Financial Conduct Authority

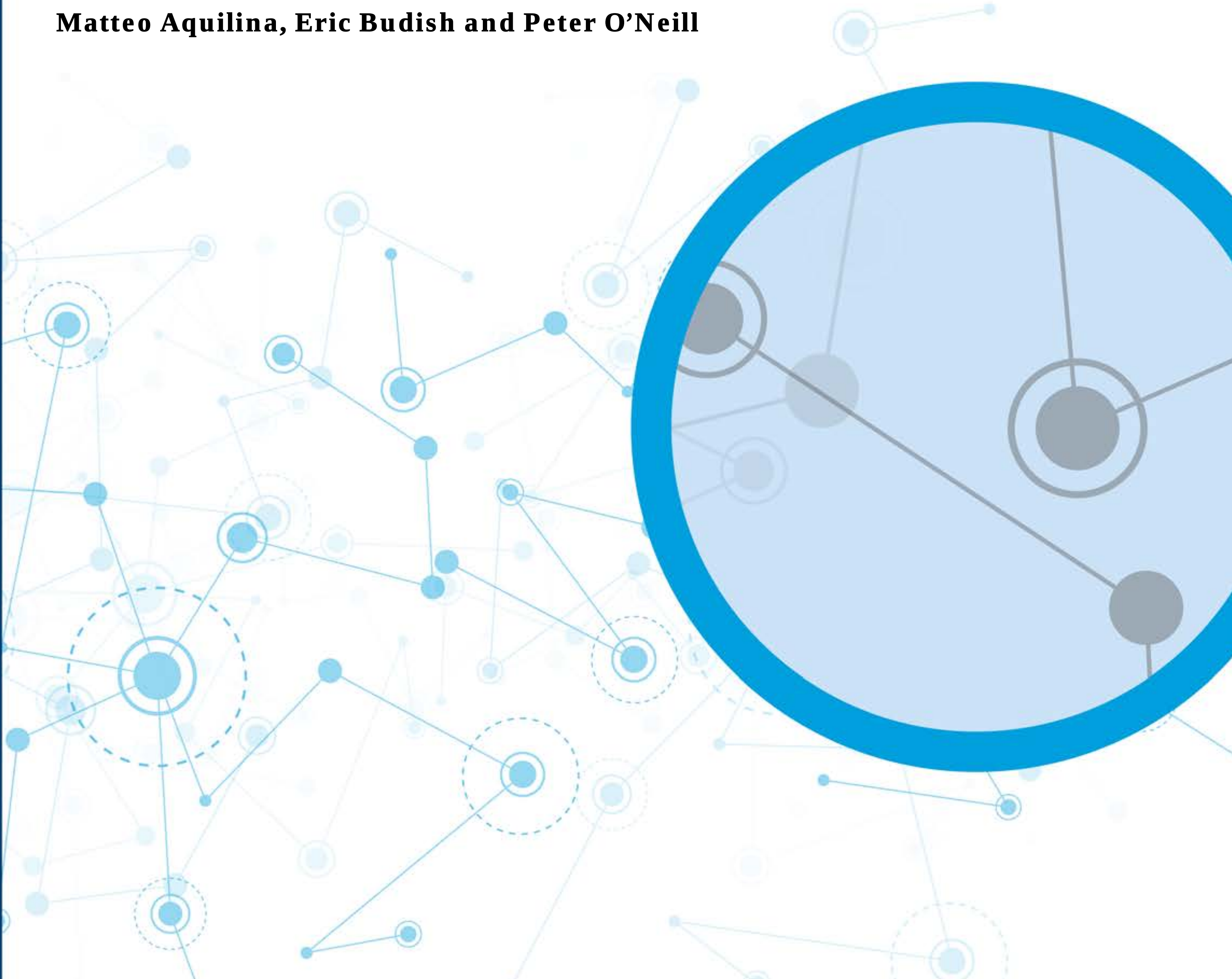
50

Occasional Paper

June 2020

Addendum to Occasional Paper 50, "Quantifying the High-Frequency Trading 'Arms Race': A Simple New Methodology and Estimates"

Matteo Aquilina, Eric Budish and Peter O'Neill



Addendum to Occasional Paper 50

Our paper "Quantifying the High-Frequency Trading 'Arms Race': A Simple New Methodology and Estimates" first circulated in January 2020 as Financial Conduct Authority Occasional Paper 50.[1] We received comments on the Occasional Paper from several academics and industry participants. In response to these comments we conducted several additional analyses which we report here.

Pattern of Takes, Cancels, and Liquidity Provision

Figure 1 Panel A shows that about 90% of races are won with a take (i.e., aggressive order or snipe attempt) with the remaining 10% won by a cancel. This finding that most races are won by aggressive orders was suggested by numbers reported in Table 5.4 of the Occasional Paper, which reported the split between take messages and cancel messages in races, but the specific figure that 90% of races are won with a take was not reported.

Figure 1 Panel B provides data on the pattern of successful takes, successful cancels, and liquidity provision across firms. The top 6 firms, as defined by the proportion of races won as shown in Figure 5.2 of the Occasional Paper, account for about 80% each of race wins, liquidity taken in races, and liquidity successfully canceled in races. In contrast, these 6 firms account for about 42% of all liquidity provided in races — that is, of all of the trading volume in races, 42% is volume where the resting order had been provided by one of the top 6 firms.

Within these top 6 firms there are two distinct patterns of race participation. 2 of the top 6 firms together account for 28% of race wins, 22% of liquidity taken, 61% of successful cancels in races, and 31% of all liquidity provided in races. These data suggest that these 2 firms engage in meaningful quantities of both stale-quote sniping and liquidity provision; their ratio of liquidity taken in races to liquidity provided in races is about 2:3. The remaining 4 of the top 6 firms together account for 54% of race wins, 57% of liquidity taken, 21% of successful cancels, and just 11% of all liquidity provided in races. These data suggest that these 4 firms engage in significantly more stale-quote sniping than liquidity provision; their ratio of liquidity taken in races to liquidity provided in races is 5:1. We therefore denote these two groups of firms as "Balanced in Top 6" and "Takers in Top 6", respectively.[2]

Market participants outside of the top 6 firms account for about 20% each of race wins, liquidity taken in races, and liquidity successfully canceled in races. Where they stand out is that they account for 58% of all liquidity provided in races; that is, they provide nearly 3 times as much liquidity in races as they take.

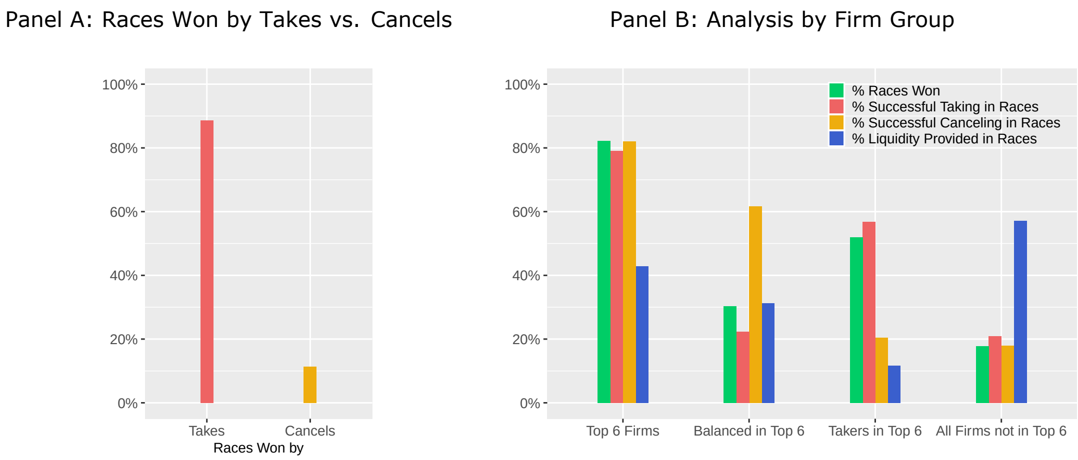
Thus, on net, much race activity consists of firms in the top 6 taking liquidity from market participants outside of the top 6. This taking is especially concentrated in a subset of the fastest firms who account for a disproportionate share of stale-quote sniping relative to liquidity provision. The modal trade in our race data consists of a Taker in Top 6 firm taking from a market participant outside the top 6 (34.3% of all race volume). There is also significant race activity that consists of the fastest firms taking from each other. This

[1] Please see https://www.fca.org.uk/publications/occasional-papers/occasional-paper-no-50-quantifying-high-frequency-trading-arms-race-new-methodology%20

[2] Previous studies that document heterogeneity across HFT firms with respect to their taking and liquidity provision behavior include Benos and Sagade (2016) and Baron et al. (2019). Benos and Sagade (2016) report that the most aggressive group of firms in their sample has an aggressiveness ratio of 82%, which means that 82% of their overall trading volume is aggressive, with the remaining 18% passive. Baron et al. (2019) report that the 90th percentile of firms in their sample has an aggressiveness ratio of 88%.

volume is especially likely to consist of a Taker in Top 6 firm sniping a Balanced in Top 6 firm (17.2%). Please see Table 1 for a matrix of race trading volume organized by such taker-provider combinations.

Figure 1: **Pattern of Takes, Cancels, and Liquidity Provision**



Notes: Panel A: For each FTSE 100 race detected by our baseline method (see Section 4.2 of the Occasional Paper for detailed description) we obtain whether the first successful message (i.e., S1) is a take or a cancel. Panel B: The first bar, % Races won, reports the data depicted in Figure 5.2 of the Occasional Paper aggregated by firm group, with the firm groups as described in the text. The second bar, % Successful Taking in Races, is computed by taking all trading volume in all FTSE 100 races detected by our baseline method, and utilizing the FirmID associated with the aggressive order in each trade. For each bar, the numerator is the total quantity taken in races by firms in that group, in GBP, and the denominator is the total quantity traded across all races in GBP. The third bar, % Successful Canceling in Races, is computed by taking all successful cancels in FTSE 100 races detected by our baseline method, and utilizing the FirmID associated with the cancel attempt. For each bar, the numerator is the total quantity canceled in races by firms in that group, in GBP, and the denominator is the total quantity canceled across all races in GBP. The fourth bar, % Liquidity Provided in Races, is computed by taking all trading volume in all FTSE 100 races detected by our baseline method, and utilizing the FirmID associated with the passive side of each trade, i.e., the resting order that was taken by the aggressive order utilized in the % Successful Taking bar. For each bar, the numerator is the total quantity provided in races by firms in that group, in GBP, and the denominator is the total quantity traded across all races in GBP.

Table 1: **Liquidity Taker-Provider Matrix**

% of Race Volume by Taker-Provider Combination

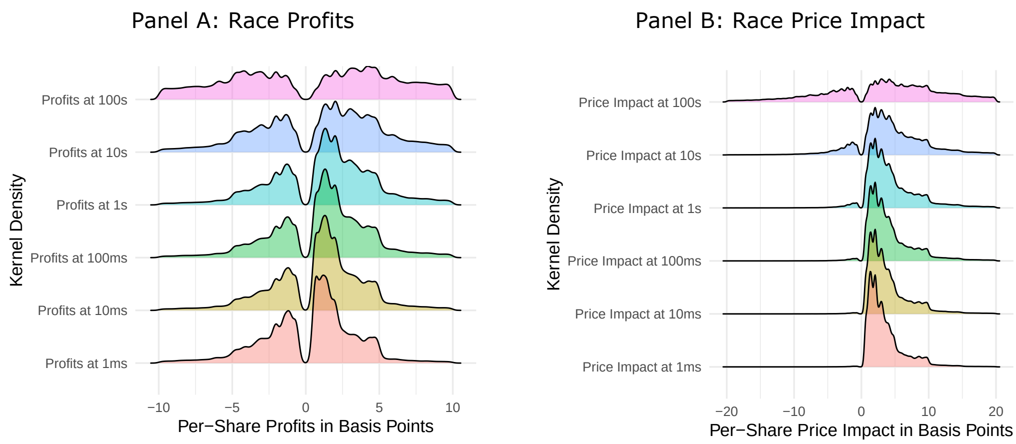
		Provider		
		Takers in Top 6	Balanced in Top 6	Non-Top 6
	Takers in Top 6	5.7	17.2	34.3
Taker	Balanced in Top 6	2.5	6.4	13.3
	Non-Top 6	3.2	7.4	10.1

Notes: For each race detected by our baseline method (see Section 4.2 of the Occasional Paper for detailed description) we obtain all executed trades, and for each executed trade we obtain the FirmID of the participant who sent the take message that executed and the FirmID of the participant whose resting order was passively filled. The FirmIDs are classified into firm groups as described in the text. Each cell of the matrix reports the percentage of GBP trading volume associated with that particular combination of taker firm group and liquidity provider firm group.

Race Price Impact Distributions at Different Time Horizons

Figure 5.3 of the Occasional Paper reported the distribution of race profits at different mark-to-market time horizons. Figure 2 in this addendum includes these race profits distributions as Panel A and the corresponding distributions of race price impact as Panel B. The difference between the two measures is that race profits are the difference between the price paid in the race and the midpoint price in the future, whereas price impact compares the midpoint at the time of the first inbound message in the race (i.e., just prior to its effect on the order book) to the midpoint price in the future (i.e., price impact does not charge the winner of the race the half bid-ask spread). The price impact panel shows that, at shorter horizons, races with negative profits almost always have weakly positive price impacts, meaning that the source of the negative profit is the aggressor in the race not recovering the half bid-ask spread.

Figure 2: **Race Profits and Price Impact Distributions at Different Time Horizons**



Notes: For each race detected by our baseline method (see Section 4.2 of the Occasional Paper for detailed description) we obtain per-share profits and price impact in basis points at different mark to market horizons ranging from 1 millisecond to 100 seconds. Profits at horizon T are defined as the signed difference between the race price and the midpoint price at time T, while price impact at horizon T is the signed difference between the midpoint price at the time of the first inbound message of the race (i.e., before that message affects the order book) and the midpoint price at time T. The figure plots the kernel density of the distribution of per-share profits (Panel A) and per-share price impact (Panel B), each in basis points, at different time horizons. To make the distributions readable, we drop all of the mass at exactly zero profits or price impact.

Modifications to the Spread Decomposition Table

Table 2 replaces Table 5.9 in the Occasional Paper with two modifications. First, we report price impact in races and not in races with different aggregation that is more interpretable. Specifically, we value-weight all trading in races and report the price impact in such trading in basis points, and similarly we value-weight all trading not in races and report the price impact in such trading in basis points. For example, for FTSE 100 symbols (Panel A), the overall price impact in our sample (value-weighted over all trades) has a mean of 3.62 basis points, price impact in races is higher on average at 5.11 basis points, and price impact in non-race trading is lower on average at 3.15 basis points.[3]

Second, we include the realized spread in race trading and non-race trading. For example, for FTSE 100 symbols, the average realized spread in in-race trading is -1.93 basis points, and the average realized spread in non-race trading is +0.15 basis points.

[3] The Occasional Paper instead decomposed the 3.62 basis points into the component from in-race trading and the component from non-race trading, such that these two components sum to the total of 3.62 basis points (of which 1.24 basis points comes from in-race trading and 2.38 basis points comes from non-race trading).

Table 2: **Spread Decomposition**

Panel A: FTSE 100 by Symbol

Description	Mean	sd	Pct01	Pct10	Pct25	Median	Pct75	Pct90	Pct99
Effective spread paid - overall (bps)	3.27	1.22	1.22	1.75	2.28	3.18	4.13	4.91	5.79
Effective spread paid - in races (bps)	3.18	1.22	0.99	1.70	2.21	3.17	4.05	4.89	5.98
Effective spread paid - not in races (bps)	3.29	1.22	1.25	1.78	2.30	3.17	4.15	4.96	5.71
Price impact - overall (bps)	3.62	1.36	1.40	1.92	2.52	3.56	4.52	5.55	6.99
Price impact - in races (bps)	5.11	1.83	2.02	2.85	3.48	4.90	6.50	7.56	8.81
Price impact - not in races (bps)	3.15	1.16	1.21	1.66	2.21	3.17	3.97	4.67	5.99
Loss avoidance (bps)	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.03
Realized spread - overall (bps)	-0.36	0.32	-1.07	-0.76	-0.55	-0.35	-0.17	0.01	0.39
Realized spread - in races (bps)	-1.93	0.70	-3.72	-2.83	-2.40	-1.79	-1.42	-1.11	-0.88
Realized spread - not in races (bps)	0.15	0.30	-0.35	-0.20	-0.05	0.08	0.34	0.56	0.90
PI in races / PI total (%)	33.16	6.09	19.99	24.88	29.53	32.13	37.23	41.72	44.72
PI in races / Effective spread (%)	36.90	7.18	19.79	27.73	33.06	36.59	41.97	46.44	51.67

Panel B: FTSE 250 by Symbol

Description	Mean	sd	Pct01	Pct10	Pct25	Median	Pct75	Pct90	Pct99
Effective spread paid - overall (bps)	8.06	3.81	2.65	4.63	5.59	7.14	9.84	13.10	19.11
Effective spread paid - in races (bps)	6.74	3.03	2.42	4.32	4.97	6.08	7.63	9.96	15.62
Effective spread paid - not in races (bps)	8.22	3.87	2.72	4.70	5.72	7.31	9.94	13.34	19.55
Price impact - overall (bps)	8.09	3.54	2.64	4.96	5.71	7.10	9.40	12.95	19.91
Price impact - in races (bps)	12.22	6.19	4.04	7.17	8.82	10.72	13.75	18.12	33.42
Price impact - not in races (bps)	7.50	3.52	2.36	4.37	5.09	6.40	8.79	12.39	19.39
Loss avoidance (bps)	0.01	0.02	-0.02	0.00	0.00	0.01	0.01	0.02	0.07
Realized spread - overall (bps)	-0.04	1.14	-2.30	-1.02	-0.53	-0.14	0.34	0.96	2.67
Realized spread - in races (bps)	-5.48	3.68	-20.22	-9.36	-6.14	-4.43	-3.44	-2.73	-1.62
Realized spread - not in races (bps)	0.72	1.07	-0.97	-0.13	0.20	0.59	1.07	1.76	3.14
PI in races / PI total (%)	21.60	9.50	1.79	6.00	14.89	22.98	28.19	32.16	39.60
PI in races / Effective spread (%)	22.50	10.92	1.58	5.62	14.84	23.57	30.44	34.79	47.67

Panel C: Full Sample by Date

Description	Mean	sd	Min	Pct10	Pct25	Median	Pct75	Pct90	Max
Effective spread paid - overall (bps)	3.17	0.27	2.74	2.92	3.06	3.12	3.22	3.38	4.52
Effective spread paid - in races (bps)	2.99	0.13	2.64	2.84	2.90	2.99	3.06	3.16	3.28
Effective spread paid - not in races (bps)	3.22	0.32	2.77	2.95	3.09	3.17	3.29	3.44	4.90
Price impact - overall (bps)	3.38	0.19	2.96	3.19	3.23	3.38	3.52	3.61	3.80
Price impact - in races (bps)	4.82	0.24	4.35	4.53	4.66	4.79	4.99	5.07	5.55
Price impact - not in races (bps)	2.99	0.19	2.57	2.79	2.86	2.95	3.13	3.29	3.38
Loss avoidance (bps)	0.01	0.00	-0.01	0.00	0.00	0.01	0.01	0.01	0.01
Realized spread - overall (bps)	-0.22	0.23	-0.62	-0.38	-0.31	-0.26	-0.15	-0.09	1.08
Realized spread - in races (bps)	-1.83	0.17	-2.43	-2.01	-1.92	-1.81	-1.74	-1.64	-1.51
Realized spread - not in races (bps)	0.23	0.26	-0.17	0.05	0.14	0.20	0.29	0.34	1.68
PI in races / PI total (%)	30.58	2.64	22.91	27.88	29.88	30.81	31.93	33.39	35.81
PI in races / Effective spread (%)	32.82	3.73	17.38	29.92	31.60	33.66	34.70	36.54	39.52

Notes: Please see the text of Section 5.5 of the Occasional Paper for definitions of Effective Spread, Price Impact (PI), Loss Avoidance, and Realized Spread. Panel A reports the distribution of these statistics by symbol, for all symbols in the FTSE 100. Panel B reports the distribution for all symbols in the FTSE 250. We only include symbols that have at least 100 races summed over all dates; this drops about one-quarter of FTSE 250 symbols and does not drop any FTSE 100 symbols. Panel C reports the distribution of these statistics by date for the full sample.

The Realized Spread is Negative in Races for Both Fast and Slow Firms

The negative realized spread in races does not appear to discriminate much by firm speed. For the top 6 firms as defined by the proportion of races won (see Figure 5.2 of the Occasional Paper) the realized spread in races is -1.699 basis points, versus -1.930 basis points for firms outside the top 6. The difference between the Takers and Balanced firms in the top 6 is small as well: -1.493 basis points versus -1.775 basis points. Please see Table 3.

Similarly, both fast and slow firms earn a modestly positive realized spread in non-race liquidity provision. For the top 6 firms the realized spread in non-race liquidity provision is 0.347 basis points versus 0.152 basis points for firms outside the top 6.

There is a more significant difference between faster and slower firms in their canceling behavior. The top 6 firms attempt to cancel in races about 35% of the time within the race horizon, and about 39% of the time within 1 millisecond of the starting time of the race. Within these top 6 firms, the maximum cancel rate is 66% within the race-horizon and 68% of the time within 1 millisecond. Firms outside of the top 6 attempt to cancel just 7.57% of the time within races and 9.47% of the time within 1 millisecond of the starting time of the race. If we look beyond 1 millisecond to include any failed cancel attempts of quotes taken in a race, the top 6 cancel attempt rate goes up to 40% and the cancel rate for firms outside of the top 6 goes up to 13.35%.[4] Thus, fast firms are about five times more likely to attempt to cancel in a race than are slower firms.

Together, these results reinforce the idea that latency arbitrage imposes a tax on liquidity provision — it is expensive to be the liquidity provider who gets sniped in a race. The fastest firms are better than slower firms at avoiding this cost, but even they get sniped with significant probability if their quotes become stale.

Table 3: **Realized Spreads in Races by Firm Group**

	Realized Spread (bps)			**Cancel Attempt Rate (%)**		
Firm Group	Overall	Non-Race	Race	In Race	Within 1ms	Ever
All Firms	-0.209	0.236	-1.833	19.29	21.89	24.53
Fast vs. Slow						
Top 6	-0.086	0.347	-1.699	35.35	38.94	39.88
All Others	-0.302	0.152	-1.930	7.57	9.47	13.35
Within Fast						
Takers in Top 6	0.016	0.455	-1.493	45.16	47.56	47.82
Balanced in Top 6	-0.120	0.311	-1.775	30.97	35.09	36.33

Notes: Firm groups are as in Figure 1. The realized spread is calculated as in Table 2. To calculate the cancel attempt rates we first compute, for each firm, the number of races in which they have a cancel attempt within the race horizon, the number of races in which they either have a cancel attempt within the race horizon or a cancel attempt within 1 millisecond of the start of the race for an order taken in the race, the number of races in which they either have a cancel attempt within the race horizon or a cancel attempt anytime after the race horizon for an order taken in the race, and the number of races in which they either have a successful cancel or provide liquidity (each is measured at the relevant price and side for the race). We then aggregate into the firm-group cancel rates by, for the numerator, summing the number of races with cancel attempts over all firms in the group (possibly counting the same race multiple times), and for the denominator, summing the number of races with either cancel attempts or liquidity provision over all firms in the group (possibly counting the same race multiple times).

[4]For firms in the top 6 essentially all of the incremental failed cancels come within 3 milliseconds after the race start (98.57% of all cancel attempts are within 3ms of the race start). For firms outside the top 6 the large majority of the incremental failed cancels come by 3 milliseconds after the race start (85.73%), and essentially all come by 1 second after the race start (99.43%).

Theory Appendix

We also received feedback to provide additional theoretical support, in the form of an appendix, for three theoretical issues discussed in the text of the Occasional Paper. First, discussion of equilibrium in the case where the firm providing liquidity is slow. Second, more detailed support for the analysis behind the bid-ask spread decomposition (5.3). Third, more detailed support for equation (5.6) and its empirical counterpart (5.7), which express the proportional reduction of the cost of liquidity if latency arbitrage were eliminated.

Equilibrium with Slow Liquidity Providers

In the equilibria of the continuous limit order book market studied in Budish, Cramton and Shim (2015), fast trading firms both engage in stale-quote sniping and provide all of the market's liquidity. There is a fringe of slow trading firms but they play no role in these equilibria (see especially Section VI.D and Proposition 3). The slow firms only play a role in equilibrium in Budish, Cramton and Shim (2015) under the frequent batch auctions market design.

In the BCS equilibria of the continuous market, fast trading firms are indifferent between liquidity provision and stale-quote sniping at the equilibrium bid-ask spread s^{CLOB}, characterized by

$$\lambda_{invest}\frac{s^{CLOB}}{2} = \lambda_{public}L(\frac{s^{CLOB}}{2}), \tag{1}$$

where λ_{invest} denotes the arrival rate of investors (i.e., liquidity traders), λ_{public} denotes the arrival rate of new public information, and $L(\frac{s^{CLOB}}{2}) \equiv \Pr(J \geq \frac{s^{CLOB}}{2})\mathbb{E}(J-\frac{s^{CLOB}}{2}|J \geq \frac{s^{CLOB}}{2})$ denotes the expected loss to a liquidity provider if there is a jump larger than their half-spread and they get sniped (J is the random variable describing the absolute value of jump sizes). In the event of a jump larger than the half-spread, stale-quote snipers are successful $\frac{1}{N}$ of the time, where N is the number of fast trading firms, and hence earn expected profits of $\frac{1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$. A fast trading firm that provides liquidity earns revenues of $\lambda_{invest}\frac{s^{CLOB}}{2}$ from providing liquidity to investors, but, if there is a public jump, they get sniped with probability $\frac{N-1}{N}$, hence incurring costs of $\frac{N-1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$. At the equilibrium spread, the revenue benefits of liquidity provision less these sniping costs net to the same $\frac{1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$ earned by snipers. This net profit can be interpreted as the fast liquidity provider earning the opportunity cost of not sniping.

Under slightly different modeling formalities, introduced in Budish, Lee and Shim (2019), there also exist equilibria in which slow trading firms provide liquidity, at exactly the same bid-ask spread $\frac{s^{CLOB}}{2}$ characterized by 1, and the N fast trading firms all engage in stale-quote sniping. The economic intuition for why this can also be an equilibrium is as follows. First, at this bid-ask spread, slow trading firms earn zero profits from liquidity provision, so slow trading firms are indifferent between liquidity provision here, and doing nothing as before. Second, with all N fast trading firms now engaged in sniping, and the bid-ask spread the same as before, the fast trading firms all earn the same profits of $\frac{1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$ as before. And, as before, at this bid-ask spread the fast trading firms are indifferent between providing liquidity or being one of $N-1$ snipers, so they do not strictly prefer to change from sniping to liquidity provision.

Formally, the configuration of play in which a slow trading firm provides liquidity at the spread characterized by 1 (or its slight generalization to include adverse selection as well, presented as equation (5.2) in the main text) is an Order Book Equilibrium as defined in

Budish, Lee and Shim (2019). The argument that this play constitutes an Order Book Equilibrium is as follows:

- If the slow TF deviates by widening their spread to $s' > s^{CLOB}$: another TF (whether slow or fast) can profitably undercut the deviation by providing liquidity at a better spread. Order Book Equilibrium requires that any deviation be robust to another TF providing better liquidity in response, so this potential deviation does not violate Order Book Equilibrium.
- If the slow TF deviates by narrowing their spread to $s' < s^{CLOB}$: they earn strictly negative profits as opposed to zero profits, so this is not a profitable deviation.
- If a fast TF undercuts the slow TF's spread to $s' < s^{CLOB}$: this is a profitable unilateral deviation for a fast TF for s' close enough to s^{CLOB}, because the fast TF gets to both earn positive expected profits from liquidity provision, of just less than $\frac{1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$, and potentially snipe the slow TF (the "have your cake and eat it too" deviation). However, the deviation is not robust to the slow TF canceling in response. Order Book Equilibrium requires that deviations are robust to other firms' responses with either cancels or price improvements ("no robust deviations").[5]
- If any other slow TF undercuts to $s' < s^{CLOB}$: this is not a profitable unilateral deviation for slow TFs, because s^{CLOB} is the bid-ask spread at which slow TFs earn zero expected profits from liquidity provision. (The reason why providing liquidity at s' close enough to s^{CLOB} is profitable for a fast TF but not a slow TF is that fast TFs get sniped with probability $\frac{N-1}{N}$, whereas slow TFs get sniped with probability 1.)

Thus there exist order book equilibria in which fast TFs provide all liquidity as well as order book equilibria in which slow TFs provide all liquidity. It follows that there also exist order book equilibria in which, proportion $\rho_{fast} \in (0,1)$ of the time, a fast TF provides liquidity at s^{CLOB}, while the remaining $1-\rho_{fast}$ of the time a slow TF provides liquidity at s^{CLOB}. Either way, the spread is the same, the profits of all fast TFs are the same ($\frac{1}{N}\lambda_{public}L(\frac{s^{CLOB}}{2})$), and the profits of all slow TFs are zero.

Support for Bid-Ask Spread Decomposition (5.3)

Equation (5.3) in the main text provides a novel bid-ask spread decomposition that includes Price Impact both in and out of races, as well as a Loss Avoidance term for the case where a liquidity provider successfully cancels in a race. In this section we provide formal support for this decomposition.

Begin with the bid-ask spread characterization presented in the main text as (5.2),

$$\lambda_{invest}\frac{s^{CLOB}}{2} = (\lambda_{public} + \lambda_{private}) \cdot L(\frac{s^{CLOB}}{2}),$$

where λ_{public} and $\lambda_{private}$ denote the arrival rate of public and private information, respectively, and $L(\frac{s^{CLOB}}{2})$ denotes the expected loss to a liquidity provider conditional on getting

[5]This case is the key technical difference between the modeling approach in Budish, Lee and Shim (2019) versus that in BCS. In the continuous-time game form considered in BCS a fast TF undercutting a slow TF in this way is a profitable deviation for the fast trading firm, because, in the small amount of time before a slow trading firm is able to respond to this deviation, the deviating fast trading firm both earns potential revenues from liquidity provision and earns potential profits from sniping the slow trading firm. In contrast, the Order Book Equilibrium concept introduced in Budish, Lee and Shim (2019) requires that the order book is at a resting point, where, if any one trading firm can profitably deviate from this resting point the deviation is no longer profitable after other trading firms respond with either price improvements or cancelations.

sniped or adversely selected. For simplicity, we assume that the jump size J is identically distributed for public and private information, and that all jumps are of size of at least the equilibrium half-spread $\frac{s^{CLOB}}{2}$, so all jumps generate attempts to trade. These assumptions can be relaxed but at considerable notational burden.[6] With these assumptions, we have $L(\frac{s^{CLOB}}{2}) = E(J) - \frac{s^{CLOB}}{2}$.[7]

As discussed in the previous subsection, there exist equilibria in which only fast TFs provide liquidity, only slow TFs provide liquidity, and in which both fast and slow TFs provide liquidity. The former case was emphasized in BCS but the latter case appears to fit the data better. Let $\rho_{fast} \in [0,1]$ denote the proportion of liquidity provided by fast TFs in equilibrium with the remaining $1 - \rho_{fast}$ provided by slow TFs. We can now formally define the terms utilized in equation (5.3).

- $\mathit{EffectiveSpread}$ is equal to $[\lambda_{invest} + \lambda_{public}(1 - \frac{\rho_{fast}}{N}) + \lambda_{private}] \cdot \frac{s^{CLOB}}{2}$. Trade occurs whenever an investor arrives (at rate λ_{invest}), whenever an informed trader arrives ($\lambda_{private}$), and whenever there is public news (λ_{public}) and the race is won by a sniper: which occurs with probability $\frac{N-1}{N}$ if the TF providing liquidity is fast, where N is the number of fast traders, and probability 1 if the TF providing liquidity is slow, hence total probability of $\rho_{fast}\frac{N-1}{N} + (1 - \rho_{fast}) = 1 - \frac{\rho_{fast}}{N}$.

- $\mathit{PriceImpact}_{Race}$ is equal to $\lambda_{public}(1 - \frac{\rho_{fast}}{N}) \cdot E(J)$: the $\lambda_{public}(1 - \frac{\rho_{fast}}{N})$ probability that a sniper wins a race, times the size of the jump $E(J)$, which will be the change in the midpoint. Using $L(\frac{s^{CLOB}}{2}) = E(J) - \frac{s^{CLOB}}{2}$ this can be rewritten as $\lambda_{public}(1 - \frac{\rho_{fast}}{N})E(J) = \lambda_{public}(1 - \frac{\rho_{fast}}{N})(\frac{s^{CLOB}}{2} + L(\frac{s^{CLOB}}{2}))$.

- $\mathit{PriceImpact}_{NonRace}$, by similar logic, is equal to $\lambda_{private}E(J)$: the $\lambda_{private}$ probability that there is an informed trader times the size of the jump $E(J)$, which will be the change in the midpoint. This can be rewritten as $\lambda_{private}E(J) = \lambda_{private}(\frac{s^{CLOB}}{2} + L(\frac{s^{CLOB}}{2}))$.

- $\mathit{LossAvoidance}$ is equal to $\lambda_{public}\frac{\rho_{fast}}{N}L(\frac{s^{CLOB}}{2})$: the $\lambda_{public}\frac{\rho_{fast}}{N}$ probability that a fast liquidity provider wins a race with a cancel, times the size of the avoided loss $L(\frac{s^{CLOB}}{2})$.

Now take the equilibrium bid-ask spread as characterized in equation (5.2),

$$\lambda_{invest}\frac{s^{CLOB}}{2} = (\lambda_{public} + \lambda_{private}) \cdot L(\frac{s^{CLOB}}{2}),$$

[6] Formally, if $J_{private}$ and J_{public} are, respectively, the jump distributions for private and public information, with cumulative distribution functions $F_{private}(x)$ and $F_{public}(x)$, respectively, then the conditional distributions of interest are $J^*_{private}$ and J^*_{public} with cdf's $F^*_{private}(x) = \frac{F_{private}(x) - F^-_{private}(\frac{s^{CLOB}}{2})}{1 - F^-_{private}(\frac{s^{CLOB}}{2})}$ and $F^*_{public}(x) = \frac{F_{public}(x) - F^-_{public}(\frac{s^{CLOB}}{2})}{1 - F^-_{public}(\frac{s^{CLOB}}{2})}$, respectively, for $x \geq \frac{s^{CLOB}}{2}$ and $F^*_{private}(x) = F^*_{public}(x) = 0$ for $x < \frac{s^{CLOB}}{2}$.

[7] In the generalization described in the previous footnote the appropriate formulas to use are $L_{private}(\frac{s^{CLOB}}{2}) \equiv E(J^*_{private}) - \frac{s^{CLOB}}{2}$ and $L_{public}(\frac{s^{CLOB}}{2}) \equiv E(J^*_{public}) - \frac{s^{CLOB}}{2}$. In the mathematics that follows it is then convenient to define $\lambda^*{}_{public} = \lambda_{public}(1 - F^-_{public}(\frac{s^{CLOB}}{2}))$ and $\lambda^*_{private} = \lambda_{private}(1 - F^-_{private}(\frac{s^{CLOB}}{2}))$ as the arrival rates of jumps that are larger than the equilibrium spread.

and add $(\lambda_{public}(1-\frac{\rho_{fast}}{N})+\lambda_{private})\cdot\frac{s^{CLOB}}{2}$ to both sides of the equation. This yields

$$\Big(\lambda_{invest}+\lambda_{public}(1-\frac{\rho_{fast}}{N})+\lambda_{private}\Big)\cdot\frac{s^{CLOB}}{2}$$
$$=\Big(\lambda_{public}(1-\frac{\rho_{fast}}{N})+\lambda_{private}\Big)\cdot\Big(\frac{s^{CLOB}}{2}+L(\frac{s^{CLOB}}{2})\Big)+\lambda_{public}\frac{\rho_{fast}}{N}L(\frac{s^{CLOB}}{2}).$$

If we substitute in terms as defined above, this in turn yields

$$\mathit{EffectiveSpread}=\mathit{PriceImpact}_{Race}+\mathit{PriceImpact}_{NonRace}+\mathit{LossAvoidance}.$$

We follow the spread decomposition literature and include $\mathit{RealizedSpread}$ as the residual in this equation for the purpose of bringing it to data, yielding equation (5.3) in the text:

$$\mathit{EffectiveSpread}=\mathit{PriceImpact}_{Race}+\mathit{PriceImpact}_{NonRace}+\mathit{LossAvoidance}+\mathit{RealizedSpread}.$$

Support for the Proportional Reduction in Cost of Liquidity Equations (5.6)-(5.7)

We start with equation (5.4) in the main text, which defines this proportional reduction theoretically:

$$\frac{\frac{s^{CLOB}}{2}-\frac{s^{FBA}}{2}}{\frac{s^{CLOB}}{2}}$$

where s^{CLOB} denotes the equilibrium bid-ask spread in the continuous limit order book market, and s^{FBA} denotes the equilibrium bid-ask spread in the frequent batch auctions market, which eliminates sniping. Next, multiply both the numerator and denominator by $(\lambda_{invest}+\lambda_{private})$:

$$\frac{(\lambda_{invest}+\lambda_{private})(\frac{s^{CLOB}}{2}-\frac{s^{FBA}}{2})}{(\lambda_{invest}+\lambda_{private})\frac{s^{CLOB}}{2}}$$

Next, use the bid-ask spread characterization (5.2) in the main text to solve out for $\lambda_{invest}\frac{s^{CLOB}}{2}$ in the numerator:

$$\frac{(\lambda_{public}+\lambda_{private})\cdot L(\frac{s^{CLOB}}{2})+\lambda_{private}\frac{s^{CLOB}}{2}-(\lambda_{invest}+\lambda_{private})(\frac{s^{FBA}}{2})}{(\lambda_{invest}+\lambda_{private})\frac{s^{CLOB}}{2}}$$

Analogously, use equation (5.1) of Budish, Lee and Shim (2019) to solve out for $\lambda_{invest}\frac{s^{FBA}}{2}$ in the numerator:

$$\frac{(\lambda_{public}+\lambda_{private})\cdot L(\frac{s^{CLOB}}{2})+\lambda_{private}\frac{s^{CLOB}}{2}-\lambda_{private}L(\frac{s^{FBA}}{2})-\lambda_{private}(\frac{s^{FBA}}{2})}{(\lambda_{invest}+\lambda_{private})\frac{s^{CLOB}}{2}}$$

Next, regroup terms to place $\lambda_{public}\cdot L(\frac{s^{CLOB}}{2})$ on the left of the numerator, and then utilize $L(\frac{s}{2})=E(J)-\frac{s}{2}$ for $\lambda_{private}L(\frac{s^{CLOB}}{2})$ and $\lambda_{private}L(\frac{s^{FBA}}{2})$:

$$\frac{\lambda_{public}\cdot L(\frac{s^{CLOB}}{2})+\lambda_{private}(E(J)-\frac{s^{CLOB}}{2})+\lambda_{private}\frac{s^{CLOB}}{2}-\lambda_{private}(E(J)-\frac{s^{FBA}}{2})-\lambda_{private}(\frac{s^{FBA}}{2})}{(\lambda_{invest}+\lambda_{private})\frac{s^{CLOB}}{2}}$$

Observe that most of the terms in the numerator cancel. Specifically, we have $\lambda_{private}(E(J)-$

$\frac{s^{CLOB}}{2}) + \lambda_{private}\frac{s^{CLOB}}{2} - \lambda_{private}(E(J) - \frac{s^{FBA}}{2}) - \lambda_{private}(\frac{s^{FBA}}{2}) = 0$. This leaves us with:

$$\frac{\lambda_{public} \cdot L(\frac{s^{CLOB}}{2})}{(\lambda_{invest} + \lambda_{private})\frac{s^{CLOB}}{2}}$$

as claimed in the text as equation (5.6). Equation (5.6)'s empirical implementation, equation (5.7), then follows immediately as described in the main text.

References

Aquilina, Matteo, Eric Budish, and Peter O’Neill. 2020. “Quantifying the High-Frequency Trading ’Arms Race’: A New Methodology and Estimates.” Financial Conduct Authority Occasional Paper No. 50.

Baron, Matthew, Jonathan Brogaard, Björn Hagströmer, and Andrei Kirilenko. 2019. “Risk and Return in High-Frequency Trading.” *Journal of Financial and Quantitative Analysis*, 54(3): 993–1024.

Benos, Evangelos, and Satchit Sagade. 2016. “Price Discovery and the Cross-Section of High-Frequency Trading.” *Journal of Financial Markets*, 30: 54–77.

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

Budish, Eric, Robin S. Lee, and John J. Shim. 2019. “A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?” National Bureau of Economic Research. NBER Working Paper No. 25855.

12 Endeavour Square, London E20 1JN
Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk