

May 15, 2014

Ms. Elizabeth M. Murphy
Secretary
Securities and Exchange Commission
100 F Street, N.E.
Washington, DC 20549-1090

Re: File No. Equity Market Structure 2013

Dear Ms. Murphy:

Thank you for the chance to comment on the SEC's new market structure portal, and congratulations to the SEC for launching such an innovative website. Not even a year old, the portal already has valuable data and analyses about our markets and is a useful tool for understanding at least some market dynamics. It's hard to think of another regulator that has reached out to the public in this way. It's also a striking contrast with the SEC of the past, where quantitative analyses, if they were done at all, were rarely and only very reluctantly shared with the public, and the data itself kept securely within the agency.

The SEC's newly launched MIDAS tool is also a good step forward. The tool takes in SIP data feeds and proprietary exchange data feeds and maintains a consolidated database of it all. SEC analysts can use that data to study aspects of the markets down to microseconds, broken out by stock and exchange.

My interest in these facilities was piqued by some recent remarks from Chair White about MIDAS. During her recent appearance before the House Financial Services Committee, she said she was "making certain that the SEC and its experts had the data they needed to *fully* understand all of the market structure issues and then respond appropriately if there's a need to respond." (Emphasis hers.) She then highlighted MIDAS as a core initiative, "The knowledge base of the Commission has been enhanced significantly by being able to bring on the MIDAS technology."

My interest sharpened further a few days later when Bloomberg reported that a senior SEC staff member had dismissed criticism the agency was behind the times. Bloomberg reported that at an industry conference in Texas the staffer pointed to "a new surveillance system, known as Midas, that collects price data from all U.S. exchanges and that helps the SEC evaluate practices such as co-location, in which traders cut trading times by placing computers inside an exchange, and proprietary feeds, or streams of data sent to clients faster than to public sources." Bloomberg paraphrased the official's bewilderment at criticism of the SEC by writing, "The agency can't be both ignorant of market behavior and have access to such expansive data"; Bloomberg then quoted him directly, "One of those statements has to be wrong," and "Those both can't be true at the same time."

I'm writing to briefly comment on the data in MIDAS and on some of that data's limitations, and to comment on the risks of drawing broad conclusions from that data without accounting for those limitations. It all suggests that, at least for now, both of those statements are true at the same time.

Data

The portal notes that MIDAS includes posted orders from the exchanges, but it's important to be precise: the exchange data in MIDAS includes only *displayed* and posted order detail, and only that from a subset of exchanges.

While order type statistics from all the exchanges aren't publicly available, some exchanges do publish those statistics. Of the exchanges that do, somewhere between 50 percent and 65 percent of all orders on those exchanges are not displayed or are at best conditionally displayed. If an order isn't displayed, its detail isn't distributed over exchange data feeds. If it isn't distributed over exchange data feeds, it isn't in MIDAS. That missing data – perhaps as much as half to two thirds of all exchange order detail - is a significant source of error unless it is well understood and all its implications are disclosed.

There are other sources of error. Among them, NYSE doesn't distribute order level detail over its proprietary feeds, so MIDAS doesn't include order details for NYSE trading, a loss of about 12 percent of the overall equities market. This is a much greater loss than its market share implies because NYSE is the primary market for so many of the country's largest companies and most active stocks. MIDAS also doesn't include dark pool and OTC market maker data, which leaves out perhaps another 40 percent of the overall equities market from its databases.

Extrapolating and then combining these limitations to understand what MIDAS databases include implies that MIDAS has order activity for only about 25 percent to 40 percent of today's lit markets and only 15 percent to 25 percent of the entire equities market. This is still valuable information, but it is far from complete information, and anyone analyzing MIDAS data should do it cautiously and with a lot of disclosure. Though in one section of the portal the SEC discusses at least one of these gaps in the data - NYSE order detail - the portal should disclose a full accounting. A subpopulation like this inevitably breeds selection bias and is far from Chair White's goal of making certain the SEC has what it needs to "*fully* understand all of the market structure issues."

MIDAS has other limitations. Of the dozens of data elements a firm can put on an order before sending it to an exchange, as a rule MIDAS receives only four of them from exchange data feeds - stock, price, size, and side. And then, of those four data elements, three are clipped before they ever get to MIDAS. Order size is clipped to contain only the displayed size of the order. The same is true for order price. The order's side is clipped to simple buy or sell and doesn't include any short sale indicators. Orders are also not marked with capacity indicators, like principal or agency. Crucially, trades and orders in MIDAS don't include account data or other participant identifiers, meaning there's no way to identify who is doing what in the markets, to isolate a firm's trading, or to identify what firms were involved in any particular market event. It also means there's no way to consolidate market activity into participant categories, like institutional or prop trading, and understand where and how these participants enter the market. These matters are important enough to deep analysis that the SEC's own recent review of high frequency trading research, available on the portal, deliberately excludes studies that don't have participant identifiers. It excludes those studies because they don't have the "trading account data that can be used to distinguish different types of market participants," making it impossible to sort out who is doing what. Presumably research based on MIDAS data has the same handicap.

MIDAS has almost no information at all about order types and order modifiers, other than the obvious fact it contains only displayed orders. It doesn't include any information about intermarket routing behavior, about

what orders are routed into and out of a market, so MIDAS can't give any insight into marketwide interactions even as the equities market becomes more fragmented and dispersed.

Altogether these add up to very significant limitations, and they are getting worse. More trading is migrating to off-exchange venues, and at the exchanges displayed order usage is shrinking fast. Last year one exchange's displayed orders declined as a proportion of all orders by almost 15 percent. At that rate the already small proportion of order data MIDAS ingests now will be cut by a third by the end of 2016, and at an extreme, and without intervention, MIDAS might then see as little as ten percent or less of all detailed order data.

The exchange feeds in MIDAS suit many purposes. They certainly do a good job for the high frequency firms for which these feeds were first designed, especially since the time and sales data they include can leak information about certain hidden orders. For more than a decade HFT firms have used the exchange feeds in many clever ways to dominate exchange markets. But a short-term algorithmic trading decision is different from a policy decision, and policy decisions should be based on data that equips regulators to *fully* understand their markets. MIDAS can't and doesn't, and even less so as the months tick by and more of the market goes dark. If regulators will ever understand the market they need to see more than one end of it, however well lit. While everyone hopes the CAT will do this, the CAT is years away. In the meantime, regulators should be completely upfront about their own data's limitations and disclose what those limitations might mean.

Judgments

Data limitations affect policy judgments, and that's why some comments in a recent speech by SEC Associate Director Gregg Berman are worth exploring. Though he pointed out he spoke for himself and not for the Commission or for other SEC staff, he is senior staff and his speech is likely an insight into policy decisions brewing at the SEC, decisions based on MIDAS data.

Mr. Berman reported that while 39 percent of all orders found in MIDAS were canceled within a half-second, 27 percent of all trades were against orders that had been resting for a half-second or less. Drilling further into that data he found that while 23 percent of all orders were canceled within 50ms, 19 percent of all trades were against orders that had been resting for 50ms or less.

For orders that were priced at or better than the inside, he found that more than half of all orders were canceled within a half-second and 25 percent of all trades were against orders resting for a half-second or less. 38 percent of all orders were canceled within 50ms, and 19 percent of all trades were against orders resting for 50ms or less.

From all this, Mr. Berman said his analyses on MIDAS data showed that "the speed of systems that take liquidity by accessing displayed quotes seems to be keeping up with the speed at which those quotes can be canceled." It is a very significant judgment and shows plain affection for the current state of things. While the market has a lot of very fast orders, it also has some trades against those orders, which he believes supports that judgment. But without better data it is no more than wishful thinking.

The first question we should ask about these results is what it means when nearly 40 percent of orders at or better than the BBO are canceled within 50ms. At least over ordinary electronic pathways, it's not nearly enough time for a quote to make a roundtrip from New Jersey, where quotes are created, to anywhere near the west coast, where about 15 percent of the country lives and nearly 20 percent of the economy is

produced. It's certainly far less time than human beings need to react to it no matter where they are. Structurally, then, nearly half of individual BBOs are something that only machines can hope to respond to, and then only if those machines are somewhere east of California.

Next, given the latencies in routing data to the SIP, and then from the SIP outward, we can make another structural judgment, namely, any machine hoping to respond to any exchange quote, fast and slow alike, needs to have exchange proprietary feeds. SIP data just isn't fast enough. Even a millisecond lost shipping data to the SIP or in processing data at the SIP means missing a significant number of quotes, especially in busy markets. The excellent data visualization tools on the portal imply that *15 percent or more* of all orders at or better than the inside are canceled within a millisecond, at least among large stocks.

The first certain judgment so far is that "accessing [all] displayed quotes" requires a co-located system and proprietary feeds. What does that mean? At current prices, a rack and data feeds at any one of the major exchanges start somewhere around \$10,000/month. It's an extraordinary amount of money, and about twice what it cost to lease a seat at the NYSE in early 2005, just before NYSE announced plans to go public and seat prices jumped.

Because so much quoting is so fast, full and competitive access to displayed quotes today at any one of the major exchanges now means spending \$100,000 a year or more to get in the door, nearly twice as much as a seat on the nation's largest and most prestigious exchange cost less than 10 years ago. It's déjà vu all over again, where privileged insiders see all the flow and act on it as outsiders never can.

And so it isn't farfetched to call these kinds of mechanics rigged, though at least one FINRA executive prefers calling them "first class."¹ The old insiders bought or leased seats, and the new insiders lease racks and buy data feeds. For all the changes of the last decade, we still suffer a two-tiered market utterly dominated by a small club of insiders, except the entry fees are much higher than they used to be while regulations meant to constrain the club have been passionately dismantled.

It's too bad Mr. Berman doesn't take his thinking in this direction, though these questions might be beyond his role or inclination. Instead he says that if a significant number of trades are against short-lived orders it's evidence the market keeps up with even the fastest quotes.

But who is it keeping up? Are they institutional traders or prop traders, or even retail, or orders from other markets, or a mix of all these, and in what proportions? We can't say because MIDAS can't tell us. What if the quickest trades are initiated almost exclusively by co-located prop firms? What if the quickest trades are initiated almost exclusively by the same small handful of prop firms? What if the fastest orders come from the same small handful of firms? What if co-located prop firms are nearly always first in price queues, squeezing out the public? Of the orders canceled within a millisecond, do any of those firms soon trade at that price, and what is their order-to-trade ratio, or is it all quote spam? Who was on the other side of those trades? What if aggressive trades by prop firms somehow neatly predict subsequent price action, and include their most profitable trades? How frequently do market making firms, armed to the teeth with an arsenal of regulatory privileges no one else has, turn aggressive, and what is the effect on prices when they do? What other trading strategies do market making firms deploy, and do they use their privileges - privileges intended only to help them provide liquidity to investors - when they deploy them? How frequently do market makers sell short aggressively? How frequently do they turn and run, cowering behind minimum quoting standards? How much inventory will a market maker hold before aggressively liquidating it, risking

¹ See <http://wealthmanagement.com/blog/high-frequency-trading-buying-first-class-says-finra-enforcement-chief>.

the kind of "hot potato" trading that slammed prices in the flash crash? What portion of displayed liquidity comes from the same firms scattered across the exchanges? When a firm trades on one exchange, does it cancel or reprice its quotes on others? Does it then trade aggressively on other exchanges, and in what direction? Does it even sell short aggressively on other exchanges? And is any of this behavior different depending on the exchange, or when markets are volatile?

Knowing at least some of this is essential before making any normative judgments. And as consequential as answers to these questions will be, what if many of those fast trades are just coincidences?

For an example of that, we can look to Mrs. Betty Johanssen in Red Lake, Minnesota as she enters an order to buy 100 shares of MMM. If her arthritis isn't acting up it takes her less than a minute to log into her broker's website, set up a marketable limit order, and click send. It might take another half-second or so for her order to make its way from her computer through her broker's systems and on its way to an exchange (let's pretend it actually does make it to an exchange).

No matter how long it took Mrs. Johanssen to enter the order and for the order to make its way to New Jersey, that order is going to trade against the inside market. Since almost 40 percent of all orders at the inside rest for no more than 50ms, Mrs. Johanssen's order has a significant chance of trading with a short-lived order completely by accident. Another way to put this is that of course a lot of short-lived orders trade if enough of them are at the inside, because the inside is where orders trade. Without first studying who is behind them, no one should think these kinds of random collisions are purposeful and then turn a coincidence into a policy judgment. You'll hit something if you throw pebbles into a fan, but that doesn't mean you aimed at what you hit.

The important question isn't whether fleeting orders at the inside or even deep in the book are tradeable. In any automated system they are tradeable, both on purpose and by accident. The important questions are what portion of exchange liquidity they provide, what costs they impose on other investors, who or what can trade with them on purpose, and what that kind of access costs. Once the SEC answers those questions we can frame new questions and perhaps begin to make normative judgments. No matter how much we might want it to, for now MIDAS can't help.

In light of all this, it's premature to call MIDAS a "surveillance system," at least as that term is used in industry. Unless there's more to MIDAS than we've been told, it's also premature to say MIDAS helps the SEC keep pace with or stay ahead of its critics, or helps the agency fully understand market structure issues. The data to achieve those goals resides inside exchange audit trails, or at FINRA, or might be collected someday by the CAT, but it's not yet in MIDAS.

It's a common failing to believe an exciting new tool will fit whatever comes along. MIDAS is a great step forward but there's much further to go. For now the SEC should think very broadly about what questions it needs to answer and not strain to answer them hastily with a tool that isn't suited for it. Doing us no good at all, that will only nurture more cynicism about the industry at a time we need to nurture goodwill.

Sincerely,

R. T. Leuchtkafer