

# Healthy Markets: An Open Data-Centric Platform For Bringing Transparency And Trust Back To Financial Markets

David Lauer, Rajarshi Das, Christopher Nagy, Dimitrios Ilias

KOR Group LLC

{dave,christopher.nagy,rajarshi,dimitris}@kortrading.com

## EXTENDED ABSTRACT (VISION STATEMENT)

### Keywords

Financial Markets, Market Structure, Open Data, Regulation, Big Data, Data Metric, Statistical Analysis, Machine Learning.

## 1. INTRODUCTION

The Financial Services Industry in the United States is in thrall of a Market Structure debate that rages from SEC to the Congress, from Twitter to the Wall Street Journal. Everybody affiliated with the industry has an opinion on the state of the market and how to fix it. Unfortunately, few can agree on how to do this. On one side, the supporters of the status quo argue that the three-way interactions between regulations, competition and technological advancements have benefited investors by providing a market structure that is easier to access and far more affordable than ever before. Others raise a contrarian voice noting that the interplay between the same three factors -- regulation, competition and technology -- have instead undermined investor confidence and trust by creating a fragmented market that is costly, complex to navigate, unstable and unfair. In the midst of such heated debate, there is an urgent need for open, data-driven facts and quantitative analysis of the recent evolution of U.S. equity trading. The aim of the Healthy Markets platform [1] is to fulfill this need.

While the process of buying and selling shares of stock may seem relatively straightforward to the average investor with a brokerage account, many aspects of the market structure governing the process of when, where and how information relating to billions of such orders are aggregated and trades are executed remains obscure not just to customers, but also to analysts, researchers and regulators. The primary goal of Healthy Markets is to increase transparency in the US stock market through data-centric approaches. In coalition with leading firms and researchers in financial services, Healthy Markets aims to provide an open platform for data gathering and data analysis not only for regulators and analysts, but also for the general public.

## 2. THE PUBLIC ASKS: IS MARKET FAIR?

Much of the debate about inefficiency, instability and uncertainty about market structure has been spurred by recent series of events and post hoc commentary that touch upon every aspects of trading. These include the ‘flash crash’ of May 6th 2010 when for a few minutes, \$1 trillion in market value vanished [2]; a study asserting that an astonishing one-fourth of public-company deals in the U.S. involve some kind of undetected insider trading [3]; widespread publicity of Michael Lewis’s book “Flash Boys” [4]

which argues how the pervasive practice of high-frequency trading can take advantage of the current market structure to jump on others in trading stocks; and the increasing proliferation of trading in opaque Dark Pools -- private trading of public stocks in off-exchange venues -- over public exchanges where stock prices and trading volumes are visible to all.

Especially in the backdrop of the Financial Meltdown in 2008, these events and associated commentaries raise questions among the general public about the U.S. stock where investors, irrespective of their financial might, can have an equal shot at a fair deal. A case in point is the U.S. federal trading rules mandating that brokers must provide “best execution,” -- finding the best stock prices for their clients who pay them to buy or sell shares. However, the complexity of the underlying market structure as well as the lack of public available data makes it difficult to prove (or disprove) conflicts of interest that under the guise of making subjective judgments about best execution, brokers are actually routing orders to exchanges that pay them the highest rebates.

## 3. CAN REGULATIONS HELP?

To better understand the range of critical issues facing U.S. equity trading and to provide better oversight, regulatory agencies such as the Securities and Exchange Commission (SEC) as well as congressional panels have scheduled several hearings and established committees. Unfortunately, the rules and metrics that the SEC designed to measure and track performance of brokers and exchanges are woefully outdated. While we are in the second decade of the new millennium, we still operate under the terms of the Securities Acts, including the Securities Act of 1933, the Securities Exchange Act of 1934 and the Amendments of 1975 that established the National Market System. Few things about our markets function as they did then, or even as they did fifteen years ago. Moreover, while the rules have been changed many times over the last fifteen years, most importantly with limit order display, decimalization, Reg ATS and Reg NMS, in an attempt to maintain pace or force changes in behavior, our capacity for understanding the results of these rule changes and technology advances remains very primitive. Out-of-date regulatory tools and inconsistent data availability and access for academics have prevented us from attaining a clear understanding of the impact of rule changes and technology. Regulators are too focused on specific events, on short-term fixes and on a narrow view of the industry. Regulators and exchanges tend to treat issues in isolation, whether the issue is order type complexity or SIP infrastructure failures, but all of these things are inextricably linked. This calls for a strategy of addressing system-wide flaws, misaligned incentives and improper transparency/disclosure,

instead of reacting to each technology failure with an endless sequence of fixes.

Market surveillance performed today by regulatory agencies is not only hampered by technological shortcomings (e.g., no timestamp synchronization across exchanges, low temporal resolution), but also by the lack of advanced analytics and machine learning tools that are needed for deep, multi-faceted analysis of high-volume, high-dimensional data produced by the markets. Such tools could also be used to monitor trading across different asset classes, but unfortunately there is complete lack of cross-asset class data. Moreover, conflicts-of-interest inherent across for-profit, publicly-traded SRO model leads to poor enforcement across biggest/best customers and lead to lack of coordination and cooperation across the industry.

## **4. THE HEALTHY MARKET APPROACH**

Healthy Market's principles focus on leveraging open, data-driven facts and quantitative analytics.

### **4.1 Transparency**

Healthy Markets primary goal is to increase transparency in the US stock market. Currently, 40% of all trading occurs in internalization systems and dark pools or Alternative Trading Systems that never interact with a public stock exchange. There are over 50 dark pools and regulatory filings for these Alternative Trading Systems are not publicly disclosed. Alternative Trading Systems need to be held to the same transparency standards as any other lit venue. Brokers route customer's orders but provide little transparency into where those orders are routed, how well they are executed, how much commission the brokers made on those order.

Transparency means that data is clearly, unambiguously received in a fair way. That means we need to reform our system of data feeds in the market, starting with the public, consolidated feed. There are several proposals on how to do this, including having multiple feeds that compete with one another, or removing control from the collection of exchanges that it now rests with. No matter what, the end result must be that the consolidated feed is built using the same level of technology as any other data feed, and that participants always receive data on the consolidated feed before any other data can be obtained.

Finally, all market centers (Exchanges, ECNs, ATS's) must synchronize timestamps across all of their servers, at microsecond-level precision. This technology is readily available, is not burdensome and would dramatically simplify research and surveillance of markets.

### **4.2 Metric**

On July 28, 2000, the SEC proposed the rules now known as SEC Rules 605 and 606 in response to increasing competition and resulting fragmentation in the market [5]. These rules were put in place in May of 2001, and functioned well to ensure Best Execution and measure outcomes. Brokers increasingly sought Best Execution and regularly published various statistics regarding execution quality.

Over time and in particular with the adoption of Regulation NMS, the rules became increasingly outdated and their usefulness, while still relevant, has diminished. In part, the rules have eroded due to the increasing complexity of order-types, speed and routing practices in today's marketplace. Rules 605 and 606 have not kept pace with these changes. The Commission even went so far to say "improved visibility could shift order-flow to those market centers that consistently generate the better prices for investors and the

Commission will assess the impact of the rules to determine whether additional action is necessary to further the Exchange Act's objectives for a National Markets System." It's time to re-assess these metrics and bring them into the electronic trading era.

In addition, before further changes in market structure are evaluated, these rules must be updated so that efficacy of changes can be objectively compared and evaluated.

### **4.3 Open Data**

Wall Street firms use open source software in nearly every part of their business. For example, many high-performance servers run on the Linux operating system and use open source analytic software packages in R or Python. The open source movement has succeeded in incredible fashion and it is time to adapt some of its principles. One of these principles is that "data yearns to be free." Healthy Markets will push to have the SEC provide open access to their MIDAS system for analyzing market data. This will spur the academic and open source communities to build incredible studies and learn far more about US markets than any team at the SEC possibly could. It will leverage the network and community effects of the open source/open data movement.

One effect of having the SEC provide access for academic research will be the burgeoning flood of independent research in an environment in which independent research has become increasingly difficult. Today, it is very difficult to gain access to the vast amounts of data needed to competently analyze market dynamics. Moreover, once academic researchers do gain access, it can be extremely difficult and costly to process it. This generally means that academics are beholden to the providers of that data, and all of the biases this implies. If researchers publish findings that their sponsor disagrees with, as most recently happened with a study conducted at Notre Dame, they risk losing access to this data. This places researchers between a rock and a hard place as they work to produce objective studies of the US stock market. We are all suffering as a result. Having the SEC serve as the intermediary will produce immense dividends for the study of our markets.

Healthy Markets will advocate in every way possible for free and open access to data. This will also include Financial Industry Regulatory Authority's release of dark pool reporting data. This data should be made freely available and in a form that can be easily processed and analyzed. It is inexplicable that a regulatory organization should look to profit from the sale of this data.

### **4.4 Displayed Liquidity**

It is often lost in the debate that the stock market is supposed to serve two purposes: capital formation (i.e. raising money via IPO) and price discovery (determining what a company is worth). Price discovery is a complicated process, but one primary component is quotes that are displayed on an exchange or ECN. Since the implementation of Regulation NMS in 2007, the volume of internalized and dark pool trading has grown every year and now stands at 40%. In addition, many stocks trade over 50% of their volume off of the standard set of exchanges. Trades can only take place off of an exchange because there are quotes being displayed, and therefore dark trading can be seen as "free riding" the public quotation. As off-exchange volume increases, it acts as a disincentive for market makers to post liquidity. Originally dark pools were built to facilitate block trades by institutional investors, but currently the average trade size in dark pools is nearly the same as on lit exchanges. In addition, all marketable retail orders are being sent to internalizers based on backroom

deals, instead of being sent to public markets where market makers can compete for those orders.

Many markets around the world have adopted some form of a trade-at rule, including Canada and Australia, and a form even exists in the US Options Market. The philosophy behind this rule is simple: publicly displayed quotes should have precedence over those in hidden systems. Moreover, if brokers are going to execute a small trade off of an exchange then they must provide a significant amount of price improvement to offset the damage to the price discovery process.

Of the critical changes advocated by Healthy Markets, this is perhaps the most vital. Indeed, in Healthy Markets has called for a pilot test to see what the impact of a trade-at rule would be. Reducing the access fee cap should be part of this test as well, to help reduce investor costs across the market and help eliminate the skewed incentives that the maker-taker exchange model has created.

5. PROTOTYPE PLATFORM

As an initial step in building a prototype for an open data-centric platform, Healthy Markets is collating data at frequent intervals from disparate and inconsistent sources into publicly available data warehouses in a fixed, predefined data formats. In addition, the prototype provides web-based, interactive visualization tools (e.g., using D3) to explore large datasets. Figure 1 shows an example of visual exploration of dark pool data from FINRA based on user's choice of report date and type of market. Using statistical software and machine learning techniques; the prototype also offers the opportunity gain insights into the relationships between different datasets. Figure 2 shows an example of overlaying user-selected Rule 605 order execution data on top of Rule 606 order flow visualization in the form of a Sankey diagram.

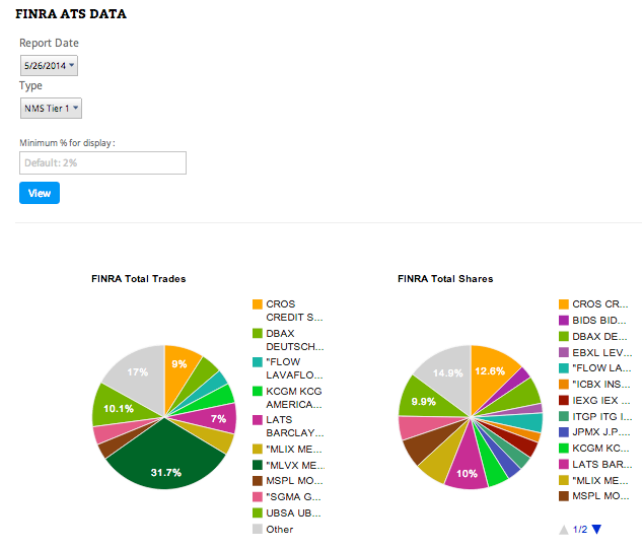


Figure 1. Interactive visualization of FINRA ATS data [5].

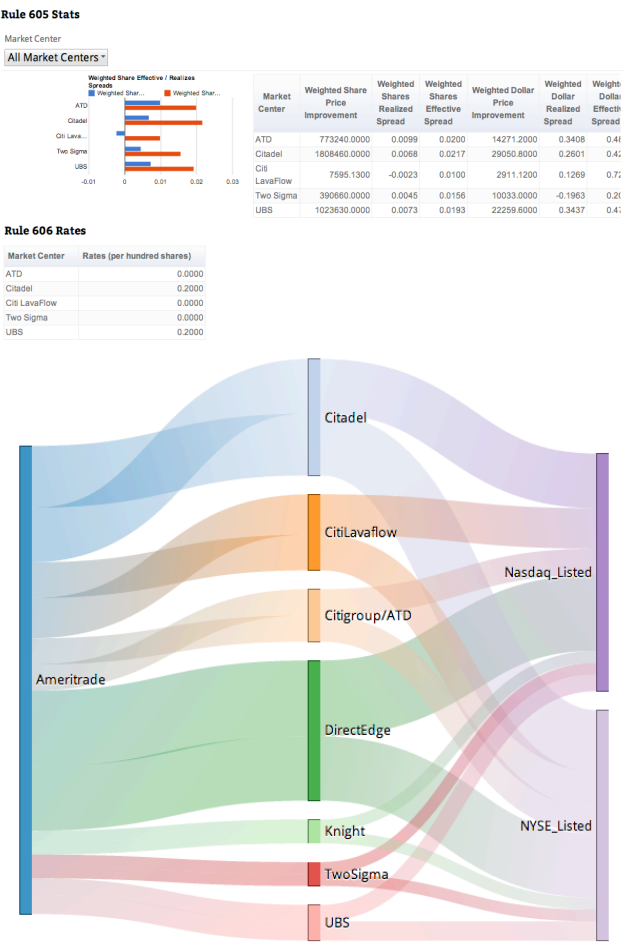


Figure 2. Interactive visualization of relationships between Rule 606 routing data and Rule 605 order execution data.

6. REFERENCES

[1] Lauer, D. and Nagy, C. <http://www.healthymarkets.org> (2014).

[2] Easley, D., Lopez de Prado, M., and O'hara, M. 1997. *THE MICROSTRUCTURE OF THE 'FLASH CRASH' Flow toxicity, liquidity crashes and the Probability of Informed Trading*. Journal of Portfolio Management 37.2 (2011): 118-128.

[3] Augustin, P., Brenner, M., B., and Subrahmanyam, M. *Informed Options Trading prior to M&A Announcements: Insider Trading?* Available at SSRN 2441606 (2014).

[4] Lewis, M. *Flash Boys: A Wall Street Revolt*. W.W. Norton & Company, New York (2014).

[5] Securities and Exchange Commission. *Disclosure of Order Execution and Routing Practices*. <http://www.sec.gov/rules/final/34-43590.htm>

[6] Financial Industry Regulatory Authority. *FINRA ATS Transparency Data*. <https://ats.finra.org/>