A technological solution to best execution and excessive market complexity

Stephane Tyc
stephane.tyc@quincy-data.com

Quincy Data, LLC

September 18th, 2014
Version 1

Introduction

The Flash Crash of May 2010, the software bug of Knight Capital, NASDAQ's glitch on the first day of trading of Facebook and the publication of *Flash Boys* all motivate the calls to improve regulation. Both the SEC and FINRA are examining ways to improve market structure and regulations to prevent any players from having an unfair advantage. Many of the proposed improvements involve reducing the number of trading venues and regulating how an order must be routed. There are additional calls for creating more comprehensive data gathering and identification of orders and executions across all trading venues. These changes would create an additional burden on brokers and traders who are already suffering from vastly increased compliance and data gathering costs due to the regulations created by Dodd-Frank.

We believe there is a way for regulation to be simplified and made more powerful at the same time. Trade publication standards can be created to support improved customer choice and to simplify and strengthen the market place. This would replace the need for more complex and costly regulation. Our suggestions apply to both the USA and to Europe but this paper will concentrate on the unique market structure of the U.S. equity markets after Regulation NMS (Reg NMS).

Our proposal would enable market participants to make rational choices on complete information about the quality of their execution. It would remove instability and complexity in the markets. It would also increase transparency for less liquid securities and keep the competition between trading venues fair and vigorous.

Best Execution and the organization of fair competition

In order to motivate our proposal, let's first revisit the two goals of Best Execution and Competition. In his 1996 paper, Lawrence Harris, discusses Best Execution: (Harris, 1996):

> "When brokers take customer orders, they assume an agency responsibility to obtain ``best execution". Unfortunately, best execution is not well defined.

Best execution means different things to different people. To unsophisticated customers, best execution may mean ``get the best price possible'' for a market order and ``trade as quickly as possible'' for a limit order.

(...) The most sophisticated customers (...) only pay for the level of execution quality that they can audit. For them, best execution means ``get me the execution that I expect you to provide given what I pay you and the limitations of my ability to audit your performance.'' These traders define best execution relative to the costs of auditing it."

The key to best execution is to empower consumers to analyze the quality of the execution they are getting. This way they can make the right decision and buy the level of service they need from the brokers. We have a proposal that would make this analysis simple, cheap and efficient.

The first stated goals of Reg NMS was to address competition, as a mechanism essential for markets (SECpg12):

NMS Principles and Objectives

Competition Among Markets and Competition Among Orders

The NMS is premised on promoting fair competition among individual markets, while at the same time assuring that all of these markets are linked together, through facilities and rules, in a unified system that promotes interaction among the orders of buyers and sellers in a particular NMS stock. The NMS thereby incorporates two distinct types of competition -- competition among individual markets and competition among individual orders -- that together contribute to efficient markets.

This goal was addressed by introducing order protection which links some of the markets and insures that, to a large extent, a market order sent to one of the protected markets will meet the best resting order of all protected markets. This goal is only partially achieved by the rule because it is simply impossible to achieve perfectly. Markets are physically separate and when the operator of a market identifies an order in a different place and sends a matching order, there is no guarantee that the order will be matched because it takes time to reach the away market.

Harris' comments on best execution were made before decimalization and before Reg NMS. They are still valid but the situation has changed as the vast majority of the trading is computerized and auditable.

Prior to the changes, order execution was primarily done by humans either in pits or "upstairs" on the phone at the desks of dealers. This presented several problems.

Trade auditing was a challenge: brokers were faced with a difficult task if they wanted to document best execution for clients. The trade information was recorded poorly and the time

stamps of the trades were approximate, the quotes were not formalized and could not be analyzed efficiently. Order handling was a lucrative business and brokers were directing their order flow to dealers on the basis of factors other than the quality of particular executions. In return dealers would offer various inducements to brokers including direct "payment for order flow". So, not only was the task difficult, but also the incentive of brokers to measure the quality of the executions provided by dealers was not very high.

The quality of the service offered by exchanges, mostly open outcry markets was also hard to assess. Pits are both inefficient and difficult to police. The famous article by Christies and Schultz that statistically demonstrated collusion between market makers led to a first swath of rule changes. Even with those changes, as long as the trading remained at the hands of humans, price and time priority were not fully enforceable and order-handling rules, in general, could not be "programmatically" defined.

In this context the average bid ask spread was large and the cost to end users was high.

Decimalization and then Reg NMS %changed this forever. It introduced competition between exchanges and made collusion between different market makers almost impossible in liquid stocks. It reduced the scope for dealers to execute outside the current bid/ask and disadvantage their clients. In addition, the introduction of the automated Small Order Execution System was first met with resistance but it is now the norm for equity trading.

The ultimate effect was to lower effective bid/ask spreads and to render markets more efficient and more traceable. But best execution is still not well defined and still not easy to assess for customers.

Despite the intention of insuring fair competition in the markets, there were unintended consequences. First the linkage between electronic order books introduced by the quote protection rule of Reg NMS has created a complex dynamic which is extremely hard to analyze and whose behavior is unpredictable. Second, most markets have introduced, and regulators have approved, many new complex order types mainly used by professionals. Those new orders have often been introduced to circumvent the difficulties introduced by quote protection and by routing between protected exchanges. There is complexity globally in the linkage and there is complexity locally in the order type. It is little wonder that there is broad agreement that the markets should be simplified.

Our Proposal

We are putting forward a very simple proposal that, in our opinion, would greatly improve market dynamics by changing the elements of Reg NMS that led to the increased complexity and would create more transparency.

The proposal has two legs. First, make data available in a format that will foster analysis and therefore enable rational choice. Second, remove the order protection rule which is the source of most of the complexity in current markets.

First, in order to make markets more efficient for market participants, they must have access to real numbers and real reporting on the quality of their executions. Mandating a particular form of data analysis is problematic. Access to the raw data should generally be preferred unless it imposes undue burden. The analyses are hard to define in a way that is sufficiently detailed to be completely trustworthy. The real test is simple. If two brokers with exactly the same executions data were to produce 605 reports it is unlikely that the reports would be identical. Defining the way to process the data is not as efficient as providing the data itself. What needs to be done is to provide access to the data free of charge to anyone who wants to do research and provide analyses.

%Second, to make markets easier to operate and more robust, the linkage imposed by order protection should be removed and arbitrage should be relied upon as a tested, trusted and visible market mechanism to keep the system in synch.

1. **Trade Data Publication:** standardize the publication of trades on all venues and make the data widely and publicly available.

2. **Linkage Removal**: rescind the order protection rules of Reg NMS (only if the first proposal is implemented)

Trade Publication can be implemented without Linkage Removal, but order protection cannot be rescinded without the availability of comprehensive trade execution data. The publication of trade data should have the following requirements:
- Apply to any trading of securities.
- Apply to all venues, including exchanges and dark pools and other forms of trading.
- Reported in a standardized way.
- Report only trades not quotes or other order messages, thereby greatly simplifying the implementation and empowering true independent analysis. This point of the proposal is key and will be discussed later.

Linkage Removal is simple to implement. It is sufficient to remove the rule making it compulsory, and exchanges and smart order routers can evolve at their own pace to stop following it.

The new regulation would require that:

1. All matching engines or individual traders effecting manual trades be registered with a unique number identifying them.
2. All matching engines be time synchronized to an accuracy that is within 10µs of the global time standard UTC and manual trades be time stamped within an accuracy of 1 minute.
3. All trade data be published, price, quantity, symbol, buy versus sell, etc, with the matching engine ID, the trade ID and with the time stamp of occurrence and time stamp of publication to a publicly accessible data stream
4. All the above trade data be made accessible free of charge and free of copyright in a common format at the end of the day with an open data license.

5. Brokers be legally required to communicate to clients, upon request, the list of all the trade IDs and matching engine IDs that constitute the execution of a given order. This report would be a simple .csv file with at least five columns including: Security identifier; matching engine identifier; date; trade identifier; fraction of the trade allocated.

Implementation issues and expected impact

How would it work?

Who produces the data, who aggregates it and how is it consumed?

The owners of matching engines or the employers of human traders produce the data. It is formatted according to a precise specification and it is sent to an organization that has a SIP-like mandate. This NMS organization makes the data available and is responsible for auditing it and reporting to regulators on the quality of the data per source. The data is then available for download free of charge and free of copyright.

How is this different from the TAQ data already available?

This only concerns the trade portion of the TAQ data, making it much smaller and easier to use. But it incorporates more information to help identify where trades occurred and the precise time stamps. It also removes any exception that may be present in the TAQ data, such as off hour trades. Perhaps it would be easy to generalize the TAQ data to include the new information. We are arguing that producing this data should be simple and that there is no good reason to delay implementation.

Why do you need two time stamps?
The time stamp of occurrence of the trade is the most important one. It is also interesting to study this impact of the dissemination of information in the trading system. If a very large block trade is done manually and reported only after some time it is interesting to know both time stamps to understand information transmission.

Who would create the matching engine ids and human ids.

In principle there are already trader ids available, so this would not need to be created. The matching engine ids would need to be registered with a regulatory body. The definition is that those ids should be sufficient to find the particular program that effected trades on a particular day. It should also include the location of the servers where the program would run. For instance, if a trade was done on the disaster recovery site of a dark pool it would need to carry a different matching engine id.

How is this new unique identifier different from the MIC?
The Market Identifier Code, defined by ISO 10383 is not precise enough. It cannot identify the actual matching engine that performed a trade. It does not have the precise location of the

matching engine. A given MIC code could correspond to matching engines which are 8000km apart. It would be useless to have a time stamping precise to better than 10μs without knowing where the matching engine is located.

How long would it take to implement?

The size of the data is very small. All the orders executed on any given day in a terse format would fit on a small thumb drive. Building applications to analyze executions would be possible at a small cost and effort. It would be open to anyone with some computer skills. The key to making this simple to use is to have a very precise specification of the data and to keep the data as simple as possible.

When should the Transaction Reporting be made accessible?

End of day reporting is adequate because this data is not supposed to be used in real time by brokers and proprietary traders. This is about enabling consumers to analyze the quality of their executions. The time scale involved in deciding to change an executing broker is closer to one year than to one day. Also, when studying the execution of illiquid securities it may be good to look at trades over many days.

Why is it cheap?

All electronic matching engines already have the necessary data on all trades. All it would take is to synchronize the data with the Institute of Electrical and Electronic Engineers Precision Time Protocol and provide the data with unique identifiers.

Why only publish trades, are quotes not also relevant?

Of course quotes are relevant. The market surveillance and abuse monitoring by regulators will still need access to quotes in some form. However, the key to empowering customers and to creating an independent cottage industry of individual trade analysis is to provide a simple system. Quotes come in a wide variety of guises. It will prove very difficult to document all quote types, including those provided by dark pools, in a single open data regime. Other data that is required to do the analysis is already available for lit exchanges and should also be available from dark pools and voice trading venues. The key insight is that trades are "sufficient" to empower customers to choose the best way to execute their trades.

Is this supposed to replace SEC Rule 605 and the publication of execution statistics?

This is not a replacement for Rule 605. It is meant to complement this rule. Current execution quality reports are centered on statistics per equity over all the executions in any given month. They provide no transparency on individual executions. The data used to compute the reports are not fully standardized and there are many exemptions. Providing open data on actual executions will help other and possibly more insightful analyses.

Why is it better than the current "best execution" reports required by Rule 605?

Some aspects of the reports could be gamed; in particular the number of shares that received price improvement can be gamed by providing an insignificant improvement to many shares in order to publish advantageous numbers. Providing the data will help better and more insightful analyses. For instance, nobody reports the quality of trade execution on "correlated assets". What about a broker who buys an ETF at the offer price on the leading cash market, in compliance with the quote protection rule, but ignores that there is a future on the same financial asset whose price was significantly below? It is impossible to define all the rules that should be applied by very good executing brokers. It is very simple to leave this to the investigation of a curious, competent and empowered public.

Why pick 10μs as the time precision?
There are two reasons: it is now routinely possible to synchronize to about one microsecond; and, the maximum error should be much less than the time information takes to go from one matching engine to another one. Ten microseconds seems to be a good compromise but 5μs or 20μs would probably be equally acceptable.

Can you introduce waivers for reporting?

Absolutely and categorically not. The whole point of the system is to provide a complete transparency and remove all the suspicion clouding the trading process. Waivers are often argued for in the case of large trades or illiquid securities. There is no reason to exclude large trades and illiquid securities as long as the reporting is done at the end of the day and not in real time.

How is this different from the consolidated tape proposals?

Consolidated tape proposals aim at improving the price formation process by providing rapid feedback on trades. Consolidated tape may be useful for market makers, professional traders and brokers. Our proposal is aimed at providing transparency for the end users, the money managers or individuals. Some money managers or individuals do not have the means or the interest to perform real-time analysis of trades and alter their real-time trading patterns even with the existence of a consolidated tape.

 Do we need to trace orders through the various internalization, give-ups and other life cycle events?

Tracing orders can be useful but it is very difficult to put in place and probably impossible to make available publicly. Publicly available trade data are sufficient to study best execution. The end result of complex order routing strategies must be made available on a private basis. Executing brokers have internal mechanisms to allocate trades to particular clients. Those allocations must be made available for their clients upon request, so that the client can see the original execution of the trade. Armed with this data clients can see which split they received from executions and have the data at hand to request explanations from brokers if needed. This should be traceable down to the subaccounts.

How can this work with internal matching of trades by brokers?

It works with any normal allocation mechanism. Let's take examples:

1) Client A sends an order to buy 10 shares of IBM and Client B sends at the same time an order to sell 7 shares of IBM. The broker will match 7 shares internally at a price of \$191.74 and will buy an extra 3 shares on the NYSE-Arca for \$191.77. Client A would get an execution report with 100\% of the first trade and 100\% of the second trade, both trades would be on different matching engines, at different prices and different times. Client B would have 100\% of the first trade.

2) Client A sends an order to buy 10 shares of IBM and client B sends and order to buy 30 shares of IBM. The broker executes a buy order for 15 shares at 191.05\$ and another buy order for 25 shares at \$191.02. Both orders are allocated pro rata of the sizes to execute to the clients. Client A would get an execution report of 1/4th (10/10+30) of 15 shares at $191.05 and 1/4th of 25 shares at $191.02.

The important point to understand is that brokers have mechanisms to determine the share allocation and to compute the execution prices before margin for their clients. The point of the reporting standard is to make these mechanisms transparent for their clients.

What securities should this apply to?

Every security. There is no reason to single out equities and leave convertible bonds, treasury bonds, municipal bonds and the whole gamut of tradable things out of this.

What about OTC trades done by humans?

This is simply a case where the "matching engine" happens to be a human. It should already have a unique ID provided by the regulator. Of course, this particular "matching engine" would not have to be synchronized to better than 10μs. Time stamping to one minute accuracy for human trades is perfectly acceptable.

What are the expected impacts of our proposal?

What impact would it have on the competition between dark pools and exchanges?
It would narrow the regulation gap between exchanges and dark pools and help lit venues compete with dark pools more fairly.

Would this change market surveillance?

Today, only regulators have access to all the data necessary to police the markets. They have access to information identifying the parties to a trade and they also have quote information. This is powerful but hard to use. With our proposal, regulators could act upon request of parties who have done a first level of analysis and can identify particular trades on particular matching engines that are the cause of their supposed problem. Regulators would then be able to drill down and judge the claims on their merit. It would be very useful if regulators could actually document and publish the cases that they have investigated, both when they impose sanctions and when they decide that a particular pattern is not problematic.The creation of case law and the documentation of the reasoning of regulators would be very useful in building a consensus on what exactly is market manipulation and what is not.

How would it work to tame the bestiary of complex order types?

Many of those orders would become useless. The well-known ISO order type was created to circumvent a difficulty introduced by Reg NMS and mandatory routing. The routing algorithms can, in theory, produce infinite loops and be very costly. ISO orders were a natural response to this. Another famous order type, "hide not slide" which is used to gain priority in the case of locked markets, would disappear because the very concept of locked markets would disappear.

How would it help end users?

Armed with the Transaction Data, clients could take their trade IDs and matching engine IDs and have a company perform analysis to see if their broker was good at executing the orders. This would have the important effect of enabling rational choice and removing suspicion in the system. This would expose the potential problems and help the good brokers shine. Only access to the raw data can give real confidence in the system.

How would it help the analysis of best execution?

Customers who want to study their executions would request from their brokers an execution report.
Armed with this report, the customer would either compare trades in the same time period or send it to an external third party who would provide independent analysis. We actually expect that there would be many companies providing independent analysis because there would be a much lower barrier to entry.

What would this have changed for *Flash Boys*?
The plot of Flash Boys revolves around the discovery of order routing to several exchanges and the reaction this triggers. Two things would have been essentially different. The mechanism which is described in the book whereby an order hits a markets and triggers trades on other markets would have been simple to read in the data. Royal Bank of Canada would have understood much earlier that it was not executing its client trades optimally by providing signals in the markets. Even if RBC had not understood the phenomenon, the clients of RBC could have sent their execution data for analysis to many different analysis companies and would have been warned to direct their flow to a more sophisticated broker.

This analysis is made possible because of the precise time stamping. Without it, the causality of the trigger mechanisms would have been obscured.

If we remove quote protection, the system becomes simpler, but will the markets stay synchronized?

Today the price synchronization of markets is excellent for liquid securities in Europe where there is no quote protection. Arbitrageurs compete to synchronize prices across markets and also across asset classes. This is the most efficient mechanism for liquid securities. For illiquid securities we believe that the trade publication mechanism is also adequate. The TRACE program in the US is the right model and would benefit from a generalisation and the removal of all the exceptions.

What is the likely impact for illiquid securities?
The introduction of Trace is credited with reducing the cost of trading corporate bonds. It is generally agreed that more publicly available data will drive trading costs to a natural level imposed by the risk and cost of holding inventory. The same will probably hold for illiquid securities.

Is the proposal specific to the U.S. market structure?

There is nothing specific to the U.S. in this proposal. What is different is the current regulation and the future regulatory process. Europe does not suffer from the high costs imposed by Reg NMS. There is no order protection rule and each market functions independently. The prices for liquid securities are kept aligned by the natural arbitrage provided by participants. Expected regulatory changes are unlikely to introduce order protection. However, brokers may be required, in the new regulatory regime, to publish in a much more detailed way their best execution policies. However, Europe is currently discussing changes and could make the functioning of the markets much more onerous for little if any benefit. The same proposal would suit the same purpose in Europe.

Works Cited

Harris, L. (1996). *The Economics of Best Execution.* Retrieved from www-bcf.usc.edu: www-bcf.usc.edu/~lharris/papers.htm

(n.d.). Retrieved from http://www.sec.gov/rules/final/34-51808.pdf - page 12