Phantom Liquidity and High Frequency Quoting

May 6, 2016

Jesse Blocher
Assistant Professor
Owen Graduate School of Management
Vanderbilt University
401 21st Avenue South, Nashville, TN 37203
p. 615-322-3687 e. jesse.blocher@owen.vanderbilt.edu

Rick Cooper
Assistant Professor
Stuart School of Business
Illinois Institute of Technology
565 W. Adams, Chicago, IL 60661
p. 312 906 6534 e. rcooper3@iit.edu

Jonathan Seddon
Associate Professor
Audencia Nantes School of Management
8 route de la Jonelière - B.P. 31222
44312 Nantes Cedex 3
p. 00 33 (0)2 40 37 46 80 e. jseddon@audencia.com

Ben Van Vliet
Assistant Professor
Stuart School of Business
Illinois Institute of Technology
565 W. Adams, Chicago, IL 60661
p. 312 906 6513 e. bvanvliet@stuart.iit.edu

Abstract

This paper examines every NASDAQ ITCH feed message for the S&P 500 stocks for 2012 and identifies clusters of extremely high and extremely low limit order cancellation activity. We find results consistent with the ideas that cancel clusters are the result of high frequency traders jockeying for queue position and reacting to information to establish a new price level. Furthermore, few trades seem to be executed during cancel clusters or even immediately after them. Low cancellation activity seems to be markedly different with many level changes all be caused by executions. Our results are consistent with high frequency trading firms behaving as agents who bring efficiency to the market without the need to have executions at intermediate prices. We also discuss the misconception that investors and low frequency trader are synonymous and its implications for policy given our results.

Key words: high frequency trading, phantom liquidity, limit order markets

"By the summer of 2013, the world's financial markets were designed to maximize the number of collisions between ordinary investors and high-frequency traders at the expense of ordinary investors and for the benefit of high-frequency traders, exchanges, Wall Street banks, and online brokerage fi rms. Around those collisions an entire ecosystem had arisen." —*Michael Lewis, Flash Boys (2014, 179)*

By some counts, high frequency trading (HFT) accounts for 70% of the volume in U.S. equities (Hoffman [2014]; Brogaard [2010]). Whether HFT helps or hurts markets is now hotly debated in many circles—the industry, among academics, the media, regulators, and the broader investing public. Some argue it makes markets more efficient and improves liquidity. Others argue it represents a systematic defrauding of other market participants. Whatever the outcome of this debate, regulators are moving to increase oversight of the activity (see BusinessWeek [2014]; New York Times [2014]). The U.S. Securities and Exchange Commission (SEC) recently enacted Regulation Systems Compliance and Integrity (Reg. SCI) to "enhance the equity market structure." The U.S. Commodity Futures Trading Commission (CFTC) recently approved "proposed rules that mark a comprehensive regulatory response to the evolution of automated trading, known as Regulation Automated Trading" (Reg. AT).

There are many well-publicized cases demonstrating the pressure regulators are bringing due to beliefs that HFTs are manipulating markets or disadvantaging other market participants. The CFTC recently fined Panther Energy Trading LLC $2.8 million for spoofing (CFTC [2013]). The SEC has cast an even wider net, targeting ten high frequency firms in a search for evidence of "abusive trading," including illegal layering and spoofing (Reuters [2014]). In its 2010 Concept Release on Equity Market Structure, the SEC asks whether high frequency quoting represents "phantom liquidity that disappears when most needed by long-term investors and other market participants (SEC [2010]).” Michael Lewis [2014] also made the claim of phantom liquidity in *Flash Boys*. Many seem to think this is a very real problem, despite the fact that little, if any, empirical evidence on the question of phantom liquidity exists. The academic literature often evaluates the market impact of HFT in terms of liquidity—measured in terms of bid-ask spreads and volume—and most studies show that it increases liquidity. See, for example, Brogaard et al. [2014] and [2015], Hendershott et al. [2011], Boehmer et al. [2013], Hasbrouck and Saar [2013], Menkveld [2013], Malinova et al. [2013]. But, if that liquidity is phantom liquidity (i.e. it disappears before longer-term traders can access it), then HFT’s impact is in question.

In this paper, we present empirical evidence about the nature of the liquidity HFTs provide. We examine whether quickly evaporating liquidity seems to be without purpose or seems to have either a beneficial or nefarious one. The market exists (at least in part) to find an equilibrium price and not necessarily fundamental value. Does HFT aid or inhibit this price discovery? Moreover, we also ask a question not addressed currently in the literature: do HFT firms use their speed to process information and aid in price discovery without the need for intermediate executions?

Similar to Conrad et al. [2015] and Hasbrouck [2015], we look at the effect of HFT quoting, focusing on HFTs as liquidity providers, rather than as liquidity takers (as in, for example, Baron et al. [2016]). In order to examine the phantom-ness of liquidity, we use “cancel clusters” as our unit of analysis. Cancel clusters arise from the combined activity of HFTs cancelling their limit orders within a small timeframe, due presumably to common private information. (Information may come from correlated assets or from the flow of orders, as in Ait-Sahala and Saglam [2013].) We use an exponentially weighted moving average (EWMA) of all messages to identify cancel clusters. If providing liquidity, an HFT will continuously adjust their quotes, cancelling 50% of them as the price moves, as they attempt to capture the bid-ask spread. This is not phantom liquidity. This is merely adjustment to discover true value.

Several theoretical works have modeled the occurrence of cancel clusters, but with competing predictions and/or explanations. For example, without the use of empirical data, Hoffman [2014] uses a sequential bargaining model of fast traders and slow traders. He shows that slower traders are disadvantaged by faster traders, and are forced to enter either more aggressively priced orders or risk lower execution probability, both of which are forms of diminished execution at the hands of fast traders. If this model is true, we would expect to see cancel clusters followed by executions of slower traders at the stale price left in their wake. Ait-Sahala and Saglam [2013] model HFTs with private information about order flow and show how HFT exploit this asymmetric information, combined with their speed, to more rapidly update their quotes to the disadvantage of other traders. Baruch and Glosten's [2013] model a dynamic limit order book with multiple, strategic, liquidity providers in an environment of both informed and noise traders. They find that what we call cancel clusters are simply a feature of HFT liquidity provision wherein strategic liquidity providers jostle for queue position in order to capture profits by executing more often against noise trades, while avoiding adverse selection by executing against informed trades. Our findings differ from these models in that we find cancel clusters are largely an HFT-only phenomena. Noise traders, or any other kind of traders, seem largely uninvolved. We explore the relationship between cancel clusters and price discovery.

Consider the following example. Assume that the current ask price is 25.01, but that HFTs have perfect information that the real price is 25.06. In one scenario, the monopolist would simply update their ask quote to 25.06, so as to not to sell at the wrong (or stale) price.

Given heterogeneous information and competition in a second scenario, however, a highly informed HFT will update their quote to 25.06 and simultaneously attempt to buy from slower market participants at 25.01 or 25.02. Lastly, in a third scenario, one cancellation simply leads to a flurry of them as HFTs all react quickly to new information. No executions occur before the new equilibrium price level is achieved. The HFTs process the information so quickly that price discovery comes from the cancelations rather than from executions. This is the more effective method, since no dollars need change hands. We provide evidence that this third scenario is what is actually occurring.

Our investigation reveals three main findings. One, cancel clusters are not a dominant feature of the trading day. They develop rapidly and end rapidly. Two, the most common behavior after a cancel cluster is that the bid or ask price reverts to its pre-cancellation level. This indicates that any phantom-ness in the liquidity is highly transient and infrequent. Phantom liquidity only can really be a problem if those who cannot get good execution find the price moving against them when they try to trade. We do not find that in our data. Three, most executions occur outside of cancel clusters, which indicates that this phenomena is largely competitive jostling in anticipation of the arrival of noise traders, as in Baruch and Glosten [2013]. In fact, we find that when cancellation behavior is low, it is *executions* that move the price. This means that investors are paying for the price discovery in executions at prices that are immediately changing. Thus, we argue, cancel clusters are a lower cost means of price discovery.

These findings lead us to conclude that cancel clusters largely appear to be HFTs sparring with one another to get to the front of the limit order queue, rather than HFTs trapping unsuspecting investors into bad executions. We do not draw conclusions about whether HFTs profit at the expense of lower frequency traders, but we do consider the duties of HFT to the marketplace and discuss the changing role of low frequency traders.

DATA

We examine the phantom-ness of HFT liquidity using data on all the S&P 500 stocks for the calendar year of 2012. Rather than simply containing transaction data (as is the case in, for example Hirschey [2013] and Brogaard [2010]), our dataset contains every message about activity in the NASDAQ limit order book. These messages include all additions, cancelations, and executions and are time-tamped to the nanosecond, so that we have an exact timing and ordering of all messages. The dataset is 5.78 terabytes of data in roughly 125,000 ticker-day files (i.e. one file per ticker per date) and was given to us by an HFT firm[1]. By having all the data about the activity in the limit order book, we are able to get a complete picture of both liquidity provision and trade executions.

Using such a rich dataset enables tracking the life of every quote. This matters because much activity in HFT occurs in very short timeframes. Others have studied high frequency quoting, but not the impact of high-speed cancellation behavior. Conrad et al. [2014] look at the effect of high speed quoting on price efficiency and find that high rates of quote activity are associated with more efficient pricing. While they look broadly at TAQ data, we focus on a more detailed dataset to identify specific events and their impacts. We find that cancel clusters are a feature of liquidity provision in high-speed markets and do not represent phantom liquidity.

IDENTIFICATION OF CANCEL CLUSTERS

Our methodology is simple. We step through the data message by message and look for clusters of cancellation activity. We assign to any message that cancels or deletes a resting limit order a value of 1. We assign all other messages a value of 0. We then keep an EWMA of this value with an average data lookback of 10 messages. We identify cancel clusters as beginning whenever the EWMA reaches a value in the top 10% of the day's values. While this threshold is arbitrary, in tests of random days and tickers, we found no differences of any qualitative significance when this number was varied. This threshold value was approximately 0.6 for almost all ticker-days. We close the cancel cluster whenever the EWMA falls to 0.1 less than the value required to enter the cluster (i.e. approximately 0.5). We also found the results to be qualitatively indifferent to the value of 10 for an average data life.

To make this clear, a typical top of book ticker feed might produce a series of actions add (A), delete (D), cancel (C), restate (R), or execute (E). Restatements were analyzed and converted to deletes and adds at a different price, or cancelations of a partial quantity depending on the exact ITCH message. Exhibit 1 depicts the beginning a cancel cluster.

nanosecond	action	bit	EWMA	
34536490012628	A	0	0.515936	
34536491050829	A	0	0.422129	
34536527284957	D	1	0.527197	
34536527295194	D	1	0.613161	← Cancel cluster identified
34536645305375	D	1	0.683495	
34536645329711	D	1	0.741012	
34536800488370	D	1	0.788125	
34536800684170	D	1	0.826648	
34536802044426	D	1	0.858166	
34537026899110	A	0	0.702136	
34537125456257	E	0	0.574475	
34537126829627	D	1	0.651843	
34537126847424	D	1	0.715144	
34537461517668	E	0	0.585118	
34537461517668	E	0	0.478733	← Cancel cluster closed

EXHIBIT 1: Example of Cancel Cluster Identification

In Exhibit 1, the *nanosecond* column represents the number of nanoseconds since midnight at which the co-located server recorder the message. The *action* column is the (possibly converted) action. The *bit* column is the 1 or 0 assigned value, and the *EWMA* is the running EWMA after the addition of bit.

Once identified, we divide the clusters for a given ticker-day into two groups: cancel clusters that clear at least one level of the book; and cancel clusters that do not clear a level before they end. The focus is on the clusters that do clear a level. This keeps down the inclusion of non-informative clusters and very short duration clusters that have no information content. Furthermore, short sequences where the EWMA happens to rise above the threshold only to quickly fall below it without clearing a level are unlikely to have a significant impact. As a comparison, we also look at clusters of low cancellation activity. These are situations where the EWMA is below the 10^{th} percentile of activity for the day (usually about 0.1). This leaves us with four quadrants of analysis as in Exhibit 2. Our primary focus is quadrant 1, though comparisons with the other three quadrants help provide context.

	Clear	**No Clear**
High EWMA	1	4
Low EWMA	2	3

Exhibit 2: Four Quadrants of Results

Exhibit 3 provides statistics on the clusters that comprise each of the four quadrants. In Exhibit 3, the columns labeled *EWMA* and *Clear* identify the quadrant from Exhibit 2. *N* is the total number of observations for all ticker-days; *pct-day* is the percentage of time per day spent in that quadrant; *ave-time* is the average time spent in a cluster; *lclr-time* is the average time from entry of a cluster to the last level clear for that cluster; *ave-clr* is the average number of level clears during a cluster; *ave-clr/C* and *ave-clr/E* break this into clears caused by cancellations and clears caused by executions.

Because the statistics from the bid book and the ask book are nearly symmetric, we focus on the ask book (without loss of generality), where over 26 million clusters cleared the book (i.e. quadrant 1). Potentially, these are events where HFTs uncovered low frequency traders and executed against them at unfavorable prices. Yet, executions that cause the level to clear occurred fewer than 0.08 times per cancel cluster, while 1.41 clears occurred on average due to cancelations. The average cancel cluster length was 5.68 seconds, though the last clear occurred at 2.2 seconds. This means several seconds elapsed after the clear, on average, where little activity occurred to drive the EWMA back down. Moreover, these situations occupied roughly 6.7% of the trading day.

			ASK BOOK					
EWMA	**Clear**	**N**	**pct-day**	**ave-time**	**lclr-time**	**ave-clr**	**ave-clr/C**	**ave-clr/E**
High (1)	Yes	26,223,610	0.0668	5.6778	2.1992	1.5027	1.4142	0.0886
Low (2)	Yes	20,702,082	0.0192	2.7063	1.7606	1.3313	0.1885	1.1427
Low (3)	No	110,648,974	0.0402	1.0196	0.0000	0.0000	0.0000	0.0000
High (4)	No	78,797,865	0.0785	2.1563	0.0000	0.0000	0.0000	0.0000
			BID BOOK					
EWMA	**Clear**	**N**	**pct-day**	**ave-time**	**lclr-time**	**ave-clr**	**ave-clr/C**	**ave-clr/E**
High (1)	Yes	26,355,895	0.0674	5.6988	2.1888	1.5013	1.4118	0.0895
Low (2)	Yes	20,729,013	0.0195	2.7399	1.7765	1.3338	0.1857	1.1481
Low (3)	No	109,922,370	0.0403	1.0314	0.0000	0.0000	0.0000	0.0000
High (4)	No	79,000,712	0.0788	2.1840	0.0000	0.0000	0.0000	0.0000

EXHIBIT 3: Cancel Cluster and Non-Cluster Statistics

Activity in the other three quadrants all lasted a much shorter time, indicating that the quadrant 1 occurrences were more purposeful. Moreover, the average level clears per execute were actually much higher in quadrant 2, the *low* EWMA quadrant, and there were nearly as many events per ticker-day. There are two points. First, HFT cancellations do not seem to result in “slower” market participants getting executed unfavorably as the “fast” move out of the way. Periods of little cancel activity actually seem more prone to executions moving the price. Second, this intense cancelation does not dominate the day. The 26 million cancel clusters that clear either book are less than 7% of the day. Most of the day seems to have a much more even split among adds, cancels, and executes.

THE EFFECT OF A CANCEL CLUSTER ON SUBSEQUENT TRADING BEHAVIOR

The question we ask here is: when a price level in the limit order book clears, what happens next? The methodology we use to answer this question is a straightforward continuation of the procedure used in the previous section. Every time a level clears during either a high cancel cluster (i.e. quadrant 1) or a low cancel cluster (i.e. quadrant 2), we look at the next message. There are five possibilities, and their frequencies are summarized in Exhibit 4.

1. The price gap caused by the level-clear may be filled in by a limit order add at the price level previously vacated. In Exhibit 4, this is shown under the row heading *Lev fill*.
2. Further cancelations may clear another level. This is the row heading *clear*.
3. The gap caused by the level-clear may be narrowed by a limit order add to the other side of the book. This is a true price change. The spread narrows and there is a new mid-spread price. This is the row heading *opp Fill*.
4. An execution occurs at the new level. This is the scenario where some trader is caught by the high frequency activity. They are put a position at an unfortunate price. This is the row heading *execute*.
5. Nothing may happen. Everything just sits as is for some time. We use five seconds, a virtual eternity in today's high-speed markets. This is the row heading *Nothing*.

As can be seen in Exhibit 4, whenever a level clears within a cancel cluster the single most common thing to occur next is the level fills back in with another add to the book in about 0.87 seconds. (Again, we cite the ask side statistics since both sides produce very similar results.) The second thing that is most likely to occur is a quote coming in from the other side, narrowing the spread, and moving the price. The most frequent occurrences seem to be those related to HFT players jostling for queue position and moving the price in response to information.

ASK BOOK	**BID BOOK**

EWMA	Result	N	Reaction T	EWMA	Result	N	Reaction T
High (1)	Lev Fill	13,599,375	0.8693	High (1)	Lev Fill	13,733,802	0.8641
	clear	5,991,658	0.4194		clear	5,883,758	0.4014
	opp Fill	13,208,252	0.6275		opp Fill	13,171,212	0.6089
	execute	1,246,375	0.9064		execute	1,479,897	0.9187
	Nothing	5,357,408	5.0000		Nothing	5,298,908	5.0000
Low (2)	Lev Fill	4,203,312	0.3019	Low (2)	Lev Fill	4,118,983	0.2997
	clear	1,699,895	0.2236		clear	1,632,997	0.2084
	opp Fill	15,130,002	0.0734		opp Fill	14,943,944	0.0687
	execute	5,904,271	0.0633		execute	6,361,095	0.0644
	Nothing	621,656	5.0000		Nothing	593,599	5.0000

EXHIBIT 4: The Next Event After a Level Clears, *By Side of Book*

This notion of HFTs jostling for position and responding to information is further supported by the fact that the least common thing occurring during a cancel cluster is that an execution occurs. Once again, there is a better chance of an execution clearing the level and the price changing during a low cancel cluster (i.e. quadrant 2) than during a high cancel cluster (i.e. quadrant 1). This is consistent with the idea that HFTs are adding liquidity and inconsistent with the notion that other market participants get worse fills when HFT activity is high.

Quite the opposite appears to be true. When the cancel activity is very low, it is most likely that little HFT activity is present. In this case, far more level-clears occur due to executions. In Exhibit 3 the average clear due to executes (*ave-clr/E*) is 0.0886 in quadrant 1 versus 1.1427 in quadrant 2. Likewise, far more executes occur after a level clear in quadrant 2 than quadrant 1, as can be seen in Exhibit 4. This indicates that price discovery is occurring mainly by executes at intermediate prices, making price discovery a tricky business involving a good many “incorrect” trades. On the other hand, when there is a great deal of cancelation activity in a short time the price discovery seems to occur before the executes.

SOME ROBUSTNESS CHECKS

In this section, we examine the statistics of Exhibit 4 by sector, by market capitalization, and by time of day. Exhibit 5 shows what happens next for high cancel clusters by Global

Industry Classification (GIC) sector. The story is the same regardless of the sector. The total number of events varies, of course, as the number and size of firms varies by sector. However, the pattern is no different. In every case. The next event after a level-clear is dominated by the level filling back in from one side of the book or the other in roughly equal measure. Following distantly behind in frequency are another level clear, then nothing, in roughly equal proportion. The least frequent occurrence is an execute at the new level without any intervening activity.

ASK BOOK				**BID BOOK**		
Result	**N**	**Reaction T**	**Sector**	**Result**	**N**	**Reaction T**
Lev Fill	2,069,799	0.8447	**10 Energy**	Lev Fill	2,088,477	0.8397
clear	791,910	0.4558		clear	788,741	0.4396
opp Fill	2,005,460	0.6369		opp Fill	2,005,028	0.6269
execute	156,405	0.9835		execute	193,006	0.9962
Nothing	654,540	5.0000		Nothing	654,725	5.0000
Lev Fill	1,003,043	0.8749	**15 Materials**	Lev Fill	1,011,168	0.8763
clear	468,047	0.4462		clear	466,968	0.4239
opp Fill	1,048,631	0.6168		opp Fill	1,052,941	0.5995
execute	78,312	0.9996		execute	94,401	1.0094
Nothing	410,842	5.0000		Nothing	410,840	5.0000
Lev Fill	1,920,272	0.8914	**20 Industrials**	Lev Fill	1,922,510	0.8864
clear	857,385	0.4917		clear	844,649	0.4651
opp Fill	1,826,764	0.6738		opp Fill	1,818,702	0.6524
execute	136,132	1.0439		execute	166,709	1.0451
Nothing	820,267	5.0000		Nothing	813,651	5.0000
Lev Fill	2,403,381	0.8958	**25 Consumer Discretionary**	Lev Fill	2,424,995	0.8908
clear	1,295,813	0.3855		clear	1,264,362	0.3695
opp Fill	2,421,289	0.6279		opp Fill	2,411,916	0.6059
execute	270,177	0.9302		execute	316,052	0.9348
Nothing	980,838	5.0000		Nothing	968,684	5.0000
Lev Fill	658,479	1.0460	**30 Consumer Staple**	Lev Fill	665,136	1.0404
clear	272,657	0.4661		clear	270,318	0.4433
opp Fill	720,610	0.6382		opp Fill	724,083	0.6138
execute	62,155	0.9393		execute	74,483	0.9389
Nothing	343,005	5.0000		Nothing	338,240	5.0000
Lev Fill	1,285,758	0.9415	**35 Health Care**	Lev Fill	1,299,138	0.9395
clear	613,516	0.4346		clear	602,033	0.4167
opp Fill	1,213,860	0.6982		opp Fill	1,210,332	0.6712
execute	118,389	0.9634		execute	144,884	0.9567
Nothing	603,767	5.0000		Nothing	594,141	5.0000
Lev Fill	1,637,710	0.8904	**40 Financials**	Lev Fill	1,647,269	0.8900
clear	625,763	0.4518		clear	625,622	0.4295
opp Fill	1,629,821	0.5961		opp Fill	1,628,948	0.5817
execute	107,263	0.8746		execute	126,989	0.9020
Nothing	652,659	5.0000		Nothing	644,391	5.0000
Lev Fill	2,130,542	0.6752	**45 Information Technology**	Lev Fill	2,181,134	0.6632
clear	932,561	0.2958		clear	884,917	0.2826
opp Fill	1,833,814	0.5444		opp Fill	1,813,364	0.5247

execute	285,506	0.7351		execute	324,913	0.7496
Nothing	591,399	5.0000		Nothing	577,427	5.0000
Lev Fill	77,875	1.0461	**50 Telecoms**	Lev Fill	79,631	1.0508
clear	28,712	0.4901		clear	29,164	0.4616
opp Fill	93,054	0.6366		opp Fill	93,026	0.6020
execute	7,573	0.8296		execute	9,008	0.8860
Nothing	38,848	5.0000		Nothing	37,616	5.0000
Lev Fill	383,823	1.1062	**55 Utilities**	Lev Fill	384,687	1.0980
clear	97,873	0.5228		clear	99,673	0.5061
opp Fill	381,907	0.6711		opp Fill	381,093	0.6663
execute	21,243	0.8812		execute	25,225	0.9274
Nothing	244,002	5.0000		Nothing	242,532	5.0000

EXHIBIT 5: The Next Event After a Level Clears, *High Cancellation Clusters, By Sector*

Exhibit 6 shows the low cancellation cluster result for Sector 10. The results from this sector are indicative of the results for all GIC sectors. The results are again very different to those the high cancel clusters would indicate. The tendency of a level-clear to result in a narrowing of the spread from the opposite side of the book is very pronounced. Furthermore, the tendency for the level-clear and execute at the next price is also at roughly the same as the level filling in from the opposite side. This behavior is consistent with a generally reduced state of liquidity in low cancel states. Whatever cancel clusters are achieving, it seems that they are not adversely affecting observed cancellations, and are improving liquidity.

ASK BOOK				**BID BOOK**		
Result	**N**	**Reaction T**	**Sector**	**Result**	**N**	**Reaction T**
Lev Fill	611,301	0.2883	**10 Energy**	Lev Fill	590,476	0.2896
clear	254,914	0.1923		clear	244,166	0.1834
opp Fill	1,999,399	0.0718		opp Fill	1,950,539	0.0684
execute	644,988	0.0640		execute	680,785	0.0679
Nothing	66,260	5.0000		Nothing	62,806	5.0000

EXHIBIT 6: The Next Event After a Level Clears, *Low Cancellation Clusters, Sector 10*

We also checked the clusters by market capitalization and by time of day. Exhibit 7 documents the largest 1/6 of the companies in the S&P 500 versus the smallest. Of course, the number of events themselves varied by market capitalization, but the pattern in the data remains the same. Exhibit 8 documents behavior from 9:00-9:30 am versus 12:00-12:30 pm. The opening half hour represents the largest number of messages. The noon half hour represents the least. Whether one looks at subsamples or the whole data, the story is the same. This documents the fractal nature of our data.

ASK BOOK					BID BOOK			
EWMA	**Result**	**N**	**Reaction T**	**cap bucket**	**EWMA**	**Result**	**N**	**Reaction T**
High (1)	Lev Fill	4445461	0.7133	Largest	High (1)	Lev Fill	4488689	0.7046
	clear	1761821	0.3386			clear	1696062	0.3303
	opp Fill	3800314	0.5817			opp Fill	3763796	0.5639
	execute	454360	0.7836			execute	521787	0.7965
	Nothing	975307	5.0000			Nothing	961303	5.0000
Low (2)	Lev Fill	1268476	0.2610		Low (2)	Lev Fill	1262476	0.2576
	clear	446851	0.1818			clear	429552	0.1717
	opp Fill	4563102	0.0656			opp Fill	4507285	0.0615
	execute	2127426	0.0517			execute	2320958	0.0519
	Nothing	115272	5.0000			Nothing	112329	5.0000
High (1)	Lev Fill	28693	1.0386	Smallest	High (1)	Lev Fill	29657	1.0541
	clear	7421	0.5993			clear	7311	0.5301
	opp Fill	33042	0.6188			opp Fill	31779	0.5748
	execute	3220	0.9563			execute	4227	1.0392
	Nothing	17241	5.0000			Nothing	16661	5.0000
Low (2)	Lev Fill	13201	0.2319		Low (2)	Lev Fill	13349	0.2471
	clear	3317	0.1882			clear	3424	0.1381
	opp Fill	70903	0.0366			opp Fill	72629	0.0336
	execute	28695	0.0241			execute	30856	0.0287
	Nothing	1775	5.0000			Nothing	1747	5.0000

EXHIBIT 7: The Next Event After a Level Clears, *by Market Capitalization*

ASK BOOK					BID BOOK			
EWMA	**Result**	**N**	**Reaction T**	**time bucket**	**EWMA**	**Result**	**N**	**Reaction T**
High (1)	Lev Fill	2,010,108	0.6559	9:30-10:30	High (1)	Lev Fill	2,027,626	0.6496
	clear	1,688,427	0.2421			clear	1,683,994	0.2291
	opp Fill	1,871,340	0.5151			opp Fill	1,860,539	0.5026
	execute	190,113	0.7530			execute	216,180	0.7644
	Nothing	423,686	5.0000			Nothing	418,921	5.0000
Low (2)	Lev Fill	768,899	0.2589		Low (2)	Lev Fill	749,943	0.2612
	clear	408,953	0.1569			clear	403,822	0.1532
	opp Fill	2,368,939	0.0683			opp Fill	2,369,214	0.0647
	execute	841,815	0.0748			execute	912,286	0.0739
	Nothing	71,266	5.0000			Nothing	69,099	5.0000
High (1)	Lev Fill	974,616	0.9161	12:00-12:30	High (1)	Lev Fill	986,658	0.9107
	clear	354,328	0.5088			clear	339,264	0.4851
	opp Fill	946,738	0.6608			opp Fill	941,520	0.6362
	execute	83,131	0.9653			execute	100,219	0.9592
	Nothing	459,494	5.0000			Nothing	451,103	5.0000
Low (2)	Lev Fill	234,091	0.3013		Low (2)	Lev Fill	231,494	0.2957
	clear	88,644	0.2545			clear	85,978	0.2300

opp Fill	918,941	0.0683	opp Fill	903,603	0.0642
execute	327,336	0.0587	execute	353,221	0.0591
Nothing	40,837	5.0000	Nothing	38,734	5.0000

EXHIBIT 8: The Next Event After a Level Clears, *by Time*

Based upon our descriptive analysis of behavior, the answer to the question of whether or not HFTs provide only phantom liquidity, only to catch other market participants in trades at bad prices, appears to be no, they do not provide only phantom liquidity. The data shows that their activity is largely due to jockeying for queue position and their engaging in price discovery.

DO HIGH FREQUENCY TRADERS HARM ORDINARY INVESTORS?

The quote from *Flash Boys* at the top of this article highlights the fundamental misconception people have when arguing about HFT. There are not two groups in the marketplace, but three. There are longer-term investors, those holding positions longer than a day. These are institutional or retail investors. There are the low frequency (intraday) traders, such as brokers and market makers. And, there are HFTs. Investors care about the bid-ask spread and the quality of their entries and exits. Much literature already shows that measures of quality improve for price takers. Our research shows there is little evidence for the idea of "collisions," where HFTs systematically defraud other market participants.

So, who is hurt by cancel clusters? One answer might be that because HFTs arrive quickly at the correct price without executions, the traditional liquidity providers, the low frequency institutional traders, no longer add any real value in the marketplace. They attempt to enter the market, but the HFTs pull away in light of new information. Our test does not catch this. But, how important is this behavior? Could it be that HFT liquidity providers have simply replaced low frequency market makers with lower cost price discovery via high-speed quote adjustment? Traditional market makers have been backing away by increasing their spreads on longer-term investors for years. Entire industries exist to reduce slippage in order from large, long-term investment firms. To draw a distinction then, the idea that low frequency traders may now encounter slippage is different from the idea that investors may have more slippage.

The bottom line is that we have uncovered no smoking gun of HFTs doing anything other than negotiating a price level at rapid speed between executions. Others have demonstrated that HFTs provide greater market liquidity, which is beneficial to ordinary investors. It would seem, then, that some definitions of market responsibility now include an obligation on the part of HFTs to ensure that slower, less efficient competitors (in liquidity provision) do not become obsolete. Cooper et al. [2016] has investigated this topic in detail.

Lastly is the issue of "crash insurance," the notion that HFTs abandon the markets rather than assuming some obligation to endure losses in times of market stress. HFT firms are highly motivated to make money by finding the correct new price in *any* environment and quickly adapting. We believe the lack of frequently occurring market crashes speaks to this point quite clearly.

CONCLUSION

We have analyzed a large data set to shed light on the behavior of HFTs surrounding high levels of cancellation activity. We have found evidence that the only thing occurring during these events is that HFT firms seek the correct price level. This is good for the market. It means that HFT firms process information and help improve price discovery without the need for intermediate executions. Intermediate executions as a means of price discovery seem to be a prevalent feature during periods of very low cancellations activity.

Furthermore, we argue that the marketplace consists of three types of participants and society as a whole should only be concerned with the plight of one of them, the long-term investor. The other two exist to make the market more efficient, and whichever one does the best job is the one that should get the profits.

This research leads to a series of other research questions. The most prominent one is what happens to the quality of the book just before and just after HFT firms move the price. That is, do prices truly move seamlessly? By the time an execution occurs at the new price level, is the limit order book just as robust as it was before the movement? These and other important questions need to be addressed to develop intelligent oversight and smooth functioning of financial markets, which benefits all participants.

REFERENCES

Aït-Sahalia, Yacine, M. Saglam. "High Frequency Traders: Taking Advantage of Speed." No. w19531. National Bureau of Economic Research, (2013).

Baron, Matthew D., J. Brogaard, B. Hagströmer, A. Kirilenko. "Risk and Return in High Frequency Trading." Available at: papers.ssrn.com/sol3/Papers.cfm?abstract_id=2433118, 2016.

Baruch, Shmuel, L. Glosten. "Fleeting Orders." Working paper, Columbia Business School, 2013.

Boehmer, Ekkehart, K. Fong, J. Wu. "International Evidence on Algorithmic Trading." Paper presented at American Finance Association 2013 San Diego Meetings, 2013.

Brogaard, Jonathan. "High Frequency Trading and Its Impact on Market Quality." Working paper, Northwestern University Kellogg School of Management, 2010.

Brogaard, Jonathan, B. Hagströmer, L. Nordén, R. Riordan. "Trading Fast and Slow: Colocation and Liquidity." Review of Financial Studies, Forthcoming, (2015).

Brogaard, Jonathan, T. Hendershott, R. Riordan. "High-frequency Trading and Price Discovery." Review of Financial Studies, 27 (2014), pp. 2267-2306.

BusinessWeek. Michaels, Dave, M. Philips, S. Brush. Accessed at: www.businessweek.com/articles/2014-04-10/sec-takes-its-time-on-high-frequency-trading-rules. 2014.

CFTC. Accessed at: www.cftc.gov/PressRoom/PressReleases/pr6649-13. 2015.

Conrad, Jennifer, S. Wahal, J. Xiang. "High-frequency Quoting, Trading, and the Efficiency of Prices." Journal of Financial Economics, 116 (2015), pp. 271-291.

Cooper, Ricky, M. Davis, B. Van Vliet. "The Mysterious Ethics of High Frequency Trading." Business Ethics Quarterly, 26 (2016), pp. 1-22.

Hasbrouck, Joel. "High Frequency Quoting: Short-term Volatility in Bids and Offers." Accessed at: papers.ssrn.com/sol3/papers.cfm?abstract_id=2237499, 2015.

Hasbrouck, Joel, G. Saar. "Low-latency Trading." Journal of Financial Markets, 16 (2013), pp. 646-679.

Hendershott, Terrence, C. Jones, A. Menkveld. "Does Algorithmic Trading Improve Liquidity?" Journal of Finance, 66 (2011), pp. 1-33.

Hirschey, Nicholas. "Do High-frequency Traders Anticipate Buying and Selling Pressure?" Accessed at: papers.ssrn.com/sol3/papers.cfm?abstract_id=2238516, 2013.

Hoffmann, Peter. “A Dynamic Limit Order Market With Fast and Slow Traders.” Journal of Financial Economics, 113 (2014), pp. 156-169.

Lewis, Michael. *Flash Boys*. New York:W. W. Norton & Company.

Malinova, Katya, A. Park, R. Riordan. “Do Retail Traders Suffer From High Frequency Traders?” Accessed at: papers.ssrn.com/sol3/papers.cfm?abstract_id=2183806, 2013.

Menkveld, Albert J. “High Frequency Trading and the New Market Makers.” Journal of Financial Markets, 16 (2013), pp. 712-740.

New York Times. Henning, Peter. Accessed at: dealbook.nytimes.com/2014/10/20/why-high-frequency-trading-is-so-hard-to-regulate/, 2014.

Reuters. McCrank, John. Accessed at: www.reuters.com/article/2014/07/17/us-sec-investigation-highfrequencytradin-idUSKBN0FM2TW20140717, 2014.

SEC. Accessed at: www.gpo.gov/fdsys/pkg/FR-2010-01-21/html/2010-1045.htm, 2010.

[1] We would like to thank Xambala, Inc., which collected the data at their co-located server, nanosecond time stamped it, and processed it into easily an usable format. Without their help, this research would not have been impossible.