IEX Group, Inc. 1 646 343 2000 tel
4 World Trade Center, 44th Floor 1 888 481 9706 tel
New York, New York 10007 www.iextrading.com

September 24, 2018

Mr. Brent J. Fields
Secretary
U.S. Securities and Exchange Commission
100 F Street, N.E.
Washington, D.C. 20549-1090

Re: Transaction Fee Pilot; Securities Exchange Act Release No. 82873; File No. S7-05-18

Dear Mr. Fields:

Investors Exchange LLC ("IEX") is pleased to submit the two attached documents by Elaine Wah and Stan Feldman to the comment file for the Commission's proposed Transaction Fee Pilot for NMS Stocks.[1] The documents include a White Paper entitled, "Gone in Sixty Seconds: the Cost of Trading in Long Queues" and a related commentary entitled "Gone in Sixty Seconds: How Much is Your Rebate Tax Bill?", published on the on-line publishing platform *Medium*. These materials were both first published on September 21, 2018.

Sincerely,

John Ramsay
Chief Market Policy Officer

cc: The Hon. Jay Clayton, Chairman
The Hon. Kara M. Stein, Commissioner
The Hon. Robert J. Jackson, Jr., Commissioner
The Hon. Hester M. Peirce, Commissioner
The Hon. Elad L. Roisman, Commissioner

Mr. Brett Redfearn, Director, Division of Trading and Markets
Mr. David S. Shillman, Associate Director, Division of Trading and Markets
Mr. Richard Holley, III, Assistant Director, Division of Trading and Markets
Dr. Chye Becker, Acting Director, Division of Economic and Risk Analysis

[1] Securities Exchange Act Release No. 82873 (March 14, 2018), 83 FR 13008 (March 26, 2018).

Gone in Sixty Seconds

How Much Is Your Rebate Tax Bill?

by Elaine Wah and Stan Feldman

One of the biggest debates in the stock market over the past decade has been over "maker-taker" rebates: the practice whereby stock exchanges pay brokers and high-frequency traders to trade on their exchange. The SEC's recent proposal for a transaction fee pilot has brought this debate to the forefront and is viewed by many as a much-needed step towards a deeper understanding of the impact of exchange fees and rebates on execution quality for investors. We applaud the SEC for their initiative and we're looking forward to the valuable data that this pilot could generate.

One common source of resistance to the SEC's proposal has been the consistent stream of "where's the harm?" questions. Some industry participants insist that rebates are a boon for investors, and that removing them would result in poorer execution quality. Most in the industry, including some of the largest investors in the world, know this is a false narrative.

To corroborate the widely held belief that rebates cause investors harm,[2] IEX has studied and quantified one specific way rebates distort competition for executions: a silent, repetitive, systemic wealth transfer where some high-speed traders – and the stock exchanges who sell them speed advantages – are benefiting at everyone else's expense. We call this the "Rebate Tax," and it imposes significant costs on long-term investors.

Investors Suffer on Maker-Taker Exchanges

Here's how it works: publicly available data shows that maker-taker exchanges (those which pay rebates for adding liquidity) have the highest market share and the longest lines to trade.

[2] "Public data now shows the negative impact on investors from the maker-taker pricing model where certain exchanges pay 'rebates' to brokers in exchange for posting quotes on the exchange" Brandes Investment Partners (https://www.sec.gov/comments/s7-05-18/s70518-3419059-162184.pdf)

Exchange	%	Add Fee / (Rebate)	Average Queue Size	Remove Fee / (Rebate)
NYSE National	0.9	5.0	6	(20.0)
Nasdaq BX	4.8	13.0	229	(18.0)
BYX	7.8	12.0	273	(17.0)
EDGA	2.0	8.0	178	(4.0)
NYSE American	0.5	0.0	47	2.0
IEX	3.9	3.0	112	3.0
NYSE	19.9	(22.0)	522	27.5
Nasdaq PSX	1.1	(30.0)	147	28.0
EDGX	8.9	(32.0)	674	30.0
BZX	9.8	(32.0)	577	30.0
NYSE Arca	14.3	(31.0)	975	30.0
Nasdaq	25.1	(30.5)	951	30.0

Top-tier exchange access fees & rebates (mils per share) for Q3 2018 for Tape A securities executed at or above $1. Exchange market share (%) is for July-August 2018, and excludes TRF volume. Average queue size (shares) is the average of aggregate size at the NBB & NBO, and is based on TAQ data from Q2 2018. Queue sizes for inverted/flat-fee exchanges are in orange; those for maker-taker exchanges are in blue.

But as public comment letters have expressed[3] and as we've previously validated with publicly available data,[4] trades on maker-taker exchanges tend to perform worse than other exchanges. This happens because these trades often occur at the worst possible time: right before the price falls for buyers (or rises for sellers). This phenomenon is known as adverse selection.

An exchange's order book is comprised of limit orders that line up at their respective buy or sell prices, forming queues/lines at each price level. Orders at the front of a line will trade first, oftentimes without immediately moving the stock price in an adverse way. As an example, a small undersized marketable order that would not impact the supply/demand of a stock would trade with

[3] "Rebates result in long queues of orders to buy and sell on the largest exchanges, which can result in delays in execution of trades for long-term investors" (Brandes Investment Partners, April 10, 2018, https://www.sec.gov/comments/s7-05-18/s70518-3419059-162184.pdf), "Better queue priority for orders leads to less adverse selection and better execution quality. Then, apart from the reduction in competition from market makers, to the extent that lowering fees and rebates reduces the amount of adverse selection on the venues where their orders are posted, investors' execution quality may benefit a second way" (Pragma, May 14, 2018, https://www.sec.gov/comments/s7-05-18/s70518-3643358-162401.pdf)

[4] See Elaine Wah, Stan Feldman, Francis Chung, Allison Bishop, and Daniel Aisen, "A Comparison of Execution Quality across U.S. Stock Exchanges," (April 19, 2017), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2955297.

an order at the front of the line, and would be deemed a high-quality trade. However, orders at the end of the line wait longer to execute and are more likely to trade with larger marketable orders. Therefore, orders at the end of a long line are more likely to be adversely selected.

Compounding this issue is another feature of maker-taker exchanges: they charge the *highest fees to orders that remove liquidity*. Consequently, smart order routers sending marketable orders will not target or prioritize maker-taker exchanges to avoid the high costs, which increases the wait time for resting orders on these exchanges. Furthermore, our prior research suggests a greater percentage of executions on maker-taker venues comes from market-wide sweeps that access multiple venues.[5] This is because firms needing to execute a large order tend to be more urgent and less cost-sensitive. Such executions are more likely to move market-wide prices.

Given the inherent nature of their business model, high-speed trading firms are much more likely to be at the front of the line than brokers representing investor orders. The ability of high-speed firms to form and quickly join any given price level within an exchange's order book usually relegates long-term investors to the back half of the line, where they tend to be the last to buy/sell at their price level. As a result, they repeatedly and systematically buy high and sell low on maker-taker exchanges.

Yet despite this, investors continue to wait in long lines to trade on maker-taker exchanges. Why? Rebates.

What's the Cost of Trading at the End of Long Queues?

Venue data on performance doesn't tell the full story – it is simply the "average" for a given exchange, blending the experience and outcomes for both high-speed traders and long-term investors. **Investors' orders are more likely to be at the back of exchange order queues, so the average doesn't reflect their experience**. Their brokers simply cannot compete with the speed and tools of high-speed trading firms, who purchase data, proximity and connectivity from exchanges to help them trade at the front of the line, consistently and reliably.

To understand how much investors may suffer because of the Rebate Tax, we need to know how joining (and trading) at the back of a long line affects performance and profitability, versus trading towards the front of the line.

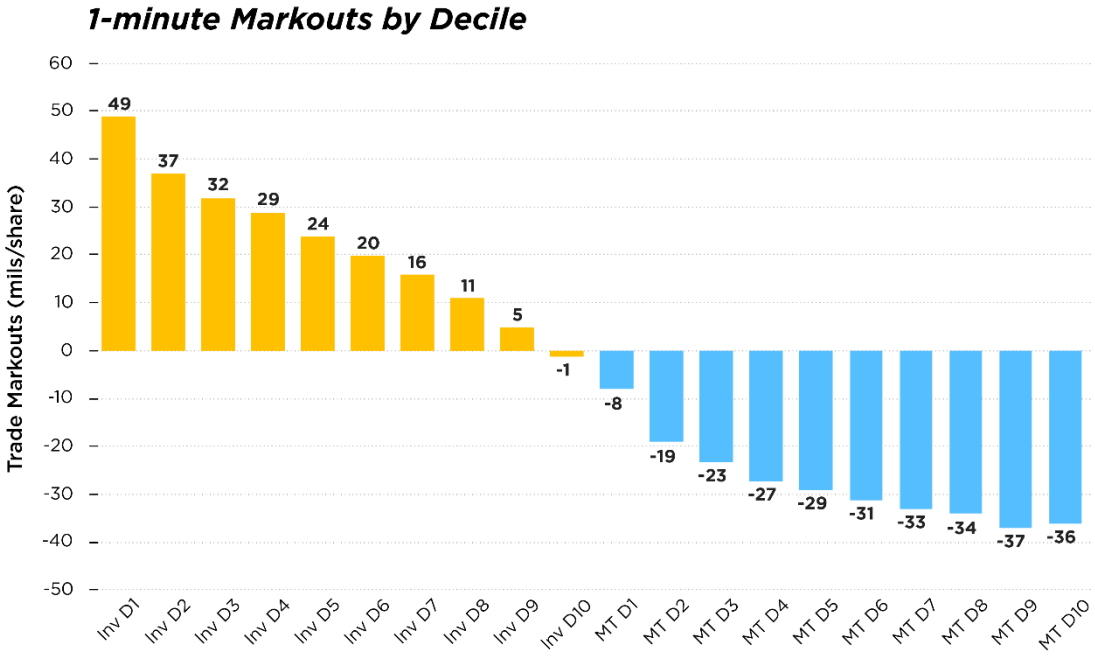
To do this we group trades at the best bid or offer based on the length of the line – as measured by aggregate quoted size – at the time of trade. Trades that happen when the quoted size is large (relative to a given symbol) are likely executing towards the front of a long line, whereas trades when the quoted size is small are more likely to be trading towards the back of the line, or when everything ahead has already been exhausted. We created symbol-specific deciles for queue

---

[5] Ibid.

position, with the first decile representing the first 10% of shares trading at the front of the queue, and the tenth decile representing the last 10% of shares trading at the back of the queue.[6]

To measure potential harm to investors, we look at 1-minute markouts,[7] a widely used metric across both industry and academia that assumes the position opened by an execution is closed out 60 seconds after the trade. Think of it as the "1-minute P&L" or a measure of how much the market moved in favor of or against the buyer/seller. Positive markouts reflect greater potential profit, whereas negative markouts reflect adverse selection, or buyer's/seller's remorse.



Markouts by decile and by exchange type, with inverted/flat-fee ("Inv") results on the left and maker-taker ("MT") results on the right. Deciles are based on the aggregate quoted size at the time of trade, and are shown left-to-right by exchange type, from the 1st decile ("D1") through the 10th decile ("D10"). Data is from TAQ for all symbols from May 2018, and excludes CHX.

---

[6] For a given symbol, we assigned executions at the NBB (NBO) to deciles based on the total quoted size at the NBB (NBO) in the symbol at the time of trade. We weighted deciles by volume to compare performance on a volume-equivalent basis. For a full description of our methodology, please see the accompanying white paper (https://iextrading.com/docs/Gone%20in%20Sixty%20Seconds%20-%20The%20Cost%20of%20Trading%20in%20Long%20Queues.pdf).

[7] We report 1-minute markouts as that is when we observe the performance curve flattening out, but our qualitative results are robust given other time intervals.

Our results above very clearly demonstrate that brokers and the investors they represent are generally worse off when trading on a maker-taker exchange. In fact, the adverse selection costs of trading in a long line on a maker-taker exchange *exceeds* the maximum rebate paid by such venues – so the rebate itself doesn't compensate for the worse performance.[8]

Moreover, performance of the back 10% of an inverted or flat-fee exchange line is often *better* than performance of the front 10% of a maker-taker exchange line. This demonstrates the extent to which inverted exchange "remove rebates" – and the intermarket queue they create – influence the routing of marketable orders.

Simply put, the US exchanges have bizarrely fragmented the markets: there is a large (add) rebate incentive to send non-marketable limit orders to maker-taker exchanges, and there is a large (remove) rebate incentive to send marketable orders to inverted exchanges – which can ultimately prevent long-term investors from otherwise trading with each other. (As reflected in our first chart above: add rebates correspond to longer queues, whereas remove fees and rebates influence the sequence in which exchanges are accessed). **This is a blatantly inefficient market for investors, but a brilliantly profitable market for middlemen like exchanges and high-speed traders.**

Investor orders being placed at the back of a maker-taker venue's order queue are ultimately subject to a significant "Rebate Tax." We present our estimated annualized costs from this Rebate

[8] We note that proprietary trading firms occupying the front of the line on maker-taker venues directly collect rebates, which offset the negative markouts of these executions. Part of the queue is also comprised of quantitative hedge funds, which typically employ "cost-plus" pricing models in which the rebate is passed through to the end client, and as such the rebate may offset any degradation in performance; however, we lack the data to estimate how much volume in each decile is due to these funds.

Tax in the table below, where deciles represent relative queue position.[9]

Decile	Markout	Daily Volume	Annualized Cost
1	(7.6)	187,512,946	**$35,909,346**
2	(18.8)	196,829,051	**$93,247,502**
3	(22.8)	197,260,955	**$113,251,416**
4	(26.9)	197,450,345	**$133,915,506**
5	(28.5)	197,445,543	**$142,004,515**
6	(30.9)	197,570,169	**$153,727,598**
7	(32.8)	197,432,795	**$163,084,550**
8	(34.4)	197,640,815	**$171,372,351**
9	(36.9)	197,966,748	**$184,228,280**
10	(36.1)	205,614,572	**$187,275,740**

Annualized costs are estimated based on the 1-minute markouts (mils per share) on maker-taker exchanges multiplied by the average daily volume (shares) on maker-taker exchanges in the given decile. Data is from TAQ for all symbols from May 2018, and excludes CHX.

We don't know where investors are exactly in the queue, so we refrain from reporting a single estimate of investor harm. We encourage investors to estimate their costs for themselves, depending on shares traded and their estimate of queue position on maker-taker exchanges. Ultimately, only the SEC and the proposed transaction fee pilot have the potential to provide more detail on this issue.

Even so, this represents significant value extraction in just *the first 60 seconds* after the trade. And it is just one measure of the harm from exchange rebates – our estimates don't include the opportunity cost of *not trading* when investor orders wait on the longest lines without trading.

What Does This Mean?

The winners in the maker-taker regime are clearly high-speed traders who have the speed and market data to systematically beat investors to the front of the queue.

Exchanges also profit handsomely from this set-up, by selling high-speed data and technology at rising, monopolistic prices to aid and abet in this race to the front of the line.

[9] We derive our cost estimates by multiplying the performance by the volume traded on maker-taker exchanges, on a decile-by-decile basis. We report annualized costs by decile as it is impossible to know the precise market participant composition in each decile.

Who's the loser in all this? Unequivocally, the investor. They are typically at the end of long lines waiting to execute and therefore are forced to pay a Rebate Tax – which is largely the reason that the largest investors in the world are fully supportive of the SEC transaction fee pilot.

* * *

About IEX

IEX is on a mission to build fairer markets. Founded in 2012 and headquartered in New York City, IEX introduced its first trading venue in 2013 and launched as a U.S. stock exchange in 2016. IEX is the stock exchange that believes that every investor has the right to trade on equal and fair terms, on every trade. Learn more at: iextrading.com.

© 2018 IEX Group, Inc. and its subsidiaries. Neither the information, nor any opinion expressed herein constitutes a solicitation or offer to buy or sell any securities or provide any investment advice or service. The information herein is believed to be reliable, but the Firm makes no representation as to the accuracy or completeness of, and undertakes no duty to update, information herein.

This document may include only a partial description of the IEX product or functionality set forth herein. For a detailed explanation of such product or functionality, please refer to the IEX Rule Book posted on the IEX website: www.iextrading.com

Gone in Sixty Seconds: The Cost of Trading in Long Queues

Elaine Wah and Stan Feldman

IEX*

September 24, 2018

Abstract

The maker-taker pricing model, which pays market participants a rebate for providing liquidity, can lead to long queues at the exchanges employing this fee structure. But some participants may be able to get better queue position than others: high-speed traders can buy speed and data advantages in order to join the queue immediately, whereas slower investor orders are relegated to the back of the line. We analyze publicly available Daily TAQ data to estimate the costs of trading near or at the end of a long queue. By using aggregate quoted size at trade time as a proxy for queue priority, we calculate the impact and scale of performance differences associated with trading in long lines, which our results suggest may impose significant costs on investors.

1. Introduction

Of the 13 U.S. equities exchanges, seven currently pay market participants a per-share rebate for providing liquidity.[1] This pricing paradigm has been shown to be connected to longer lines to trade [Battalio et al., 2016, Wah et al., 2017]. But some traders may be able to get better queue position than others: faster market participants can exploit their speed advantages to join the queue at a new price level with near immediacy. By the time slower investors get in line, they have to wait for the orders ahead of them to execute first, and as a result their performance suffers. In this paper, we analyze Daily TAQ data to demonstrate that performance of orders trading near or at the end of a long line is substantially worse, which suggests that maker-taker rebates may impose significant costs on investors.

U.S. stock exchanges employ a number of different pricing paradigms based on which party is adding or removing liquidity. Every trade involves two participants: a "maker" who posts a buy or sell order (and in doing so gets in line at the back of the queue at the given price level), and a "taker" who trades against the order posted by the maker, either as a seller or a buyer. *Maker-taker*

*© 2018 IEX Group, Inc. and its subsidiaries, including Investors Exchange LLC and IEX Services LLC. IEX Services LLC, Member FINRA/SIPC. All rights reserved. This document may include only a partial description of the IEX product or functionality set forth herein. For a detailed explanation of such product or functionality, please refer to the IEX Rule Book posted on the IEX website. `www.iextrading.com`. Corresponding author: Elaine Wah. Mailing address: IEX Group, Inc., 4 World Trade Center, 150 Greenwich Street, 44th Floor, New York, NY 10007. Email: `elaine.wah@iextrading.com`.

[1] Based on the top-tier access fees/rebates for adding liquidity in Tape A securities executed at or above $1.

exchanges pay a rebate for adding liquidity and charge an access fee for removing liquidity, whereas *inverted* exchanges charge a fee for adding and pay a rebate for removing. *Flat-fee* exchanges charge fees for both adding and removing liquidity, and do not pay rebates.

Access fees and rebates have come under increased scrutiny recently: In March 2018, the U.S. Securities and Exchange Commission proposed a Transaction Fee Pilot[2] to study the impact of fee structure on routing behavior and market quality, an initiative already endorsed by the U.S. Treasury [2017]. These initiatives are important because exchange pricing models have the potential to not only create conflicts of interest between brokers and their customers, but also spawn longer lines at the National Best Bid and Offer (NBBO), which represents the best prices at which one can buy or sell across all exchanges [Wah et al., 2017].

To illustrate how these long lines might arise: A market participant who posts an order on a maker-taker exchange receives a rebate for doing so (if and when the trade happens). However, a participant trading against a posted order on a maker-taker exchange is charged a fee for this trade, whereas she would be paid a rebate for doing the same thing on an inverted exchange. As such, makers (or liquidity adders) inherently have an incentive to post orders on maker-taker exchanges, and takers (or liquidity removers) who need to complete a trade have an incentive to go to inverted exchanges. This misalignment of incentives can contribute to long lines on the maker-taker exchanges, because the potential counterparty is incentivized to trade elsewhere.

Due to the price-time priority rules prevalent in today's equity markets, establishing a position in these queues early is inherently advantageous [Moallemi and Yuan, 2017]. To compete accordingly requires the ability to receive and respond to information as quickly as possible, which a market participant may attain by paying for direct data feeds, co-locating their servers within an exchange's data center, or investing in more sophisticated technology. Acquiring an informational edge then allows these faster participants (i.e., high-frequency traders and market makers) to respond near instantaneously to market events such as quote changes. Lacking the speed advantages to respond immediately to NBBO changes, slower market participants (i.e., investors), have to wait at the back of long lines.

But what is the potential cost ultimately borne by investors of these long lines? To answer this question, we analyze Daily TAQ data to estimate the cost of trading near the end of a long line. We describe our dataset in Section 2 and our model and methodology in Section 3. We discuss our results in Section 4, and we conclude in Section 5.

2. Data

For our analysis, we use publicly available Daily TAQ data, which comprises trade and quote data with microsecond-level timestamps. Our dataset consists of trades and quotes from May 2018. We include data from 12 U.S. equities exchanges in our analysis. The maker-taker venues included are Arca, Cboe BZX, Cboe EDGX, Nasdaq, New York Stock Exchange (NYSE), and Nasdaq PHLX (PSX); the inverted or flat-fee exchanges are Nasdaq BX, Cboe BYX, Cboe EDGA, Investors Exchange (IEX), NYSE American (MKT), and NYSE National (NSX). As with Wah et al. [2017], we exclude the Chicago Stock Exchange (CHX) due to sample size and data robustness concerns.

We include only trades at the NBBO, which reflect the executions of the market participants in line at the best prices across all venues. As for the TAQ quote data, we apply the same filters as Wah et al. [2017], which include excluding locked and crossed markets, as well excluding abnormal

[2]See `https://www.sec.gov/rules/proposed/2018/34-82873.pdf`.

quotes (i.e., where the NBO is outside the range $\left[\frac{1}{3}NBB, 3\,NBB\right]$, where NBB is the National Best Bid and NBO is the National Best Offer).

3. Methodology

Since TAQ data comprises only publicly reported trade and quote data, and does not include any information about the original underlying orders, it is impossible to ascertain actual order priority for an observed execution from TAQ. However, we can use aggregate quoted size for a given symbol at trade time as a proxy for queue priority:

- Trades when the aggregate quoted size is large (relative to the symbol's typical queue size) are likely executing at the front of a long line.
- Trades when the aggregate quoted size is small (relative to the symbol's typical queue size) are likely executing at the end of the line, after everything ahead in the line has already been exhausted.

To estimate the costs to investors of trading in long lines, we group trades based on the aggregate quoted size at the NBBO during the time of trade. More specifically, we group executions of resting buy (sell) orders at the NBB (NBO) based on the total size available across all exchanges at the NBB (NBO).

We discuss how we assign trades to weighted deciles based on aggregate quoted size in Section 3.1. To compare the performance of trades within each group across different types of exchanges, we compute trade markouts as described in Section 3.2.

3.1. Weighted deciles

Since traded volume can vary significantly by execution, we group trades into weighted deciles by volume in order to ensure that we can compare performance on a volume-equivalent basis. Standard deciles ensure only the same total number of observations per group; in contrast, weighted deciles ensure the same total weight—in our case, volume—per group.

To avoid implicit bias from preexisting orderings such as time of day, we analyze a random permutation of the trades dataset, analogous to shuffling a deck of cards. Note that this does not alter the decile to which a trade belongs, except in the instances where multiple trades near and around the volume threshold between two deciles have the same total quoted size.

Given N trades in symbol s on a given exchange on a given date, we first sort the trades in descending order by aggregate quoted size on the side in question, that is, from high to low based on the total size available at the NBB (NBO) for a trade at the NBB (NBO). If the sorted trades have trade sizes $q_1, q_2, q_3, \ldots, q_N$, let Q_s represent the total executed volume for symbol s, where $Q_s = \sum_{i=1}^{N} q_i$.

There are ten deciles, so each decile should comprise $\frac{1}{10}Q_s$ shares. We assign the kth trade to the first decile as long as the cumulative volume up to the kth trade is less than or equal to the volume per decile, or $\frac{1}{10}Q_s$. That is, every trade up to and including the kth trade is assigned to the first decile if the following holds:

$$\sum_{i=1}^{k} q_i \leq \frac{1}{10}Q_s \tag{1}$$

More generally, we assign the kth trade to the nth decile if the following condition holds:

$$\frac{n-1}{10}Q_s < \sum_{i=1}^{k} q_i \leq \frac{n}{10}Q_s \tag{2}$$

Our method does not split up trades that straddle the threshold between deciles, so the total volume per decile is not necessarily exactly the same.

3.2. Markouts

We measure performance via trade markouts, or realized spread, which are a standard in both industry and the academic literature. Markouts compare the price at execution to the midpoint of the market at some specified future time after the trade. More positive markouts reflect greater potential profit after the trade, whereas negative markouts reflect buyer's or seller's remorse—i.e., when the price has gone down (up) after the trade for a buyer (seller). These adverse selection costs arise when informed traders sell to (buy from) a resting buy (sell) order right before prices fall (rise).

Trade markouts are typically restricted to executions at the NBB or NBO because it is not always possible to determine the direction of an execution (i.e., whether the liquidity remover was a buyer or seller) happening inside the NBBO. This constraint does not affect our model, however, as we are only concerned with executions at the NBB or NBO, as these reflect trades of the participants waiting in line at the inside quote.

Trade markouts are measured from the perspective of the resting order, or the participant waiting in the line to trade. Given the NBBO midpoint $M_t = \frac{1}{2}(NBB_t + NBO_t)$ for a given symbol at time t, we define the markout for a trade i that executed at price $p_{i,t}$ at time t as follows:

$$\delta_{\text{markout}} = \begin{cases} M_{t+\tau} - p_{i,t} & \text{for buy orders} \\ p_{i,t} - M_{i,t+\tau} & \text{for sell orders} \end{cases} \tag{3}$$

where $\tau > 0$ is some fixed time interval. We volume-weight markouts by the shares executed.

We also compute volume-weighted relative markouts. A relative markout is the trade markout defined above divided by the NBBO midpoint at the time of trade. More formally, we define the relative markout $\delta_{\%\text{markout}}$ for a trade i that executed at time t as follows:

$$\delta_{\%\text{markout}} = \begin{cases} \dfrac{M_{t+\tau} - p_{i,t}}{M_t} & \text{for buy orders} \\ \dfrac{p_{i,t} - M_{t+\tau}}{M_t} & \text{for sell orders} \end{cases} \tag{4}$$

4. Results

Overall, our results demonstrate the drastic difference in performance between trading at the front of the line versus the back. This is consistent across all exchanges, as evidenced in Figure 1. Notably, the most negative markouts are on the largest maker-taker exchanges such as NYSE, Arca, and Nasdaq, reflecting the greatest degree of adverse selection; in contrast, the inverted/flat-fee venues have universally positive markouts for the first 5 deciles. The plot only shows trade markouts at the 1-minute mark, but our qualitative results are robust given other settings of τ (including 1ms, 10ms, 100ms, 1s, 5s, 10s, 20s, 30s, and 5min).

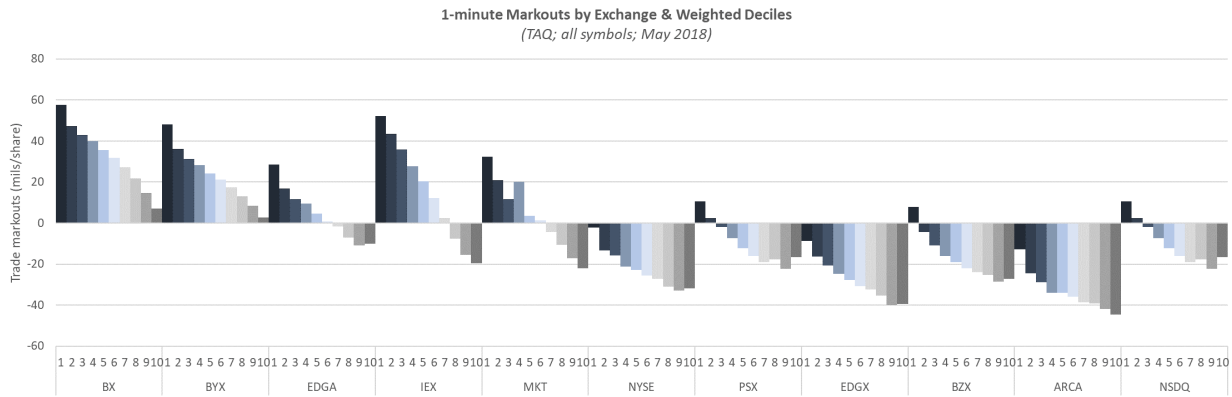


Fig. 1. Markouts by exchange and decile, over all symbols. Deciles are labeled from 1 through 10 for each exchange. The first decile for each exchange is in dark blue; the tenth decile is in gray. NYSE National is excluded here due to low volume during the period in question.

When we group the results by exchange type, as in Figure 2, a clear pattern emerges. Trading on inverted/flat-fee exchanges is associated with better performance, whereas trading on the maker-taker exchanges is worse regardless of place in the queue. The slight uptick in the 10th decile relative to the 9th decile on maker-taker exchanges may be because relatively low aggregated quoted size at the inside can potentially also reflect instances before the queue has fully formed, when only the fastest market participants are in line. Trades during such instances are likely rare, however, given that the potential counterparties would need to be equally fast (in order to have seen the queue start to form) and willing to cross the spread to execute.

We observe that performance in the last decile of an inverted/flat-fee venue is better than performance in the first decile of a maker-taker venue. This suggests that exchanges are generally accessed in order of the cost to remove liquidity, since the inverted/flat-fee venues pay a rebate (or charge a low fee) to liquidity removers whereas the maker-taker venues typically charge the maximum fee allowed (30 mils per share). In other words, a non-trivial portion of order routing is done on a cost-effective basis. Furthermore, our results indicate that markouts associated with the last 5 deciles on maker-taker venues are higher in magnitude than the highest-tier rebate paid by these exchanges. This suggests that the sub-optimal outcomes associated with being among the last to trade on these venues are not necessarily counterbalanced by the rebate payment.

So what do these results mean in terms of the costs imposed on investors for trading near or at the end of long queues? To estimate the cost to investors, we multiply the by-decile performance on maker-taker exchanges by the volume traded on those venues. We assess performance on a decile-by-decile basis to capture the relative differences in speed and access across brokers, and to err on the conservative side. An investor order placed at the end of a long line on a maker-taker exchange is potentially costing the investor a better execution elsewhere, but orders at the back of the line (represented by trades in the 10th decile) on maker-taker exchanges are also less likely to be able to compete for optimal queue position on other venues.

Our results by decile are in Table 1, which shows the annualized cost estimates on a per-decile basis. We expect the market participant composition of each decile to vary significantly: the lower deciles (the front of the queue) are more likely to include high-speed traders, whereas the higher deciles (the back of the queue) are more likely to include long-term investors. However, we cannot determine with certainty how much of each decile is comprised of investor orders. Another

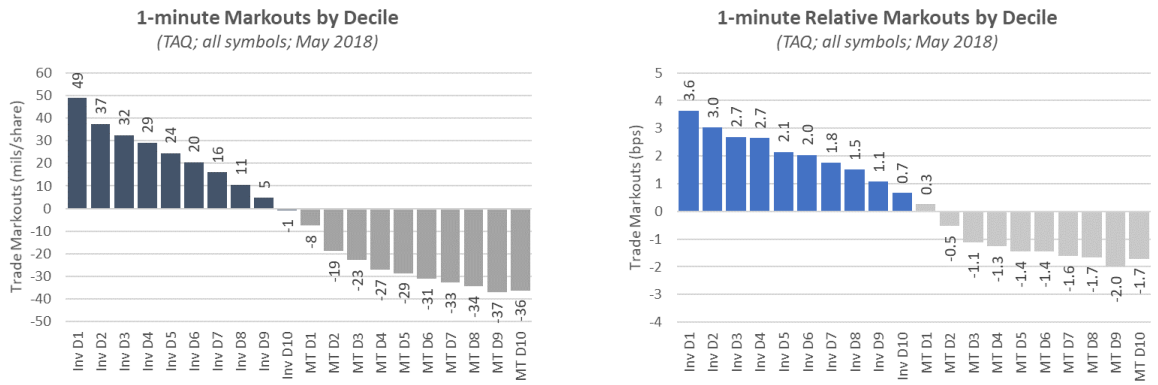


Fig. 2. Markouts by decile and by exchange type, over all symbols. Trade markouts (mils per share) are on the left, with relative trade markouts (bps) on the right. Deciles are labeled from D1 (1st decile) through D10 (10th decile). Inverted/flat-fee ("Inv") results are shown in blue and maker-taker ("MT") results are in gray.

consideration in evaluating these results is whether the rebate is passed back to a firm adding liquidity on maker-taker exchanges, as the rebate payment may partially or fully compensate any performance degradation.[3] For these reasons, we do not provide a singular estimate of "total harm." Nonetheless, our results unequivocally show that performance on maker-taker venues is substantially worse at the back of the queue—in just the first 60 seconds after trading! As such, any investor orders sent to maker-taker exchanges are subject to significant potential cost, especially if the rebate is not passed through to the end investor.

We note that our results presented here are generally in line with prior estimates.[4] A KCG analysis found a 4.5 basis points difference in performance between orders at the very front versus the back of the queue on Nasdaq and Nasdaq BX [Mackintosh, 2014]. Using our methodology, the difference in 30-second relative markouts between the 1st decile for inverted/flat-fee exchanges and the 10th decile for maker-taker exchanges is 5.3 bps, as seen in Figure 2.

[3]We note that quantitative hedge funds typically employ "cost-plus" pricing models in which the rebate is passed through to the end client; however, we lack the data to estimate how much volume in each decile is due to these funds.

[4]We also validated our results by analyzing a day's worth of order-by-order data: we assigned trades into weighted deciles by each trade's original order priority, and the qualitative results were consistent with those presented here. However, due to data distribution restrictions, we are unable to report the order-by-order results here.

Table 1: Estimated costs of trading by decile. We annualize based on the 1-minute markouts (mils per share) on maker-taker exchanges multiplied by the average daily volume (shares) on maker-taker exchanges in the given decile.

Decile	Maker-Taker	M-T Volume	Daily Cost ($)	Annualized Cost ($)
1	(7.6)	187,512,946	$142,497	$35,909,346
2	(18.8)	196,829,051	$370,030	$93,247,502
3	(22.8)	197,260,955	$449,410	$113,251,416
4	(26.9)	197,450,345	$531,411	$133,915,506
5	(28.5)	197,445,543	$563,510	$142,004,515
6	(30.9)	197,570,169	$610,030	$153,727,598
7	(32.8)	197,432,795	$647,161	$163,084,550
8	(34.4)	197,640,815	$680,049	$171,372,351
9	(36.9)	197,966,748	$731,065	$184,228,280
10	(36.1)	205,614,572	$743,158	$187,275,740

5. Conclusion

In this study, we analyzed Daily TAQ data to estimate the impact to investors of trading in the back half of the queue at the National Best Bid and Offer. We use aggregate quoted size as a proxy for queue priority: trades executing when the quoted size is large are likely occurring at the front of a long line, whereas trades executing when the quoted size is small are likely occurring near the back of the line, when the rest of the displayed quote has already been exhausted. Lacking the speed and data advantages purchased by faster market participants, investors are unable to join the queue immediately. By the time investors get in line, the queue is already long, which could ultimately result in substantial and unnecessary losses.

Our model only captures the costs of trading near or at the end of the line, but investors also suffer the opportunity cost of either canceling or simply *not* trading, due to the length of the line. By the time slower investors get in line, they have to wait for the orders ahead of them to execute first—which reduces their likelihood of trading. As such, our estimate likely understates the total cost of trading in long queues. Since TAQ data only includes trades and quotes, an estimate of opportunity cost to investors necessitates more granular data.

Maker-taker and inverted pricing models have made it more difficult for buyers and sellers to meet, because they create economic incentives for different—but mutually dependent—types of activity to be allocated to different exchanges, which can lead to unnecessary intermediation. Oftentimes, makers will only meet takers when the urgency to trade is great enough to incentivize forgoing the rebate (or once orders on venues paying rebates to liquidity removers have been exhausted). For instance, a market participant may want to trade against all displayed shares available, and will thus execute against posted orders at any venue with available liquidity, despite the fee for taking on maker-taker venues.

The exchange pricing models entrenched in today's U.S. equity market structure have ultimately created a system in which investor orders are being placed at the end of long lines, and trading is fragmented across multiple venues. And the significant performance disparity between trading at the front of the line versus the back has placed a premium on high-speed market data and connectivity. To remain competitive, market participants have no choice but to purchase data and connectivity from multiple exchanges—thereby perpetuating an ecosystem in which exchanges profit at the expense of investors.

Further work will be necessary to determine the full impact of access fees and rebates on investors, but the SEC's proposed Transaction Fee Pilot, most notably with its zero-rebate bucket, is a much-needed step towards eliminating the conflicts of interest present today and further safeguarding investors.

References

Robert Battalio, Shane A Corwin, and Robert Jennings. Can brokers have it all? On the relation between make-take fees and limit order execution quality. *Journal of Finance*, 71(5):2193–2238, 2016.

Phil Mackintosh. The need for speed: It's important, even for VWAP strategies. Technical report, KCG, 2014.

Ciamac C. Moallemi and Kai Yuan. A model for queue position valuation in a limit order book. Technical report, Columbia Business School Research Paper No. 17-70, 2017.

U.S. Treasury. A financial system that creates economic opportunities: Capital markets. Available at: `https://www.treasury.gov/press-center/press-releases/Documents/A-Financial-System-Capital-Markets-FINAL-FINAL.pdf`, 2017.

Elaine Wah, Stan Feldman, Francis Chung, Allison Bishop, and Daniel Aisen. A comparison of execution quality across U.S. stock exchanges. *SSRN Electronic Journal*, pages 1–55, 2017.