

Market Structure: What we Know, and What we Need to Know

Presented by Gregg E. Berman
12th Annual SIFMA Market Structure Conference
September 21, 2011, New York

Good Morning. It is a pleasure to be here today and I very much appreciate the opportunity to speak with you once again about market structure.

When I was here last year, the topic was of course the May 6th flash crash. We were just putting the finishing touches on the final report which was subsequently released at the end of September. The results of our analysis show how a cascade of liquidity crises led to remarkable price swings as securities trading at twenty, thirty, or forty dollars suddenly traded down to a penny or less before bouncing back to normal pricing levels.

To recap what happened: The markets were trading in negative territory most of the day due to fresh concerns about the European debt crisis (it seems like the more things change, the more they remain the same). By 2:30 in the afternoon the S&P 500 had already declined 3%, only to lose another few percent by 2:40p.m. During that period of time volume was increasing, sell pressure was mounting, and many stocks trading on the New York Stock Exchange were triggering so-called liquidity replenishment points designed to temporarily suspend the auto-execution of trades when prices move very fast.

It was also at this time that a very large sell order in the S&P 500 E-Mini futures contract was placed to be executed by an automated volume-driven algorithm. Though the bid-ask spread in the E-Mini was relatively tight, the depth of book had been thinning all day, and detailed analysis by the staffs of the CFTC and SEC revealed that the steady pressure from this large order interacted with other automated algorithms leading to a chain reaction in which the order book for the E-Mini was effectively drained. We refer to this as the first liquidity crisis.

In just about 5 minutes the E-Mini fell a further 5 percent until trading was paused for 5 seconds by automated stop-logic employed at the CME. And as the E-Mini fell so did the broader markets and individual stocks. Arbitrageurs reported buying the E-Mini as it declined, and sold stocks or broad-based exchange-traded funds such as the S&P 500 "Spyder."

The rapid moves of individual stocks led to further selling as many retail stop-loss orders were triggered. Volume continued to increase, and for a period of time there were even lengthy delays in the public reporting of trade executions to the consolidated tape of a significant subset of stocks trading on the New York Stock Exchange.

The Securities and Exchange Commission disclaims responsibility for any private publication or statement of any SEC employee or Commissioner.

This speech expresses the author's views and does not necessarily reflect those of the Commission, the Commissioners, or other members of the staff.

The speed at which prices were moving triggered stops in many of the automated trading systems used by market makers. These stops are designed to prevent algorithms from continuing to trade when risk limits are breached, or prices are moving so fast that there is a question about data quality. From a market maker's perspective, if the rapid decline of prices was an indication of data integrity problems then they couldn't reliably trade. And if such volatility was indeed real, as it turned out to be, then some market makers may have not wanted to trade.

Particularly sensitive to such rapid changes in prices were market makers in equity-based exchange-traded funds, or ETFs. The price of an ETF generally tracks the value of the stocks underlying the ETF. If many of these stocks are moving very fast or erratically, it is difficult to get a "fix" on the aggregate value of the underlying portfolio, and therefore very hard to determine at what price one should bid for an ETF.

As a result, not only did some market makers in individual stocks widen their quotes, but many market makers of ETFs pulled back entirely. We refer to this as the second liquidity crisis. Unfortunately, even though market makers had temporarily withdrawn their quotes from over 300 securities, 70 percent of which were ETFs, orders to sell continued to hit the exchanges. As we reported, before market makers resumed quoting and prices returned to normal, over 20,000 trades were executed under these highly disordered conditions, and at prices more than 60 percent away from their pre-flash levels. These trades were subsequently deemed clearly erroneous and canceled by the exchanges and FINRA. With respect to the equity markets, it is the sudden evaporation of liquidity that I find to be the most alarming element of the flash crash.

So that, in a nutshell, is what happened well over a year ago in a thirty minute period on the afternoon of May 6th.

But what have we learned since then? What do we actually know about market structure? What data is available to help investors, professional market participants, and, yes, regulators, make decisions? More so, what do we need to know in order to address many of today's questions about market structure? How are we going to acquire such knowledge?

These are the topics I would like to explore this morning. As always, my comments are my own and do not necessarily reflect the views or opinions of my colleagues at the SEC, any of the Commissioners, or the Commission itself.

I'd like to start with two actions that were taken by the exchanges and FINRA in the immediate aftermath of the flash crash – namely the creation of a pilot program for single-stock circuit breakers, and more formal rules defining clearly erroneous trades. Since the first panel today will be discussing these topics in further detail, I don't want to spend much time on the mechanisms themselves. However, the implementation of these rules had an interesting side-effect, which was to highlight to all market participants, the media, and academics, every instance of a large, rapid move in the prices of stocks that were covered.

What seems to be the “common wisdom” among market participants, the media, and academics studying the market, is that each of these “stock events” represents a single-stock version of what happened on May 6th. I often read quotes by market professionals, or statements in academic papers, saying that though we haven’t had a repeat of May 6th, so-called “mini flash crashes” happen all the time, providing evidence of deep-seated problems in equity market structure.

Well, that’s quite a statement. And it’s been said so many times, that it’s almost become passé. When the first single-stock circuit breakers were triggered it was serious news. Today, there is barely a mention when these events occur.

If indeed the market continues to experience mini flash crashes, we need to understand this and perhaps do something about it. Recall that during the flash crash many securities traded in a very disorderly fashion as a result of the sudden, wholesale withdrawal of liquidity from the market. Does this really regularly occur in single stocks? Do liquidity providers simply withdraw their quotes under certain circumstances?

A review of notifications publicly issued by the exchanges and FINRA shows that from June 2010 through May 2011 there were 43 instances across 29 stocks in which single-stock circuit breakers were triggered. In addition, from August 2010 through May 2011 there were 61 instances in which price spikes in the trading of one or more securities did not trigger a circuit breaker, but did lead to a review by the markets to determine if these trades should be cancelled. It appears these events were not caught by the single-stock circuit breakers either because they occurred in securities that were not included in the pilot program, or because they took place outside the 9:45 a.m. to 3:35 p.m. operative period of the pilot.

My own observations and understanding of these events is that they are generally not caused by the sudden withdrawal of liquidity.

In 17 cases these events coincided with news about the effected company published during the trading day. Note that the markets currently have procedures for temporarily halting the trading of a security in the middle of the trading day in preparation for an important news announcement by a company. But in 16 of those 17 events news-related information was released by the media, on the web, or in rumors, and not by the company itself.

I believe that many of the other cases were the result of so-called “fat-finger errors,” such as mistyped security symbols or incorrectly specified limit prices. The simple transposition of a single digit can inadvertently turn a passive order into a highly aggressive one. Sometimes you can see from the public tape that a clearly-outsized order was placed, dwarfing immediately available market depth. In a few instances, it appears that off-market trades were incorrectly reported to the public tape itself. And in some cases it appears blocks of immediately executable orders were inadvertently generated, perhaps by a system, or by a human directing the system.

There are two important lessons to be learned from these events. The first is that there are a myriad of reasons for their occurrence, and classifying them simply as mini-flash crashes is not very informative.

Secondly, market structure itself may only be tangential to many of these events. It's certainly difficult for me to see how market structure creates unexpected news items. Similarly, it's not obvious how market structure causes fat-finger errors where the wrong symbol is typed into an order screen.

Don't get me wrong, these are important events that impact the markets. I am not trying to minimize their effects. Orders that are placed in error can be very disruptive to the market place. And in fact, the Commission last year adopted rules regarding market access and pre-trade risk management to directly address some of these issues.

However, none of these events appear to be caused by the sudden, yet temporary withdrawal of liquidity by market makers. It would seem that characterizing these as mini-flash crashes, in the sense of trying to make a direct comparison with the May 6th flash crash, is not only incorrect, but misleading.

But what is the big deal? Does it matter whether or not we call these events mini-flash crashes or not. Absolutely. And here's why: First, it's not exactly good for investor confidence. Would you want to invest in a market wrought with mini-flash crashes?

Second, and maybe even more importantly, if we simply want to describe and theorize about the markets without taking any action, then it probably doesn't matter if our characterizations are actually correct. But if we want to consider potential changes or new rules designed to alter market structure, we need to be confident in our assessment of cause-and-effect relationships. This requires a lot of care and precision in how we characterize and classify market effects. It also requires looking deeper than just the surface data. Standard economic tools that solely rely on statistical descriptions, correlations, and regressions do not often lend themselves to revealing causality. But it is causality that we must seek.

If these events are actually mini-flash crashes, then in order to address this problem we would have to look for solutions that deal with what would seem to be the random withdrawal of liquidity by market makers in individual stocks. But other fact patterns might lead to other types of solutions.

Understanding causes of actual events thus helps us understand how to address issues. Recall that circuit breakers were designed to halt the trading of a security during periods of uncertainty, such as during the flash crash, so that market makers and other participants can assess the situation and provide liquidity as needed. But sometimes other actions, such as a mistyped limit price, can trigger a large market move. In that case we need a mechanism that delays the triggering of a full circuit breaker so that the market can correct itself without the need for a full halt.

The proposed limit-up/limit-down plan contains such a mechanism. As stated in the proposing release, "...limit up-limit down requirements would be coupled with Trading Pauses [circuit breaker halts]... to accommodate more fundamental price moves (as opposed to erroneous trades or momentary gaps in liquidity)."

The solution is therefore informed by what we know and what we have learned, not simply by what we might hypothesize.

So far I've spent a good deal of time describing just one, somewhat small and specific issue related to market structure and the flash crash. As you can see, there is a lot of ground to cover in order to arrive at a possible solution. But what does this imply about the bigger and broader issues facing the markets today – fragmentation, the impact of high-speed computer-based trading, concerns about investor confidence, dark pools, new products, new trading mechanisms, and even swaps and other derivatives?

It means that for each of these issues we need to go beyond simple statistical descriptions to try to understand causal relationships whenever possible. Just as we may be able to discern the various causes for the instances of circuit breakers and clearly erroneous trades that occurred between June 2010 and May 2011, we need to know, or at least better understand, the causal relationships related to any questions or concerns about market structure. This is not an easy task.

Take fragmentation as an example. One form of fragmentation comes from the ability of market participants to trade the exact same stock via multiple venues. But we have rules governing how this may occur – they form the basis of Regulation NMS.

However, a much more complex form of fragmentation comes from the ability of market participants to trade the same stock via multiple products. For instance, an investor who wants exposure to stocks composing the S&P 500 index can do so by:

- Purchasing those individual stocks;
- Purchasing call options or selling put options on those stocks (both on exchanges and over-the-counter);
- Purchasing call options or selling put options on the index itself (also both on exchanges and over-the-counter);
- Purchasing futures contracts on the index, or on individual stocks;
- Investing in shares of a mutual fund that holds individual S&P 500 stocks, or otherwise tracks the index;
- Buying a structured note that is linked to the performance of the S&P 500 index or some complex formula involving a combination of one or more individual stocks;
- Engaging in equity swaps, volatility swaps, or dividend swaps, on individual stocks or the index as a whole;
- Purchasing shares of an ETF that tracks the S&P 500 or some subset of S&P 500 stocks; and even
- Selling shares of an ETF that inversely tracks the S&P 500 index. In this case though two “wrongs” may not make a “right,” two shorts do make a long.

What's remarkable is not the incredible number of ways in which one can trade the same set of stocks, but that market participants expect – no, they demand – that regardless of which path they choose, similar exposures should all be priced the same.

This is a tremendous ask. We often forget that there are no rules of nature that force prices of related products to be the same. No fundamental law of attraction pushes the price of an S&P 500 futures contract towards the price of a related exchange-traded fund. In fact, given isolated markets these

products would naturally drift apart. It is only through the action of market participants who continuously place simultaneous and opposing trades in both products that prices are kept in sync. We commonly refer to this as arbitrage. But note that arbitrage is not passive – prices don't "know" that they have to stay in sync, lest there be an opportunity to profit from arbitrage. Prices stay in sync because market participants are actively performing that arbitrage to capture those profits.

Consider a basic, stock-based exchange traded fund – not necessarily a broad index ETF like the Spyder, but let's say an ETF that simply tracks a particular sector or sub-sector. Investors seeking to gain exposure to a particular sector may find such an ETF to be an efficient way of implementing a long-term fundamental investment strategy, such as saving money for college or retirement. And who wouldn't want individual investors to be able to implement their long term holding strategies in an efficient manner? That would be like criticizing Mom and apple pie.

But how can an investor in an ETF be assured that the price he or she pays for this product accurately reflects the value of the stocks underlying the ETF at the very moment the trade is placed? After all, investors wouldn't want to pay a premium over the current value of the underlying shares.

The answer is of course arbitrage. We expect that professional market participants will keep the price of an ETF in line with the prices of its underlying components by constantly buying and selling shares of the ETF and its underlying stocks. Every time one of the stocks underlying the ETF changes price, the ETF price needs to update as well. More so, it's not just trades that need to be kept in line, but potential trades – in other words, quotes need to be constantly adjusted for the ETF and the underlying stocks. Even if just one fundamental investor during an entire business day decides to buy shares of an ETF, the market doesn't know when that order will be requested, and the ETF prices must be kept in line with dozens or even hundreds of other stocks throughout the entire day, regardless of trading volume.

Keeping the price of an ETF in sync with the prices of the stocks underlying the ETF can require rapid re-pricing, very fast algorithms, low-latency data feeds, and continuous re-quoting of the many securities linked through the ETF. These are of course some of the characteristics of high-frequency trading strategies.

Before I continue, a few important points: my example is not intended to call out ETFs. The exact same analysis applies to options, futures, swaps, and any other product that derives its value from other products. Also, I am not passing any judgment, either positive or negative, on the broader market impacts of high-frequency trading.

What I'm trying to show is that you can draw a fairly straight line between at least some of the activities of, and products used by, "slow-trading" individual long-term investors, and at least some of the resulting activities of high frequency trading systems. These activities are inextricably linked.

This is because our present market structure is itself the product of evolutionary advancements in regulations, technologies, products, venues, news, investor sentiment, and probably even twitter. It is not a simple mosaic of different actors operating in isolation. The interdependencies of every participant and every system has led to an exponential growth in complexity.

This is not necessarily bad. My I-Phone is vastly more complex than the computer systems used to get astronauts to the moon, but yet my 8-year old nephew can reliably download an episode of the Simpsons anytime he desires.

But if the download doesn't work, or it's slow, or the video is coming across in the wrong format, he can't debug the problem. And frankly, neither can I. The problem with very complex systems is that they are very hard to debug.

And that, I think, is a perfectly valid way of looking at the markets. We are trying to debug what has evolved into a very complex structure that generally works, and has provided many new benefits over the years; but any actions taken to address an issue or concern in one part of the system necessarily affects the entire system.

The implication is that we cannot simply hope to make changes to one part of market structure in isolation, or without consideration, of other aspects of the markets. We need to know the causes of the effects we are trying to address. We must be careful of the unintended consequences of changing rules and regulations. This is something that industry participants remind regulators of quite frequently.

But there is another side of this same coin.

In the nearly two years I've been at the SEC, I've seen an overwhelming number of requests by the industry to bring forth new products, new ways to trade, new venues, new margin methodologies, and new technologies. Each of these requests is generally accompanied by explanations of the business purpose and intent for the new innovation.

However, every one of these requests has the potential to impact market structure and lead to unintended consequences by creating new linkages, new relationships, and new complexities in the same way that new rules and regulations can. It behooves us to consider how all types of changes can affect the markets – innovations proposed by industry as well as rules proposed by regulators.

As such, some of you may have noticed that over the past year or so my colleagues and I have become quite inquisitive as we consider industry-driven requests to offer new types of products or services. Note that being inquisitive does not necessarily translate to being conservative. It simply means that our markets are so interlinked, and so complex, that we need to consider the broader impacts that any new product or service might have on market structure and ultimately, the protection of investors.

This is an area in which you can directly help. We have a formal process to gather and consider comments from market participants regarding rules that are proposed by the Commission. And in general we do receive lots of informative, thoughtful, and evidence-based comments. But the comment process is not limited to only Commission-proposed rules.

Market proposals regarding new products, new trading mechanisms, and other types of filings are also released for public feedback. Though we do receive some comments on particular filings, I am surprised that we do not receive the same level or depth of response as we do for Commission rule proposals. To re-iterate, the market structure we have today is, in large part, a reflection of evolving industry practices

accompanied by continuous change. We need to know how every proposed change, regardless of the source, might impact the marketplace as a whole. Market structure is as much of a function of exchange rules as it is Commission rules. Your input in this process is extremely valuable. It's part of how we learn what we need to know.

In contrast, way on the other side of the spectrum, there is indeed a certain, very large topic for which we always receive a tremendous amount of feedback. I am of course referring to high frequency trading.

I have recently recognized that in parallel with the rise of high frequency trading another HFT phenomenon has developed: high-frequency theorizing. Under this form of HFT, it seems that each day for every new theory explaining why high frequency trading is beneficial, another theory is proposed about why it is not. The truth is that this is very complex issue, and there is a surprising dearth of detailed, causal-based analysis.

It's important to recognize that regardless of how trading is done, there are already rules and regulations to address a host of illegal or prohibited activities such as fraud, insider trading, cyber attacks, and manipulation. These activities can be, and have been, done by humans at a slow pace. And they can also be done by computers at the direction of humans at a very fast pace, in which case I guess I might call that high-frequency violating. But what's illegal is illegal at any speed. Rather than hypothesize about the potential for illegal activities that are already covered under existing rules, I'd like to focus on the more complex issues regarding HFTs and market structure.

As I hope I've demonstrated through my example, the development of certain types of new products, new ways of trading, and new venues, can create linkages between existing securities with an expectation by market participants that prices will always be aligned and fair. Satisfying this expectation may require arbitrage strategies that rely on the rapid trading, quoting and re-quoting of lots of securities. High frequency trading does not sit outside of, or parallel to, the rest of the market. Any analysis of high frequency trading is necessarily a consideration of market structure itself.

For example, we know that during the second week in August, the markets experienced an extraordinary amount of volatility, under somewhat extraordinary circumstances. It's not every day that an NRSRO downgrades the rating of U.S. government securities. And it's also not every day that securities rally when they are downgraded. What role, if any, did high frequency trading play in the wild swings observed that week?

I had an opportunity to speak with a variety of market participants to get their perspective on these events. There were also quite a number of articles in the press summarizing and quoting what market professionals were saying about such volatility. I also spent the better part of some days watching the markets react to public speeches, news, and announcement as they came across the TV.

My own observations are that though the markets were extremely volatile, and volume was very high, the markets themselves functioned in an orderly fashion – there were no signs of flash-crash like issues

or disorderly trading. There were a few small system glitches, as one might expect during such conditions, but nothing like the market data issues that occurred on May 6th.

Curiously, though there were many wild swings, there were only a handful of S&P 500 and Russell 1000 stocks that triggered their circuit breakers. There were however, a number of very low-priced stocks that triggered their circuit breakers – ironically these stocks were just added to the pilot program that very week. In addition, on some days thousands of individual stocks triggered the Reg SHO price test and were subject to restrictions on aggressive short-selling.

So, did high-frequency trading create, contain, minimize, or exacerbate volatility? Those I have heard from and read had differing opinions, some quite passionate. But I've not yet seen any evidence one way or another, and clear, identifiable causal relationships are not immediately obvious. However, I have heard that even though volume approached record highs, liquidity did not seem to follow. The difference between liquidity and volume is something we explored in our studies of the flash crash, and though the concept of liquidity can be a bit nebulous, we do know that it is not simply the same as volume.

If we are to fully understand market structure, we need to know the answer to these types of questions.

But lo and behold, no sooner had we asked the question than the United Kingdom's Government Office for Science released the results of their 9 month study titled the *Future of Computer Trading in Financial Markets*. Sponsored by the UK Treasury, a panel of leading market experts was assembled to both review and generate data and studies related to what they call computer-based-trading and its implications.

I've just finished reading through their high-level 50-page working document that comprises three summary papers, and in case you haven't read through the papers, consider this fair warning – spoiler ahead.

Rather than paraphrase the key findings of their first summary paper, it might be best to simply quote them:

Key Findings

Economic research thus far provides no direct evidence that high frequency computer based trading has increased volatility.

Wow, that sounds pretty conclusive. But wait, there's more:

However, in specific circumstances, a key type of mechanism can lead to significant instability in financial markets with computer based trading (CBT): self-reinforcing feedback loops (the effect of a small change looping back on itself and triggering a bigger change, which again loops back and so on) within well-intentioned management and control processes can amplify internal risks and lead to undesired interactions and outcomes.

The feedback loops can involve risk-management systems, and can be driven by changes in market volume or volatility, by market news, and by delays in distributing reference data.

A second cause of instability is social: a process known as normalisation of deviance, where unexpected and risky events come to be seen as ever more normal (e.g. extremely rapid crashes), until a disaster occurs.

Hmmm, that doesn't sound as conclusive. In fact, I'm not sure what to make of these findings. They seem a bit contradictory. HFT activities don't increase volatility, except under some circumstances, when they might create significant instability.

What's a regulator to do?

To be fair, this is a very important study, undertaken by experts with significant experience in market structure. The papers contain terrific discussions of the issues surrounding computer-based trading, and include an excellent overview of the current literature. But I'm not sure that when all is said and done, a definitive or conclusive roadmap emerged that suggests what next steps, if any, should be taken. With one exception: the need to continue probing, to keep uncovering facts, analyzing data, and asking questions.

And in that I think we are in complete agreement.

A few months ago the Commission adopted a Large Trader Reporting rule that greatly enhances the utility of the existing Electronic Blue Sheets System by assigning a common and unique identifier to all trades by the same large trader. However, as the name implies, this only cover trades, not quotes or orders. To track quotes and orders as well as trades we must look towards the proposed Consolidated Audit Trail. Both of these initiatives are intended not only to augment the ability of regulators to better oversee the markets, but also to provide a wealth of continuous data that can inform broader discussions about market structure, including the role and impact of high-frequency trading.

As discussed in the Commission's proposing release, the Consolidated Audit Trail would for the first time provide regulators with a linking of not only all orders, trades, routes, and cancellations, but also with information on which parties were responsible for each of these events.

In commenting on both the Large Trader and CAT in public testimony of last December, Chairman of the SEC Mary Schapiro stated:

These proposals would tremendously enhance regulators' ability to identify significant market participants, collect information on their activity, and analyze their trading behavior. Both of these initiatives seek to address significant shortcomings in the agency's present ability to collect and monitor data in an efficient and scalable manner and to address discrete market structure problems.

And indeed SIFMA itself through one of its comment letter stated that "...a consolidated audit trail is a much-needed improvement over today's fragmented audit trail platforms."

Though I've been in finance for nearly two decades, my formal training and degree is actually in experimental physics. And as an experimentalist, I can tell you that there is no such thing as too much data. In fact, there are often good reasons to gather similar types of data from different perspectives – different view point sometimes provide for different insights or conclusions. So I like to conclude today by demonstrating how this practice can be applied to the present topic.

In the spirit of the old adage "it takes one to know one," we've recognized that the best way to understand high frequency trading might be to become a high frequency trader.

Though that would also be an interesting way of supplementing the agency's budget, even a non-lawyer like me knows that would violate countless rules, regulations, and laws. But really understanding the world of high frequency trading does not necessarily mean we actually have to trade – more likely we need to further immerse ourselves in market structure, study the ebbs and flows of trading, bid-ask spreads, quoting...and quote cancelling.

Last year we reached out to the industry in a request for information regarding the availability of tools that might be used by market professionals who trade in real time. We were quite pleased to find that it is possible for us to procure systems, databases, and applications that will enable us to read, monitor, and analyze the same low-latency trade-and-quote feeds that many of the most sophisticated market participants use today.

If you are a market maker or high frequency trader, I wouldn't worry about the Commission competing with you for order flow. However, such systems would allow us to observe every trade, every quote, and every cancel, at just about the same time, and in the same way, that other market participants do today. Personally, I am quite excited about possibilities like these.

The markets may be complex, but they are not inscrutable. As such, I hope that this morning I've been able to provide you with some color on what we currently know, and, perhaps more importantly, how we will continue learning what we need to know.

Thank you.