



RESEARCH@BESTEXRESEARCH.COM
BESTEXRESEARCH.COM

THE GOOD, THE BAD & THE UGLY OF PAYMENT FOR ORDER FLOW

MAY 3, 2021

Hitesh Mittal

Founder & CEO

Kathryn Berkow

Senior Researcher

INTRODUCTION

Payment for order flow (PFOF) and the market structure related to how retail equity orders execute has always been a controversial topic but has recently become a center of debate among more market participants than ever before in light of the Robinhood-GameStop controversy.

In simplest terms, retail brokers like Robinhood send their clients' orders to market makers also known as wholesalers. Wholesalers commit to providing retail brokers liquidity at the current national best offer (when the investor is buying) or national best bid (when the retail investor is selling)—or better. Wholesalers make money by earning much of the difference between the offer and the bid—the bid-offer spread. Wholesalers pay part of that profit in form of rebates, PFOF, to the retail broker who sent them the orders (frequently called “flow”).

The practice of PFOF began decades ago and was mainstreamed by the controversial figure Bernie Madoff himself, who once justified paying retail brokers by comparing the practice to hiring salespeople on commission to bring order flow to a broker. But a lot has happened since then. Markets have become almost completely electronic, minimum tick size has gone from $1/16^{\text{th}}$ of a dollar to $1/100^{\text{th}}$ (\$.01), and retail brokerage commissions have declined to almost zero. As the markets evolved, so has the relationship between wholesalers and retail brokers. In the post-decimalization era, some retail brokers, such as Fidelity and E*TRADE created their own wholesalers to trade against their retail flow, but later shut down those operations in favor of sending that order flow to external wholesalers. A new breed of wholesalers was born that specialized in electronic trading and could profit in an environment with tighter bid-offer spreads. Electronic market making firms like Knight, ATD, and Citadel quickly became some of the largest wholesalers. Some banks also attempted to get into the wholesaling business, for example UBS through a multi-year deal with Schwab and Citi by purchasing ATD, but their wholesale businesses have diminished over a period of years. Today, there are a few market making firms that control the majority of order flow. The largest of these firms is Citadel, controlling about half of all the retail order flow, while Virtu controls just over one quarter of retail order flow through its acquisition of Knight.

Over the last ten years, dozens of charges have been brought by the SEC against wholesalers and retail brokers related to best execution, insufficient disclosures, and misleading clients. But the practice has only continued to grow. Now more than 20% of the entire US equity market trades with a few wholesalers who receive order flow from retail brokers directly rather than interacting with the volume trading on exchanges.

PFOF is front and center again as a result of the “GME-gate” controversy. Retail investors are fuming at the prospect that wholesaling market makers may have influenced Robinhood into stopping trading in GME. Those against PFOF equate it to “legal bribery” when retail brokers sell order flow to wholesalers in exchange for revenues and allege that wholesalers are front running their orders. Others say that PFOF improves outcomes for retail investors, reducing their trading costs more than ever. As is almost always the case, the truth is likely somewhere in the middle. In this paper, we unpack the market structure surrounding wholesaling, analyze its impact on investors' execution quality, and recommend ways to improve market structure for retail investors and all market participants.

Contents

INTRODUCTION	2
MARKET MAKERS AND THEIR PROFIT	4
THE “GOOD” OF PAYMENT FOR ORDER FLOW	5
THE BAD & THE UGLY	7
Real Price Improvement Statistics are Overestimated by at Least 8% of the NBBO Spread.....	8
Moving Retail Flow to Exchanges Would Narrow NBBO Spreads by 25%.....	11
Retail Limit Orders Receive Poorer Execution in the Current Market Structure.....	14
Information Asymmetry Creates a Monopolistic Environment and Increases Costs.....	14
CONCLUSION	18
ACKNOWLEDGMENTS	19
APPENDIX	20

MARKET MAKERS AND THEIR PROFIT

To understand the implications of PFOF, one must first understand the role of market makers and how they compete in public and private markets.

Investors can access liquidity in two ways: they can place limit orders when they do not need immediate execution, or market orders or marketable limit orders (orders to buy at the best offer price or sell at the best bid price) if they want immediacy. Those desiring immediacy pay a premium for it, “crossing the spread”—the distance between the bid and offer—to trade, and their counterparties earn the corresponding premium for supplying their liquidity.

One of the nice properties of public exchanges is that they allow a complete disintermediation of middlemen by allowing investors to submit limit orders that have the same priority as limit orders from market makers, and these limit orders can interact with other investors’ market orders. But limit orders placed by investors on the opposite side may not always be sufficient to satisfy the demand for immediacy of other investors, and investor limit orders may not offer the best price for the investor market orders willing to cross the spread.

Market makers on public exchanges solve both of these problems by providing bids and offers that may be priced better than other investors’ limit orders. Such quotes from market makers provide immediacy to investors who need it and in return earn the bid-offer spread (also referred to as the “spread” or “NBBO spread”, which we will use going forward). Since public exchanges are anonymous, there is no way for investors to distinguish between a limit order from a market maker and that of another investor. Similarly, a market maker placing a limit order cannot distinguish one type of counterparty from another.

Earning the spread is not easy for market makers. To earn the full spread, prices must not change between the time they put on a position (e.g., buying from a seller’s incoming market order) and the time they offload that inventory (e.g., selling what they purchased). Since prices often change, market makers take the risk that the prices may go down (or up) after a seller sells to them at their bid price. Market makers would be less worried about price risk if it were symmetric—if the odds were the same that the price goes up after buying at the bid as the potential for the price to go down. In that case, they would sometimes make additional profit on top of the full spread and sometimes less, but on average they would earn the full spread.

Unless a market maker can choose their counterparty (which they cannot in a public exchange), the price risk they face is asymmetric. Prices are more likely to go down after a market maker buys from an investor and up after a market maker sells to an investor. This phenomenon is described as “adverse selection” as market makers fill more orders at the bid when the prices are about to go down and at the offer when the prices are about to go up. This adverse price movement—known as adverse selection cost—directly reduces the profit market makers earn for providing liquidity and sometimes even creates losses.

Adverse selection happens because market makers (or other investors submitting limit orders) only control the timing of order submission; the timing of their actual *execution* is controlled by liquidity takers—traders submitting the market order (or marketable limit order) that trades against that limit order. There are a variety of liquidity takers in the marketplace, and they can be categorized into three segments—HFT liquidity takers, institutional traders with long-term investment horizon, and retail traders. “Toxicity” is a term often used to categorize the flow that adversely selects the market makers, with HFT traders considered “high” in toxicity and retail investors considered “low” or “no toxicity”.

Each investor has a different investment time horizon planned for holding their investment. The longer the investment horizon, the less likely prices will move against the market maker in the short time horizon they need to offload their position. Retail order flow is often nontoxic because their trades are smaller and there is no short-term alpha in their trades, meaning they do not adversely select the market makers providing their liquidity.

Institutional investor flow is somewhat more toxic because they tend to break up their large orders into smaller ones, sending the smaller pieces for execution on exchanges and in off-exchange venues. Market makers trading against these

smaller slices of a larger order stand to lose money as the additional slices continue to move the market against the market maker.

But when liquidity taking HFT firms and other short-term, event-driven investors send their market orders to exchanges, prices are likely to move against market makers more quickly than they can offload their positions. HFT liquidity taking activity is likely to have short-term alpha, but such traders are also likely to trade more aggressively than others due to their conviction in short-term prices, creating price impact and even bigger losses for those providing their liquidity.

Market makers view the attractiveness of a counterparty trading against their limit order as shown in Figure 1. Flow from retail investors—“retail flow”—is the least toxic and the flow from short-term professional traders is the most toxic, as depicted in Figure 1.

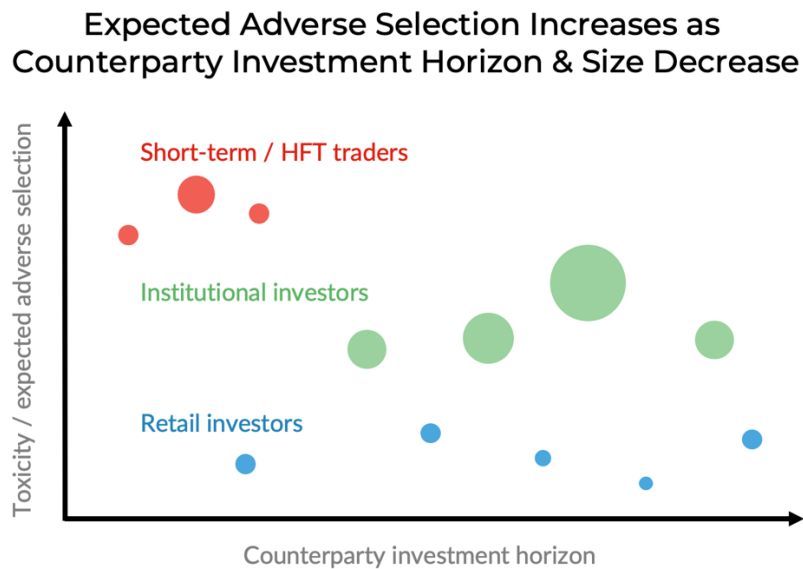


Figure 1. On average, retail investors represent a mixed bag of investment strategies and horizons. For this reason, their trade flow is not directional, and counterparties can expect very little adverse selection after execution. As investors' horizons become shorter and/or trade volume becomes larger, counterparties fear the price momentum associated with execution. Market makers who unknowingly execute trades against short-term HFT traders expect prices to move against their trades, reducing the associated profit margin for providing liquidity. This image is a generalization, as there is wide variety in the investment horizon of each type of investor and adverse selection is highly variable.

THE “GOOD” OF PAYMENT FOR ORDER FLOW

The terms “payment for order flow” and “selling order flow” tend to have a negative connotation, as if the order information itself is being sold to wholesalers. Critics of PFOF think wholesalers are “front-running” retail investor orders and that markets are rigged against Main Street investors. But there is simply nothing incendiary about the wholesaler business model¹ in general. Since retail investors generally execute their trades in a single order, concerns about “front running” don’t really apply.

Think of market making as a simple business, where the product the market maker is selling is access to immediate liquidity. The price that a market maker charges for its goods is the NBBO spread, and its cost-of-goods sold is any corresponding adverse selection². To maximize profits, market makers work to maximize the quantity of goods sold times their profit margin. In a competitive market there is tension between profit margin and quantity of goods sold; the tighter the profit margin (i.e., the narrower the spread they are willing to offer) the more goods they must sell (i.e.,

¹ For more discussion of the wholesaler business model, readers can refer to [this 2014 article](#) by Cliff Asness, Aaron Brown, Michael Mendelson, and Hitesh Mittal.

² Of course, there are other costs associated with a market making business, for example extensive technology costs, the cost of capital, regulatory costs, and people costs, but here we are focusing on adverse selection as it is the least understood aspect of cost and a cost directly associated with the function of the service provided.

increased volume). Adverse selection costs faced by market makers create a natural lower bound for spreads, since they must earn enough to cover this cost and still earn some profit.

Since public exchanges are anonymous, all liquidity taking investors receive the same prices and NBBO spread. More toxic flow on public exchanges is subsidized by less toxic flow since market makers set their quoted spread using information about the **average** adverse selection they face³.

Since they expect to retain the entire spread when interacting with this less-toxic retail flow, some market makers, also known as wholesalers, have side-stepped exchanges by going directly to retail brokers representing such flow. They offer a portion of their increased profit trading against retail flow to the retail broker who provided it as “payment for order flow” (PFOF). Over time, retail brokers began asking wholesalers to split PFOF into a smaller PFOF and a price improvement for retail investors, improving the optics of this arrangement since retail investors were getting better prices than were available on exchanges. Figure 2 illustrates the interaction between the retail investor and the market maker and the additional portion of the NBBO spread paid to the retail broker as PFOF.



Figure 2. Market makers offer tighter NBBO spreads to retail investors, forgoing some portion of their expected profit in exchange for reduced adverse selection. For the opportunity to trade against this less-toxic flow, market makers pay a portion of their profit to the retail brokers for sourcing liquidity—PFOF.

While the actual price improvement provided for each trade varies, it averaged at about 24.5%⁴ of the spread across all trades reported by the top 5 wholesalers—Citadel, Virtu, Susquehanna, Two Sigma, and Wolverine—in their December 2020 Rule 605 Reports⁵. The average price improvement provided by each of these wholesalers is illustrated in Figure 3.

³ Each market maker faces different adverse selection based on the proprietary short-term price prediction models they utilize to price their limit orders on exchanges.

⁴ We have removed all “negative” price improvement from statistics presented in this paper, for both wholesalers and exchanges. Negative price improvement often comes from trades reported to the consolidated tape out of order, implying execution “outside the NBBO”, so we have removed all such cases from our calculations (about 4% of shares reported in the 605 Reports).

⁵ Each market center (i.e. market makers, exchanges, and ATSs) is required to file a monthly report called a Rule 605 Report, that contains information about aggregate price improvement and volume associated with limit orders and market orders for each stock executed in that market center during that month.

Price Improvement by Wholesaler

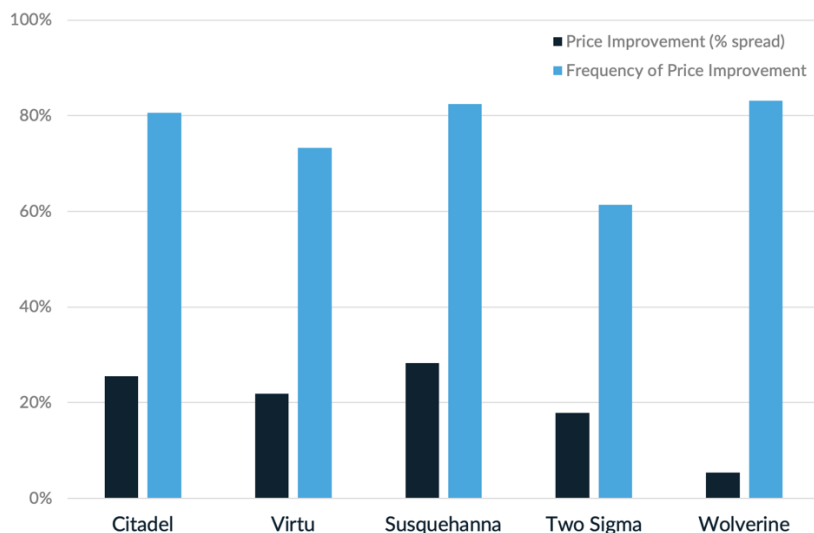


Figure 3. Average net price improvement and frequency of price improvement of the top 5 wholesalers, as presented in December 2020 Rule 605 Reports. Note the average price improvement is a net price improvement incorporating shares experiencing positive or zero price improvement. Frequency of price improvement is the portion of executed shares receiving positive price improvement.

In this arrangement, the market maker, retail broker, and retail investor all “win”. Without PFOF, each would have paid more (or lost additional profit) for the same transaction. The PFOF arrangement results in a tighter spread paid by retail investors for their immediate liquidity demand, revenue for the retail broker, and a greater portion of the spread earned by market makers—an improved outcome for all.

In absence of this kind of arrangement, all parties would appear to pay higher prices. Market makers would provide their service only on exchanges to anonymous counterparties—unable to provide price improvement to those with less toxic flow—often picked off by large institutional orders and short-term traders, reducing their earned spread for the service of providing liquidity. Retail traders would be getting the same price as anyone else on the exchanges rather than earning that price improvement. Retail brokers would not receive payment for order flow and instead pay exchanges a fee (though they would also have the option of earning rebates for liquidity taking at inverted venues when possible). This lack of revenue for retail brokers would likely lead to increased commissions charged to retail investors.

It is important to note that in exchanges and ATSs, executions at fractions of tick size (one penny for most stocks) are largely limited to pricing at the NBBO midpoint⁶ due to Rule 612, which prohibits submitting orders with limit prices at non-tick size increments. This rule was designed to prevent market participants from getting priority over other limit orders by providing an economically insignificant price improvement⁷. Market makers, however, are permitted to provide execution prices representing fractional tick size in the over-the-counter market. For example, a buy order for a stock with NBBO of \$5.01 and \$5.02 may receive an execution price of \$5.0195 and a buy order for a stock with NBBO of \$50.01 and \$50.10 may receive an execution price of \$50.0852. The regulation only prevents the pricing of orders in fractions of tick size, but the execution price itself could be anything.

THE BAD & THE UGLY

In the tri-party arrangement of retail investor / retail broker / wholesaler, all seem to be benefiting from PFOF. But there are many more implications of PFOF than appear in the arrangement described above. Retail brokers and wholesalers market their price improvement statistics, and there is no doubt that retail investors receive better prices relative to the

⁶ Some ATSs also allow executions priced at a quarter of one tick if a liquidity taker interacts with a midpoint liquidity supplier.

⁷ The only exception to this rule in exchanges and ATSs are hidden orders priced at midpoint that may receive an execution that is not a multiple of tick size, depending on the NBBO. For example, a hidden midpoint order may receive an execution of \$5.015 if the NBBO is \$5.01 and \$5.02.

NBBO prices on the exchanges. But the question remains whether retail investors would further benefit from an alternative market structure that does not involve private arrangements between retail brokers and wholesalers.

In marketing the benefits of the PFOF arrangement for retail investors, wholesalers and retail brokers implicitly assume that if not for wholesalers, retail investors' cost would be the current NBBO spread. This rests on two assumptions:

1. No liquidity exists **inside** the NBBO on exchanges and other liquidity sources
2. The NBBO spread would remain the same even after retail flow moved to exchanges

Each of these assumptions is incorrect. First, there is substantial liquidity hidden within the NBBO that firms providing execution services to institutional brokers frequently locate to improve execution. And second, the NBBO itself would significantly narrow if non-toxic retail order flow moved to exchanges and the information asymmetry between wholesalers and "regular" HFT market makers were eliminated. In the sections to follow, we estimate how much price improvement retail investors could expect to receive in an alternative market structure and how those savings compare to the current savings provided by wholesalers.

Real Price Improvement Statistics are Overestimated by at Least 8% of the NBBO Spread

NBBO prices do not always represent the best bid and best offer for a stock in the US equity market. In addition to standard limit orders, there are both hidden orders—limit orders placed by investors with the instruction not to publish publicly—and odd lot⁸ orders—placed for fewer than 100 shares—that are placed inside the NBBO but not published in exchange limit order books from which the NBBO is formulated. In addition, there is a substantial amount of unpublished liquidity available in off-exchange ATSs priced at the NBBO midpoint.

For example, if the current NBBO is \$10.90 to buy and \$11 to sell, it is possible a hidden limit order is placed at \$10.99 to sell, but market participants remain unaware that there is a better offer available. If a market order to buy is sent to the exchange where this hidden order is posted, it will automatically receive price improvement of \$.01 when it trades at \$10.99—an improvement of 10% of the spread⁹, as illustrated in Figure 4. Similarly, a hidden order placed at the NBBO midpoint on this exchange would provide price improvement of 50% of the spread.

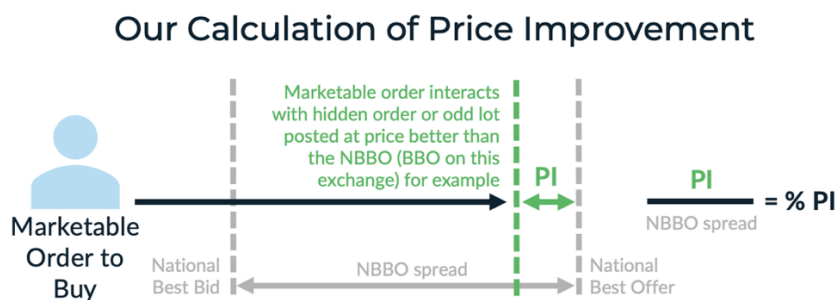


Figure 4. In all calculations of price improvement—those for wholesalers and on exchanges—we use the full NBBO spread in the denominator, and the dollar price improvement for a single trade of a market order against its counterparty in the numerator, as depicted here.

Hidden orders are not a rarity on exchanges, but rather the norm; they tend to represent about 16.7% of daily volume in liquid stocks and 20.8% for illiquid stocks¹⁰. Like hidden orders, odd lot orders are extremely common—especially for high-priced stocks. The narrower spread they create is not captured by the NBBO spread, a metric used as a benchmark by retail brokers to market their total price improvement.

⁸ Recent updates to RegNMS calling for a change in the definition of the round lot size included in the NBBO and PBO (Protected Best Bid and Offer) were finalized in November 2020, but these changes have not yet been implemented.

⁹ In all calculations of price improvement—from both wholesalers and orders on exchanges—we represent the portion of the full NBBO spread saved in a single trade of a market order against its counterparty, as depicted in Figure 4.

¹⁰ Referencing statistics from June 2020 provided by SEC.gov, as included in our earlier paper [Queue-Jumping & Strategic Limit Order Routing](#).

These two types of unpublished orders can tighten the spreads experienced by investors on exchanges. In fact, the price improvement over the published NBBO spread on exchanges is substantial, though not the same as what retail investors currently receive from wholesalers. While retail investors currently receive an average of 24.5% price improvement from wholesalers, other market participants are receiving an average of 8.7% of the spread on exchanges^{11,12,13}. A comparison of price improvement aggregated across the top five wholesalers and that available on exchanges in the same time period is shown in Figure 5.

The calculations in Figure 5, and later in Figure 6AB, come from analysis of TAQ data for all executions in the US equity market during the week of December 7-11, 2020 (though we do not find that the results change much from week to week). Each trade is matched with the prevailing NBBO at the time of execution, and the price improvement is measured as a ratio of the difference between the NBO and trade price for market orders to buy, or NBB and trade price for market orders to sell divided by the NBBO spread at the time of execution. Trades are partitioned by execution venue, grouping exchanges together and off-exchange venues separately. We remove trades that “occur outside the NBBO”, which represent trades reported out of order; for fairness, we also remove negative price improvement from the calculation of price improvement offered to retail investors by wholesalers for the same reason. We average price improvement across trades, weighted by shares traded, in calculations for both exchanges and wholesalers.

Price Improvement as a Portion of the Spread

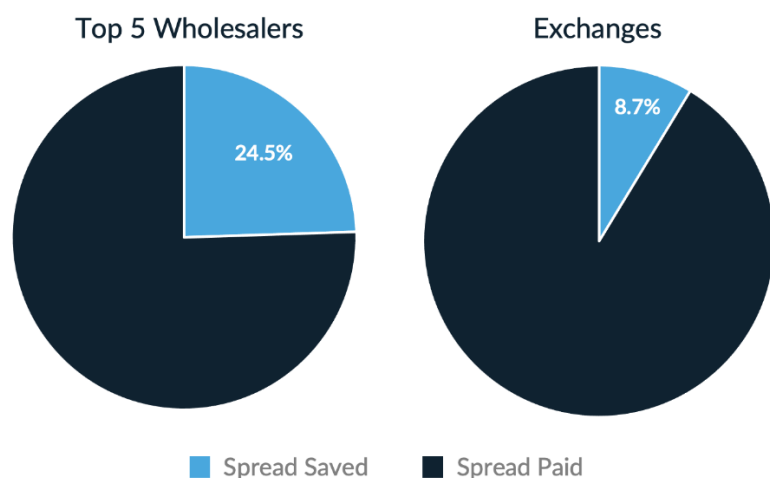


Figure 5. A comparison of the spread savings experienced by retail investors via marketable flow sent to wholesalers and the current price improvement experienced by all investors' market orders on exchanges as a result of hidden orders and odd lot liquidity. Spread savings for the top 5 wholesalers is aggregated from their December 2020 Rule 605 Reports. Note the average price improvement is a net price improvement incorporating shares experiencing positive or zero price improvement. Our exchange calculation does include odd lots (as they represent a major source of price improvement) while the wholesaler calculation does not (odd lots are not included in 605 Reports).

Figure 6AB takes the analysis of price improvement on exchanges further to illustrate price improvement according to individual stocks' liquidity and price. Figure 6A illustrates that the price improvement available on exchanges creeps up a bit to more than 10% for the less liquid stocks in the Russell 2000 Index, while the price improvement offered by

¹¹ Our calculation of price improvement on exchanges does include odd lots, because they represent a major source of price improvement on exchanges, as discussed in the text. However, the wholesalers' price improvement calculation does not include odd lots, as they are not currently required to be included in monthly 605 reporting. Information about price improvement on exchanges excluding odd lots is available in the Appendix.

¹² Our calculation of price improvement for wholesalers includes only marketable order flow—market orders and marketable limit orders.

¹³ Our calculation of price improvement on exchanges aggregates behaviors of all trades receiving zero or positive price improvement, as discussed in the text, weighted by shares traded. However, we also considered the same calculation aggregating price improvement weighted by the shares traded by retail investors in the same stocks (as reported in the top 5 wholesalers' December 2020 Rule 605 reports). The resulting overall average price improvement was 7.9% rather than 8.7%, and similar differences existed in each liquidity and price bucket represented in this analysis.

wholesalers declines a bit for these stocks. Retail investors tend to trade more than the rest of the market in these less liquid stocks, totaling more than 37% of their executed shares, as is illustrated in Figure 7. Figure 6B shows that for higher priced stocks, wholesalers are indeed providing increased price improvement as a percent of [generally wider] spreads, but we do see increases on exchanges as well, where there is often increased hidden and odd lot liquidity available for higher priced stocks.

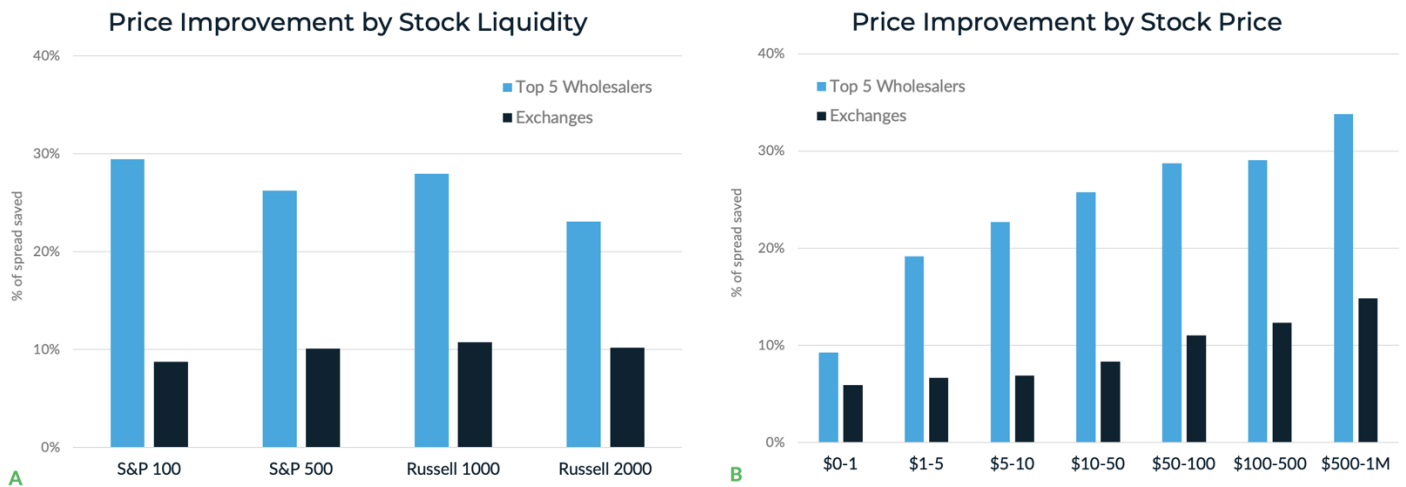


Figure 6AB. Average net price improvement of the top 5 wholesalers in aggregate compared to price improvement experienced on exchanges by stock price by liquidity group (index) (6A) and by price (6B). As before, this calculation is a net price improvement, meaning that it incorporates shares experiencing no price improvement according to the 605 Reports, but it excludes shares receiving negative price improvement.

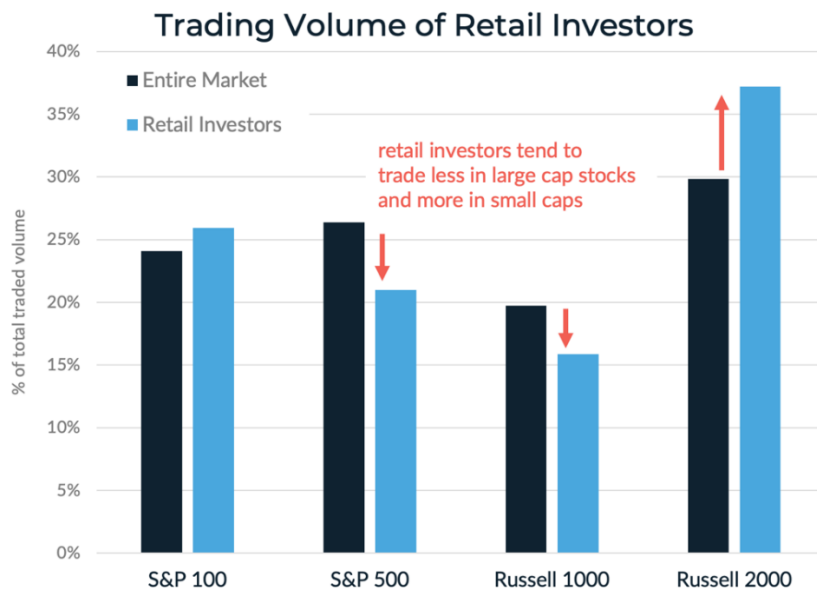


Figure 7. Comparison of the behavior of investors overall to that of exclusively retail investors in December 2020. On average, retail investors tended to trade as many of the highly liquid stocks (S&P 100 Index constituents), but fewer of the remaining S&P 500 and Russell 1000 constituents and relatively more of the less liquid Russell 2000 constituents than other investors. The entire market does include retail investors. Retail investor flow here includes both marketable flow and limit orders.

It is important to keep in mind, however, that our estimates of price improvement are conservative. Our calculations do not, for example, include the executions at the midpoint price that can be received in broker ATSs if retail brokers were to send IOC orders to these ATSs before trading on exchanges. For instance, our execution algorithms receive midpoint executions nearly 10% of the time using this strategy. These statistics do not include the midpoint executions we receive when resting orders in ATSs, because those executions are not applicable for a market order.

Despite all this, there is no denying that retail brokers receive better execution than other market participants for market orders, though the savings may be less than what is marketed. By our estimate, the savings wholesalers are

providing is about 15% above what is available on exchanges (rather than the advertised 24.5% over the NBBO). And we believe this number is far less than it would be if retail brokers also utilized ATSS to check for midpoint liquidity before sending retail market orders to wholesalers. Of course, the reported savings apply only to retail investors' marketable orders and make the assumption that the NBBO itself would not contract if the flow were to move to the exchanges.

Moving Retail Flow to Exchanges Would Narrow NBBO Spreads by 25%

Not only are the price improvement statistics touted by wholesalers and retail brokers overestimated because they do not account for hidden orders and odd lots of liquidity inside the NBBO, but they also rely on the assumption that the NBBO spread itself would remain constant if retail flow moved to public markets.

Since wholesalers intercept the highest quality (least toxic) flow before it arrives on public exchanges, market makers (and other participants) on public exchanges are left to provide liquidity to more toxic order flow. As discussed above, adverse selection is the largest cost for market makers and they must earn enough spread to compensate for the adverse selection they experience just to break even.

As a result, quantifying the difference in adverse selection experienced by limit orders on exchanges trading against non-retail market orders versus that of wholesalers trading against retail market orders will allow us to estimate the spread reduction experienced by market participants if retail flow moved to exchanges. We use three steps to estimate this expected reduction in spread:

1. Estimate the average adverse selection faced by limit orders on exchanges and observe the relationship between NBBO spread and adverse selection
2. Estimate the average adverse selection faced by wholesalers' limit orders trading against retail market orders
3. Using the adverse selection of these types of flow and their relative volumes, calculate the weighted average of adverse selection if retail volume moved to exchanges and corresponding reduction in NBBO spread

Step 1: Estimated Adverse Selection on Exchanges is 61% of Spread

Using TAQ data, we routinely estimate the adverse selection costs¹⁴ incurred by limit orders by comparing the midpoint price at the time of execution to the midpoint price after a series of trades in that stock. Here, we will also use it to verify the direct relationship between the adverse selection costs faced by limit orders on exchanges and NBBO spread costs.

Figure 8 illustrates the adverse selection on average for limit orders on exchanges, where adverse selection is calculated as the difference in the midpoint price at the time of execution and that after a series of trades—1, 5, 15, etc.—as a percentage of the spread in that stock. This average represents only round lot executions at the NBBO on exchanges. Each price change is adjusted for the side of the order, such that a negative price change indicates that the price declined following a sell market order or increased following a buy market order. A negative price change in the chart represents the loss of profit due to adverse selection for the limit order crossing against the market order.

As shown in Figure 8, there is clearly adverse selection for limit orders, but it levels off after 30 trades in the stock. For this reason, we will use a 30-trade time horizon for analyzing adverse selection in the paragraphs to follow, as we compare retail market orders to others. Thirty trades serves as a proxy for a market maker's inventory horizon, yielding a longer time horizon for illiquid stocks that trade less frequently and a shorter time horizon for liquid stocks trading more frequently. Figure 8 shows that the 30-trade adverse selection associated with market orders across all exchanges during continuous trading hours is 60.64% of the NBBO spread at the time of the trade.

¹⁴ This paper cites earlier results based on TAQ data from August 2020, though we have found that adverse selection estimates aggregated over full trading days tend to remain roughly unchanged over time.

Adverse Selection on Exchanges

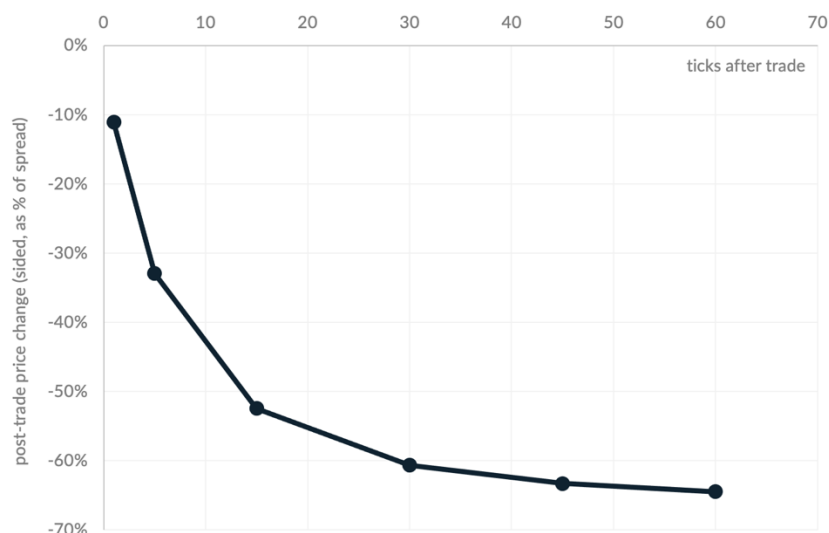


Figure 8. Illustration of price movement following a market maker's execution of a limit order against an incoming market order, on average. Our calculation is sided, such that a negative value means prices went down following a market order to sell or up after a market order to buy. We represent the value as a percent of the NBBO spread at the time of each order's execution. After 30 trades, the price change levels off; as a result, we use a 30-trade time horizon for evaluating adverse selection.

The adverse selection in Figure 8 is represented as a fraction of the NBBO spread at the time of execution because adverse selection and spreads are correlated. To illustrate, Figure 9 compares the average adverse selection for stocks of varying liquidity. Figure 9A shows the average adverse selection as a fraction of price, and Figure 9B shows the adverse selection on the same trades as a fraction of spread. For the most liquid stocks—S&P 100 constituents—the adverse selection is -.8 basis points as a fraction of price, while for the least liquid—Russell 2000 constituents—the adverse selection is -6.1 basis points as a fraction of price. However, when measured as a fraction of spread, for all categories the adverse selection is very similar (as shown in Figure 9B), thus underscoring the direct linear relationship between adverse selection and NBBO spread. This relationship implies that a 1% reduction in adverse selection will similarly translate to a 1% reduction in spread.

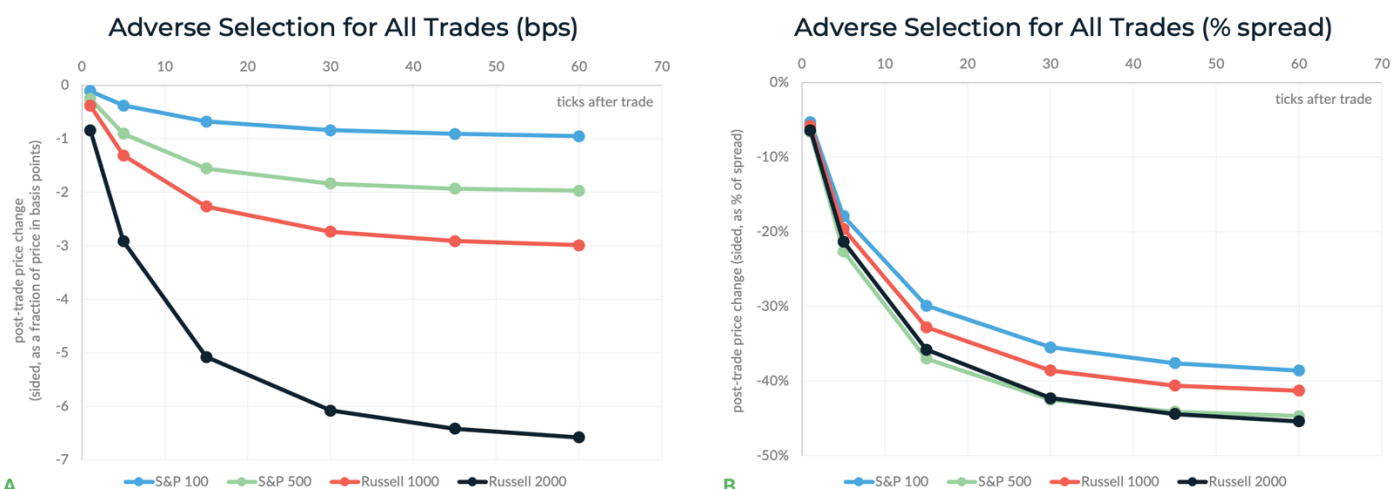


Figure 9. Comparison of average adverse selection for stocks in various equity indices. The calculation is sided, where a negative value indicates prices declining after a market order to sell is executed. Figure 9A shows adverse selection as a fraction of price, while Figure 9B shows the adverse selection on the same trades calculated as a fraction of the NBBO spread at the time of the trade. When measured as a fraction of price, the adverse selection appears to vary dramatically by stock liquidity, but when measured as a fraction of spread, the adverse selection across liquidity groups is roughly constant.

Step 2: Estimated Adverse Selection for Retail Flow is 12% of Spread

Now that we have a measure of average adverse selection on exchanges for Russell 3000 constituent stocks (60.64% of the NBBO spread), we must estimate the adverse selection for retail market orders. TAQ reports all off-exchange trading as a single category, so retail orders submitted to wholesalers cannot be separated from other off-exchange trades, for example, trades in broker-operated ATSs. Therefore, we use sub-penny executions (e.g., a trade priced at \$10.0122) in off-exchange venues (excluding midpoint executions) as a proxy for retail orders. These sub-penny executions are not allowed in ATSs, and hence serve as a good proxy.

Figure 10 compares the adverse selection due to retail market orders (by proxy described above) with that due to exchange market orders for all stocks in the Russell 3000 universe. After 30 trades, the adverse selection generated by retail market orders is only 11.92% of the spread as opposed to 60.64% of the spread on exchanges—a difference of almost 50% of the spread.

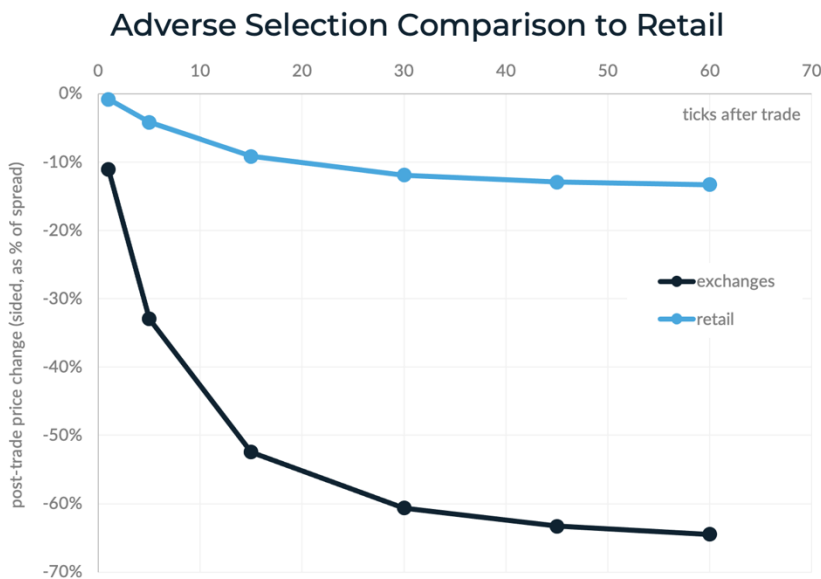


Figure 10. Comparison of adverse selection generated by retail market orders versus market orders on exchanges. The calculation is sided, where a negative value indicates prices declining after a market order to sell is executed. Here, we include all Russell 3000 constituent stocks. Clearly, market orders on exchanges generate much higher adverse selection for their counterparties' limit orders than do retail market orders—60.64% of spread as compared to 11.92% of spread, respectively.

Step 3: Estimated Spread Reduction is 25% if Retail Volume Moved to Exchanges

Analysis of wholesalers' Rule 605 Reports shows that the five largest wholesalers traded over 53 billion shares in December 2020. To put this in context, their total volume represents almost 47% of the volume traded across all sixteen exchanges during continuous market hours¹⁵.

If retail volume were traded on exchanges instead of being received directly by wholesalers, the net adverse selection faced by limit orders on exchanges would be the weighted average of adverse selection from the two types of flow. Using the numbers discussed above, the expected reduction in adverse selection would be over 25%, implying that the NBBO spreads would decline proportionally given their linear relationship—also by over 25%. Our calculations are shown below:

$$\begin{aligned} \text{new adv sel} &= \frac{\text{exchange adv sel} * \text{exchange volume} + \text{retail adv sel} * \text{retail volume}}{\text{exchange volume} + \text{retail volume}} \\ &= \frac{.6064 * 113.21e9 + .1192 * 53.14e9}{113.21e9 + 53.14e9} = 45.08\% \text{ as a fraction of spread} \end{aligned}$$

¹⁵ Volume from the top 5 wholesalers' Rule 605 Reports in December 2020 equates to 21.94% of overall market volume, which is 46.94% of the total volume on exchanges during continuous trading hours (excluding opening and closing auction volume). Our denominator in this calculation comes from Cboe Global Markets US Equity Volume Summary data.

$$\% \text{ change in adv sel} = \frac{\text{new adv sel} - \text{old adv sel}}{\text{old adv sel}} = \frac{.4508 - .6064}{.6064} = -25.66\%$$

Considering a reduction in the NBBO spread by over 25%, the price improvement currently received by retail investors (about 15% of NBBO spread as discussed earlier) is not the “deal” it is advertised to be.

Retail Limit Orders Receive Poorer Execution in the Current Market Structure

Not only are retail investors’ market orders likely to be much better off if they were traded on public exchanges, their limit orders are subject much worse outcomes because they do not get to interact with other retail market orders.

Wholesalers only execute against retail investors’ marketable orders; they route non-marketable retail limit orders to exchanges. Most exchanges provide rebates on limit orders but charge for market orders¹⁶. With this mechanism in place, retail brokers receive rebates on limit orders from exchanges and payment for order flow on market orders from wholesalers.

As a result of this bifurcation of retail flow, limit orders from retail investors lose the opportunity to interact with retail market orders and instead only interact with the more toxic flow on exchanges, yielding the same higher adverse selection and lower fill rates other investors face.

Information Asymmetry Creates a Monopolistic Environment and Increases Costs

To this point in our discussion of the adverse selection experienced by limit orders on exchanges, we have only considered the toxicity of marketable order flow. But the adverse selection faced by market makers, institutional investors, and retail investors using limit orders on exchanges does not stem only from the toxicity of marketable order flow. It also comes from an asymmetry in information between the various market participants placing limit orders on exchanges. A trader placing a limit order can choose to price their limit orders based on public information (e.g., limit order book imbalance, prices of correlated securities) or private information only they have. Consumption of public information leads to natural price discovery in securities as all market makers react to new public information as quickly as possible to not get picked off by a liquidity taker who is processing the same information.

On the other hand, market makers with private information clearly have an edge over other market makers, as they can price quotes more efficiently than others. For example, if based on private information, a market maker believes the price is going down, they will quote their bid on public exchanges below the NBB, and other market makers and limit order providers will face disproportionately higher adverse selection when their orders execute at the NBB price. Over time, this can put market makers relying exclusively on public information out of business and increase concentration among a small selection of market makers on public exchanges.

It may also reduce the incentives for institutional execution algorithms to provide liquidity using limit orders, as the earned spread benefit declines after accounting for adverse selection and execution risk increases as limit orders are less likely to fill. Thus, asymmetric information among liquidity providers reduces competition in public markets and leads to a higher NBBO spread, increasing costs for all investors.

Wholesaling market makers are not making markets to retail investors in a vacuum; the same wholesalers are also the largest market makers trading on exchanges. Wholesalers, by virtue of being sole participant trading against retail order flow, have a wealth of private information giving them a significant edge over other market makers on public exchanges.

¹⁶ Approximately 13.7% of retail execution volume in December were limit orders, according to the top 5 wholesalers’ 605 Reports.

The private information itself is not obtained nefariously, as wholesalers are entitled to use information about the order flow they receive to predict the direction of future order flow and prices.

Given the amount of private information they have and the role they play in public exchanges, the concentration of the large wholesalers is concerning. Analysis of wholesalers' Rule 605 Reports shows that the five largest wholesalers traded volume against retail brokers is 47% of the total volume traded across all sixteen exchanges during continuous market hours¹⁷. The largest wholesaler—Citadel—controlled 50% of the total flow from the top five, while the second largest—Virtu—controlled another 26%. Their volume as a portion of total US equity market volume is illustrated in Figure 11.

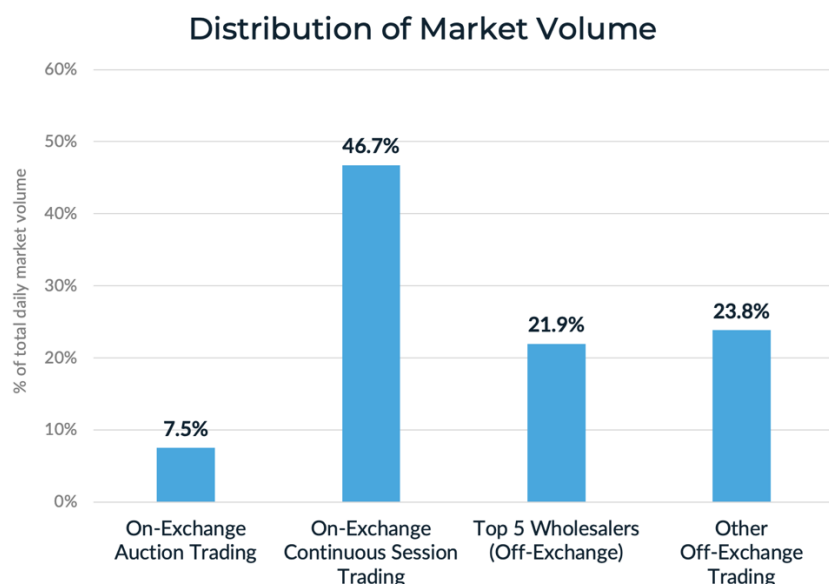


Figure 11. The top five wholesalers in the US equity market traded almost 22% of the total market volume in December 2020. The opening auction averaged about 1.5% of total market volume per stock in December, while the closing auction averaged about 6%. Wholesalers' 22% of market volume is equivalent to 47% of the total volume traded on exchanges during continuous trading hours.

We expect wholesalers' actual volumes are much greater because odd lot orders are not required to be included in Rule 605 reporting—but they likely represent a large portion of retail volume. While a small group of wholesalers' off-exchange volume growing to 47% of on-exchange continuous limit order book volume is alarming on its own, their concentration in lower liquidity stocks makes it even more concerning. To illustrate this high concentration, Figure 12 below shows the largest wholesaler Citadel's highest participation rates in Russell 2000 stocks. Among the top 10 of these high participation rate stocks, for the entire month of December, Citadel's retail executions alone represented 69% to 188% of the total volume traded on all exchanges during continuous trading hours.

¹⁷ Volume from the top 5 wholesalers' Rule 605 Reports in December 2020 equates to 21.94% of overall market volume, which is equivalent to 46.94% of the total volume on exchanges during continuous trading hours (excluding opening and closing auction volume).

Citadel's Top Participation in Russell 2000 Stocks

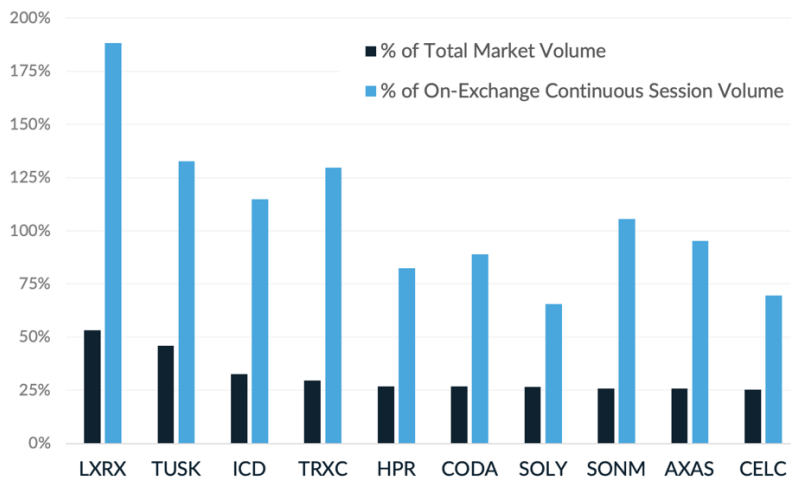


Figure 12. The top wholesaler in the US equity market, Citadel, represented more than 25% of volume in 1% of Russell 2000 names. The specific participation rates of those stocks are illustrated here. In these stocks, the total participation of Citadel alone is more than 50% of the volume traded on exchanges during continuous trading hours, sometimes representing well over 100% of that volume. Information about shares executed comes from Citadel's published Rule 605 Reports.

For these stocks, it is likely that the largest wholesalers are also making markets on exchanges. And if 50% of the order flow in a stock is received privately by a single wholesaler, other market makers are at an informational disadvantage and unlikely to continue providing liquidity. Clearly, in the stocks shown in Figure 12, Citadel has a distinct advantage over other market participants, holding substantially more private information to drive their pricing on exchanges. To underscore this point, this statement was recently made by Citadel's CEO:

"On Wednesday, January 27, we executed 7.4 billion shares on behalf of retail investors. To put this into perspective, on that day Citadel Securities **executed more shares for retail investors than the average daily volume of the entire U.S. equities market in 2019.**"¹⁸

The wholesalers, on the other hand, are perversely incentivized in those cases to offer wider spreads because the improvement they offer retail investors is measured against those spread metrics; the higher the spread, the more price improvement they can provide¹⁹.

Aside from their high volume, it is also important to note that wholesalers are not competing for retail order flow in the traditional sense, winning orders in real time because they offer the best price. When market makers compete on exchanges, the Order Protection Rule of RegNMS allows the market maker with the best price to trade against the next incoming market order, creating good outcomes for liquidity takers. While some market makers may have more capital and are able to offer better spreads on average, this is not true all the time for all stocks. Market makers are incentivized to quote a better bid than offer if they are short a stock, and a better offer than bid if they are long a stock, creating an opportunity for smaller market makers to step in front of large market makers when their inventories are skewed against incoming order flow. If the criteria to be eligible for making markets was to be the best on average, it would be impossible for other market makers to step in and the largest two or three would dominate the market.

But this is exactly what we see in retail market structure. Retail brokers do not compare prices from various wholesalers at the time of routing; rather, they evaluate routing preferences over a longer horizon (multiple days to weeks to months). If there were a competitive marketplace where wholesalers compete for each order, new market makers

¹⁸ Quote from Citadel CEO, Ken Griffin, as reported by [Business Insider](#).

¹⁹ Of the stocks in the Russell 2000, there were 54 (3%) in December where a wholesaler held over 20% of total market volume in a particular name. According to our preliminary analysis, the average spread in December for those stocks was 88bps as compared to 63bps for stocks where no wholesaler held more than 20% of volume—a statistically significant difference according to our two-sample t-test (though not accounting for other potential differences among these stocks).

would join, and spreads would likely decline given the low toxicity of retail flow. But narrower spreads for retail would mean reduced profits for wholesalers and no PFOF for retail brokers. Wholesalers have the most power in the existing structure and no incentive to opt for a more competitive marketplace with reduced profits. In the absence of a change in regulation, competitive pricing for retail order flow is unlikely.

But information asymmetry does not come exclusively from retail market orders. There are several other sources of private information wholesalers have that other market makers and investors placing limit orders on exchanges do not. First, wholesalers have access to limit orders from retail investors in addition to marketable orders. But there is no logical explanation for routing retail limit orders through wholesalers. Since wholesalers do not trade against non-marketable limit orders—as they would earn no spread—they route them to exchanges. Until recently the practice for retail brokers was to route limit orders to exchanges—typically those offering the highest rebate—and marketable orders to wholesalers. The only explanation we can imagine for this change in practice is optics. Perhaps sending all limit orders to exchanges and all market orders to wholesalers makes it obvious that routing is largely based on rebates. Regardless of the reason, routing limit orders through wholesalers to exchanges increases information asymmetry, as wholesalers have information about which limit orders reside on which exchanges while other market participants do not. And as we discussed above, information asymmetry reduces competition and increases spreads for all investors.

Another source of private information is ELPs (electronic liquidity providers) that the largest wholesalers like Citadel and Virtu run for trading orders from institutional brokers. Wholesalers solicit institutional brokers to route marketable limit or market order child order slices from their execution algorithms to ELPs as IOC orders. Wholesalers then have the option to fill the orders at current NBBO prices or return them back. Unlike a wholesaler program, an institutional ELP does not guarantee execution; fill rates are often less than 5%. Wholesalers often ask the institutional brokers to branch out their flows in ways so that the toxic flows can be separated from the less toxic flows, and some institutional brokers accommodate that request in order to get better fill rates from the wholesalers. Wholesalers experience multiple benefits from running ELPs. Institutional order flow, while more toxic than retail order flow, is less toxic than the aggregate order flow on exchanges. Second, even though they only fill a small portion of that order flow, they know that unexecuted flow is likely a portion of a larger order, allowing them to paint a good picture of future order flow—including its expected direction and toxicity.

With intimate, segmented knowledge of flows from retail and institutional brokers, wholesalers have a clear picture of the supply and demand in a stock—far clearer than market makers trading only on public exchanges and privy only to public information. As a direct result of asymmetric information, wholesalers can price stocks more accurately than other high-frequency market makers, stepping in front with a higher bid or lower offer when they expect the price to go their way and backing away when they expect high adverse selection. This advantage creates even higher adverse selection for less-informed market makers and all other market participants trading on public exchanges.

This statement made by the CEO of one of the largest wholesalers on the value of asymmetric information makes this point clear:

“The reason the strategies are successful is because we have this enormous kind of cornucopia of orders that we’re getting from retail brokers, but we’re also getting from other broker-dealers, ...and we are also acquiring on an exchange or a dark pool **and all those get kind of thrown into our central risk book...It’s not a coincidence** that when Knight and Virtu combined...we’ve seen improvements in our strategies and our performance.”²⁰

²⁰ Quote from Virtu Financial CEO, Douglas Cifu, from the transcript of the Virtu Financial earnings call for Q4 2020.

CONCLUSION

The existing wholesale market structure and PFOF have significant implications on all investors, both retail and institutional. On one hand, there is nothing evil about allowing retail market order flow to go to wholesalers. Retail order flow is less toxic than institutional order flow and high-frequency liquidity taking flow, which makes it more profitable for wholesalers to provide liquidity. They share their higher profit with retail investors in the form of price improvement—24.5% of the NBBO spread—and with the retail broker as PFOF. But while this is the advertised improvement, the actual price improvement is likely less than 15%, considering the hidden orders and odd lots placed within the NBBO on exchanges. And we believe 15% is a conservative estimate, considering that retail brokers also have the opportunity to ping ATSS at midpoint before going to exchanges, and that is not accounted for in this statistic. While superficially it may seem retail investors benefit from routing their order flow to wholesalers, the reality is quite different—especially given the recent rise in the total amount of retail trading volume. According to our estimate, if retail trading moved to a public forum the NBBO itself would decline dramatically, eclipsing the 15% difference between what is currently available on exchanges and what wholesalers offer in the form of price improvement.

Some spread reduction would come from the reduced toxicity of flow in exchanges if retail investors were present. Market makers and limit order providers face higher adverse selection in exchanges because the less toxic flow is intercepted by wholesalers. Retail volume now represents just under half of the total trading volume in the displayed limit order books of 16 exchanges in the US equity market. We estimate that spreads would decline by approximately 25% if this less toxic retail flow moved to public exchanges.

Additional reduction in spreads would come from reduced information asymmetry. Currently, retail volume is controlled by a small group of very large wholesalers. PFOF incentives, network effects, and lack of order-by-order competition for retail orders makes it hard for new wholesalers to compete with incumbents. Large wholesalers have private information from the retail market orders, retail limit orders, and from their own institutional ELP programs, allowing them to price orders much more efficiently than other market makers and limit order providers on exchanges. We believe this reduces competition on exchanges and artificially widens spreads. Although it is hard to accurately quantify the increase in spreads due exclusively to this information asymmetry, spreads are likely to reduce even further than we have projected.

The issues surrounding retail order execution also affect institutional investors, who ultimately represent retail investors as well. Institutional investors receive poorer execution because of increased toxicity of liquidity and increased information asymmetry among participants on public exchanges. While retail market orders do receive slightly better execution than institutional investors, their limit orders face much higher adverse selection and lower fill rate because they are routed to public exchanges. If institutional investors have the ability to interact with marketable retail flow, they will likely experience improved fill rates and reduced adverse selection and make more use of limit orders and hidden midpoint orders, reducing effective spreads even further.

For these reasons, we liken price improvement on retail market orders is akin to getting a 30% discount on an item after the shopkeeper raises the price by 40%. Retail investors end up paying 10% more for their market orders. Institutional investors—who often represent retail investors—do not receive the discount and pay the new, higher price for the item, as do retail limit orders. With the recent rise in retail flow, we believe there is an opportunity for a dialog to improve the US equity market structure in a way that could lead to significant improvement in liquidity and significant reduction in execution costs for retail and institutional investors alike, as well as healthier competition among liquidity providers.

Moving retail flow to a public forum would likely require regulatory changes. One option may be increased regulations, for example, curtailing PFOF or limiting the amount of flow a wholesaler can accept each day based on a stock's average daily volume to allow for natural price discovery to occur on public exchanges. A second option could be to allow for a market-based solution to this problem by making regulations around OTC market making and exchanges or ATSS more homogeneous. Currently there are two sets of rules; one set applies to exchanges and ATSS and the other applies to OTC

market making. On exchanges and ATSS, for example, orders cannot be priced at fractional ticks, but OTC market makers can execute orders at fractional ticks. For another example, the Order Protection Rule of RegNMS requires brokers to route orders to the exchange with the best price, but retail brokers are not required to compare the prices from market makers on an order-by-order basis. In addition, market participants cannot segment order flow on exchanges based on toxicity, but OTC market makers can when they receive flow directly. And these are just some examples of how regulatory changes could contribute to improving the execution experience of all investors.

The goal of this paper is to cut through the noise and help all market participants to see that the current market structure is far from optimal for retail investors and institutional investors alike and that there is an opportunity to improve it through disintermediation and increased competition.

ACKNOWLEDGMENTS

We would like to thank Larry Tabb, Head of Market Structure Research at Bloomberg Intelligence and Founder of TABB Group, and Richard DeLayo of Raymond James for their feedback on this work. In addition, we would like to thank Nigam Saraiya and David Conner of BestEx Research for their contributions.

APPENDIX

Our calculations of price improvement on exchanges do include odd lot trades, as they are a significant source of price improvement. However, wholesalers' price improvement statistics do not include odd lots, which may represent a large portion of investor volume. Wholesalers are not currently required to include odd lot orders in their monthly 605 reports.

In order to present a fair comparison, we have included the exchange price improvement statistics including and excluding odd lots below.

Venue Type	Price Improvement as % of NBBO Spread (Including Odd Lot Trades)	Price Improvement as % of NBBO Spread (Excluding Odd Lot Trades)
Exchanges	8.7%	8.2%
Off-Exchange	17.9%	17.8%

Liquidity / Index Group	Price Improvement as % of NBBO Spread (Including Odd Lot Trades)	Price Improvement as % of NBBO Spread (Excluding Odd Lot Trades)
S&P 100	8.7%	8.2%
S&P 500	10.1%	9.5%
Russell 1000	10.7%	10.1%
Russell 2000	10.2%	9.6%

Price Group	Price Improvement as % of NBBO Spread (Including Odd Lot Trades)	Price Improvement as % of NBBO Spread (Excluding Odd Lot Trades)
\$0-1	5.9%	5.9%
\$1-5	6.6%	6.6%
\$5-10	6.9%	6.8%
\$10-50	8.3%	8.1%
\$50-100	11.0%	10.5%
\$100-500	12.4%	11.3%
\$500-1M	14.8%	12.6%