\right). \tag{B18}$$

Condition (B17) is equivalent to $\widetilde{N}_0\geq M_0$. Condition (B18) implies that $\widetilde{N}_0$ is weakly decreasing in $\pi_I$. So, when $\pi_I$ decreases, $\widetilde{N}_0$ increases and condition (B17) is more likely to hold. Then, there exists a $\pi_2$ such that, for all $\pi_I\in(0,\pi_2]$, there exists an equilibrium such that $\widetilde{N}=0$ and $\widetilde{N}_0\geq 2$. When $\pi_I=\pi_2$, this is the case when just $M_0$ institutional traders operate in equilibrium. So, $\pi_2=\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0}+c_2\right)$.[17] When $\pi_I\leq\pi_2$, there exists an equilibrium such that $\widetilde{N}=0$ and $\widetilde{N}_0\geq 2$. In this equilibrium, only institutional traders are active. If $\pi_I\geq\pi_W$, we must have $\widetilde{N}_0\leq\hat{N}$, and thus, retail investors' welfare must be weakly lower when the entry of institutional traders is allowed; if $\pi_I\leq\pi_W$, we must have $\widetilde{N}_0\geq\hat{N}$, and thus, retail investors' welfare must be weakly higher when the entry of institutional traders is allowed.

If in equilibrium $\widetilde{N}\geq 1$ and $\widetilde{N}_0\geq 2$, then the equilibrium conditions are

$$\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0}+c_2\right)}{\widetilde{N}+\widetilde{N}_0+1}\geq\pi_W>\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0+1}+c_2\right)}{\widetilde{N}+\widetilde{N}_0+2} \tag{B19}$$

and

$$\begin{array}{c}\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0}+c_2\right)\\+\frac{1}{2}\frac{1}{\widetilde{N}_0}\frac{1}{\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}_0}+c_2\right)\end{array}\geq\pi_I>\begin{array}{c}\frac{1}{2}\frac{1}{\widetilde{N}+\widetilde{N}_0+1}\frac{1}{\widetilde{N}+\widetilde{N}_0+2}\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0+1}+c_2\right)\\+\frac{1}{2}\frac{1}{\widetilde{N}_0+1}\frac{1}{\widetilde{N}_0+2}\left(\frac{c_1}{\widetilde{N}_0+1}+c_2\right)\end{array}. \tag{B20}$$

Condition (B19) is equivalent to $\widetilde{N}+\widetilde{N}_0=M_0$. Condition (B20) then becomes

$$\begin{aligned}&\frac{1}{2}\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0}+c_2\right)+\frac{1}{2}\frac{1}{\widetilde{N}_0}\frac{1}{\widetilde{N}_0+1}\left(\frac{c_1}{\widetilde{N}_0}+c_2\right)\\&\geq\pi_I>\frac{1}{2}\frac{1}{M_0+1}\frac{1}{M_0+2}\left(\frac{c_1}{M_0+1}+c_2\right)+\frac{1}{2}\frac{1}{\widetilde{N}_0+1}\frac{1}{\widetilde{N}_0+2}\left(\frac{c_1}{\widetilde{N}_0+1}+c_2\right).\end{aligned}$$

Since $M_0$ is independent of $\pi_I$, it is clear that $\widetilde{N}_0$ is weakly decreasing in $\pi_I$ in this equilibrium. Thus, in this equilibrium, we require $2\leq\widetilde{N}_0\leq M_0-1$. These

[17] Notice that $\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0}+c_2\right)\geq 2\pi_W>\pi_W$.

two constraints together imply

$$\widetilde{N}_0 \geq 2 \Longleftrightarrow \pi_I \leq \pi_3 = \frac{1}{2}\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0}+c_2\right)+\frac{1}{2}\frac{1}{2}\frac{1}{2+1}\left(\frac{c_1}{2}+c_2\right)$$
$$= \frac{1}{2}\frac{1}{M_0}\frac{1}{M_0+1}\left(\frac{c_1}{M_0}+c_2\right)+\frac{1}{12}\left(\frac{c_1}{2}+c_2\right).$$

Moreover, we have that

$$\widetilde{N}_0 \leq M_0 - 1 \Longleftrightarrow \pi_I > \pi_1 = \begin{array}{c} \frac{1}{2}\frac{1}{M_0+1}\frac{1}{M_0+2}\left(\frac{c_1}{M_0+1}+c_2\right)+ \\ \frac{1}{2}\frac{1}{M_0-1+1}\frac{1}{M_0-1+2}\left(\frac{c_1}{M_0-1+1}+c_2\right) \end{array}$$
$$= \frac{1}{2}\frac{1}{M_0+1}\frac{\left(\frac{c_1}{M_0+1}+c_2\right)}{M_0+2}+\frac{1}{2}\frac{1}{M_0}\frac{\left(\frac{c_1}{M_0}+c_2\right)}{M_0+1}.$$

In this equilibrium, condition (B19) implies

$$\frac{1}{\widetilde{N}+\widetilde{N}_0}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0}+c_2\right)}{\widetilde{N}+\widetilde{N}_0+1} \geq 2\pi_W > \frac{1}{\widetilde{N}+\widetilde{N}_0+1}\frac{\left(\frac{c_1}{\widetilde{N}+\widetilde{N}_0+1}+c_2\right)}{\widetilde{N}+\widetilde{N}_0+2},$$

while $\hat{N}$ satisfies

$$\frac{1}{\hat{N}}\frac{1}{\hat{N}+1}\left(\frac{c_1}{\hat{N}}+c_2\right) \geq \pi_W > \frac{1}{\hat{N}+1}\frac{1}{\hat{N}+2}\left(\frac{c_1}{\hat{N}+1}+c_2\right).$$

We thus have $\widetilde{N}+\widetilde{N}_0 \leq \hat{N}$. For a high-cost order, $\widetilde{N}+\widetilde{N}_0$ market makers compete for the order, while for a low-cost order, $\widetilde{N}_0 < \hat{N}$ market makers compete for the order. It is then easy to show that retail investors' welfare is strictly lower when the entry of institutional traders is allowed in this case. Since $M_0 \geq 3$, it is clear that $\pi_1 < \pi_2 < \pi_3$.

In summary,

1. when $\pi_I > \pi_3$, there does not exist an equilibrium such that $\widetilde{N}_0 \geq 2$;
2. when $\pi_I \in (\pi_1, \pi_3]$, there exists an equilibrium such that $\widetilde{N} \geq 1$ and $\widetilde{N}_0 \geq 2$, and retail investors' welfare is strictly lower when the entry of institutional traders is allowed; and
3. when $\pi_I \in (0, \pi_2]$, there exists an equilibrium such that $\widetilde{N} = 0$ and $\widetilde{N}_0 \geq 2$, and if $\pi_I \geq \pi_W$ ($\pi_I \leq \pi_W$), retail investors' welfare is weakly lower (higher) when the entry of institutional traders is allowed.

Note that since $\pi_1 < \pi_2$, when $\pi_I \in (\pi_1, \pi_2]$, multiple equilibria may exist.

*Proof of Proposition 12*

For the first result, note that under brokers' routing, we have that $N$ wholesalers compete for the order, and under order-by-order auctions, we have that $N$ wholesalers compete for the low-cost order and $N$ wholesalers and $N_0$ institutional traders compete for the high-cost order. In addition, the order is always allocated to the most efficient bidder under order-by-order auctions, but this is not necessarily the case under brokers' routing. Since total welfare only concerns allocation efficiency, total welfare under order-by-order auctions must be higher. For the second result,

$$
\begin{aligned}
&\widetilde{W}_I^{OBO} > \widetilde{W}_I^{BR}\\
\Longleftrightarrow & -c_0 - \frac{1}{2}\left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2\\ +\frac{1}{N(N+1)}c_1 - \frac{N-3}{2(N+1)}c_2\end{array}\right] > -\left[c_0 + p\frac{2c_1-(N-3)Nc_2}{2N(1+N)}\right]\\
\Longleftrightarrow & \frac{1}{2}\left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2\\ +\frac{1}{N(N+1)}c_1 - \frac{N-3}{2(N+1)}c_2\end{array}\right] < p\frac{2c_1-(N-3)Nc_2}{2N(1+N)}\\
\Longleftrightarrow & \left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)}c_1 - \frac{N+N_0-3}{2(N+N_0+1)}c_2\\ +\frac{1}{N(N+1)}c_1 - \frac{N-3}{2(N+1)}c_2\end{array}\right] < p\frac{2c_1-(N-3)Nc_2}{N(1+N)}\\
\Longleftrightarrow & \left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)} + \frac{1}{N(N+1)}\\ -\frac{2p}{N(1+N)}\end{array}\right]\frac{c_1}{c_2} < \left[\frac{N+N_0-3}{2(N+N_0+1)} + \frac{N-3}{2(N+1)}\right] - p\frac{(N-3)}{(1+N)}\\
\Longleftrightarrow & \left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)}\\ +\frac{1}{N(N+1)}(1-2p)\end{array}\right]\frac{c_1}{c_2} < \frac{N+N_0-3}{2(N+N_0+1)} + \frac{N-3}{(N+1)}\left(\frac{1}{2}-p\right)\\
\Longleftrightarrow & \left[\begin{array}{c}\frac{1}{(N+N_0)(N+N_0+1)}\\ +\frac{1}{N(N+1)}(1-2p)\end{array}\right]\frac{c_1}{c_2} < \frac{1}{2}\left(\frac{N+N_0-3}{N+N_0+1} - \frac{N-3}{N+1}\right) + \frac{N-3}{N+1}(1-p).
\end{aligned}
$$

## REFERENCES


Adams, William James, and Janet L. Yellen, 1976, Commodity bundling and the burden of monopoly, *Quarterly Journal of Economics* 90, 475–498.

Baldauf, Markus, Joshua Mollner, and Bart Zhou Yueshen, 2024, Siphoned apart: A portfolio perspective on order flow segmentation, *Journal of Financial Economics* 154, 103807.

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, *SSRN Electronic Journal* 10.2139/ssrn.4304124.

Battalio, Robert, and Robert Jennings, 2023, On the potential cost of mandating qualified auctions for marketable retail orders, Working Paper.

Battalio, Robert, Robert Jennings, and Jamie Selway, 2001, The relationship among market-making revenue, payment for order flow, and trading costs for market orders, *Journal of Financial Services Research* 19, 39–56.

Bernhardt, Dan, and Eric Hughson, 1997, Splitting orders, *Review of Financial Studies* 10, 69–101.

Biais, Bruno, David Martimort, and Jean-Charles Rochet, 2000, Competing mechanisms in a common value environment, *Econometrica* 68, 799–837.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya, 2023, Retail trading in options and the rise of the big three wholesalers, *Journal of Finance* 78, 3465–3514.

Chen, Yi, and Zhe Wang, 2023, Optimal sequential selling mechanism and deal protections in mergers and acquisitions, *Journal of Finance* 78, 2139–2188.

Corwin, Shane A., and Jay F. Coughenour, 2008, Limited attention and the allocation of effort in securities trading, *Journal of Finance* 63, 3031–3067.

De Vries, Sven, and Rakesh V. Vohra, 2003, Combinatorial auctions: A survey, *INFORMS Journal on Computing* 15, 284–309.

Dyhrberg, Anne Haubo, Andriy Shkilko, and Ingrid M. Werner, 2022, The retail execution quality landscape, Working Paper, Fisher College of Business.

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.

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

Ernst, Thomas, Andrey Malenko, Chester S. Spatt, and Jian Sun, 2024, What does best execution look like?, Working Paper.

Ernst, Thomas, and Chester S. Spatt, 2022, Payment for order flow and option internalization, NBER Working Paper 29883.

Ernst, Thomas, Chester S. Spatt, and Jian Sun, 2024, Why did retail liquidity programs fail?, Working Paper.

Foley, Sean, Anqi Liu, Katya Malinova, Andreas Park, and Andriy Shkilko, 2022, Cross-subsidizing liquidity, Working Paper, Macquarie University.

Gensler, Gary, 2021, Prepared remarks at the global exchange and fintech conference, Remarks at the Global Exchange and FinTech Conference.

Gorbenko, Alexander S, and Andrey Malenko, 2024, Auctions with endogenous initiation, *Journal of Finance* 79, 1353–1403.

Harris, Milton, and Robert M. Townsend, 1981, Resource allocation under asymmetric information, *Econometrica* 49, 33–64.

He, Zhiguo, Jing Huang, and Cecilia Parlatore, 2024, Bank competition with specialized lending, Working Paper.

Hendershott, Terrence, Saad Khan, and Ryan Riordan, 2022, Option auctions, Working Paper.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2015, Are institutions informed about news?, *Journal of Financial Economics* 117, 249–287.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2021, All-to-all liquidity in corporate bonds, Swiss Finance Institute Research Paper.

Hu, Edwin, and Dermot Murphy, 2022, Competition for retail order flow and market quality, Working Paper.

Hubbard, Timothy P, and René Kirkegaard, 2015, Asymmetric auctions with more than two bidders, Working Paper, Department of Economics and Finance, University of Guelph.

Jain, Pankaj K, Suchismita Mishra, Shawn M O'Donoghue, and Le Zhao, 2024, Trading volume shares and market quality: Pre-and post-zero commissions, *Journal of Empirical Finance* 79, 101564.

Jehiel, Philippe, Moritz Meyer-Ter-Vehn, and Benny Moldovanu, 2007, Mixed bundling auctions, *Journal of Economic Theory* 134, 494–512.

Kaplan, Todd R., and Shmuel Zamir, 2012, Asymmetric first-price auctions with uniform distributions: Analytic solutions to the general case, *Economic Theory* 50, 269–302.

Kim, Jinwoo, 2008, The value of an informed bidder in common value auctions, *Journal of Economic Theory* 143, 585–595.

Klemperer, Paul, 2018, *Auctions: Theory and Practice* (Princeton University Press, Princeton, NJ).

Lebrun, Bernard, 1999, First price auctions in the asymmetric n bidder case, *International Economic Review* 40, 125–142.

Liu, Tingjun, 2016, Optimal equity auctions with heterogeneous bidders, *Journal of Economic Theory* 166, 94–123.

Maskin, Eric, and John Riley, 2000, Asymmetric auctions, *Review of Economic Studies* 67, 413–438.

Maskin, Eric, and John Riley, 2003, Uniqueness of equilibrium in sealed high-bid auctions, *Games and Economic Behavior* 45, 395–409.

McAfee, R. Preston, John McMillan, and Michael D. Whinston, 1989, Multiproduct monopoly, commodity bundling, and correlation of values, *Quarterly Journal of Economics* 104, 371–383.

Menezes, Flavio M., and Paulo K. Monteiro, 2004, *An Introduction to Auction Theory* (OUP, Oxford).

Milgrom, Paul R., and Robert J. Weber, 1982, A theory of auctions and competitive bidding, *Econometrica* 50, 1089–1122.

Myerson, Roger B., 1979, Incentive compatibility and the bargaining problem, *Econometrica* 47, 61–73.

Parlour, Christine A., and Uday Rajan, 2003, Payment for order flow, *Journal of Financial Economics* 68, 379–411.

Pekeč, Aleksandar, and Michael H. Rothkopf, 2003, Combinatorial auction design, *Management Science* 49, 1485–1503.

Povel, Paul, and Rajdeep Singh, 2006, Takeover contests with asymmetric bidders, *Review of Financial Studies* 19, 1399–1431.

Schwarz, Christopher, Brad M. Barber, Xing Huang, Philippe Jorion, and Terrance Odean, 2023, The 'actual retail price' of equity trades, Working Paper.

Securities and Exchange Commission, 2022, Order competition rule, Proposal for Rule 615.

van Kervel, Vincent, and Bart Zhou Yueshen, 2023, Anticompetitive price referencing, Working Paper.

Yang, Liyan, and Haoxiang Zhu, 2020, Back-running: Seeking and hiding fundamental information in order flows, *Review of Financial Studies* 33, 1484–1533.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Appendix S1:** Internet Appendix.