Eric Budish
Paul G. McDermott Professor of Economics and Entrepreneurship
The University of Chicago Booth School of Business
[redacted]

March 31st, 2023

Ms. Vanessa A. Countryman, Secretary
U.S. Securities and Exchange Commission
100 F Street NE
Washington, DC 20549-1090

Re: Order Competition Rule Proposal (Release No. 34-96495; File No. S7-31-22)

Dear Ms. Countryman,

I appreciate the opportunity to comment on the Commission's proposed Order Competition Rule. I am an economics professor at the University of Chicago Booth School of Business who researches market design – designing the "rules of the game" for markets, to borrow Milton Friedman's phrasing – with a specific focus on the design of financial exchanges. Market design research assumes that participants in a market act optimally in their rational self-interest with respect to market rules, but takes seriously the possibility that the market rules themselves may be sub-optimal. I believe that this approach brings a useful perspective to debates over equity market structure. I write independently and have no financial conflicts of interest.

In this letter I will cover the following 4 topics:

Topic 1. I review the economic reasons why pricing in the proposed retail auctions should be expected to be significantly better for retail investors than pricing in the "lit market", i.e., on continuous order book exchanges. This comparison is not exactly the question at hand but helps set up the relevant economic issues. The two sources of savings are: reduced adverse selection and elimination of latency arbitrage.

Topic 2: I discuss the economic reasons why pricing in the proposed retail auctions should be expected to be better for retail investors than pricing on wholesale execution venues. The wholesale execution venues also benefit from reduced adverse selection and the elimination of latency arbitrage, as compared to the lit market. The sources of potential savings for the retail auctions relative to the wholesale execution venues are: reduced payment-for-order-flow expense to brokers, and reduced net revenue (i.e., net of PFOF expense) to liquidity providers due to broader and more open competition.

Topic 3: I briefly discuss the relationship between the proposed retail auctions and the financial market auction design proposed in my own past research, frequent batch auctions. The key difference is that the proposed retail auctions are designed to complement on-exchange trading on a continuous order book market, whereas frequent batch auctions are designed as an exchange design alternative to or competitor to the continuous order book market design. The proposed retail auctions have many of the same benefits as frequent batch auctions, but for off-exchange trading for a set of participants (i.e., retail investors) rather than on-exchange trading for the whole market.

Topic 4: I discuss some of the specific auction design details for the proposed retail auctions.

Topic 1. The economics of why pricing in the proposed retail auctions should be expected to be significantly better for retail investors than pricing in the "lit market", i.e., on continuous order book exchanges.

A. Background on Spread Decompositions

The SEC's analysis makes frequent use of bid-ask spread decompositions that are widely used in the empirical market microstructure literature. In this decomposition, originally suggested by Glosten (1987) and closely related to the famous Glosten and Milgrom (1985) model, the spread paid by the investor ("Effective Spread"), which is a measure of the cost of accessing liquidity,[1] can be split into two terms, "Price Impact" and "Realized Spread." Price Impact is measured by looking at the signed change in the midpoint price between the time the trade occurred and some modest distance out into the future (e.g., the SEC analysis uses 60 seconds). For any one trade this measure will contain both signal and noise, but on average this measure tells us how much adverse selection there is in the market. That is, conditional on selling at the offer, how much on average will the price subsequently go up, or, conditional on buying at the bid, how much on average will the price subsequently go down. Realized Spread is then just defined as the difference between the measured Effective Spread and the measured Price Impact. The interpretation is that it represents a liquidity provider's profits after accounting for adverse selection. In the Glosten and Milgrom (1987) model, if there is free entry into liquidity provision, the realized spread will be zero in equilibrium. This is a stylized benchmark, in the way that price equals marginal cost is a stylized benchmark in economics more generally, but it is a useful one to keep in mind and that I will come back to.

A complication in bringing this spread decomposition formula to data is that U.S. stock exchanges mostly use a two-sided markets fee structure, in which takers of liquidity pay a fee while providers of liquidity earn a rebate. Conceptually, therefore, the Effective Spread paid by investors should be adjusted upwards for the fee, and the Realized Spread received by liquidity providers should be adjusted downwards for the rebate. The SEC's data provides the latter adjustment, i.e., the average rebate in

---

[1] More precisely, the effective spread is a measure of the cost of liquidity in a single trade, whether at a single price level (e.g., taking at the offer) or as a weighted sum across multiple price levels. For retail investors trading relatively small sums in a single trade, this is a complete measure of the cost of liquidity. For institutional investors trading larger amounts over multiple trades, on the other hand, it is just a proxy for the cost of liquidity. For such investors, better measures of the cost of liquidity take into account the available depth and the impact their trading has on prices as they trade over time. To do this properly requires non-public datasets as one needs to know the institutional investors' desired total trade (usually called the "parent" trade) and the path they took to get there (called "child" trades). Moreover, any such measurement of this cost is sensitive to whether the institutional investor traded in a way that didn't inadvertently tip off the market to their trading intentions. For these reasons, effective spread remains the simplest and most widely used measure of the market's cost of liquidity in empirical research, despite its limitations for measuring liquidity costs for larger investors.

basis points. For simplicity, I will use the same adjustment for the average fee paid in basis points as well; this is imperfect but reasonable for the purposes of discussion.[2] Using the SEC's CAT data from Table 7, focusing on the "All Symbols" column and the "EX" row ("EX" = on-exchange trading), and using the 0.82bps difference between "EX Realized Spread" and "EX Realized Spread Adj Rebate" for the fees/rebates adjustment, we have:

<u>On-Exchange Spread Decomposition, Traditional</u>

EffectiveSpread =	*PriceImpact* +	*RealizedSpread*
(4.00bps)	(4.40bps)	(-0.40bps)

(Please note: all numbers reported in this comment letter are intended as credible back-of-envelope magnitudes, mostly interpretations of quantities reported by the SEC in its proposal, as opposed to a rigorous academic report from original data analysis.)

In words, the data tell us that the average on-exchange trade in the SEC's CAT data (for orders <$200k) pays a spread of about 4 basis points, and, as in the Glosten and Milgrom (1985) model, this spread is entirely accounted for by adverse selection. In fact, the residual Realized Spread is slightly negative, meaning that adverse selection slightly exceeds the spread; this finding has shown up in other empirical studies, including my own, and seems to depend on the horizon used for measuring price impact. O'Hara (2015) provides some discussion of the phenomenon, and Conrad and Wahal (2020) has the most comprehensive evidence of which I am aware.

In my own research, together with Matteo Aquilina and Peter O'Neill, I have found it useful to further decompose Price Impact into two components: a first which reflects the cost of getting "picked off" in high-frequency trading races (i.e., "latency arbitrage"), and a second which reflects traditional adverse selection in which a single informed trader trades in isolation. Measuring the race component of price impact requires a kind of data called "message data", in which the researcher can observe the full back-and-forth message traffic between market participants and the exchange; including, crucially, attempts to trade or cancel that fail, which are the empirical signature of a trading race.[3] In Aquilina, Budish, and O'Neill (2022), we find that the split is about 1/3-2/3: that is, about 1/3 of price impact occurs in latency-arbitrage races, while the remaining 2/3 of price impact occurs in non-race trading. (Races constitute about 20% of all trading volume in our data, with the modal race lasting between 5-15 microseconds). If we extrapolate this 1/3-2/3 split to the spread decomposition above based on the SEC's CAT data, then we have:

---

[2] In Budish, Lee and Shim (2022), we report that on average the difference between the taker fee and the maker rebate, for the five largest exchanges by market share, is about $0.0002 per share traded. This corresponds to a taker fee of $0.0030 and a maker rebate averaging $0.0028. The SEC's proposal filing reports similar magnitudes reported by industry contacts. This means that the rebate is about 7% lower than the fee, on average, so we could use, as a ballpark, a take fee of 0.88bps versus a rebate of 0.82bps. This would change the equation below to 4.06bps = 4.40bps + (-0.40bps) + 0.06bps, where this last 0.06bps is an exchange net fee capture term.

[3] Other researchers and industry participants are encouraged to conduct their own studies of exchange message data, if they can obtain such data from an exchange. Code from our study is available with extensive documentation at: https://github.com/ericbudish/HFT-Races

On-Exchange Spread Decomposition, Split Between Race and Non-Race Price Impact

$$\underset{(4.00\text{bps})}{\textit{EffectiveSpread}} = \underset{(1.47\text{bps})}{\textit{PriceImpact\_Races}} + \underset{(2.93\text{bps})}{\textit{PriceImpact\_Non-Race}} + \underset{(-0.40\text{bps})}{\textit{RealizedSpread}}$$

In words: extrapolating the Aquilina, Budish and O'Neill (2022) price-impact results to the SEC CAT data reported in Table 7, indicates that of the average effective spread of 4 basis points (inclusive of fees), just under 3 basis points is accounted for by traditional adverse selection, just under 1.5 basis points is accounted for by the cost of getting picked off in latency arbitrage races, and the remainder is a slightly negative realized spread term (-0.40bps).

B. Implications for Retail Auctions' Cost of Liquidity

This background in hand, we can discuss the economic reasons why the proposed Retail Auctions should be expected to lower the cost of liquidity relative to on-exchange trading via the continuous order book. While this is not the policy comparison directly at hand, this is the best way to set up the relevant economic issues.

There are two reasons why the Retail Auctions will have a lower cost of liquidity than the lit market.

First, the retail auctions are not vulnerable to latency arbitrage. This is true of many off-exchange trading protocols, including wholesale execution and midpoint matching venues, reflecting that off-exchange trading often does not require participants to post orders that are vulnerable to getting "picked off" in a latency arbitrage race. This first savings is 1.47bps in the decomposition just above, which I reiterate is based on extrapolation from one study to another (it would be great if researchers or regulators could obtain exchange message data from a U.S. stock exchange, facilitating an analogous study to Aquilina, Budish and O'Neill (2022), using our publicly available code).

Second, retail investors are less likely to be informed, and hence a source of traditional adverse selection, than institutional investors trading on exchange. In the same Table 7 referenced above, the figure for Wholesale venue Price Impact is 1.26bps. This is (1.26bps)/(2.93bps)=43% of the non-race price impact in on-exchange trading, suggesting that a single retail investor order has about 43% of the adverse selection of a single institutional investor order.[4]

It is hard to make a concrete prediction about realized spread in the proposed Retail Auction; zero is probably the best conceptual benchmark, and -0.40bps would be the figure that matches the observed Realized Spread on exchange in the data in Table 7. Table 6, which uses the Rule 605 data, has a realized spread adjusted for rebates of about zero (-0.001bps). In the range of Realized Spreads from 0.00 to

[4] 43% may feel high, but keep in mind that retail investors tend to trade in a single order whereas institutional investors tend to trade in a much larger number of orders over a period of time. Again, as emphasized in fn. 1, effective spread is a proxy for the cost of liquidity, not the full cost, and it is a better proxy for retail investors trading a single time than for institutional investors trading a large quantity over a period of time.

-0.40bps, this implies that the Effective Spread for retail investors on the Retail Auctions should be in the range 0.86bps to 1.26bps. This is a 68-78% reduction versus spreads on-exchange of 4.00bps.

Next, we will turn to the comparison versus wholesale execution.

Topic 2. The economic reasons why pricing in the proposed retail auctions should be expected to be better for retail investors than pricing on wholesale execution venues, though by a smaller margin than in the comparison to on-exchange trading.

It is widely understood that wholesale execution venues are able to provide better prices to retail investors than they would receive on the lit market because retail investors are less informed (i.e., less "toxic" or less adversely selecting) than institutional investors. Not as widely appreciated is that wholesale execution venues also benefit, relative to liquidity provision on-exchange, from not having to post quotes that are vulnerable to getting picked off in latency-arbitrage races. Intuitively, providing liquidity via quotes on the lit market is more expensive than providing liquidity via a wholesale venue <u>both</u> because the quote on the lit market is more likely to be consumed by an informed investor <u>and</u> because the quote on the lit market is vulnerable to getting picked off in a high-frequency-trading race if it becomes stale.

The data discussed above indicate that, in total, the savings from these two channels is

PriceImpact_On-Exchange - *PriceImpact_Wholesale* = *PriceImpact_Delta*
(4.40bps) (1.26bps) (3.14bps)

and that of this 3.14bps of combined lower price impact, about 2.93-1.26=1.67bps is the lower price impact of retail investor trades (i.e., lower toxicity) and 1.47bps is the avoidance of latency arbitrage costs.

The difference in effective spreads between the lit market and the wholesale market, using the Table 7 data, is:

EffectiveSpread_On-Exchange - *EffectiveSpread_Wholesale* = *EffectiveSpread_Delta*
(4.00bps) (2.11bps) (1.89bps)

In words, of 3.14 basis points of reduction in price impact comparing wholesale execution of retail orders versus on-exchange trading, 1.89 basis points are passed on to retail investors in the form of lower effective spreads. This leaves 1.25 basis points of reduced price impact that is not passed on to retail investors.

A slightly different way of viewing these same data is to focus on the Realized Spread on wholesale markets. This is the difference *EffectiveSpread_Wholesale* - *PriceImpact_Wholesale* = 2.11bps - 1.26bps = 0.85bps. This 0.85bps can be interpreted as liquidity provision revenue in the Wholesale market. This revenue is split in practice between brokers (via PFOF) and wholesale execution venues.

The difference between this +0.85bps of *RealizedSpread* in the wholesale market and the (-0.40bps) of *RealizedSpread* in on-exchange trading reconciles to the 1.25bps discussed in the previous paragraph.

The proposed Retail Auctions also would have the benefits currently enjoyed in the wholesale market of the reduced adverse selection of retail order flow relative to on-exchange trading and the avoidance of latency arbitrage costs due to price competition in the auction. In addition, there is credible reason to believe that competition in the auction would lead to lower realized spreads, or equivalently, greater pass through of the reduction of price impact costs, than the current wholesale market. There are two reasons. First, the auctions would eliminate a large expense faced in the current wholesale market, payment for order flow (PFOF) to retail brokers. Second, the auction, by being open to all market participants who can respond to the auction message within a reasonable period of time, would have broader competition, which should lower liquidity-provider net revenue. In particular, institutional investors could compete directly for the right to trade against a particular retail investor's order.

If we use a realized spread of zero as the competitive benchmark, based on both the logic of Glosten and Milgrom (1985) and the observed low or even negative realized spreads in on-exchange trading, this implies an Effective Spread in the auction of 1.26bps. This constitutes a 40% reduction versus the 2.11bps in the wholesale market.

Topic 3: The relationship between the proposed Retail Auctions and the financial market design proposed in my own past research, Frequent Batch Auctions.

My past research, together with collaborators Peter Cramton and John Shim, showed that the continuous order book exchange design used widely in financial markets around the world, including the U.S. stock market, has an important design flaw. Because requests to trade are processed serially (i.e., one-at-a-time) in a continuous process (i.e., instant by instant), displayed limit orders are vulnerable to getting "picked off" in trading races every time there is new public information about any instrument's value, e.g., a change in the price of a correlated asset. The rents in these trading races create a never-ending race for trading speed. If, instead of the continuous order book, exchanges used very frequently conducted batch auctions (e.g., once per millisecond or potentially even finer), then new public information would create competition on price, to win the auction, rather than competition on speed, to be fastest to pick off the stale quote. This reduces the cost of liquidity provision, because liquidity providers don't have to account for the cost of getting picked off in the price they charge for liquidity. Hence, spreads would be narrower and markets deeper.

There are some important design differences between Frequent Batch Auctions as we proposed in Budish, Cramton and Shim (2015) and the SEC's proposed Retail Auctions. These differences reflect different design goals. We designed Frequent Batch Auctions as a potential alternative to the continuous order book design. Time is discrete (e.g., once per millisecond) and orders are processed in batch using auctions, but otherwise all design details are very similar to the continuous market. In particular, orders remain outstanding until executed or canceled, and the state of the order book is displayed publicly, updating in discrete time. So, there is a notion of the best-bid-and-offer just like on a continuous order book market, there is a notion of taking at the offer or selling at the bid just like on a continuous order book market, etc. The only substantive difference is that if there is a burst of activity in the discrete time interval – e.g., there is a change in the price of a correlated asset that many trading algorithms are responding to – then this burst of activity is processed in batch, using an auction, rather than serially in order of arrival.

The SEC's proposed Retail Auctions, on the other hand, are not intended as an alternative to the continuous order book. Rather, they are intended to facilitate price competition for retail investor orders within a range of prices that references the prices discovered by the continuous order book. In this way, the proposed Retail Auctions have some similarity to other forms of off-exchange trading, such as midpoint matching venues, even though they technically would be conducted by exchanges.

This difference in intent drives the key difference in auction design, which is that the SEC's proposed Retail Auctions are "order driven" rather than occurring in frequent, discrete, time intervals. Specifically, if a retail investor order arrives at time *t*, then the SEC's proposed Retail Auction would take place in a window of time starting at time *t*. The window in time is long enough to allow multiple different kinds of trading counterparties time to respond (I will comment on the time window below as part of Topic 4).

I think this difference in design is appropriate given the difference in intended use case. The SEC's proposed Retail Auction is not intending to transforms "bursts of competition" from competition on speed into competition on price. Rather, its goal is that when a single retail order arrives to market – which occurs idiosyncratically, since retail investors do not trade using high-frequency trading algorithms and direct price feeds – the order benefits from the vigorous competition of an auction.

We can use the SEC's data analyzed earlier in this letter to compare the SEC's proposal, of auctions specifically for retail investors, to the alternative of Frequent Batch Auctions for all investors. I do this not as an alternative policy proposal but to illuminate the economics in play.

As discussed above, the Effective Spread on the lit market is 4.00bps, and the Effective Spread on the proposed Retail Auction, using the high end of the range, would be 1.26bps.

Suppose there were a single integrated Frequent Batch Auction market in which all investors interacted, including both retail and institutional. We can create a blended non-race price impact figure by using the data from Table 1, which says that Exchanges constitute 59.7% of share volume and Wholesalers constitute 23.9%. Weighting the retail price impact of 1.26bps and the on-exchange non-race price impact of 2.93bps by their respective weights (i.e., by (23.9%/(23.9%+59.7%))=28.6% and by (59.7%/(23.9%+59.7%))=71.4%) gives a blended, non-race price impact of:

$$\textit{Blended\_PriceImpact\_Non-Race} = 28.6\%*1.26\text{bps} + 71.4\%*2.93\text{bps} = 2.45\text{bps}$$

Thus, the comparison is as follows. Under the SEC's proposed Retail Auctions, we can expect retail investors to enjoy an effective spread of 1.26bps, while on-exchange trading using the continuous order book continues to have an effective spread of 4.00bps. Alternatively, if both retail and institutional investors participated in an integrated Frequent Batch Auction exchange, we could expect the effective spread to be 2.45bps.[5] This would be worse for retail investors, who would no longer be able to benefit from their reduced adverse selection in a segmented market. It would be better for institutional investors and other investors who trade on-exchange, because they would benefit both from the elimination of latency arbitrage and from the opportunity to blend their own trading with less-informed retail investor order flow.

Topic 4. Comments on specific design details of the SEC's proposed Retail Auctions

Last, I would like to make a few comments on specific aspects of the proposed auction design.

First, I think it is reasonable to structure the auctions as "on-demand" as opposed to on a pre-announced schedule (whether every 1ms, every 100ms, etc.). As discussed above, retail orders are likely to arrive to the market relatively idiosyncratically as opposed to in bursts, so having an on-demand auction every time a new order arrives for a particular symbol seems reasonable.

Second, my instinct is that the 100ms time horizon may be longer than is technologically necessary to create a level playing field across a broad set of industry participants. I would have guessed, from my own discussions with industry participants over the years, that 1 millisecond might be long enough for this purpose, but the SEC and industry commenters will certainly have better information than I do. Whether 1ms or 100ms, we are talking about algorithmic decision making as opposed to human decision making. So, the relevant inquiry is how long it will take different kinds of market participants to algorithmically respond to a retail auction message.

Third, and related to the first two points, I would like to discuss batching of retail orders. The reason batching is so important to the FBA proposal in my previous work is precisely to reengineer competition when algorithms are responding to a new public signal. Here, the algorithmic responses to a retail order are indeed batched, using the proposed auction, whereas the retail orders themselves need not be. I suspect that this is fine. For most symbols, the probability of multiple retail investors arriving within the same 1-100ms interval is likely very small. But, that is an empirically measurable quantity, and if such coincidences are more common than I anticipate, then it might be worth exchanges thinking about how to handle this case. I think it is wise that the SEC proposal language does not specifically mandate or prohibit that exchanges deal with multiple retail investors arriving at around the same time in a specific way.

Fourth, I would like to discuss information leakage. I have seen some observers talk about information leakage and I agree it is an important thing to think about. Information leakage is one reason why on-demand auctions for institutional investors may give such institutional investors pause – initiating the auction reveals one's intent to trade a large quantity, which could be problematic even if the direction of trade isn't revealed. FBAs as proposed in my research would allow institutional investors to trade

[5] I want to repeat the caveat from earlier that all numbers in this letter are meant to be credible back-of-envelope magnitudes as opposed to a rigorous academic report from original data analysis. In particular, this number does not account for the equilibrium effects that a significant reduction in the cost of trading on-exchange would have on the composition and quantity of trade.

much as they currently do, breaking their trading into small orders spread out over time, attempting to blend in with other trading. Recent research on "Flow Trading" with Cramton, Kyle, Lee and Malec (2023) indeed builds this capability directly into the market design. Here, though, the fact of a retail investor wanting to trade a small amount of some stock (<$200k) does not seem like information leakage that should move the needle on price. That said, it seems worth explicitly clarifying that using the indication of retail interest as a signal to trade on before the retail auction executes is against the rules. I will note as well that a faster auction interval, such as 1ms, would help mitigate this concern of information leakage.

Concluding Remarks

Auctions are a classic form of market competition, and auctions for retail investor orders should help reduce costs and broaden competition. As I have written elsewhere, whether the potential savings sound large or small depends on the vantage point from which they are viewed. The potential cost savings per transaction are quite small: relative to trading on the wholesale market, at most about 1 basis point; relative to trading on-exchange, a few basis points. But, over a large number of transactions, a basis point or two adds up to significant sums.

Sincerely,

Eric Budish

<u>Bibliographical Details of Academic Papers Referenced</u>

Aquilina, Matteo, Eric Budish and Peter O'Neill (2022). "Quantifying the High-Frequency Trading 'Arms Race'." *The Quarterly Journal of Economics*, 137(1), 493-564. Data and Code Appendix at https://github.com/ericbudish/HFT-Races.

Budish, Eric, Peter Cramton and John Shim (2015). "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." *The Quarterly Journal of Economics*, 130(4), 1547-1621.

Budish, Eric, Robin Lee and John Shim (2022). "A Theory of Stock-Exchange Competition and Innovation: Will the Market Fix the Market?" NBER Working Paper No. 25855.

Budish, Eric, Peter Cramton, Albert S. Kyle, Jeognmin Lee, and David Malec (2023). "Flow Trading." Becker Friedman Institute Working Paper.

Conrad, Jennifer and Sunil Wahal (2020). "The Term Structure of Liquidity Provision." *Journal of Financial Economics*, 136(1), 239-259.

Glosten, Lawrence R. (1987). “Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices.” *The Journal of Finance*, 42(5), 1293-1307.

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

O’Hara, Maureen (2015). “High-Frequency Market Microstructure.” *Journal of Financial Economics*, 116(2), 257-270.