Editor: **Michael Wooldridge,** University of Oxford, mjw@cs.ox.ac.uk

High-Frequency Trading: The Faster, the Better?

Rahul Savani, *University of Liverpool*

Between 2:31 pm and 2:51 pm on 6 May 2010, the Dow Jones Industrial Average of major US stock prices fell by an astonishing 998 points (approximately 9 percent), its largest ever same-day point decline. This drop, subsequently known as the Flash Crash, caused a temporary loss of more than US$1 trillion in market value, with some major stocks briefly falling to $.01 per share. Prices rebounded quickly, and the loss in market value was regained in the following days. One of the causes behind the Flash Crash is now believed to be high-frequency trading (HFT): automated trading by computer programs that buy and sell stocks in trades that often last only seconds.

The official report on the Flash Crash by regulators[1] blamed a single large algorithmic trade in E-mini S&P 500 futures by a mutual fund that Reuters identified as Waddell & Reed. The regulators didn't blame HFT directly for causing the sell-off, but they did claim that some HFT firms might have exacerbated the decline by withdrawing from the market during the turmoil. However, recent evidence indicates that the trades of HFT firms *did* contribute to or even cause the crash. In-depth studies by data firm Nanex dispute the regulators' report that the crash was triggered by a single, bad algorithm—rather, Nanex's study suggests that HFT firms initiated market orders that removed liquidity from the volatile market, which in turn triggered the crash. And according to Nanex, HFT is continuing to change the trading landscape. The firm reports that on 24 April 2012, in one second, an HFT firm submitted 47,138 quotes for a single equity listed on the NASDAQ exchange—to put that in perspective, during the height of the Flash Crash, the highest number of quotes in any one second for all NASDAQ stocks was 78,000 (www.nanex.net/aqck/3247.html). It also claims that mini Flash Crashes in individual stocks happen all the time (www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html).

Does HFT deserve this criticism? Markets have always required short-term traders—market makers willing to buy or sell to provide liquidity to retail investors—because they help the market function smoothly. HFT supporters argue that, as algorithms have replaced human floor traders and specialists, markets have become more efficient for retail investors than ever before—but others argue that HFT firms have an unfair advantage and increase systemic risk.

The History of HFT

Computerized trading became a significant part of Wall Street in the 1980s; it was already blamed for exacerbating the market crash in October 1987. Both technological advances and regulatory changes drove the emergence of HFT. In 1998, the Securities and Exchange Commission (SEC) brought in what it called Regulation of Exchanges and Alternative Trading Systems, which allowed electronic trading platforms to compete with the primary exchanges. Around 2000, the exchanges started quoting prices in cents rather than sixteenths of a dollar, causing spreads between bid and offer prices to narrow and destroying the trading business models of some market makers. Finally, the SEC's Regulation National Market System of 2005 required that orders be posted nationally instead of only at individual exchanges. This opened up possibilities for venue arbitrage, where fast-moving traders could profit from discrepancies in the prices of a stock at different exchanges.

Consequently, today's markets are fragmented, comprising primary exchanges, such as the New York Stock Exchange, secondary exchanges, such as the Better Alternative Trading System (BATS), and numerous "dark pools of liquidity." HFT helps these fragmented markets work smoothly as one, but it also exacerbates problems at times of market stress by dramatically speeding up the transmission of shocks and reinforcing them.

HFT now accounts for about 50 percent of the trading volume on the Chicago Mercantile

1541-1672/12/$31.00 © 2012 IEEE
Published by the IEEE Computer Society

Order-Driven Markets

Order-driven markets are the primary way in which financial securities, including equities, futures, and options, change hands. Bids and offers, collectively known as *quotes*, are stored in an order book that is available to market participants.

Order-driven markets have two types of orders, limit orders and market orders. A *limit order* specifies a limit price and a volume. The limit price is the worst price at which the order will be filled. If a limit order cannot be filled immediately (against existing limit orders), then it is stored in the limit order book; see Figure A for an example. A *market order* can be thought of as a limit order with an extreme price. It specifies only a volume, and will be filled at the best available price(s) if liquidity is available in the book. A limit order guarantees price, but not execution; a market order guarantees execution (in liquid markets), but not price.

Buy orders are known as bids and sell orders as offers. The best bid is the bid with the highest price, which is $99.50 in the example; the best offer is the offer with the lowest price, which is $100.50 in the example. If a market order arrived to buy 20 shares, the average execution price would be 1/20 ($100.50 * 13 + $101.00 * 7). The resulting spread, the difference between the best offer and best bid, would have widened from $100.5 – $99.50 = $1 to $101.00 – $99.50 = $1.50, and there would be volume of 5 at the best offer. The spread can be seen as measuring the disagreement between buyers and sellers, so small spreads are seen as characteristic of efficient markets. The tick size is the minimum price increment, which in the example is $0.25.

The central role of market makers is to provide liquidity by posting competitive quotes. For example, posting bids provides more liquidity available to sellers, because they can sell more without impacting the market. Taking out bids with sell market orders removes the liquidity available to sellers, because they can sell less before impacting the market. High-frequency trading (HFT) firms often make their money by market making. If they post the best bid and best offer, and market orders are executed against these, they make the spread as profit. In addition, many exchanges offer liquidity rebates (given to a trader when one of its liquidity-providing limit orders is executed).

Bids	Price	Offers
	101.50	5
	101.25	11
	101.00	12
	100.75	
	100.50	13
	100.25	
	100.00	
	99.75	
35	99.50	
3	99.25	
11	99.00	

Figure A. An example of a limit order book.

Here's an example where the presence of a large order in the book may be exploited by HFT rebate traders. Suppose there's a large bid to buy 10,000 shares at $65.10. HFT rebate traders will take note and "lean" on this bid as the price approaches it: they'll place buy orders in front of the bid at $65.11, knowing that if the market suddenly drops, they can utilize their speed and quickly take out the $65.10 bid and lose only a cent. If the market stays stable, they can continue to buy at $65.11 and sell at $65.12 or $65.13, making a couple of cents and collecting the liquidity rebates as well.

Dark Pools of Liquidity

Large trades in order-driven markets generally suffer a poor execution price. A large market order will likely take out limit orders at multiple price levels, with each new level worsening the average price of execution. A large limit order might not be executed, and would indicate to other market participants the trading intention. Thus, large orders are often split into smaller ones and traded so as to disguise the larger "hidden" order. Nonetheless, many market participants use sophisticated statistical models to detect such hidden orders. Dark pools of liquidity arose in response to the problem of executing large trades without market impact.

Dark pools are trading venues in which the liquidity isn't visible to the public. They operate by matching buyers and sellers who have placed their hidden liquidity in the dark pool. When a match is made, the trade is usually executed at the mid-price of the corresponding lit (order-driven) market at the time of the match. Dark pools give rise to interesting algorithmic questions.[1]

Reference

1. K. Ganchev et al., "Censored Exploration and the Dark Pool Problem," *Comm. ACM*, vol. 53, no. 5, 2010, pp. 99–107.

Exchange, the world's largest commodities and futures exchange. On equities exchanges, HFT now accounts for as much as 70 percent of trading volume on some days. Although banks and hedge funds do take part, independent firms account for the bulk of HFT, many of them founded in the last decade and a half.

The Good, the Bad, and the Ugly

HFT firms, via their investment in technology, get early and privileged access to market data and the ability to act on it before others. Evidence suggests that they exploit this to the detriment of other investors: empirical research shows that market orders from HFT firms convey more information than human orders,[2] and that human traders' limit order executions are (to some extent) bad news, while for computers they're profitable.[3]

A central argument in HFT's favor is that it adds liquidity to markets, making them more efficient by lowering spreads and helping price discovery. Certainly HFT generates a large number of quotes and trades on an intraday basis, and in principal, more market quotes should translate into more market liquidity. Some argue that this ensures that there is always a buyer or seller available when one wants to trade.

Joel Hasbrouck and Gideon Saar challenge this traditional view of quotes patiently supplying liquidity.[4] They showed that in trading of 100 NASDAQ-listed stocks, one-third of quotes are canceled within two seconds. Other evidence suggests that many HFT-generated orders expire in milliseconds, before any other investor could possibly utilize them. According to Lasair Capital,[5] HFT firms often dominate limited market bandwidth with uncompetitive quotes during periods of market stress. By doing so, they clog up the system and hamper other investors. Lasair Capital's study suggests that HFT creates only the illusion of liquidity.

Several HFT practices that utilize the ability to generate and cancel vast number of quotes quickly have come to light:

- Layering/spoofing is the placement of quotes to give the impression of market depth (demand to buy or sell). The quotes are typically placed in the opposite direction to the desired trade. For example, a trader who wants to buy might post offers to sell, hoping to push down the price. If accomplished, the offers would be canceled.
- Smoking is where attractive quotes are placed and then rapidly revised to less favorable prices with the hope of executing profitably against an incoming stream of market orders that arise in response to the original quotes.
- Stuffing is the submission of quotes with immediate cancellation to generate congestion. The flood of quotes must be processed by their competitors, which costs them valuable time.
- Momentum ignition events are caused by HFT strategies that seek to cause short-term market disruptions by initiating or enhancing a trend. If successful, the instigator can profitably reverse its position.

These practices don't aid price discovery, produce smaller spreads, or increase liquidity. In addition, the vast volume of HFT-generated quotes greatly increases the cost of receiving, storing, and analyzing market data. HFT appears to be generating noise and disrupting genuine economic trading, which leaves serious questions for markets regulators: What types of behavior should be allowed? How can you detect illegal behavior?

The Future of HFT

Declining trading volumes loom over HFT. Volumes in US equity markets have continued to fall since the 2008 financial crisis. In April 2012, the average daily volume of US equities trades on all exchanges stood at nearly half of its peak in 2008, according to Credit Suisse Trading Strategy. This isn't good news for HFT firms or exchanges—in particular, it gives HFT firms incentives to get even faster to compete for limited volume. However, regulatory uncertainty could discourage HFT firms from making the investments in the technology and expertise needed to stay competitive.

In recent years, regulators and governments have started paying serious attention to HFT. Since the Flash Crash, it has largely escaped regulation, but now some regulatory change seems inevitable. Regulators face a difficult question: How, or to what extent, can they ensure the stability of trading systems while at the same time ensuring global competitiveness and maximum economic efficiency? Numerous solutions have been proposed, such as a transaction tax, minimum latency restrictions, and quote cancellation fees.

In fact, many European states are currently pushing for a financial transaction tax. These Tobin taxes were originally proposed for currency exchange markets, and they have the effect of making smaller or more speculative transactions less profitable, and so serve to dampen down trading activity. There are obvious difficulties with such taxes, however. They might reduce the competitiveness of a market. And, because they're "throwing sand in the wheels" of the economy (to use Tobin's own phrase), such taxes would reduce the market's economic efficiency.

Simulating Markets

In 1998, the Bios Group and NASDAQ built an agent-based model of the NASDAQ exchange.[1] They explored the effects, including unintended consequences, of changes to market microstructure and rules on the behavior of market participants.

The project was motivated by the pending decimalization and associated decision of the minimum tick size. Based on the results of the Bios Group's research, NASDAQ implemented decimalization on 12 March 2001. It was subsequently found that the data overwhelmingly supported the agent-based model's six predictions. For example, the model correctly predicted that parasitic strategies would become more prominent after the tick size was reduced, and that spread and quote-clustering patterns would appear more often.

This project was a compelling example of the use of agent-based models in understanding structural change in complex financial systems. However HFT firms have such wide reach in today's fragmented, highly interconnected markets that a useful agent-based model for studying impacts on HFT would need to be orders of magnitude more complicated than the NASDAQ simulation.

Reference

1. V. Darley and A. V. Outkin, *Nasdaq Market Simulation: Insights on a Major Market from the Science of Complex Adaptive Systems*, World Scientific, 2007.

There's no doubt that HFT firms play a crucial market-making role, but they seem to be engaged in a socially wasteful arms race for speed. Perhaps to mitigate some of the distortions created by HFT, regulators should impose a minimum latency, for example, one-tenth of a second. Some oppose minimum latency requirements on the grounds that such limits are a backward and hopeless attempt to avoid technological progress. The same criticism, however, applies to speed limits on the roads. And it's hard to believe that going from a latency of a millisecond to a latency of one-tenth of a second would significantly hinder the information aggregation function of the market.

Can we design marketplaces that function effectively with respect to stability, competitiveness, and economic efficiency?

A major difficulty in developing effective interventions is that it's hard to understand their impact. Without a realistic simulation testbed, exchanges and regulators have no means to try out actions in advance of using them in the real market. Agent-based models have been used previously to good effect in understanding and predicting the impact of structural change at the exchange level. It would be an immense challenge to do this at the scale and complexity required to capture the reach of a typical HFT firm, but developing our ability to model complex adaptive systems like these is a challenge worth pursuing.[6] As HFT firms use machine learning and other AI techniques to develop trading strategies, so too should we use these techniques to make our models of markets realistic and thus useful.

One feature of existing markets that gives HFT firms an advantage is that trades can happen at any time. As a remedy, Michael Wellman of the University of Michigan has advocated call markets, where trades can't happen at any time but instead are cleared at regular intervals, for example, every minute. This is just one example of alternative market mechanisms that are worth studying. How might such mechanisms be implemented, and would they have any unintended consequences?

HFT is highly strategic, but little academic work studies it from a game-theoretic perspective. Order-based markets, also known as continuous double auctions, have already received some attention in the AI community. How should agents behave in interconnected high-frequency order-driven markets? It would be interesting to study models that explicitly capture that significant speed advantages that some agents enjoy. ■

References

1. "Findings Regarding the Market Events of May 6, 2010," Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues, 2010; www.sec.gov/news/studies/2010/marketevents-report.pdf.
2. T. Hendershott and R. Riordan, "Algorithmic Trading and Information," Working Paper No. 09-08, NET Inst., 2011; http://dx.doi.org/10.2139/ssrn.1472050.
3. A. Chaboud et al., "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," Discussion Paper No. 980, FRB Int'l Finance, 13 June 2011; http://dx.doi.org/10.2139/ssrn.1501135.
4. J. Hasbrouck and G. Saar, "Technology and Liquidity Provision: The Blurring of Traditional Definitions," *J. Financial Markets*, vol. 12, no. 2, 2009, pp. 143–172.
5. "The Impact of High Frequency Trading on Asset Managers," Lasair Capital, Feb. 2012; www.lasaircap.com/pdf/commentary/The-Impact-of-High-Frequency-Trading-on-Asset-Managers-February-2012.pdf.
6. J. Doyne Farmer and D. Foley, "The Economy Needs Agent-Based Modeling," *Nature*, vol. 460, 2009, pp. 685–686.

Rahul Savani is a lecturer in the Department of Computer Science at the University of Liverpool. Contact him at rahul.savani@liverpool.ac.uk.