Competition and Privacy in Off-Market Trading

Markus Baldauf[*] Joshua Mollner[†] Basil Williams[‡]

November 18, 2025

[PRELIMINARY AND INCOMPLETE]

Abstract

In U.S. equity markets, brokers route retail orders to wholesalers who execute them off-exchange at prices beating the NBBO. We propose a new theory for why wholesalers offer these discounts: by executing retail flow, a wholesaler privately learns, implying information rents from subsequent on-exchange market making. These rents motivate improved retail pricing, potentially generating midpoint-or-better executions, a commonly observed pattern that classic adverse-selection or inventory models struggle to explain. Policy implications include: (1) enhanced trade reporting erodes information rents and can worsen retail prices; (2) order-by-order auctions, despite increasing competition, may also worsen prices by eroding information rents; (3) a request-for-stream mechanism increases competition while preserving information rents, improving retail pricing.

1 Introduction

In U.S. equity markets, the vast majority of retail orders are not executed on public exchanges but instead are routed to off-exchange wholesalers. In these arrangements, brokers typically receive payments from wholesalers in exchange for directing retail order flow, and wholesalers in turn commit to executing orders at prices at least as favorable as those available on exchanges. Regulations ensure that retail orders executed off-exchange cannot be filled at prices worse than the National Best Bid and Offer (NBBO). Despite these protective rules, however, these arrangements have recently come under scrutiny by the Securities and Exchange Commission (SEC). Central to this debate is whether wholesalers genuinely compete for retail order flow in a way that delivers the best execution quality, or whether the current structure dampens competition relative to a public auction for retail orders. The SEC estimates that opening up individual investor orders to order-by-order competition in the form of an auction would lead to additional gains for retail investors in the amount of $1.5 billion per year (Securities and Exchange Commission, 2023b). Yet, we argue that this mechanism might, despite enhancing competition, nevertheless backfire by harming retail investors instead.

Whether retail traders are better off under order-by-order auctions compared to the status quo requires a theory for *why* wholesalers willingly pay for retail order flow in the first place. One

[*]University of British Columbia, Sauder School of Business: baldauf@mail.ubc.ca, https://sites.google.com/site/mbaldauf/.

[†]Northwestern University, Kellogg School of Management: joshua.mollner@kellogg.northwestern.edu, https://sites.google.com/site/joshuamollner/.

[‡]Imperial College Business School

common explanation is that retail traders are among the least informed among equity traders, implying low adverse selection costs for the wholesalers who execute their orders. This paper proposes a new theory: the execution of retail orders gives wholesalers private information which is valuable for on-market trading. We formalize this theory in a model, and we show that it explains otherwise puzzling pricing behavior. Furthermore, we show that order-by-order auctions are a blunt tool that may harm retail traders. We propose an alternative mechanism that unambiguously improves upon the status quo.

We make two main contributions. First, we develop a theoretical model of market making with learning about order flow, which rationalizes stylized facts about retail order execution that conventional models cannot. The price improvement offered by wholesalers in off-exchange retail execution is substantial relative to the NBBO benchmark. Indeed, the effective retail prices investors obtain often imply a negative spread. Schwarz et al. (forthcoming, Table IV and Figure 2) documents that TD Ameritrade provided an average price improvement of 47.2% relative to the NBBO (where 50% means midpoint execution), with 69% of trades executed at prices equal to or better than the NBBO midpoint.[1] In addition, it received 0.099 cents per share in payment for order flow. Price improvement of these magnitudes, combined with payment for order flow, imply that market makers earn negative spreads from a large fraction of the retail orders they handle. These outcomes cannot be reconciled with traditional models of liquidity provision. In adverse selection models (e.g., Glosten and Milgrom, 1985), the spread can never be negative because it must compensate market makers for their losses to informed traders. In inventory models (e.g., Ho and Stoll, 1983), the spread can never be negative because it must compensate market makers for holding costly inventory positions. Our model, in contrast, can rationalize market making with negative spreads for retail traders.

Second, we use the model to study the consequences for retail transaction costs under alternative mechanisms for retail order execution. We show that order-by-order auctions are not necessarily superior to the status quo, even if they succeed in intensifying competition between wholesalers for retail orders. Indeed, they may be dominated by the status quo if the learning channel is sufficiently strong. This is because order-by-order auctions dissipate the learning benefits that accrue from intermediating retail order flow. We also propose a novel mechanism where wholesalers stream quotes to brokers, essentially a request-for-stream (RFS) arrangement that is commonplace in various over-the-counter (OTC) markets, e.g., in foreign exchange markets (International Organization of Securities Commissions, 2012). We show formally that the RFS mechanism dominates the status quo: wholesaler competition is increased while the benefits derived from privately observing retail execution are unharmed.

We start with a general model that features off-market (retail) and on-market (exchange) trading. Two market makers compete to trade an asset with uncertain fundamental value $\tilde{v}$. At date 1, representing off-market retail execution, competition among market makers may be imperfect, modeled via a differentiated demand system à la Hotelling. At date 2, representing on-market trading, competition is perfect. However, the market makers differ in their information sets: only the market maker who traded at date 1 can condition on the date-1 trade when setting date-2 quotes. The model is general in the sense that it does not impose structure on, e.g., the probability that a liquidity-demanding trader is informed at each date, what precisely an informed trader knows, and so on. (Subsequent sections specialize the model in various ways, precisely by adding

[1] Similarly, Battalio et al. (2023) also find evidence of a substantial number of trades being filled at prices equal to or better than the NBBO midpoint. According to their Table 2, out of 12.9 million retail buy orders of tick-constrained stocks, at least 39.3% are filled at prices at or below the mid.

this sort of structure.) Rather, we work directly in conditional expectation space, that is, with the quantities $\mathbb{E}[\tilde{v}|o_1, o_2]$, where $o_1 \in \{B_1, S_1\}$ and $o_2 \in \{B_2, S_2\}$ represent whether the date-1 and date-2 orders were buys or sells. These conditional expectations differ if there is an informed component to trading.

In setting their date-2 asks, both market makers condition on the event that $o_2 = B_2$ in estimating the value of the asset. The informed market maker also conditions on the date-1 order that he privately observed, implying an estimated asset value for him of $\mathbb{E}[\tilde{v}|o_1, B_2]$. We say that a buy at date 1 functions as the "high signal" if a date-1 buy (when considered in conjunction with a date-2 buy) implies a higher expected value for the asset, i.e., that $\mathbb{E}[\tilde{v}|B_1, B_2] > \mathbb{E}[\tilde{v}|S_1, B_2]$. Conversely, a *sell* at date 1 functions as the high signal if $\mathbb{E}[\tilde{v}|S_1, B_2] > \mathbb{E}[\tilde{v}|B_1, B_2]$. Either case is possible, which we illustrate in Section 3 with an example of each.

We solve the model by backward induction, first characterizing the unique date-2 (on-market) equilibrium. The informational asymmetry between market makers at date 2 creates information rents, which influence the prices they offer to retail traders at date 1. We characterize the date-1 (off-market) equilibrium, showing that the equilibrium spread for retail traders at date 1 can be decomposed into three components, capturing adverse selection, competition intensity, and information rents due to private learning about order flow. If the information rents exceed the adverse selection cost and the markup due to imperfect competition, then the equilibrium spread is negative.

Next, we specialize the general model to two microfoundations. First, in Section 3.1, we consider a standard specification in which, at each date t, a fraction $\alpha_t \in (0, 1)$ of traders are perfectly informed about the fundamental value $\tilde{v}$, with the remainder trading for uncorrelated liquidity reasons. By handling date-1 order flow, a market maker therefore learns about fundamentals. Learning takes the usual form: a buy order is the high signal, in the sense that $\mathbb{E}[\tilde{v}|B_1, B_2] > \mathbb{E}[\tilde{v}|S_1, B_2]$.

In Section 3.2, we consider a second type of learning. In particular, we introduce correlation of date-1 and date-2 uninformed trade. By handling date-1 order flow, a market maker therefore learns about the directionality of uninformed traders. To isolate the implications of learning about uninformed directionality, we also shut down the learning about fundamentals that Section 3.1 considers, by supposing that date-1 trade is purely uninformed. Learning about date-1 order flow is again valuable in this setting, but for a more subtle reason: a *sell* order at date 1 is now the high signal, in the sense that $\mathbb{E}[\tilde{v}|S_1, B_2] > \mathbb{E}[\tilde{v}|B_1, B_2]$. The intuition is that, because date-1 trade is uninformed and because uninformed trade is positively correlated, the direction of the date-1 order serves as a signal of what the date-2 trader will do conditional on being uninformed. If, instead, the date-2 order is in the opposite direction, then it is more likely to have been placed by an informed trader. Therefore, the sequence $\{S_1, B_2\}$ is indicative of informed buying at date 2, which suggests that the asset value is high.

Next, we turn to different mechanisms for handling off-exchange order flow, and compare their outcomes for retail investors against the status quo. In Section 4.2 we consider order-by-order auctions, akin to the SEC's proposal. Upon the arrival of a date-1 retail order, a public auction is triggered where market makers can bid for handling that order. To stack the deck in favor of order-by-order auctions, we suppose that this change succeeds in intensifying competition—so much so that it leads market makers to compete perfectly for that order. This force benefits retail traders. However, there is a more subtle countervailing force: the public nature of the auction event eliminates the information rents that would have been present under the status quo. This force harms retail traders, because without these information rents market makers have less reason to compete for retail orders. As a result, the auction outcome and the status quo cannot be ranked unambiguously. Whether order-by-order auctions improve upon the status quo is an empirical question that requires careful balancing of information rents against the markups stemming from imperfect competition.

We consider two alternative protocols that can be ranked against outcomes under status quo and order-by-order auctions. In Section 4.3 we consider a request-for-stream protocol, whereby the retail trader's broker requests a stream from each market maker and routes the incoming order to the best-priced market maker. Theoretically, this should intensify competition among market makers and eliminate any markup due to imperfect competition. At the same time, the direction of the incoming order would remain private, preserving the information rents that prevailed under the status quo. As a result, the request-for-stream protocol is unambiguously better than both the status quo and order-by-order auctions, because it combines the benefits of each: it entails informations rents like the status quo, and it enhances competition like order-by-order auctions. Finally, in Section 4.1 we analyze a modification of the status quo with enhanced trade reporting: immediately after the order gets filled, the order's direction is made public. As a result, enhanced trade reporting is unambiguously worse than both the status quo and order-by-order auctions, because it combines the shortcomings of each: it preserves any markups permitted by the status quo, and it eliminates information rents like order-by-order auctions.

1.1 Literature Review

Our paper connects to a literature that examines why retail orders often receive price improvement. One explanation is that retail traders are less informed than institutional traders, reducing adverse-selection risk for liquidity providers. Battalio and Holden (2001) develop a model of this, demonstrating that price improvement, payment for order flow, and internalization can arise when orders' informedness are correlated with their verifiable characteristics (such as being retail or not). This is also consistent with empirical evidence from Easley, Kiefer and O'Hara (1996), who show that purchased order flow executed on the Cincinnati Stock Exchange contains significantly less information than orders executed on the New York Stock Exchange. A complementary view developed in Baldauf, Mollner and Yueshen (2024) emphasizes that retail flow is less directional, making it more attractive to internalize or intermediate in the presence of inventory costs. In contrast, our paper highlights a distinct mechanism: retail flow generates information rents for market makers. Because internalizing dealers can extract valuable information from the pattern of incoming retail orders, they are willing to offer price improvement as compensation for the informational advantage this flow provides.

Our paper relates to a literature that examines drawbacks of auctions. Perhaps closest to our work is Ernst, Spatt and Sun (2025) who, like us, study the SEC's order-by-order proposal and argue that the welfare consequences for retail traders are theoretically ambiguous relative to the status quo. In their model, inventory costs have both private-value and common-value components. Owing to the private-value component, wholesalers differ in their inventory costs, and, on the one hand, a public auction does a better job of allocating each order to the wholesaler with the lowest inventory cost. On the other hand, the common value-component of the inventory cost introduces a winner's curse, leading to less aggressive bidding. Glebkin, Yueshen and Shen (2023) compare simultaneous multilateral search, essentially a first price auction among dealers, to bilateral bargaining, and find ambiguous welfare effects. Bilateral bargaining may be preferable to traders to the extent that they may improve their bargaining power. In settings where information leakage is a concern, auctions may also lead to worse outcome compared to bilateral trade (Burdett and

O'Hara, 1987; Hendershott and Madhavan, 2015; Baldauf and Mollner, 2024). In our paper, order-by-order auctions also have ambiguous welfare consequences for retail traders relative to the status quo, but for a different reason: auctions erode markups due to imperfect competition, but they erode future information rents at the same time.

Pinter, Wang and Zou (2025) study a setting where ex-ante identical dealers compete for an order by an informed trader to subsequently benefit from trading with liquidity traders. Similarly, in Röell (1983), dealers offer narrow spreads to privately observed orders because they enable information rents in later trades in the same market. Our paper differs from these in two key respects. First, negative spreads are possible in our paper. And second, we show that learning need not be about asset fundamentals, as in the case of adverse selection models, but that it may involve learning about order flow characteristics such as the correlation of uninformed trade.

2 Model

There are three dates: 1, 2, and 3. There are two market makers and a continuum of traders. These agents are interested in trading an asset with an uncertain fundamental value, which we denote $\tilde{v}$, a real-valued random variable. At dates 1 and 2, as in Glosten and Milgrom (1985), market makers post bids and asks and trade with a randomly drawn trader, and at date 3, the value of the asset is revealed publicly. Date 1 represents off-market execution of retail orders, and date 2 represents on-market trading. At both dates, trading reflects a mixture of information and liquidity motives.

To capture some key differences between off- and on-market trade, our setup makes three main departures from Glosten and Milgrom (1985). First, we allow for differences between the date 1 and date 2 trader characteristics. For example, because retail traders empirically tend to be less informed (e.g., Easley, Kiefer and O'Hara, 1996; Chakrabarty, Cox and Upson, 2024), it may be natural to think that informed trading is relatively less likely at date 1 than at date 2.

Second, off-market (that is, date 1) trade is private, which leads to asymmetrically informed market makers at date 2. In particular, the date-1 order is observed only by the market maker chosen to execute it. As a result, this market maker's date-2 priors about the asset value are informed by the date-1 order, giving him an informational advantage over the other market maker, and leading to information rents at date 2.

Third, we allow for the possibility that off-market (that is, date 1) trade is imperfectly competitive. In particular, at date 1, market makers compete for the trader's order in the style of Hotelling (1929). This allows us to capture concerns that some regulators have voiced about how off-market execution of retail orders is not currently as competitive as it might be under alternative mechanisms (e.g., order-by-order auctions). For on-exchange (that is, date 2) trade, in contrast, we retain the standard assumption of perfect competition, in the sense that a market maker can win the trader's order only by posting the best quote.

2.1 Parameters

We denote the date-1 order by $o_1 \in \{B_1, S_1\}$ and the date-2 order by $o_2 \in \{B_2, S_2\}$. Throughout Section 2, we work directly with the quantities $P(o_1, o_2)$ and $\mathbb{E}[\tilde{v}|o_1, o_2]$. We impose ex-ante bid-ask symmetry by assuming the existence of a parameter $\phi \in [0,1]$ such that

$$\begin{aligned} P(B_1, B_2) = P(S_1, S_2) &= \frac{\phi}{2}, \\ P(S_1, B_2) = P(B_1, S_2) &= \frac{1-\phi}{2}, \end{aligned} \tag{1}$$

and furthermore that

$$\begin{aligned} \mathbb{E}[\tilde{v}|B_1, B_2] &= -\mathbb{E}[\tilde{v}|S_1, S_2], \\ \mathbb{E}[\tilde{v}|S_1, B_2] &= -\mathbb{E}[\tilde{v}|B_1, S_2]. \end{aligned} \tag{2}$$

Thus, there are three free parameters: $\mathbb{E}[\tilde{v}|B_1, B_2]$, $\mathbb{E}[\tilde{v}|S_1, B_2]$, and ϕ. Later, in Section 3, we discuss various potential micro-foundations for these parameters by imposing structure on (i) the fraction of traders who are informed at each date, and (ii) intertemporal correlation in the trading directions of uninformed traders.

Remark 1. It follows from the assumptions above that $\mathbb{E}[\tilde{v}] = 0$. This is done primarily for convenience. It would be straightforward to relax this assumption by shifting each $\mathbb{E}[\tilde{v}|o_1, o_2]$ by a fixed constant.

It is natural that—owing to the possibility of informed on-market trading at date 2—we have both $\mathbb{E}[\tilde{v}|B_1, B_2] > 0$ and $\mathbb{E}[\tilde{v}|S_1, B_2] > 0$. As we will demonstrate with two different micro-foundations in Section 3, either $\mathbb{E}[\tilde{v}|B_1, B_2] > \mathbb{E}[\tilde{v}|S_1, B_2]$ or $\mathbb{E}[\tilde{v}|B_1, B_2] < \mathbb{E}[\tilde{v}|S_1, B_2]$ might naturally arise. To simplify some of the subsequent arguments, we also assume away the knife-edge case in which we have the equality $\mathbb{E}[\tilde{v}|B_1, B_2] = \mathbb{E}[\tilde{v}|S_1, B_2]$.

2.2 On-Market Equilibrium

We solve for the equilibrium using backward induction. At the beginning of date 2, only one market maker has observed the date 1 order, which was either a buy or a sell order. We refer to this market maker as *informed* and the other market maker as *uninformed*.

Without loss of generality, we consider date 2 asks, the prices set for a date 2 buy order $o_2 = B_2$. In setting their date 2 asks, both market makers condition on the event that $o_2 = B_2$ in estimating the value of the asset. The informed market maker also conditions on the date 1 order that he privately observed, implying an estimated asset value for him of $\mathbb{E}[\tilde{v}|o_1, B_2]$. Note that either $o_1 = B_1$ or $o_1 = S_1$ might lead to a higher estimate.

To encompass both cases, we define H_1 and L_1 as the date 1 order that leads to a higher and lower estimated value for the informed market maker, respectively. It follows from equation (2) that we obtain the same definition for both a date-2 buy ($o_2 = B_2$) and a date-2 sell ($o_2 = S_2$). Formally,

$$H_1 := \underset{o_1 \in \{B_1, S_1\}}{\arg\max}\ \mathbb{E}[\tilde{v}|o_1, B_2], \qquad L_1 := \underset{o_1 \in \{B_1, S_1\}}{\arg\min}\ \mathbb{E}[\tilde{v}|o_1, B_2],$$

or, equivalently,

$$H_1 := \underset{o_1 \in \{B_1, S_1\}}{\arg\max}\ \mathbb{E}[\tilde{v}|o_1, S_2], \qquad L_1 := \underset{o_1 \in \{B_1, S_1\}}{\arg\min}\ \mathbb{E}[\tilde{v}|o_1, S_2].$$

Although H_1 and L_1 are orders, we sometimes refer to them as a high signal and low signal, respectively.[2]

[2] To reiterate, what it would mean for S_1 to be the "high signal" is that $\mathbb{E}[\tilde{v}|S_1, B_2] > \mathbb{E}[\tilde{v}|B_1, B_2]$, or equivalently that $\mathbb{E}[\tilde{v}|S_1, S_2] > \mathbb{E}[\tilde{v}|B_1, S_2]$. This might hold in applications (e.g, Section 3.2). It does *not* mean that $\mathbb{E}[\tilde{v}|S_1] > \mathbb{E}[\tilde{v}|B_1]$, which is quite unlikely to hold in applications.

The informed market maker's conditional expected value, therefore, will be either $\mathbb{E}[\tilde{v}|H_1, B_2]$ or $\mathbb{E}[\tilde{v}|L_1, B_2]$. In setting his date-2 ask, the uninformed market maker, on the other hand, conditions only on the date 2 buy order B_2. He does, however, use the realization of the date 2 order to form a belief about the date 1 order. The uninformed market maker's estimated value, therefore, is

$$\mathbb{E}[\tilde{v}|B_2] = P(H_1|B_2)\mathbb{E}[\tilde{v}|H_1, B_2] + P(L_1|B_2)\mathbb{E}[\tilde{v}|L_1, B_2].$$

Note that, by (1), the conditional probabilities $P(H_1|B_2)$ and $P(L_1|B_2)$, which appear in the equation above, will be either ϕ or $1-\phi$. In particular, if $H_1 = B_1$, then $P(H_1|B_2) = \phi$ and $P(L_1|B_2) = 1-\phi$, and vice versa if $H_1 = S_1$.

The following proposition characterizes the Bayes-Nash equilibrium of the date-2 subgame, and shows that the market maker who has private information may earn information rents, which we denote by $r(o_1, o_2)$.

Proposition 1. *There exists a unique equilibrium of the date-2 subgame, the ask-side of which takes the following form:*

1. *If the informed market maker observed a low signal L_1,*
   - *he posts ask a_I according to a mixed strategy over the interval $\big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\big]$ with CDF given by equation (10) in the appendix, and*
   - *earns expected profit*

   $$r(L_1, B_2) = P(H_1|B_2)(\mathbb{E}[\tilde{v}|H_1, B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]) > 0.$$

2. *If the informed market maker observed a high signal H_1,*
   - *he posts ask $a_I = \mathbb{E}[\tilde{v}|H_1, B_2]$, and*
   - *earns expected profit*

   $$r(H_1, B_2) = 0.$$

3. *The uninformed market maker*
   - *posts ask a_U according to a mixed strategy over the interval $\big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\big]$ with CDF given by equation (11) in the appendix, and*
   - *earns zero expected profits.*

A symmetric result holds for date 2 bid-side equilibrium. In that case, the roles played by the high and low signals are reversed. Thus, the informed market maker earns no information rents if he observed the low signal, i.e., $r(L_1, S_2) = 0$. And he earns positive information rents if he observed the high signal, which can be expressed as

$$r(H_1, S_2) = P(L_1|S_2)\big(\mathbb{E}[\tilde{v}|H_1, S_2] - \mathbb{E}[\tilde{v}|L_1, S_2]\big).$$

2.3 Off-Market Trade

At date 1, market makers post bids and asks, and a trader randomly drawn from the off-market pool observes the posted prices and selects one market maker to execute his order. At the beginning of date 1, the two market makers are symmetric. To denote them, we use M' and M''.

For date 2, we had assumed perfect competition—that the order would go to the market maker who posted the lowest ask or the highest bid. Here, for date 1, however, we allow for the possibility that market makers compete only imperfectly. A tractable, reduced-form way of modeling this is to adopt the framework of Hotelling. In particular, we assume that the two market makers have a metaphorical "distance" 1 between them. The "location" λ of the off-market trader is randomly drawn from a uniform distribution with support $[0, 1]$. We let $c > 0$ denote the Hotelling "travel cost", so that higher values of c imply less competition.[3]

We again focus on date 1 asks (buy orders), without loss of generality. Suppose the two market makers post asks a' and a''. Then a trader at location λ will select M' if and only if

$$a' + c\lambda \leq a'' + c(1 - \lambda),$$

that is, if and only if

$$\lambda \leq \frac{1}{2} + \frac{a'' - a'}{2c}.$$

So the trader selects M' with probability $\frac{1}{2} + \frac{a''-a'}{2c}$ and M'' otherwise.

If M' is selected to execute the buy order, his direct profit from the trade is $a' - \mathbb{E}[\tilde{v}|B_1]$.[4] But he also now has an informational advantage over the other market maker, which may yield information rents at date 2. To calculate these information rents, we consider the two cases separately.

- First, suppose $H_1 = B_1$. In this case, the informed market maker obtains information rents following a date 1 buy order if the date 2 order is to sell. This happens with probability $P(S_2|B_1) = 1 - \phi$. Conditional on this happening, the informed market maker's information rents are

  $$r(B_1, S_2) = P(S_1|S_2)\big(\mathbb{E}[\tilde{v}|B_1, S_2] - \mathbb{E}[\tilde{v}|S_1, S_2]\big) = \phi\big(\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big),$$

  where the last equality follows from equations (1) and (2). It follows that the informed market maker's overall expected information rents following a date 1 buy are

  $$(1 - \phi)\phi\big(\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big).$$

- Second, suppose $L_1 = B_1$. In this case, the informed market maker obtains information rents following a date 1 buy order if the date 2 order is to buy. This happens with probability $P(B_2|B_1) = \phi$. Conditional on this happening, the informed market maker's information rents are

  $$r(B_1, B_2) = P(S_1|B_2)\big(\mathbb{E}[\tilde{v}|S_1, B_2] - \mathbb{E}[\tilde{v}|B_1, B_2]\big) = \phi\big(\mathbb{E}[\tilde{v}|S_1, B_2] - \mathbb{E}[\tilde{v}|B_1, B_2]\big),$$

  It follows that the informed market maker's overall expected information rents following a date 1 buy are

  $$\phi(1 - \phi)\big(\mathbb{E}[\tilde{v}|S_1, B_2] - \mathbb{E}[\tilde{v}|B_1, B_2]\big).$$

[3] Although the subsequent derivations use the fact that $c > 0$, the ultimate result, given by equation (4), applies also to the case of $c = 0$, which corresponds to Bertrand competition, as in Glosten and Milgrom (1985).

[4] Note that we can express $\mathbb{E}[\tilde{v}|B_1]$ in terms of the parameters ϕ, $\mathbb{E}[\tilde{v}|B_1, B_2]$, and $\mathbb{E}[\tilde{v}|S_1, B_2]$ as follows:

$$\mathbb{E}[\tilde{v}|B_1] = P(B_2|B_1)\mathbb{E}[\tilde{v}|B_1, B_2] + P(S_2|B_1)\mathbb{E}[\tilde{v}|B_1, S_2] = \phi\mathbb{E}[\tilde{v}|B_1, B_2] - (1 - \phi)\mathbb{E}[\tilde{v}|S_1, B_2].$$

Thus, regardless of which case we are in, the informed market maker's expected information rents following a date 1 buy order (or in fact, by symmetric arguments, following a date 1 sell order) can be written

$$R := \phi(1-\phi)\big|\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big|. \tag{3}$$

Whereas we previously used $r(o_1, o_2)$ to denote information rents *conditional* on the order sequence (o_1, o_2), we now use R to denote *unconditional* expected information rents.

Taking the date 1 and date 2 profits together, the expected profit to M' for posting ask a' is:

$$\left(\frac{1}{2} + \frac{a'' - a'}{2c}\right)\big(a' - \mathbb{E}[\tilde{v}|B_1] + R\big).$$

The first order condition is:

$$0 = -\frac{1}{2c}\big(a' - \mathbb{E}[\tilde{v}|B_1] + R\big) + \left(\frac{1}{2} + \frac{a'' - a'}{2c}\right),$$

which implies that the best response of M' to M'' is:

$$a' = \frac{1}{2}\big(\mathbb{E}[\tilde{v}|B_1] - R + c + a''\big).$$

By symmetry, the best response of M'' to M' is

$$a'' = \frac{1}{2}\big(\mathbb{E}[\tilde{v}|B_1] - R + c + a'\big).$$

So in equilibrium,

$$a' = a'' = \mathbb{E}[\tilde{v}|B_1] - R + c.$$

We therefore obtain the following result:

Proposition 2. *There exists a unique subgame perfect equilibrium, in which at date 1 both market makers quote the ask*

$$a = \mathbb{E}[\tilde{v}|B_1] - R + c, \tag{4}$$

where

$$\mathbb{E}[\tilde{v}|B_1] = \phi\mathbb{E}[\tilde{v}|B_1, B_2] - (1-\phi)\mathbb{E}[\tilde{v}|S_1, B_2],$$
$$R = \phi(1-\phi)\big|\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big|.$$

By symmetry, the bids b_1 and b_2 are simply the negative of the asks given by Proposition 2.[5]

The proposition shows that three forces affect transaction costs for off-market traders. First, adverse selection, as is standard in Glosten-Milgrom models: the possibility of informed trading implies a gap between $\mathbb{E}[\tilde{v}|B_1]$ and $\mathbb{E}[\tilde{v}]$. Second, information rents, as measured by R. Third, competition: greater competition among market makers would be captured by a smaller value for c. Any reduction in adverse selection, or increase in information rents, or intensification of competition would translate into a lower date-1 spread. In fact, as we will see in Section 3.2, information rents can be high enough to induce negative spreads.

[5] In more detail, the equilibrium bid is given by $b = \mathbb{E}[\tilde{v}|S_1] + R - c = -\mathbb{E}[\tilde{v}|B_1] + R - c = -a$.

Remark 2. In the U.S., the trade-through rule prohibits off-exchange trades at prices worse than those prevailing on-exchange. Although the model does not explicitly account for this rule, here is a potential way of extending the model to do so. Without loss of generality, focus on the ask side. Now, the model does not provide a clear way to think about the date-1 *on*-exchange ask. (Proposition 2 only provides the date-1 *off*-exchange ask.) That said, a potential way to conceive of the date-1 on-exchange ask within the model is as $\mathbb{E}[\tilde{v}|B_2]$. So, if the model were to explicitly account for the trade-through rule, the date-1 off-exchange ask would, rather than (4), be instead

$$a = \min\{\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|B_1] - R + c\}. \tag{5}$$

In many cases, including the two special cases of the model considered in Section 3.1 (assuming $\alpha_1 < \alpha_2$) and in Section 3.2, we will have $\mathbb{E}[\tilde{v}|B_1] < \mathbb{E}[\tilde{v}|B_2]$. Moreover, $R \geq 0$. It then follows that (4) and (5) coincide for sufficiently small values of $c > 0$. In that sense, abstracting away from the trade-through rule was without consequence, so long as c is not too large.

2.4 Contrast with Classical Models

The fact that negative spreads can arise in this model represents an interesting contrast against classical models of adverse selection (e.g., Glosten and Milgrom, 1985) and inventory (e.g., Ho and Stoll, 1983). To further explain that point, this section reviews why negative spreads cannot arise in those two models.

Why negative spreads cannot arise in Glosten and Milgrom (1985). In Glosten and Milgrom (1985), Proposition 1 states that the ask is weakly greater than the bid, so that spreads are nonnegative. Intuitively, informed traders buy only when they value the asset high enough, so a buy order causes the market maker to revise his expectation upward. Likewise, sell orders cause the market maker to revise his expectation downward. The ask and bid are precisely those revised expectations, implying nonnegative spreads. That is,

$$a_t = \mathbb{E}_t[\tilde{v}|B_t] \geq \mathbb{E}_t[\tilde{v}] \geq \mathbb{E}_t[\tilde{v}|S_t] = b_t.$$

What is key for Glosten and Milgrom (1985) is that spreads reflect only an order's impact on expectations. In contrast, the off-exchange spread in our setting reflects not only an order's impact on expectations, but also the information rents it generates on-exchange. If those rents are large enough, market makers can afford to post negative off-exchange spreads.

Why negative spreads cannot arise in Ho and Stoll (1983). In Ho and Stoll (1983), Proposition 3 states that the spread is nonnegative (for a version of their model with a one period horizon, and two or more dealers with homogeneous preferences). We sketch the argument here. Dealer i's reservation buying and selling fees are shown to be $b_i = \frac{1}{2}\sigma^2 R(Q + 2I_i)$ and $a_i = \frac{1}{2}\sigma^2 R(Q - 2I_i)$, respectively, where σ^2 is the variance of the stock's return, R is the coefficient of absolute risk aversion, Q is the transaction size, and I_i is the dealer's inventory. Let a^0 be the reservation selling fee for the dealer with the second-largest inventory, and let b' be the reservation buying fee for the dealer with the second-highest inventory. By standard Bertrand logic, the market bid-ask spread is $s = a^0 + b' = R\sigma^2[Q - (I^0 - I')]$. They moreover argue that $I^0 - I' \leq Q$ must hold in equilibrium, for otherwise dealers would trade in the interdealer market and reduce the gap in their inventories. It follows that $s \geq 0$.

As mentioned in the introduction, retail spreads are indeed often negative—a fact that, as we have just seen, classical adverse selection and inventory models struggle to explain. One of the contributions of this paper is to propose a mechanism that can explain these negative spreads.

3 Two Microfoundations

In this section, we consider two plausible instances of the general framework presented in Section 2.

3.1 Informed Off-Market Trade

Each trader is either informed or uninformed. We denote the fraction of informed traders at date $t \in \{1, 2\}$ by $\alpha_t \in (0, 1)$. Traders who are informed observe $\tilde{v}$ perfectly, where we assume that $\tilde{v} = 1$ or $\tilde{v} = -1$ with equal probability. Of the traders who are uninformed, half will buy and half will sell, both at date 1 and at date 2.

Remark 3. We have assumed $\alpha_1 > 0$, meaning that at least a small fraction of off-market traders are informed. That there is an informed component to retail trading is consistent with empirical evidence (Kaniel, Liu, Saar and Titman, 2012; Kelley and Tetlock, 2013; Boehmer, Jones, Zhang and Zhang, 2021). Note, moreover, that this assumption is perfectly compatible with $\alpha_1 < \alpha_2$, in which case off-market traders are less likely to be informed than on-market traders—a property that is likely to hold under natural parametrizations of the model, in light of empirical evidence that retail traders tend to be relatively less informed (Easley, Kiefer and O'Hara, 1996; Chakrabarty, Cox and Upson, 2024).

The general model provided in Section 2 contained three free parameters: $\mathbb{E}[\tilde{v}|B_1, B_2]$, $\mathbb{E}[\tilde{v}|S_1, B_2]$, and ϕ. We next derive explicit formulas for these quantities within this microfoundation, by expressing them directly in terms of (α_1, α_2). To that end, we first compute

$$P(B_1, B_2|\tilde{v} = 1) = \alpha_1\alpha_2 + \frac{\alpha_1(1-\alpha_2)}{2} + \frac{(1-\alpha_1)\alpha_2}{2} + \frac{(1-\alpha_1)(1-\alpha_2)}{4} = \frac{(1+\alpha_1)(1+\alpha_2)}{4},$$
$$P(B_1, B_2|\tilde{v} = -1) = \frac{(1-\alpha_1)(1-\alpha_2)}{4},$$

which imply that

$$P(B_1, B_2) = P(B_1, B_2|\tilde{v} = 1)P(\tilde{v} = 1) + P(B_1, B_2|\tilde{v} = -1)P(\tilde{v} = -1) = \frac{1 + \alpha_1\alpha_2}{4},$$
$$\mathbb{E}[\tilde{v}|B_1, B_2] = \frac{P(B_1, B_2|\tilde{v} = 1)P(\tilde{v} = 1) - P(B_1, B_2|\tilde{v} = -1)P(\tilde{v} = -1)}{P(B_1, B_2)} = \frac{\alpha_1 + \alpha_2}{1 + \alpha_1\alpha_2}. \tag{6}$$

Analogously, we have

$$P(S_1, B_2|\tilde{v} = 1) = \frac{(1-\alpha_1)\alpha_2}{2} + \frac{(1-\alpha_1)(1-\alpha_2)}{4} = \frac{(1-\alpha_1)(1+\alpha_2)}{4},$$
$$P(S_1, B_2|\tilde{v} = -1) = \frac{\alpha_1(1-\alpha_2)}{2} + \frac{(1-\alpha_1)(1-\alpha_2)}{4} = \frac{(1+\alpha_1)(1-\alpha_2)}{4},$$

implying that

$$P(S_1, B_2) = P(S_1, B_2|\tilde{v} = 1)P(\tilde{v} = 1) + P(S_1, B_2|\tilde{v} = -1)P(\tilde{v} = -1) = \frac{1 - \alpha_1\alpha_2}{4}$$
$$\mathbb{E}[\tilde{v}|S_1, B_2] = \frac{P(S_1, B_2|\tilde{v} = 1)P(\tilde{v} = 1) - P(S_1, B_2|\tilde{v} = -1)P(\tilde{v} = -1)}{P(S_1, B_2)} = \frac{\alpha_2 - \alpha_1}{1 - \alpha_1\alpha_2}. \tag{7}$$

Finally, note that

$$\phi = 2P(B_1, B_2) = \frac{1 + \alpha_1\alpha_2}{2}.$$

One feature of this microfoundation is that a date-1 buy order B_1 plays the role of the high signal H_1 (with a date-1 sell order S_1 being the low signal L_1). This is because $\mathbb{E}[\tilde{v}|B_1, B_2] > \mathbb{E}[\tilde{v}|S_1, B_2]$, which we can see by comparing equations (6) and (7):

$$\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2] = \frac{2\alpha_1(1-\alpha_2^2)}{(1+\alpha_1\alpha_2)(1-\alpha_1\alpha_2)} > 0.$$

The intuition is standard. Because there is some possibility that the date-1 order is informed ($\alpha_1 > 0$), a buy B_1 signals a higher fundamental value, while a sell S_1 signals a lower fundamental value. This is true even conditioning on the direction of date-2 order, because date-2 orders are not perfectly informative by themselves ($\alpha_2 < 1$).

Finally, we use Proposition 2 to derive the equilibrium date-1 ask. To that end, we first compute

$$\begin{aligned}
R &= \phi(1-\phi)\big|\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big| \\
&= \frac{1+\alpha_1\alpha_2}{2}\frac{1-\alpha_1\alpha_2}{2}\left(\frac{\alpha_1+\alpha_2}{1+\alpha_1\alpha_2} - \frac{\alpha_2-\alpha_1}{1-\alpha_1\alpha_2}\right) = \frac{\alpha_1(1-\alpha_2)^2}{2}, \\
\mathbb{E}[\tilde{v}|B_1] &= \phi\mathbb{E}[\tilde{v}|B_1, B_2] - (1-\phi)\mathbb{E}[\tilde{v}|S_1, B_2] \\
&= \frac{1+\alpha_1\alpha_2}{2}\frac{\alpha_1+\alpha_2}{1+\alpha_1\alpha_2} - \frac{1-\alpha_1\alpha_2}{2}\frac{\alpha_2-\alpha_1}{1-\alpha_1\alpha_2} = \alpha_1.
\end{aligned}$$

From this we conclude that the equilibrium date-1 ask is

$$\begin{aligned}
a &= \mathbb{E}[\tilde{v}|B_1] - R + c \\
&= \alpha_1 - \frac{\alpha_1(1-\alpha_2)^2}{2} + c = \frac{\alpha_1(1+\alpha_2^2)}{2} + c.
\end{aligned}$$

Hence, the date-1 spread is unambiguously positive within this microfoundation.

The second microfoundation, which we consider below, will differ in two key respects. First, a date-1 buy order B_1 will instead play the role of the *low* signal L_1 (with a date-1 sell order S_1 being the high signal H_1). Second, negative date-1 spreads will be possible.

3.2 Correlated Uninformed Trade

Each trader is either informed or uninformed. We denote the fraction of informed traders at date 2 by $\alpha_2 \in (0, 1)$. We furthermore assume that no traders are informed at date 1 (i.e., $\alpha_1 = 0$). Traders who are informed observe $\tilde{v}$ perfectly, where we assume that $\tilde{v} = 1$ or $\tilde{v} = -1$ with equal probability. Of the traders who are uninformed, a share $\frac{1}{2}(1+\tilde{\delta})$ will buy and $\frac{1}{2}(1-\tilde{\delta})$ will sell, both at date 1 and at date 2, where $\tilde{\delta}$ is an independent random variable that is not observed by the market makers. In particular, we assume that $\tilde{\delta} = \delta$ or $\tilde{\delta} = -\delta$ with equal probability, for some parameter $\delta \in (0, 1)$.

Remark 4. Thus, we assume positive correlation between on- and off-exchange uninformed order flow, with δ parametrizing the extent of this correlation. Here, we explain why this might occur in reality. As an initial observation, the trades of retail investors may be correlated either due to herding or sentiment. For example, they may pursue common trading strategies (e.g., risk on/risk off, or thematic investing). This, in turn, may lead to correlation between off- and on-exchange order flows for a number of reasons. First, some retail orders may be filled on-exchange and others off-market. Second, retail investors may purchase securities, some of which may be filled as institutional order flow on-exchange, and some that may be filled as retail order flow off-market.

As an example of the latter, consider thematic investing around technology and AI. An investor may buy a stock such as Nvidia together with a technology mutual fund. The Nvidia order is likely to be filled off-market. However, the cash inflow to the mutual fund may lead to subsequent on-exchange purchases of its constituents, which may include Nvidia. This would generate positive correlation between off- and on-exchange order flow in Nvidia.

The general model provided in Section 2 contained three free parameters: $\mathbb{E}[\tilde{v}|B_1, B_2]$, $\mathbb{E}[\tilde{v}|S_1, B_2]$, and ϕ. We next derive explicit formulas for these quantities within this microfoundation, by expressing them directly in terms of (α_2, δ). To that end, we first compute

$$\begin{aligned}
P(B_1, B_2|\tilde{v}=1) &= P(B_1, B_2|\tilde{v}=1, \tilde{\delta}=\delta)P(\tilde{\delta}=\delta) + P(B_1, B_2|\tilde{v}=1, \tilde{\delta}=-\delta)P(\tilde{\delta}=-\delta) \\
&= \frac{1+\delta}{2}\left[\alpha_2 + (1-\alpha_2)\frac{1+\delta}{2}\right]\frac{1}{2} + \frac{1-\delta}{2}\left[\alpha_2 + (1-\alpha_2)\frac{1-\delta}{2}\right]\frac{1}{2} \\
&= \frac{\alpha_2}{2} + \frac{1-\alpha_2}{4}(1+\delta^2), \\
P(B_1, B_2|\tilde{v}=-1) &= P(B_1, B_2|\tilde{v}=-1, \tilde{\delta}=\delta)P(\tilde{\delta}=\delta) + P(B_1, B_2|\tilde{v}=-1, \tilde{\delta}=-\delta)P(\tilde{\delta}=-\delta) \\
&= \frac{1+\delta}{2}\left[(1-\alpha_2)\frac{1+\delta}{2}\right]\frac{1}{2} + \frac{1-\delta}{2}\left[(1-\alpha_2)\frac{1-\delta}{2}\right]\frac{1}{2} \\
&= \frac{1-\alpha_2}{4}(1+\delta^2),
\end{aligned}$$

which imply that

$$\begin{aligned}
P(B_1, B_2) &= P(B_1, B_2|\tilde{v}=1)P(\tilde{v}=1) + P(B_1, B_2|\tilde{v}=-1)P(\tilde{v}=-1) = \frac{1}{4} + \frac{1-\alpha_2}{4}\delta^2, \\
\mathbb{E}[\tilde{v}|B_1, B_2] &= \frac{P(B_1, B_2|\tilde{v}=1)P(\tilde{v}=1) - P(B_1, B_2|\tilde{v}=-1)P(\tilde{v}=-1)}{P(B_1, B_2)} = \frac{\alpha_2}{1+(1-\alpha_2)\delta^2}. \qquad (8)
\end{aligned}$$

Analogously, we have

$$\begin{aligned}
P(S_1, B_2|\tilde{v}=1) &= P(S_1, B_2|\tilde{v}=1, \tilde{\delta}=\delta)P(\tilde{\delta}=\delta) + P(S_1, B_2|\tilde{v}=1, \tilde{\delta}=-\delta)P(\tilde{\delta}=-\delta) \\
&= \frac{1-\delta}{2}\left[\alpha_2 + (1-\alpha_2)\frac{1+\delta}{2}\right]\frac{1}{2} + \frac{1+\delta}{2}\left[\alpha_2 + (1-\alpha_2)\frac{1-\delta}{2}\right]\frac{1}{2} \\
&= \frac{\alpha_2}{2} + \frac{1-\alpha_2}{4}(1-\delta^2), \\
P(S_1, B_2|\tilde{v}=-1) &= P(S_1, B_2|\tilde{v}=-1, \tilde{\delta}=\delta)P(\tilde{\delta}=\delta) + P(S_1, B_2|\tilde{v}=-1, \tilde{\delta}=-\delta)P(\tilde{\delta}=-\delta) \\
&= \frac{1-\delta}{2}\left[(1-\alpha_2)\frac{1+\delta}{2}\right]\frac{1}{2} + \frac{1+\delta}{2}\left[(1-\alpha_2)\frac{1-\delta}{2}\right]\frac{1}{2} \\
&= \frac{1-\alpha_2}{4}(1-\delta^2),
\end{aligned}$$

implying that

$$\begin{aligned}
P(S_1, B_2) &= P(S_1, B_2|\tilde{v}=1)P(\tilde{v}=1) + P(S_1, B_2|\tilde{v}=-1)P(\tilde{v}=-1) = \frac{1}{4} - \frac{1-\alpha_2}{4}\delta^2 \\
\mathbb{E}[\tilde{v}|S_1, B_2] &= \frac{P(S_1, B_2|\tilde{v}=1)P(\tilde{v}=1) - P(S_1, B_2|\tilde{v}=-1)P(\tilde{v}=-1)}{P(S_1, B_2)} = \frac{\alpha_2}{1-(1-\alpha_2)\delta^2}. \qquad (9)
\end{aligned}$$

Finally, note that

$$\phi = 2P(B_1, B_2) = \frac{1}{2} + \frac{1-\alpha_2}{2}\delta^2.$$

One feature of this microfoundation is that a date-1 buy order B_1 plays the role of the low signal L_1 (with a date-1 sell order S_1 being the high signal H_1). This is because $\mathbb{E}[\tilde{v}|B_1, B_2] < \mathbb{E}[\tilde{v}|S_1, B_2]$, which we can see by comparing equations (8) and (9):

$$\frac{\mathbb{E}[\tilde{v}|B_1, B_2]}{\mathbb{E}[\tilde{v}|S_1, B_2]} = \frac{1-(1-\alpha_2)\delta^2}{1+(1-\alpha_2)\delta^2} < 1.$$

Intuitively, because uninformed trade is positively correlated, on-market orders that go against the off-market sentiment are more likely to have been placed by informed traders, and are therefore more informative about $\tilde{v}$. For example, consider the extreme case of $\delta = 1$. In this extreme case, a date-1 sell order S_1 reveals that all uninformed traders seek to sell, meaning that a date-2 buy order B_2 can only come from an informed trader who has observed a high fundamental value $\tilde{v} = 1$. It follows that, in this case, $\mathbb{E}[\tilde{v}|S_1, B_2] = 1$. In contrast, a date-1 buy order B_1 reveals that all uninformed traders seek to buy, meaning that a date-2 buy order B_2 could come from either an informed trader or an uninformed one, implying $\mathbb{E}[\tilde{v}|B_1, B_2] < 1$.

Finally, we use Proposition 2 to derive the equilibrium date-1 ask. To that end, we first compute

$$\begin{aligned}
R &= \phi(1-\phi)\big|\mathbb{E}[\tilde{v}|B_1, B_2] - \mathbb{E}[\tilde{v}|S_1, B_2]\big| \\
&= \left(\frac{1}{2} + \frac{1-\alpha_2}{2}\delta^2\right)\left(\frac{1}{2} - \frac{1-\alpha_2}{2}\delta^2\right)\left(\frac{\alpha_2}{1-(1-\alpha_2)\delta^2} - \frac{\alpha_2}{1+(1-\alpha_2)\delta^2}\right) \\
&= \frac{(1-\alpha_2)\alpha_2\delta^2}{2}, \\
\mathbb{E}[\tilde{v}|B_1] &= \phi\mathbb{E}[\tilde{v}|B_1, B_2] - (1-\phi)\mathbb{E}[\tilde{v}|S_1, B_2] \\
&= \left(\frac{1}{2} + \frac{1-\alpha_2}{2}\delta^2\right)\frac{\alpha_2}{1+(1-\alpha_2)\delta^2} - \left(\frac{1}{2} - \frac{1-\alpha_2}{2}\delta^2\right)\frac{\alpha_2}{1-(1-\alpha_2)\delta^2} = 0.
\end{aligned}$$

That $\mathbb{E}[\tilde{v}|B_1] = 0$ is because date-1 traders are exclusively uninformed. Although date-1 orders are not informative about the fundamental on their own, they do, as we have seen above, become informative when they are considered in conjunction with date-2 order flow.

From the calculations above, we conclude that the equilibrium date-1 ask is

$$\begin{aligned}
a &= \mathbb{E}[\tilde{v}|B_1] - R + c \\
&= -\frac{(1-\alpha_2)\alpha_2\delta^2}{2} + c.
\end{aligned}$$

Hence, the date-1 spread might be negative within this microfoundation.

4 Off-Exchange Protocols

We now compare the status quo to a variety of alternative protocols for off-exchange execution. One of the alternatives we consider—order-by-order auctions—had been proposed by the SEC, with the objective of reducing transaction costs for retail traders (Securities and Exchange Commission, 2023b). The focus of our analysis will, therefore, be on retail transaction costs. To that end, use Proposition 2 to define

$$a^{SQ} = \mathbb{E}[\tilde{v}|B_1] - R + c$$

as the date-1 ask prevailing under the status quo (the superscript SQ stands for "status quo"). Recall that $c \geq 0$ is a parameter, and that $R > 0$ is a function of the parameters defined by equation (3). In what follows, we derive corresponding expressions for the date-1 asks under three alternative protocols: enhanced trade reporting, order-by-order auctions, and request-for-stream. We then discuss how they compare.

4.1 Enhanced Trade Reporting

In practice, what exactly do wholesalers privately learn from the retail order flow that they handle? In principle, the private information may take several forms.

- First, the wholesaler who handles the retail order would have short-horizon private knowledge about its *existence* (or likewise, its *size*), during the lag between when it receives the order and when the execution is reported to a Trade Reporting Facility (TRF). In practice, however, these lags tend to be extremely short: Upson, Chakrabarty and Cox (2025) estimate that they are typically in the neighborhood of 2 milliseconds. Thus, this source is likely to be a relatively minor contributor of private information. The next two forms are likely to play more major roles in practice.

- Second, the wholesaler who handles the retail order might have long-horizon private information about its *direction*. Indeed, trade reporting data does not distinguish between buyer-initiated and seller-initiated trades. Moreover, many retail orders are filled at midpoint-or-better prices (e.g., Schwarz et al., forthcoming), rendering unreliable the traditional methods of trade signing that are based on the algorithm of Lee and Ready (1991). Wholesalers who did not handle the order might therefore struggle to reliably sign its direction, whereas the wholesaler who did handle it would, of course, observe direction directly.

- Third, the wholesaler who handles the retail order would have long-horizon private knowledge about its *source*. For example, this wholesaler would know the retail brokerage from which the order originated, which is informative because brokerages differ in the characteristics of their clientele (e.g., their likelihood of being informed, or their susceptibility to herding on sentiment).[6] Wholesalers who did not handle the order would not know the retail brokerage of origin. In fact, wholesalers who did not handle the order might even struggle to reliably identify the order as retail in the first place (Boehmer, Jones, Zhang and Zhang, 2021; Battalio, Jennings, Saglam and Wu, 2023). Although they would observe some characteristics of the trade—including that it was executed off-exchange—this does not immediately imply a retail trade, because other types of trades sometimes also occur off-exchange and are reported to the same TRFs (e.g., midpoint dark pools that cater to institutions).

To keep our analysis focused and clear, we model private information about retail order *direction*, effectively assuming that order existence and source are common knowledge. However, our results could be adapted to accommodate richer settings, and the main point of our paper would continue to hold: if off-exchange orders contain valuable information for on-market trade, then the value of that information will be reflected in better off-exchange prices.

One intervention that regulators might contemplate is to enhance the reporting of retail trades. For example, regulators might require the direction of retail orders to be reported directly.

To capture this in the model, suppose that the direction of the date 1 order $o_1 \in \{B_1, S_1\}$ becomes public before date 2 quotes are set. As a result, information rents would disappear, i.e., $R = 0$.

[6] For empirical evidence about informativeness of the originating retail brokerage, see, for example, Table 12 of Securities and Exchange Commission (2023b).

Other than that, the arguments used to derive Proposition 2 continue to apply. Hence, the date-1 ask under this regime would be

$$a^{ETR} = \mathbb{E}[\tilde{v}|B_1] + c,$$

where the ETR superscript stands for "enhanced trade reporting."

Clearly, we have $a^{ETR} \geq a^{SQ}$. The intuition is simple. Under the status quo, the possibility of obtaining information rents for subsequent on-exchange market making incentivized wholesalers to compete down the bid-ask spread that they offered for off-exchange retail trades. Enhanced trade reporting would eliminate these information rents, remove this extra incentive to compete for retail orders, and consequently worsen prices for retail traders.

4.2 Order-by-Order Auctions

The SEC once proposed a rule that would require retail orders to be exposed to *order-by-order auctions* (Securities and Exchange Commission, 2023b). Specifically, each incoming retail order would trigger an auction. Each auction would be widely disseminated in consolidated market data feeds, collect quotes from market participants during an interval lasting 100 to 300 milliseconds, then award the order to the market participant submitting the best quote, with execution occurring at the quoted price.

The rule was motivated by analysis suggesting that the current system—in which retail orders are routed to wholesalers—leads to less than perfect competition. They write:

> The routing of individual investors' marketable orders to wholesalers occurs because the orders impose lower costs on liquidity providers than the order flow routed to national securities exchanges. These costs are known as "adverse selection costs" and reflect the extent to which prices move against the liquidity provider after executing an order. While the lower adverse selection order flow of individual investors allows wholesalers to execute their orders with price improvement, Commission analysis of trading data indicates that the level of price improvement offered by wholesalers does not fully reflect the much lower cost. (Securities and Exchange Commission, 2023a)

The thinking was that order-by-order auctions would enhance competition for retail orders, which we can capture in our model by a reduction in the parameter c (i.e., the Hotelling "travel cost"). Perfect competition is captured by reducing this parameter all the way to zero, and for simplicity, we assume that order-by-order auctions would achieve this.

There would, however, be a further effect, similar to the implications of enhanced trading reporting. By exposing retail orders to order-by-order auctions, the direction of each retail order would be observed by *all* market participants—and not solely by the one who wins the auction and executes the order.[7] To capture this in the model, suppose that the direction of the date 1 order $o_1 \in \{B_1, S_1\}$ becomes public before date 2 quotes are set. As a result, information rents would disappear, i.e., $R = 0$.

[7]Indeed, see paragraph (c)(1)(i) of the SEC's proposed rule:

> An auction message announcing the initiation of a qualified auction for a segmented order shall be provided for dissemination in consolidated market data pursuant to §242.603(b). Each such auction message shall invite priced auction responses to trade with a segmented order and shall include the identity of the open competition trading center and the symbol, side, size, limit price, and identity of the originating broker for the segmented order. (Securities and Exchange Commission, 2023b)

Other than the two points above, the arguments used to derive Proposition 2 continue to apply. Hence, the date-1 ask under this regime would be

$$a^{OBO} = \mathbb{E}[\tilde{v}|B_1],$$

where the *OBO* superscript stands for "order-by-order auctions."

4.3 Request-for-Stream

A natural question is: does there exist a mechanism that would achieve the intensification of competition aimed at by order-by-order auctions, without eliminating information rents? Such a mechanism would, in the setting of our model, represent "the best of both worlds."

A *request-for-stream* mechanism would achieve this. More precisely, suppose that market makers continuously stream binding quotes. These quotes would be observable to retail brokers (and, perhaps also, to other market makers), and brokers would be required to route to the best quote. Crucially—just as in the status quo—only the market maker who receives an order would learn of its existence and direction. For that reason, information rents would be preserved. However, the observability of quotes puts us in the standard Bertrand setting, and we should therefore expect intensified competition—just as under order-by-order auctions—which, as above, we can capture within the model by setting $c = 0$.

Other than this, the arguments used to derive Proposition 2 continue to apply. Hence, the date-1 ask under this regime would be

$$a^{RFS} = \mathbb{E}[\tilde{v}|B_1] - R,$$

where the *RFS* superscript stands for "request-for-stream."

4.4 Summary: Off-Exchange

The observations of this section can be summarized in the following result.

Proposition 3. *Among the four protocols considered above, off-exchange transaction costs are lowest under the request-for-quote protocol and highest under enhanced trade reporting:*

$$a^{RFS} \leq a^{OBO}, a^{SQ} \leq a^{ETR}.$$

Relative to the status quo, the introduction of order-by-order auctions might either raise ($a^{OBO} > a^{SQ}$) or lower ($a^{OBO} < a^{SQ}$) transaction costs, depending on parameters.

The comparison in Proposition 3 summarizes the tradeoffs among the four protocols. The status quo (SQ) and order-by-order auctions (OBO) cannot be ranked unambiguously, owing to two opposing forces. On the one hand, order-by-order auctions intensify wholesalers' competition for retail flow. On the other hand, order-by-order auctions eliminate wholesalers' information rents, reducing their incentive to offer price improvement. Depending on the relative strength of these two effects, these auctions can either improve or worsen retail execution quality relative to the status quo. By contrast, enhanced trade reporting (ETR) is unambiguously worse, as it combines the drawbacks of the SQ and OBO mechanisms: like OBO, it removes the information rents that incentivize wholesalers to offer price improvement; like SQ, it preserves imperfect competition. Finally, the request-for-stream (RFS) mechanism is unambiguously better, as it combines the *benefits*: like SQ, it preserves information rents; like OBO, it enhances competition. Taken together, these results underscore the importance of market design, and in particular the necessity of careful attention to information structure.

5 Conclusion

This paper develops a new theory of off-exchange retail trading that highlights the role of *private learning* by wholesalers as a driver of retail price improvement. By linking wholesalers' willingness to offer midpoint-or-better executions to the informational rents they earn in subsequent on-market trading, the model explains a set of empirical facts that classic adverse-selection and inventory-based frameworks cannot. The model's structure provides a unified way to interpret how competition and information together shape retail execution quality.

The analysis also clarifies that retail traders benefit from these information rents. When wholesalers' private learning is valuable, the rents act as a competitive subsidy that narrows or even reverses the effective spread faced by retail investors. However, if regulation or market design eliminates these rents, e.g., through enhanced transparency or public auctions, then wholesalers' incentives to compete diminish, leading to higher transaction costs.

The results therefore have direct implications for ongoing regulatory debates surrounding retail order execution. In particular, they suggest that reforms, such order-by-order auctions of the form once proposed by the SEC could, depending on parameter values, lead to worse outcomes for retail traders.

Finally, the paper proposes a practical mechanism that is in some sense the best of both worlds: a *request-for-stream mechanism*, which fosters intense competition while preserving the informational privacy that sustains price improvement. By allowing wholesalers to compete for retail orders without disclosing the orders themselves, this protocol combines the efficiency of auctions with the incentive benefits of private learning. More broadly, our framework suggests that the optimal design of retail trading systems requires balancing competition with information rents—recognizing that policies which erode one may unintentionally weaken the other, to the detriment of the very investors they aim to protect.

A Proof of Proposition 1

Our setting is a first-price common value auction with two bidders, one of whom privately observes a binary-valued signal of the common value. This is a specific case of Engelbrecht-Wiggans et al. (1983), which allows for many uninformed bidders and an arbitrarily distributed private signal for the informed bidder. Although their results establish existence and uniqueness of equilibrium, they do not characterize the equilibrium form for our specific setting. So as to have a complete characterization of equilibrium, we proceed by showing directly that the following CDFs of the informed and uninformed pricing strategies constitute an equilibrium:

$$F_I(a_I|H_1,B_2) = \begin{cases} 0 & \text{if } a_I < \mathbb{E}[\tilde{v}|H_1,B_2] \\ 1 & \text{if } a_I \geq \mathbb{E}[\tilde{v}|H_1,B_2], \end{cases}$$

$$F_I(a_I|L_1,B_2) := \begin{cases} 0 & \text{if } a_I < \mathbb{E}[\tilde{v}|B_2] \\ 1 - \frac{P(H_1|B_2)}{P(L_1|B_2)} \cdot \frac{\mathbb{E}[\tilde{v}|H_1,B_2]-a_I}{a_I-\mathbb{E}[\tilde{v}|L_1,B_2]} & \text{if } a_I \in [\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1|B_2]) \\ 1 & \text{if } a_I \geq \mathbb{E}[\tilde{v}|H_1,B_2], \end{cases} \tag{10}$$

$$F_U(a_U|B_2) := \begin{cases} 0 & \text{if } a_U < \mathbb{E}[\tilde{v}|B_2] \\ 1 - \frac{\mathbb{E}[\tilde{v}|B_2]-\mathbb{E}[\tilde{v}|L_1,B_2]}{a_U-\mathbb{E}[\tilde{v}|L_1,B_2]} & \text{if } a_U \in [\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1,B_2]) \\ 1 & \text{if } a_U \geq \mathbb{E}[\tilde{v}|H_1,B_2]. \end{cases} \tag{11}$$

We show below that under this strategy profile, neither market maker has a profitable deviation.

As a preliminary, note that $F_I(\cdot|L_1,B_2)$ is continuous, whereas $F_U(\cdot|B_2)$ is continuous except for a mass point at the upper bound of its support. To see this, recall that $\mathbb{E}[\tilde{v}|B_2] = P(H_1|B_2)\mathbb{E}[\tilde{v}|H_1,B_2] + P(L_1|B_2)\mathbb{E}[\tilde{v}|L_1,B_2]$, which implies

$$\begin{aligned} \mathbb{E}[\tilde{v}|H_1,B_2] - \mathbb{E}[\tilde{v}|B_2] &= P(L_1|B_2)\Big(\mathbb{E}[\tilde{v}|H_1,B_2] - \mathbb{E}[\tilde{v}|L_1,B_2]\Big) \\ \mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1,B_2] &= P(H_1|B_2)\Big(\mathbb{E}[\tilde{v}|H_1,B_2] - \mathbb{E}[\tilde{v}|L_1,B_2]\Big). \end{aligned}$$

Using these expressions to evaluate (10) and (11) at the support boundaries, it is clear that $F_I(\cdot|L_1,B_2)$ is continuous, but $F_U(\cdot|B_2)$ is continuous everywhere except at the point $a_U = \mathbb{E}[\tilde{v}|H_1,B_2]$, which has mass

$$\begin{aligned} P(a_U = \mathbb{E}[\tilde{v}|H_1,B_2]|B_2) &= F_U(\mathbb{E}[\tilde{v}|H_1,B_2]|B_2) - \lim_{a_U \uparrow \mathbb{E}[\tilde{v}|H_1,B_2]} F_U(a_U|B_2) \\ &= \frac{\mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1,B_2]}{\mathbb{E}[\tilde{v}|H_1,B_2] - \mathbb{E}[\tilde{v}|L_1,B_2]} > 0. \end{aligned}$$

We now show that under this strategy profile, the informed market maker does not have a profitable deviation.

Suppose that the informed market maker observed H_1. If he posts $a_I < \mathbb{E}[\tilde{v}|H_1,B_2]$, any trade he wins gives him negative profits. If he posts $a_I = \mathbb{E}[\tilde{v}|H_1,B_2]$, the trade is worth zero, so his expected profit is zero. If he posts $a_I > \mathbb{E}[\tilde{v}|H_1,B_2]$, his ask exceeds the support of the uninformed market maker, so he loses the trade with certainty and receives zero expected profit. So he does not have a profitable deviation from $a_I = \mathbb{E}[\tilde{v}|H_1,B_2]$, and his expected profit if he observed H_1 is zero.

Next, suppose that the informed market maker observed L_1. If he posts ask $a_I \in \mathbb{R}$, the value of the trade is $a_I - \mathbb{E}[\tilde{v}|L_1, B_2]$, and the probability that he wins the trade is

$$\begin{aligned}
&P(a_U > a_I|B_2) + 0.5P(a_U = a_I|B_2) \\
=&1 - F_U(a_I|B_2) + 0.5P(a_U = \mathbb{E}[\tilde{v}|H_1, B_2]|B_2)\mathbb{1}[a_I = \mathbb{E}[\tilde{v}|H_1, B_2]],
\end{aligned}$$

which follows from the fact that $F_U(\cdot|B_2)$ is continuous everywhere except the mass point $a_U = \mathbb{E}[\tilde{v}|H_1, B_2]$.

As a result, the informed market maker's expected profit for posting $a_I \in \mathbb{R}$ is

$$\begin{aligned}
&\Big(a_I - \mathbb{E}[\tilde{v}|L_1, B_2]\Big) \cdot \Big(1 - F_U(a_I|B_2) + 0.5 \cdot P(a_U = \mathbb{E}[\tilde{v}|H_1, B_2]|B_2) \cdot \mathbb{1}\Big[a_I = \mathbb{E}[\tilde{v}|H_1, B_2]\Big]\Big) \\
=&\Big(a_I - \mathbb{E}[\tilde{v}|L_1, B_2]\Big) \cdot \begin{cases} 1 & \text{if } a_I < \mathbb{E}[\tilde{v}|B_2] \\ \frac{\mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]}{a_I - \mathbb{E}[\tilde{v}|L_1, B_2]} & \text{if } a_I \in \Big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\Big). \\ 0.5 \cdot \frac{\mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]}{\mathbb{E}[\tilde{v}|H_1, B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]} & \text{if } a_I = \mathbb{E}[\tilde{v}|H_1, B_2] \\ 0 & \text{if } a_I > \mathbb{E}[\tilde{v}|H_1, B_2]. \end{cases} \\
=&\begin{cases} a_I - \mathbb{E}[\tilde{v}|L_1, B_2] & \text{if } a_I < \mathbb{E}[\tilde{v}|B_2] \\ \mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1, B_2] & \text{if } a_I \in \Big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\Big) \\ 0.5 \cdot \Big(\mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]\Big) & \text{if } a_I = \mathbb{E}[\tilde{v}|H_1, B_2] \\ 0 & \text{if } a_I > \mathbb{E}[\tilde{v}|H_1, B_2]. \end{cases}
\end{aligned}$$

Clearly, the informed market maker is indifferent over the interval $a_I \in \Big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\Big)$ and is strictly worse off otherwise. So he does not have a profitable deviation, and earns expected profits $\mathbb{E}[\tilde{v}|B_2] - \mathbb{E}[\tilde{v}|L_1, B_2] = P(H_1|B_1)\Big(\mathbb{E}[\tilde{v}|H_1, B_2] - \mathbb{E}[\tilde{v}|L_1, B_2]\Big) > 0$.

We conclude by showing that the uninformed market maker does not have a profitable deviation. If he posts an ask $a_U \in \mathbb{R}$, the value of his trade and the probability of winning the trade depend on whether the informed market maker observed H_1 or L_1. Given $o_1 \in \{H_1, L_1\}$, the uninformed market maker's value of the trade is $a_U - \mathbb{E}[\tilde{v}|o_1, B_2]$, and his probability of winning the trade is

$$\begin{aligned}
&P(a_I > a_U|o_1, B_2) + 0.5P(a_I = a_U|o_1, B_2) \\
=&1 - F_I(a_U|o_1, B_2) + 0.5 \cdot \mathbb{1}\Big[a_U = \mathbb{E}[\tilde{v}|H_1, B_2], o_1 = H_1\Big],
\end{aligned}$$

where the substitution in the second term follows from the fact that $F_I(\cdot|o_1, B_2)$ is continuous if $o_1 = L$ and has mass 1 at the point $a_I = \mathbb{E}[\tilde{v}|H_1, B_2]$ if $o_2 = H$.

So the uninformed market maker's expected profit for posting $a_U \in \mathbb{R}$ is

$$\sum_{o_1 \in \{H_1, L_1\}} P(o_1|B_2)\Big(a_U - \mathbb{E}[\tilde{v}|o_1, B_2]\Big)\Big(1 - F_I(a_U|o_1, B_2) + 0.5 \cdot \mathbb{1}\Big[a_U = \mathbb{E}[\tilde{v}|H_1, B_2], o_1 = H_1\Big]\Big)$$

$$= \begin{cases} P(H_1|B_2)(a_U - \mathbb{E}[\tilde{v}|H_1, B_2])(1) + P(L_1|B_2)(a_U - \mathbb{E}[\tilde{v}|L_1, B_2])(1) & \text{if } a_U < \mathbb{E}[\tilde{v}|B_2]. \\ P(H_1|B_2)(a_U - \mathbb{E}[\tilde{v}|H_1, B_2])(1) & \\ \qquad +P(L_1|B_2)(a_U - \mathbb{E}[\tilde{v}|L_1, B_2])\frac{P(H_1|B_2)}{P(L_1|B_2)}\frac{\mathbb{E}[\tilde{v}|H_1,B_2]-a_U}{a_U-\mathbb{E}[\tilde{v}|L_1,B_2]} & \text{if } a_U \in [\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]) \\ P(H_1|B_2)(0)(0.5) + P(L_1|B_2)(\mathbb{E}[\tilde{v}|H_1, B_2] - \mathbb{E}[\tilde{v}|L_1, B_2])(0) & \text{if } a_U = \mathbb{E}[\tilde{v}|H_1, B_2] \\ P(H_1|B_2)(a_U - \mathbb{E}[\tilde{v}|H_1, B_2])(0) + P(L_1|B_2)(a_U - \mathbb{E}[\tilde{v}|L_1, B_2])(0) & \text{if } a_U > \mathbb{E}[\tilde{v}|H_1, B_2] \end{cases}$$

$$= \begin{cases} a_U - \mathbb{E}[\tilde{v}|B_2] & \text{if } a_U < \mathbb{E}[\tilde{v}|B_2]. \\ 0 & \text{if } a_U \geq \mathbb{E}[\tilde{v}|B_2]. \end{cases}$$

Clearly, the uninformed market maker is indifferent over the interval $\Big[\mathbb{E}[\tilde{v}|B_2], \mathbb{E}[\tilde{v}|H_1, B_2]\Big]$ and is weakly worse off otherwise. So he does not have a profitable deviation, and makes zero profits in equilibrium. □

B Proof of Proposition 2

The arguments of the proof are contained in the text leading up to the statement of the proposition.

C Proof of Proposition 3

The arguments of the proof are contained in the text leading up to the statement of the proposition.

References

Baldauf, Markus and Joshua Mollner, "Competition and Information Leakage," *Journal of Political Economy*, 2024, *132* (5), 1603–1641.

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

Battalio, Robert and Craig W. Holden, "A simple model of payment for order flow, internalization, and total trading cost," *Journal of Financial Markets*, 2001, *4* (1), 33–71.

Battalio, Robert H., Robert H. Jennings, Mehmet Saglam, and Jun Wu, "Difficulties in obtaining a representative sample of retail trades from public data sources," *Available at SSRN 4579159*, 2023.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, "Tracking retail investor activity," *The Journal of Finance*, 2021, *76* (5), 2249–2305.

Burdett, Kenneth and Maureen O'Hara, "Building Blocks: An Introduction to Block Trading," *Journal of Banking & Finance*, 1987, *11* (2), 193–212.

Chakrabarty, Bidisha, Justin Cox, and James E. Upson, "The information content of retail order flow: Evidence from fragmented markets," *Journal of Banking & Finance*, 2024, *167*, 107275.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, "Cream-skimming or profit-sharing? The curious role of purchased order flow," *The Journal of Finance*, 1996, *51* (3), 811–833.

Engelbrecht-Wiggans, Richard, Paul R. Milgrom, and Robert J. Weber, "Competitive bidding and proprietary information," *Journal of Mathematical Economics*, 1983, *11* (2), 161–169.

Ernst, Thomas, Chester Spatt, and Jian Sun, "Would Order-By-Order Auctions Be Competitive?," *The Journal of Finance*, 2025, *80* (4), 1879–1927.

Glebkin, Sergei, Bart Zhou Yueshen, and Ji Shen, "Simultaneous Multilateral Search," *The Review of Financial Studies*, 2023, *36* (2), 571–614.

Glosten, Lawrence R. and Paul R. Milgrom, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 1985, *14* (1), 71–100.

Hendershott, Terrence and Ananth Madhavan, "Click or Call? Auction versus Search in the Over-the-Counter Market," *Journal of Finance*, 2015, *70* (1), 419–447.

Ho, Thomas and Hans R. Stoll, "The Dynamics of Dealer Markets Under Competition," *Journal of Finance*, September 1983, *38* (4), 1053–1074.

Hotelling, Harold, "Stability in Competition," *The Economic Journal*, 1929, *39* (153), 41–57.

International Organization of Securities Commissions, "Follow-On Analysis to the Report on Trading of OTC Derivatives," Technical Report, IOSCO January 2012. IOSCO Report OR02/12.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, "Individual investor trading and return patterns around earnings announcements," *The Journal of Finance*, 2012, *67* (2), 639–680.

Kelley, Eric K. and Paul C. Tetlock, "How wise are crowds? Insights from retail orders and stock returns," *The Journal of Finance*, 2013, *68* (3), 1229–1265.

Lee, Charles M. C. and Mark J. Ready, "Inferring trade direction from intraday data," *The Journal of Finance*, 1991, *46* (2), 733–746.

Pinter, Gabor, Chaojun Wang, and Junyuan Zou, "Information chasing versus adverse selection," *Working Paper*, 2025.

Röell, Ailsa, "Regulating information disclosure among stock exchange market makers," *Working Paper*, 1983.

Schwarz, Christopher, Brad M. Barber, Xing Huang, Philippe Jorion, and Terrance Odean, "The "Actual Retail Price" of Equity Trades," *Journal of Finance*, forthcoming.

Securities and Exchange Commission, "Fact Sheet: Proposed Rule to Enhance Order Competition," https://www.sec.gov/files/34-96495-fact-sheet.pdf 2023. [Online; accessed Nov-08-2025].

_ , "Order Competition Rule," *Federal Register*, January 2023, *88* (128), 128–245.

Upson, James, Bidisha Chakrabarty, and Justin Cox, "A (Sub-) penny for your thoughts: Using trade reporting lags to identify retail trades," *Available at SSRN 5125899*, 2025.