Testimony of Andrei Kirilenko
Professor of the Practice of Finance
Sloan School of Management
Massachusetts Institute of Technology
Before the Senate Committee on Agriculture, Nutrition & Forestry
Hearing on High Frequency and Automated Trading in Futures Markets
Tuesday, May 13, 2014

Chairwoman Stabenow, Ranking Member Cochran, I am honored to appear before you today at the hearing on high frequency and automated trading in futures markets.

Not so long ago, futures were traded by human traders in face-to-face markets. While to an outsider a pit at a futures exchange looked like a chaotic crowd of agitated people, a trained eye would see structure behind the frenzy. Some of the people stood at certain posts and "made markets." Others were floor brokers who formed circles around the market makers to get the best prices for a broad range of their customers - from farmers hedging their crops to hedge funds taking a view on where the stock market might be going. Yet others were scalpers or spreaders or opportunistic position takers, who wandered around the trading floor looking for opportunities to exploit.

The ecosystem of an open outcry market was well-known, visible to the human eye, and rigidly prescribed and regulated: traders had designated functions, used common gestures to trade, wore jackets of certain colors, and could be found in specific locations on the trading floor.

Today, trading floors have been replaced by server farms, prescribed gestures have been replaced by message protocols, and automated trading is not visible to the human eye. The traders themselves have been replaced by anonymous algorithms that often operate with little or no human oversight.

These days in order to understand what's going on in automated markets, one needs to be fluent not only in regulation and finance, but also in technology and data processing.

In 2008, I joined the Commodity Futures Trading Commission with the goal of developing analytical tools for surveillance and enforcement in automated markets. For a number of years prior to that, I had a vague notion that as markets became automated, financial theories of market structure developed three decades ago ceased to be valid. I just did not know how far things had gone. The events of May 6, 2010 brought it all to a head.

The Flash Crash[1]

On May 6, 2010, in the course of about 36 minutes starting at 2:32pm ET, U.S. financial markets experienced one of the most turbulent periods in their history. Broad stock market indices – the S&P 500, the Nasdaq 100, and the Russell 2000 – collapsed and rebounded with extraordinary velocity. The Dow Jones Industrial Average (DJIA) experienced the biggest intraday point decline in its entire history. Stock index futures, options, and exchange-traded funds, as well as individual stocks experienced extraordinary price volatility often accompanied by spikes in trading volume. Because these dramatic events happened so quickly, the events of May 6, 2010 have become known as the "Flash Crash."

In the aftermath of the Flash Crash, the media became fascinated with the blend of high-powered technology and hyperactive market activity known as high frequency trading (HFT). To many investors and market commentators, high frequency trading has become the root cause of the unfairness and fragility of automated markets. In response to public pressure, government regulators and self-regulatory organizations around the world have come up with a variety of measures to address HFT. Most of these measures proposed in one way or another to "slow things down" or to remove the "speed advantage" of HFT.

Within hours after the Flash Crash, my colleagues and I began conducting an empirical analysis of trading in the E-mini S&P 500 stock index futures market several days before and during May 6, 2010. Based on the analysis of regulatory transaction-level data, we discovered that HFTs did not cause the Flash Crash, but contributed to extraordinary market volatility experienced on May 6, 2010.

We also discovered how high frequency trading can contribute to flash-crash-type events by exploiting short-lived imbalances in market conditions. We argued that in the ordinary course of business, high frequency traders (HFTs) employ strategies that use their technological advantage to aggressively remove the last few contracts at the best bid or offer and then establish new best bids and offers at adjacent price levels.

This type of trading activity, which we call "immediacy absorption", imposes a cost on slower traders, including traditional market makers. Even the small cost of maintaining continuous market presence makes traditional market makers adjust their inventory holdings to levels that can be too low to offset temporary liquidity imbalances. As a result, because the inventory levels of traditional market makers are low, a large enough sell order can lead to a liquidity-based crash accompanied by high trading volume and large price volatility in times of market stress.

---

[1] This section is based on a study (joint with Albert S. Kyle, Mehrdad Samadi and Tugkan Tuzun) entitled "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market." The study is here: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686004. On February 21, 2014, the CFTC authorized the study for public distribution after a lengthy review process. A draft of the study was previously authorized for public distribution prior to the release of the joint report of the staffs of the CFTC and Securities and Exchange Commission (SEC) entitled "Findings Regarding the Market Events of May 6,

On September 30, 2010, the staffs of the CFTC and Securities and Exchange Commission (SEC) issued a report on the events of May 6, 2010. The 104-page report described how an automated execution program to sell 75,000 contracts of the E-Mini S&P 500 futures, algorithmic trading activity, and obscure order submission practices all conspired to create the Flash Crash.

Flash-crash-type events shake the confidence of market participants and raise questions about the optimal market structure of automated futures markets. With that in mind, my colleagues and I decided to undertake a study of the ecosystem of market participants in automated futures markets over a period of two years. What we found is rather hard to explain.

Concentration of the HFT Industry[2]

According to many anecdotes circulating in the media, high frequency traders seem to possess an uncanny ability to profit in all circumstances. Is that actually empirically true? If yes, why aren't competitive market forces driving those profits to zero through competition? Can anyone actually compete with HFTs?

My colleagues and I set out to answer these questions empirically by looking at two years of regulatory, transaction-level data in the E-mini S&P 500 stock index futures market.

We found, among other things:

- HFTs who specialize in liquidity-taking (aggressive) strategies generate substantially more trading profits than those who specialize in liquidity-providing (passive) strategies.
- Trading profits persistently and disproportionally accumulate to a handful of HFTs. This evidence is consistent with a winner-takes-all industry structure. In a winner-takes-all environment, the trader who is first able to identify and respond to a profitable opportunity will capture all the gains. Traders who are just nanoseconds late simply miss out.
- The concentration of revenue among top-performing HFTs did not decrease over our two-year sample, after adjusting for volatility and non-HFT trading volume.
- New entrants, who are trying to break into the HFT space, earn substantially fewer profits and are more likely to exit.

Taken together, over the two-year period we studied, the HFT industry remained dominated by a small number of fast, opaque (most of the HFTs we identified were not

[2] This section is based on a study (joint with Matthew Baron and Jonathan Brogaard) entitled "The Risk and Return in High Frequency Trading." The study is here: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2433118. On February 19, 2014, the U.S. Commodity Futures Trading Commission (CFTC) authorized the study for public distribution after a lengthy review process. The study was previously authorized for public distribution by the CFTC's Office of General Counsel on September 27, 2012 and has previously circulated under the title "The Trading Profits of High Frequency Traders."

registered with the federal regulators), and aggressive incumbents who earn high and persistent returns.

Economists generally think that competition from new entrants will improve markets: there will be more liquidity, greater price efficiency, lower transaction costs for investors, and less potential for any one firm to influence markets. The idea behind automated markets was to use technology to reduce execution costs to fundamental traders – farmers, manufacturers and asset managers. However, the competitive environment in which HFTs interact may actually prevent improvements in market quality from being fully realized. With limited competition from new entrants to engage incumbent HFTs, we may not be seeing the gains in market quality that we would otherwise see. A concentrated environment can also lead to socially inefficient investment in faster technology, as small increases in trading speed lead to large payouts, driving an arms race for seemingly small reductions in latency.

This prompted us to look into latency a lot more carefully.

Latency

Latency is the delay between when something happens and when it is recorded. Latency is not unique to automated trading. It is observed in Physics, Systems Engineering, and Computer Science.

In automated financial markets, there are three main types of latency that affect the trading process: communication latency, market feed latency, and trading system latency. Reducing each of these latencies is the first order of business for a trader who wants to trade faster. It is also critical for a trader to be able to predict how much latency its trades and orders will experience in practice.

Communication latency is the time it takes for a message – a standardized packet of data that traders use to communicate with exchanges – to travel between a trader's computer and an automated exchange. In order to reduce communication latency, traders can, for a fee, locate their trading servers as close as possible to the "matching engine" of a trading venue and use the fastest data processing technology inside their co-located cage.

Market feed latency is the time it takes for an automated trading venue to disseminate market data – bids and offers, as well as executed transactions – to all market participants. Market feed latency can be reduced by subscribing, for a fee, to data services provided by exchanges. A "free" market feed data is typically too slow to be used for trading on a regular business; it is just fast enough to meet the regulatory requirements, if any.

Trading system latency is the time it takes for a message to travel inside an automated exchange. Trading system latency can be reduced by investing in technology that moves packets of data inside the exchange faster. There is usually nothing a trader can do about this type of latency. It is a risk factor. Yet, high frequency traders can still fare better than everyone else other than the exchange itself.

Based on preliminary results from estimating latency inside an automated futures exchange, my colleagues and I are able to show that system latency actually is significantly more volatile than many believe. System latency varies a lot and is not described by a bell-shaped distribution. This means that when an exchange announces that its average system latency is 3 milliseconds, it is not very meaningful. In practice, it can take an exchange 100 microseconds to process a trader's message or 25 milliseconds. System latency is just that random.

We hypothesize that this randomness could be the very reason why the HFT industry can remain immune from competitive market pressures for so long. Incumbent HFTs already transact a huge fraction of the trading volume – by some estimates 50 percent of the total. This gives them significantly more data to estimate latency than anybody else in the market. Over time, this advantage could become the very factor that keeps possible entrants at bay because the challengers would have to sustain losses on their trades long enough to get the latency estimates right. With a limited amount of proprietary capital to deploy, this could well become prohibitive to most entrants.

Regulators and policy makers intuitively noticed that HFTs ability to exploit latency and proposed many measures to "slow things down" or to remove the "speed advantage." Most of these proposals effectively add latency to the trading process as a whole, to the specific actions of market participants, or to the trading activity of a certain group of market participants. For example, a measure called "minimum quote life" proposes to add latency to the time a resting order must be available for trading before it could be cancelled. Another measure proposes to scramble the time priority of market participants by adding random latencies to their original order submission times. Yet another measure proposes to give latency priority to certain market participants ahead of others.

While some of these measures have an intuitive appeal, market participants and the public needs to be aware that they are about adjusting latency in various ways. Then, the public would be better engaged in a debate on how to properly calibrate, test, and evaluate the effectiveness of these measures. Otherwise, a well-intentioned measure can be misunderstood, or worse, give those whose influence it wishes to erode yet another advantage over others.

Recommendations

High frequency and automated trading lies in the intersection of regulation, finance, technology, and data processing. I recommend improvements along each of these lines.

Regulation

Regulation needs to reflect the shift of trading activity towards opaque hyperactive algorithms. In this regard, I recommend creating a broad definition of "automated brokers and traders" that would be similar to what used be called "floor brokers and traders." The definition needs to be broad enough to cover the activity of all active proprietary traders. For example, if a trader is co-located directly or through a technology vendor (i.e., has the ability to be "present on the automated floor") and uses more than a certain (small) number of messages to communicate with the exchange (i.e., it is "active"), it should register as an automated broker and trader. In my opinion, the registration process does not need to set capital requirements or offer trading privileges. It should, however, require the new registered entities to keep books and records, as well as to implement consistent policies, procedures and safeguards. It is time to go at least this far, so when the next flash crash or technological malfunction happens, the regulators could go deeper into the market ecosystem to piece things together.

Finance

Concentrated industry structure leads to inefficient behavior including arms races and rent seeking. I recommend that regulators closely examine why market forces are not eroding the high concentration of the HFT industry. At the end, the reasons for such high concentration might be benign, but the regulators should not just believe it to be so. They should get a solid understanding of why competition may not be working among the black boxes and take the steps to encourage it.

Technology

Knowledge about system latency can be behind the uneven playing field. To this end, automated exchanges should report system latency indicators to all market participants. Latency for messages for submitted, cancelled, modified, and executed orders should be reported on a periodic basis. This would greatly improve the transparency of the trading process in automated exchanges and level the playing field between those market participants that can estimate how long a bid or offer is likely to be available for trading and those that cannot.

Data processing

Automated trading venues are large systems that generate enormous amounts of data. Algorithms of all sorts – from a slow-moving automated execution program to a market maker to an opportunistic arbitrageur – run on data. Unlike humans who process information at roughly similar speeds, some algorithms are much slower than others. Slower moving "fundamental" algorithms might be optimizing their trading strategies over days, arbitrageurs over hours and minutes, market makers over seconds and high frequency traders over microseconds. On a typical day, differences in speed among different traders contribute to the strength and liquidity of the market. But at a time of market stress, algorithms might need to be aligned with each other, so the entire ecosystem of automated traders functions as a whole.

To this end, I recommend that automated futures exchanges broaden the use of short trading pauses and re-opening auctions. This type of functionality seems to have helped arrest and reverse the flash crash in the E-mini futures market. It was, however, tailored to kick in only after a giant "gap" has developed in the central limit order book. We don't need to wait for the gaps to become very big or even for the gaps to appear at all. Markets could kick into short trading pauses followed by re-opening auctions for a variety of reasons: too many messages, too long of a time to process, prices moving too fast. This functionality is not without a cost, but the benefits to public confidence – especially for the slower public – might be substantial.

Conclusion

For the public to remain confident that there are no stealthy predators lurking inside our automated futures markets, regulators need to demonstrate that they have drastically upgraded their skills. In the past, the public believed that a regulator is able to spot trouble by monitoring the movements of a designated human trader. Now, the public needs to know that a regulator has the tools and personnel to look for persistent patterns in the data.

The age when a regulator could rely on an overheard conversation to begin an investigation is over. Algorithms don't brag on the phone that they just "hammered the market" or send text messages to their girlfriends about how "fabulous" they are. To catch a manipulative or disruptive behavior of an algorithm, regulators need to have the technological tools to sift through communication and trading patterns among the new inhabitants of the market place – the machines.

This would require not only a substantial investment in new technology, but an equally, if not more, substantial investment in human talent. Since effective regulation of automated markets requires expertise in technology, finance, and data processing, regulators need to develop capacity along all of these lines. Regulators should also ask academia for intellectual guidance and help with building capacity. One of the main reasons I went to MIT after leaving the CFTC is to build this capacity for the regulators and the public. Efficient markets are a public good. We all have a stake in making them better, free of manipulation, abuse, and rent-seeking behavior.

Thank you very much the opportunity to testify before you.