OFFICE OF FINANCIAL RESEARCH
STAFF DISCUSSION PAPER

No. 2014-02 | August 27, 2014

The Role of Visual Analysis in the Regulation of Electronic Order Book Markets

Mark E. Paddrik
Office of Financial Research
Mark.Paddrik@treasury.gov

Richard Haynes
U.S. Department of the Treasury
Richard.Haynes@treasury.gov

Andrew E. Todd
University of Virginia
aet6cd@virginia.edu

Peter A. Beling
University of Virginia
pb3a@virginia.edu

William T. Scherer
University of Virginia
wts@virginia.edu

The Role of Visual Analysis in the Regulation of Electronic Order Book Markets

Mark E. Paddrik[a], Richard Haynes[b], Andrew E. Todd[c], Peter A. Beling[c], and William T. Scherer[c]

***Abstract*— Electronic markets and automated trading have resulted in a drastic increase in the quantity and complexity of regulatory data. Reconstructing the limit order book and analyzing order flow is an emerging challenge for financial regulators. New order types, intra-market behavior and other exchange functionality further complicate the task of understanding market behavior at multiple levels. Data visualizations have proven to be a fundamental tool for building intuition and enabling exploratory data analysis in many fields. In this paper, we propose the incorporation of visualizations in the workflow of multiple financial regulatory roles, including market surveillance, enforcement, and supporting academic research.**

***Index Terms*—Visualization, Visual Databases, Financial Markets, Law Enforcement**

I. Introduction

REGULATORY analysis of behavior in financial markets has traditionally focused on consummated actions within a market environment such as completed trades and the resulting inventory. For example, regulators need to identify self-trades and monitor accounts that are taking outsized positions in specific financial instruments. However, the prevalence of automated trading and related market anomalies, such as those informally described as "Flash Crashes",[1] highlight the need for additional attention to the underlying details of order flow and the evolving order book. This need is especially acute for those entities with surveillance or oversight authority. As a result, effective regulatory surveillance and enforcement now require tools to closely examine the order book and the detailed interaction of participants with exchanges.

Recent improvements in the regulatory audit trails available from financial markets now allow a far more complete and accurate reconstruction of the order book. Regulators are now able to create a more detailed market picture consisting not only of market depth, but also detailed information on order status, history and priority. Where this data may be less complete, efforts are also underway by some regulators, such as the Securities and Exchange Commission,[2] to further standardize and consolidate audit trails.

Processing and analyzing order flow data, however, presents major challenges. The variety of order types and intra-market capabilities offered by exchanges complicate the task of order book reconstruction. In addition, careful verification and validation are necessary to ensure the accuracy of any aggregations of the raw order data, such as for the reconstruction of the limit order book or the audit trail of a specific account. The complexity and volume of regulatory data, and the general need for rapid synthesis related to order activity, calls for strong and flexible tools. Such tools should facilitate quick analysis of changes in participant and market behavior and subsequent dissemination of this information to relevant parties (including the exchange, the clearing firm, or the participating firm itself).

The dissemination and analysis of high dimensional or otherwise complex data is often accomplished or enhanced through visualizations. Visualizations can provide high level, compact graphical summaries of large or complex sets of data such as a snapshot of a stock market portfolio. Examples of these high-level summaries include popular financial websites, like the "Map of the Market", which provide users with a bird's eye view of relative activity of hundreds of rapidly trading stocks [1].

Visualizations in the context of financial markets are traditionally associated with forecasting and market analysis for the purpose of trading, also known as technical analysis [2]. Sophisticated visualization systems are created by market participants to support trading decisions, designed to depict derived market variables and dynamics (e.g. candlesticks, Bollinger Bands, Elliott Waves, Fibonacci Levels). In this paper, however, we focus on the goals of a specific, and often under-represented, market observer, the market regulator.[3]

Regulators are tasked with ensuring fair, orderly and efficient markets. Fundamental to this mission is the ability to process and analyze market data generated from exchanges both nationally and internationally. Advances in technology at

[a] Office of Financial Research, Mark.Paddrik@treasury.gov
[b] U.S. Treasury Department, Richard.Haynes@treasury.gov
[c] University of Virginia, {aet6cd}{pb3a}{wts}@virginia.edu

[1] A phrase used to describe unusually large price falls occurring over very short periods of time.

[2] For more detail on the SEC's audit trail efforts see: http://*www.sec.gov/rules/final/2012/34-67457.pdf*

[3] By "market regulator" we implicitly encompass a number of different parties, the most obvious being that of public government market regulators like the SEC, CFTC or FSA. In addition, however, this could include SRO's such as FINRA or NFA within the US, and exchanges tasked with market oversight.

market venues and the associated growth in regulatory data have reinforced the need to design methods of viewing and communicating data in a way that are both information rich (e.g. from which complex market connections may be inferred) and easily digestible (e.g. understandable within a required time frame or without overly burdensome prior knowledge).

In this paper, we propose the incorporation of visualizations into the workflow of regulators engaged in a variety of activities. Visualizations are valuable tools for supporting a range of activities within the regulatory sphere including: real-time and day after market monitoring, regulatory enforcement review and, more abstractly, academic research supporting regulatory policy. Each of these tasks requires quick and effective analysis of extremely large quantities of data (often on the order of millions of trades); in those tasks that are focused on imminent risks to the market system, the resulting analysis might also be required within constrained time periods. As market technologies continue to advance, the resulting increase in order flow and transaction volume will lead to further growth in regulatory data, further heightening the demand for efficiency and innovation in the creation of supporting visualizations.

To make the case that visualizations are indeed valuable to regulators, and to suggest some potentially powerful visualization techniques, this paper discusses the scope and structure of financial data and how regulatory objectives have been affected by them in sections 2 and 3. Section 4 discusses the concepts underlying visualizations and their application to data. Section 5 develops criteria and a set of visualization techniques that can support both exploratory data analysis and core regulatory tasks. Section 6 goes through several example cases in which visualizations have been developed to accomplish specific tasks, and finally section 7 provides a summary of conclusions.

II. Regulatory Objectives

With the technological advances in the past decade leading to an explosion of market data, financial exchange operators in particular have needed to adapt to a new environment in which the storage, management and transmission of data are at the very forefront of their activities. Resulting advances have enabled exchanges to disseminate market information over high-speed networks to automated trading systems that then respond with follow-up data within milliseconds. Transforming these vast quantities of data into understandable and actionable information is the focus of many different groups in both the private and public sectors. Financial regulators, in particular, face the challenge of managing and interpreting the fullest scope of available data for oversight, enforcement, and research purposes at a scale much greater than that which they are traditionally accustomed.

Today's financial regulators face the challenge of managing and interpreting data on a scale far greater than even the very recent past in order to meet their objectives including market oversight, enforcement, and research relevant to regulatory policy.

Regulators tasked with market oversight must verify the integrity of market conditions and, conversely, identify those periods where integrity has been compromised. Volatile events, such as instances similar to what is now known as the "Flash Crash," are extreme examples of timeframes where market integrity and orderly execution of market expectations failed at a dramatic level. In cases such as these, it is the role of market regulators to help to prevent similar occurrences in the future, manage the effects during those periods when they occur, and, finally, determine the proximate and ultimate causes of the instability. Each of these, to a greater or lesser degree, requires the construction of accurate, reliable, and replicable information regarding market conditions around these periods. Although in many, or all, cases, the results of volatile events can be detected in multiple market feeds, determining the ultimate cause of many market failures can require data only included within sets of private feeds.

As described, initial steps in developing these causal links may incorporate the volume and the price of a given order or set of orders. These two dimensions can often be usefully incorporated within static visualizations such as order book heat-maps and dynamic representations like order book updating (described further below). Through these techniques, even using this limited subset of information can identify quite complex, "emergent" market behavior.

One demonstration of this utility can be seen through an analysis of the May 6th, 2010 Flash Crash [3], where depth on the bid side of the market was far overwhelmed by depth requested by the ask side. Charts providing information regarding market depth can isolate the initiation of the occurrence, and potential mitigates of price spikes. During periods of this nature, where the imbalance between buying and selling interest over abbreviated timeframes is great, market movements can be especially abrupt. The rapid disappearance or elimination of one side of the order book is a necessary condition for high intraday volatility. In the case of the Flash Crash, this was initiated by a request for liquidity on the bid side, but it could come as a result of unexpected passive liquidity loss (e.g. a high volume of order cancellation), unexpected aggressive liquidity loss (e.g. high execution levels), or expected aggressive conditional liquidity (e.g. high levels of stop orders at a specific price point).

The above discussion makes clear that unexpected and anomalous, though market accepted, behavior is of importance to market oversight. In cases such as this, regulators often have to acknowledge the potential of an occurrence, while simultaneously limiting the impact during an event. However, in other cases, similar market reaction may be initiated for the sole reason of inducing a heightened market impact; often market behavior of this type is considered illegal either by the market venue and/or its regulatory body. In these cases, retrospective investigation and economic analysis of the event would be passed to a market enforcement team, or related regulatory body.

In the case of illegal behavior and subsequent enforcement, it is often important to study the behavior of an individual entity, or set of entities that have chosen to collude. It is this behavior which requires enforcement investigation and then clear communication to an adjudicatory body like a court of law. To do this successfully, two categories of information must be clarified: that the identified individuals acted in an exceptional manner and that those exceptional actions should be considered illegal activity. Behavior becomes exceptional if it deviates from traditional market accepted practices; it is useful, then, to be able to quickly assess the behavior of a market participant compared with that of other market participants and of the market as a whole. This implies the need for isolating the orders for one account or firm and contrasting them with the market. In addition, it emphasizes the use of a dynamic portrait of the orders as they are entered and cancelled, providing a window into the possible intentions and goals, in real time, of a given market actor.

Finally, regulators often need to take into consideration what might be characterized as the "long view", the market effects of policies and procedures created by the regulatory body or by market structure changes; this can often be clarified through formal academic research. Recent changes of this type include such things as the introduction of dark pools [4], the rise of high-frequency trading, market-making programs, and the increased effort to move over-the-counter financial instruments to exchanges and clearinghouses. Each of these may have intended and unintended consequences for market liquidity, volatility or participation by various market groups. Investigation of these effects may require the use of years, or even decades, of market data, necessitating sets of visualizations, which summarize millions of data points within one field of view. The creation of such visualizations would also have to be responsive to highly sophisticated analytical techniques designed by experts and academic researchers.

III. Scope of Financial Data

A. Structure of Financial Data

The rapid adoption of technology by exchanges has quickly driven the vast majority of financial trading activity off the human face-to-face physical trading floors and into computer-based electronic order book systems.

As a result, the ability of exchanges to communicate quickly, simply and in a secure manner with market participants and regulators that are located in disparate geographical regions is a fundamental requirement. Given the sensitivity of the data, these communication channels are often divided into both public and private data feeds. Order flow between individual participants and the exchange are transmitted through single information channels, whereas public feeds such as updates to the order book are sent equally to all connected market participants. The following paragraphs provide a more detailed background to this diverse set of market data systems.

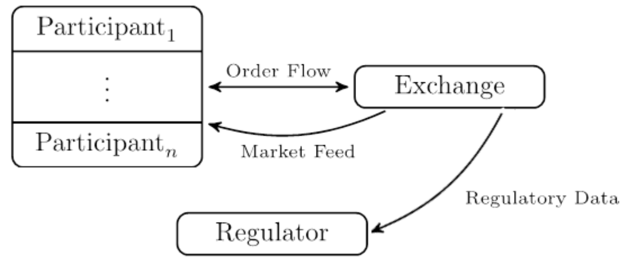

Fig. 1. Data flow between Market Stakeholders

Order flow data is the aggregation of bidirectional private communications between individual participants and a financial exchange. Order flow data consists of requests for new orders, the modification and cancellation of extant orders as well as confirmation notifications from the exchange when an order is successfully created, modified, canceled or executed.[4] These messages make it possible for the exchange to fairly execute trades using a matching engine, an order matching system which follows a set of publicly known prioritization rules with which it identifies appropriate trade counterparties.

Typically, exchanges use a price-time based system for matching and a limit order book as their mechanism for structuring the rule and simplifying the queuing of orders to be executed at set price points.[5] The limit order book contains both visible and invisible (e.g. "iceberg") interest in purchasing or selling a given financial instrument at a given price. Prior to a standing order's modification or cancellation within the order book, other market participants are able to execute against that order. As trade executions and order cancellations occur, the volume of indicated interest at a given price often changes, with this dynamic order behavior through the trading day representing the "market" as a unified system.

In order to efficiently transmit this high volume information set, roughly 80-100 million messages per day [5], the exchange is aided by a messaging format designed specifically for financial data, the Financial Information Exchange or FIX. The FIX Protocol format is the typical language for communication between exchanges and market participants, designed and updated over a number of years and containing built in data fields helpful for concise order flow.

Distinct from the private feeds which make up the individual parts of order flow data, there exist public aggregations of these private communications. These order aggregations are known as market feeds. These feeds traditionally include the price, trade and market depth data that result from net order flow and trade executions. Table 1 provides a sample of data which might be provided to all market participants in such a data feed (here, the top ten best bid and ask prices and quantities;[6] in the figure, the raw market feed is displayed in a more easily digestible, and

[4] Errors and general status notifications from the matching engine are also part of order flow data.

[5] One notable exception is the Eurodollar futures contract, which follows a pro-rata matching system.

[6] The bid is the price that you can sell an asset for. The ask is the price that you can buy an asset for.

commonly used, format). With both private and public feeds, market participants not only need to manage their individual communication with the exchange, but also to determine how those actions interact with the actions of others within the trading system.

TABLE I
PUBLIC MARKET FEED AS SEEN BY MARKET PARTICIPANTS

Bid Quantity	Bid Price	Ask Price	Ask Quantity
50	$10.56	$10.57	140
36	$10.55	$10.58	67
142	$10.54	$10.59	89
32	$10.53	$10.60	52
49	$10.52	$10.61	103
100	$10.51	$10.62	40
110	$10.50	$10.63	205
65	$10.49	$10.64	178
258	$10.48	$10.65	245
178	$10.47	$10.66	90
Last Trade:	**Price**	**Quantity**	**Time**
	$10.57	10	12:56:24.047

Finally, in order for regulators to verify that rules and processes are being followed by both market participants and exchanges, regulatory data sets have been created for oversight purposes; these are often an augmented version of order flow data and exchange trade matching logic. These data sets come in a variety of different types, but informally one can consider the underlying, primary regulatory data set as the aggregation of all the individual private data feeds. With such a data set, the regulator should be able to pinpoint and summarize the activity of an individual market actor, or the combined activity of a related group or other market subset. In other words, regulatory data is both broad and granular, with added unique identifiers that allow regulators to associate to each action its source entity.

As a result, the data set provided to regulators is extremely large and not easily parsed, stored and managed. Consider that at least 10 million contracts are traded per day within the CME Group exchanges alone, executions which result from an order of magnitude larger incidence of overall order book activity [3]. Because of this size, regulators are often not able to fully analyze the full breadth of activity in the market environment; storage constraints, the lack of analytical technology and an ever-changing marketplace increase the difficulty in extracting a significant and useful understanding of activities behind specific market movements.

B. Structure of Financial Regulatory data

In examining data, it is fundamental to first understanding the fields that make up structure of the data. These building blocks give the data the communicative properties necessary to transmit the information and basic functions throughout the financial market system. Simple features included within this set of fields might be the time a new order is placed, the type of order, and the identity of the trader who placed it. Starting from these basics, one can design analytical systems incorporating more complicated structures, such as where a given order is relative to the entire market supply/demand curve, what actions it follows or precedes in the order book, how much it adds (or subtracts) to that price level in the queue, and how other actors respond to that order entry.

The order flow, a data type often used as a building block, traditionally arrives at a matching engine or market participant with little advance processing and minimal structure, allowing for quick, automated processing by the exchange. Examples of orders include new orders, modification and cancellation messages sent by participants, and confirmations and executions sent by exchanges. Using this data's inherently simple structure allows a regulator the ability to perform simple aggregation analysis like: identifying the most active users and markets within a given day or specified time period, or isolating accounts with aberrant modification or cancellation levels. That the data has a universal structure also provides a regulator enough contexts to act as a building block for higher dimensional aggregations, for example, reconstructing an order book or determining the relative risk metrics associated with given traders.

For a regulator, the first useful layer of structure to be added to this atomic data is often temporal, where an outside reference point is imposed on the data stream, most commonly wall-clock time. By doing this, it is possible to weight orders by trade time, by volume, or relative to the timing of another series of events. By incorporating this metric, a number of visualizations can be then be overlaid on the data stream. A common example of this for financial data is a price chart that depicts price movements throughout a day, month, or year.

A second dimension to consider incorporates the matching logic of the market mechanism. This creates what we will denote as spatially oriented data that associates coordinates to orders, based on the price-time priority within the order book. The spatial dimension provides a user the ability to track a given order as an element of the order book and determine its relevance to the market as a whole. Regulators could use this additional structure to analyze such questions as:

- What is an order's relative size compared to other orders at similar price points?
- What is the likelihood (over time) that this order might be executed or cause a trade?
- Is this order visible in the public market feed data and how might it impact other firms/accounts' decisions?

Combining both these data structures in a spatial-temporal system, an analyst is able to navigate the order book and recreate a prior sequence of events, similar to that seen by a theoretical participant that experiences no market feed latency. This ability is vital to accurately reconstructing market activity, and associating a causal chain of activity to events.

IV. VISUAL ANALYSIS CONCEPTS

A. Features of Visual Communication

Describing and analyzing large data sets is not an easy task, and it can often be difficult to effectively communicate large amounts of information through words and summary statistics. This task is made increasingly more difficult when the data being examined is part of a much larger system, such as the case of order flow data in financial markets, as using words and statistics to describe the system, by necessity, must incorporate linear, often, one-dimensional descriptions. Contrasting this, individual activity and the evolution of the system happen continuously and are affected by not just one dimension, but in multiple dimensions simultaneously [6].

As an example, we can consider a simple trade execution on a market platform. Here a trade denotes an account actively choosing to execute against a standing order sitting at the top of the order book (originally entered by a second account). This execution may have exhausted the liquidity at a given price point, changing the prevailing bid-ask levels in the public market feed. The execution has also moved new orders to the top of price-time priority. If price did in fact change during execution, the order may have also have triggered additional market liquidity contingent on the prevailing price point. Finally, with this execution, the contract holdings of the two relevant accounts have changed. In order to encapsulate these changes, all of these pieces of information set need to be updated simultaneously at the time of the trade.

In order to depict all these types of information, then, it is necessary to use a language that shares some of the same properties as the phenomena under observation [7]. Visualizations are one, especially powerful, method which provide that multi-dimensional summary, and allow an analyst to describe financial market evolution. Visualizations are additionally useful because they provide a balanced portrait of the dynamic and static components of the market system.

Good visualizations take advantage of graphical language to describe the syntactic and semantic properties of data, by capturing the expressive and effective features of the data. A visualization should embed important information while effectively communicating with, and capture the attention of, the human visual system [8].

B. *Cognitive Processes served by Visuals Analysis*

Visualizations often have acted as means of communicating results, by translating multidimensional data into a form that is visually accessible to users. This ease of communication comes from the ability of visuals to both help externalize the memory associated with the data, and to more closely represent a user's mental model of the data [9]. This efficiency can free the user's memory to support further cognitive operations or tasks. An example would be a city map for first time visitors, since it stores unfamiliar information in an easy to retrieve format. Thereby allowing visitors to spent time seeing attractions rather than learning all street names/intersections or even worst getting lost.

This ease of cognitive load on a user leaves room for insights which iteratively explore the data. In order for visualizations to do this, their design and use must be appropriate to the task at hand [10,11]. Categorizing tasks allow us to classify visualizations into three broad categories:

- *Information Retrieval* operations for exploring the data space through overview, browsing, navigation, zooming, observing derived quantities such as data ranges, distributions, errors, certainty and sensitivity of those value. For spatial and temporal data sets it involves inspecting features via viewing animated or sequential representations.
- *Information Analysis* serves as a method for gaining further insight, by fostering the constructive evaluation, correction and rapid improvement of model and hypothesis that provide for enhanced decision-making. Figure 2, shows how the analysis process makes use of visualization and model construction to deliberate and build upon current knowledge. This includes a wide range of analytical tasks, such as identifying extremes, anomalies and clusters, exploring data to make comparisons and identifying inherent correlations, so as to evaluate the truth of initial hypotheses.

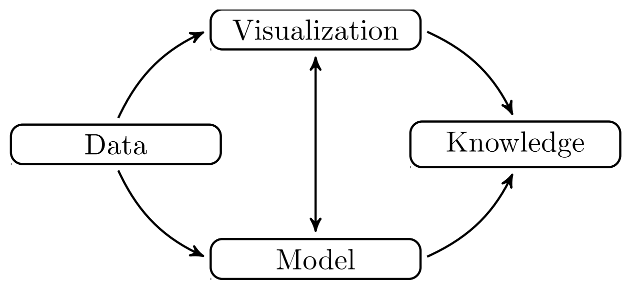

Fig. 2. Model of Data Visualization Analysis [12]

- *Information Dissemination* falls under the system where visuals are meant to act as aids in presenting information to others, so as to allow for easy data comprehension. In this case, the visualization should summarize, annotate, and illustrate analytics which support or reject a hypothesis.

V. Applying Visualizations to Financial Regulation

The full transaction and order histories that are available from electronic markets contain information about the intentions of the participants and market responses. Because of this, the information gained from data analysis is a promising approach to effectively combat unlawful activities and understand market anomalies. However, since the patterns used by abusive firms change whenever they are made aware of a detection mechanism, new patterns have to be reviewed and identified through adaptive means [13]. Therefore, while effective designs of summary visualizations must provide storylines for market events, their ability to stay flexible enough to adapt to changing market behavior or structure makes them useful in ever changing market environments.

A. *Visualizations for Market Regulatory Data*

In the following section, we cover five visualization techniques that can be used individually or in unison as part of integrated process for satisfying regulatory needs. Each one is meant to capture different aspects of structure, hierarchy, and information that exist in regulatory data (in this case using market-simulated data [14].

1) Market Tree-Maps

Tree-maps are a more recent design for organizing complex tree structured data and, in contrast to the traditional treeing approach, use a two-dimensional space-filling approach where rectangles represent individuals. The area of a given rectangle would be proportional to a chosen attribute [15]. Further attributes can be represented by color coding (or shading) of regions; this planar representation provides users with a chart that clearly indicates relative comparisons between individuals across multiple dimensions.

A traditional approach has been to select and display reference points through a tree structure, depicted graphically with a root node at the top of the page and children nodes below the parent node (selection criteria) with lines connecting them. This manner of breaking down a large set of unique individuals is perfect for creating branches, or groups of individuals, but it does require a prior understanding of the ordering variables and, in the case of numerical data, appropriate group cut-offs. Non-financial uses of tree structures such as family trees, evolutionary trees, or organization charts, have found that beyond a certain point a large wall is necessary to capture the entire picture [16]. Even in these cases, only the structural relationship is shown; additional information, such as the size or importance of each node, was often ignored or included in a summary external to the visualization.

Fig. 3. Tree-map of Market Participant contrition to Order book Depth and Executions

Figure 3, above, is an example of a tree map which depicts the contribution accounts make to the overall liquidity for an hour on a fictional market. The size of a "node" (individual rectangles in the diagram) denotes the average relative amount of contracts that they offer to buy and sell in this market. The level of liquidity provision is seen as an important variable in measuring the market effects of a given participant [17].

A participant's importance to liquidity can also be measured using the participant's number of trades, denoted by the color scale at the top of the figure. This representation gives us a better picture as to who (denoted by the number in the boxes) in the market is truly important to liquidity (and so may cause important changes to market quality if their behavior changes). While doing this, the representation also provides a level of flexibility, since it relies on a relative scale and so can quickly incorporate comparisons across market participants. It is, however, limited by the number of dimensions it can represent and its ability to organize data points into clusters.

2) Classification Scatterplot

Regulators are tasked with the objective of trying to create policy to help better markets, which in part requires a basic understanding of the stakeholders (market participants) within the markets they regulate. Market participants, however, are wary about directly divulging too much information about their trading strategies to regulators. This requires regulators to gain some understanding about their market of interest using data that is available to them in order to develop an understanding of the implications that regulations might have on the present actions and behaviors of market participants. This requires them to classify market participants to understand the stakeholder base they are influencing.

Whenever organization and classification is the objective of a chart, the most common and effective graphic is the scatterplot. It is a tool used by data analysts before trying other forms of analysis, and the insights gained may stimulate the production of more complicated variations or may guide the choice of a model [18]. Linear or nonlinear relations are easy to discern and the human eye is robust to the effects of outliers and other aberrations in the data [19].

The scatterplot is the predominate graphing tool used in the physical, biological, and social sciences, making up an estimated that 75% of the graphs used in science [20]. It has been useful in financial markets in helping to classify traders in the market according to objectives and trading behavior. A notable example includes the May 6th Flash Crash investigation done by the CFTC and SEC to understand who was responsible for the event [21]. In figure 4 below, trading volume and end of day trading positions are used to classify traders into five groups based on clustering. This process allowed investigators to determine what cut of data points they might use in investigating traders of the high frequency traders group.

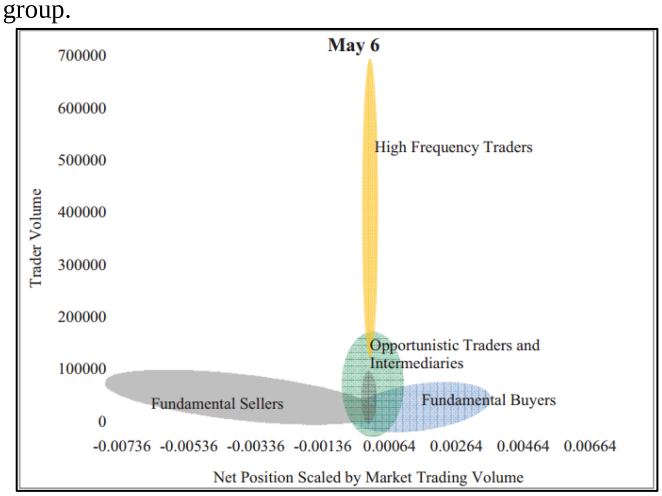

Fig. 4 Classification of Traders in E-Mini Futures Market [21]

However, there is a significant amount of overlay between groups making it difficult to split them apart completely. Taking this same graphing technique and adding a third dimension could help further divide groups. For example adding the amount of liquidity a trader on average offers to the market through resting limit orders could further separate traders by another observable behavior. In figure 5, you can see how the layered clustering of Opportunistic and Intermediaries cluster no longer are mixed as they were in the two dimensional example.

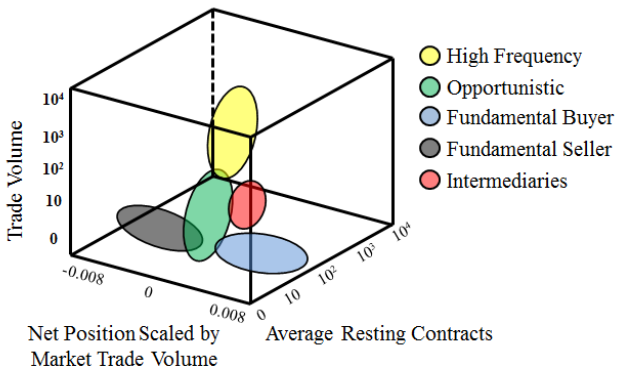

Fig. 5. Example of Possible Classification of Traders in E-Mini Futures Market

3) Order Book Heat Map

A heat map is constructed using rectangular tiling of a data matrix; this facilitates inspection of three dimensional data using a row and columns structure to show transition of a third attribute as it changes in respect to the other two. This allows large data matrices (several thousand rows/columns) can be displayed effectively on a high-resolution color images [22].

Heat maps, such as figure 6, are constructed to depict the state of a financial markets over many years as a method of monitoring the prices of sectors over time [23]. For this application, using the rectangular tiling of a data matrix facilitates the expression of spatial-temporal data using the price of a group of assets and time to structurally describe the transition of a set of market, the colors indicating a growing (green), stagnant (yellow), or declining (red) market. This allows for the heat map to apply the spatial relationship to price, a typically not well represented set of elements, in a manner that can organize information, facilitate memory and empower spatial inference [24,25].

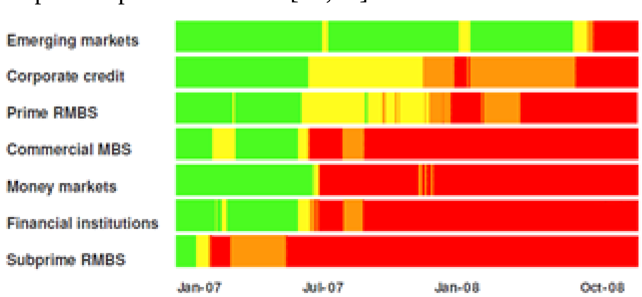

Fig 6. Example of Sector Heat Map (IMF, 2008) [26]

The order book heat map, seen in figure 7 below, visually allows for the examination of liquidity expansion and contraction in the market. The colors are applied at specific price levels and colored by resting limit order contracts depth (for the 100 ms interval), to show the depth of the order book. The depth is color coding to represent buy (violet to dark blue) and sell (yellow to red) orders based on the number of resting contracts.

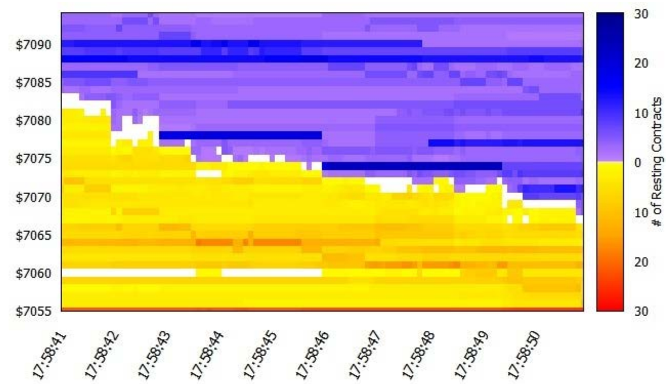

Fig. 7. Example of Heat Map of Order Book Depth

From figure 7 we can, extract simple information about the direction of price over this time. Inferences however can also be made about the supply and demand of this market by examining market depth over time. A large selling depth relative to buying depth can be observed above current best bid/ask spread of the market which drives price downward over time.

4) Order Book Animations

In communicating the events of a market, it can be difficult to fully capture the complexity of ever changing orders in the order book. The number of structural dimensions that graphics can offer can make expressing complexity to permit extraction of useful information difficult. Animation is the rapid display of a sequence of images to create an illusion of movement and it can be used to better express such complex processes such as the behavior of market or the impact of single individuals by portraying changes over time.

The effectiveness of animation is still a matter of debate. Animation does overcome some of the difficulties that research on static graphics has shown concerning their design and limitations for conveying complex systems. Other research however has shown, though, that the effectiveness of animated over static graphics does not exist in testing or learning [27].

Animations have been successful in conveying extra information and producing interactivity, rather than the animation of the information per se [28]. The ability for user to be interactive with the animations, combining technology together with a user interface, have been good in rapidly filtering and facilitate deeper comprehension of content through interactivity [29,30,31].

Considering the high dimensionality of regulatory data, animation tools have been constructed to enable regulators to step through time to examine a market or an individual participant’s orders. Using a histogram framework, figure 8 below gives an instantaneous snap shot of the limit and stop-loss order book along with a historical trading volume chart. The snap shots are put together sequentially in a video format that allows users to select a time intervals between shots and the ability to play at different speeds or directions.

Fig. 8. Example of Market Order Book Animation Snapshot[1]

Animations do not preclude the inclusion of more standard methods of information summation. As noted above, one of the strengths of an animation is the ability to convey large amounts of information within a compact space[7]. However, there may be cases when the animation designer wishes to emphasize individual items within that information set. For example, an animation may depict all elements within a statistical distribution, but the designer may wish to highlight within this the mean and extreme values of the distribution. Because this subset of information is of a limited nature, one can often return to the simpler method of inserting the values in text. By doing this, the user is able to move between the holistic view given by the animation and the targeted information set within the associated text. This also simplifies a user's task by avoiding the need to estimate values, by eye, attributed to certain participants. Thus, the addition of text is meant to complement the holistic view of the market through specified values of potential interest, and likely keeps the visualization from being overly complicated [32]. This hybrid approach attempts to incorporate multiple methods of information transfer to allow for multiple analytical responses.

Items of information potentially of value within the text summary are dependent on the visualization use. In the case of market monitoring, market resiliency has been emphasized; resiliency often is highly dependent on the level of liquidity provided in a market, relative to the level of liquidity demanded. A negative imbalance between the two can cause instability in the market, resulting in sudden, large price moves. Liquidity provided is given by the number of limit orders added to the order book during a specified time interval. Liquidity demanded is given by the number of market, or marketable, orders during the same period (with some perhaps resulting from the activation of contingent orders). A third, related category is the velocity of order cancellation, which itself reduces the prior level of liquidity provision. The combination of the three indicates to the monitor the change in liquidity levels over time. To summarize these categories, one can include information about the most important accounts within the groups. In other words, the text metrics can show the account with largest liquidity provision (perhaps divided into bid and offer sides), the account with largest liquidity demand, and the account that cancels the largest number of contracts, during the interval. If unusual price movements were experienced within a known period of time, the unusual movement can likely be most commonly attributed to those accounts with the highest velocities in the groups outlined above.

Within much of our discussion, we have highlighted the common regulatory need to understand both the actions of market participants as a whole (often in the context of market oversight) and the actions of a single participant within this larger system (often when trying to categorize the intentions, whether benign or malicious, of the actor). A single participant animation tool, with information mirroring the above, built for examining the practices of a traders helps to break down the impact a single trader can have on a market's behavior. The animation in figure 9 above shows the total number of limit orders provided contracts by a specified individual (in this case a simulated participant ("Anonymous 23") together with a reference order book of the entire market to compare the relative liquidity importance (i.e. resting orders) that the individual plays in reference to the rest of the market. Similar to the above, in addition to the information provided by the animations, below the charts are included more targeted metrics, in text, regarding orders, trades, inventory, and profit/loss.

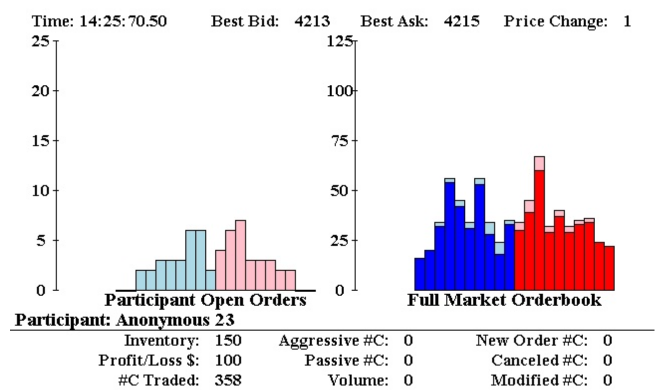

Fig. 9. Example of Participant Order Book Animation Snapshot[1]

Figure 10 gives an example of the extensive degree of information that a visual can prove especially in multipolar systems such as the equities market where there are numerous exchanges trading the same securities. The figure on the top depicts the structure of each exchange and the figure on the bottom depicts the structure of classes of traders (Fundamental Traders, Market makers, etc.), while segregating their decisions and giving investigators perspective of the joint systems through a consolidated limit order book depth chart and a moving average heat map of trade volume. The figure provides insights into two core areas of regulatory focus: intermediation and concentration. The number of off-white non-diagonal blocks speaks to the degree of intermediation, while the intensity of the diagonal is a measure of within-ECN/trader class concentration.

[7] Colloquially described as "A picture is worth a thousand words."

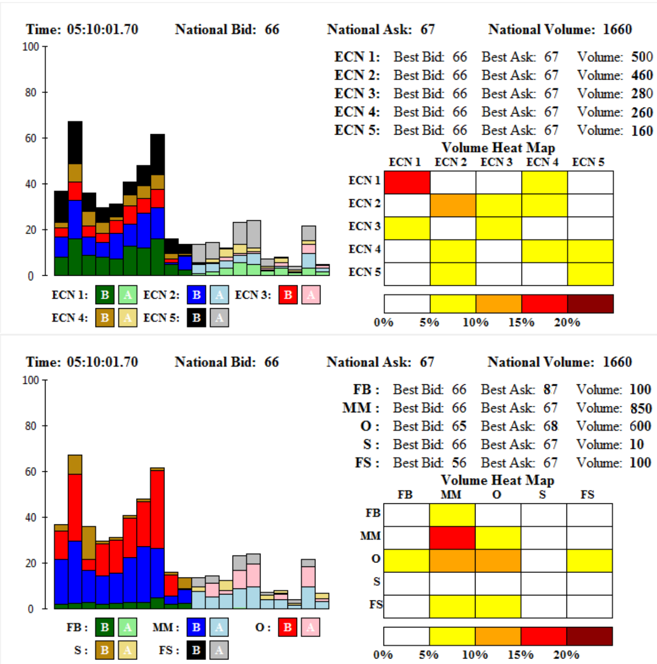

Fig. 10. Consolidated Order Book in respect to ECNs (top) and Participants (bottom)

5) *Order Tracing Graph*

The most granular element within the market system is the order and its evolution over time. It can be difficult to depict the modifications made to an order as it moves around the order book depth, especially since, at a relative level, an order may “change” because the status of the order book changes around it. An individual order can change in type, quantity, and price throughout its life before being either traded or cancelled (in part or in full). A visualization which we present here, the Order Trace Graph in figure 11, depicts an order’s lifecycle.

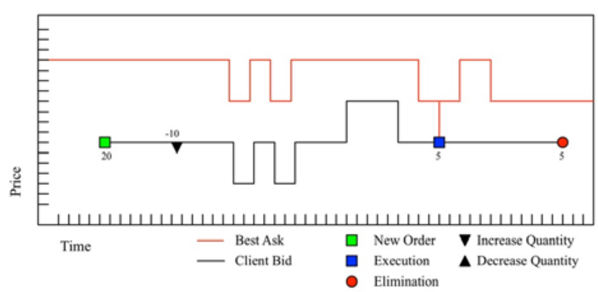

Fig. 11. Example of Participant Order Trace Graph

The Order Trace Graph allows a user to depict individual orders over time, and visually process the events during its lifecycle through a spatial-temporal framework of price level and clock time. In figure 11, an individual order of a client bid, is tracked from its inception to its final elimination. The order is modified a number of times throughout its life span, both in price, seen in the line’s vertical movement six times, and the size of the order, illustrated with shapes specifying the altered levels. Through this, an analyst can construct, within a single visualization, the “storyline” of an order and relate it to the entire order book’s activity.

B. *Visualization Selection*

One notable problem with visualizations however is their scalability. After reaching a certain sizes or levels of complexity, they can become too large for efficient examination. Too many elements and tendencies towards visual clutter may reduce information transfer when more data is included. The critical threshold of diminishing returns depends on the graphical density of the data’s dimensions of structures and hierarchy, while still effectively achieving task or goal of the visual. The proper selection of a visual framework, similar to that of statistician selecting a statistical test, is critical to answering the question of an investigator while objectively communicating the information gained.

In order to take advantage of the visual frameworks introduced, a proper understanding of the question an investigator is trying to answer must be assessed so that both data and task orientation can be selected accordingly. This requires selecting objective for the investigation that can be properly defined such that a metrics of measure can be constructed that proves or demonstrates the objectives. An examples of such objective metrics might be the modification and cancellation rates per minute when trying to investigate ‘quote stuffing’ in a market, a practice of slowing down a markets matching system with large quantities of messaging traffic.

The initial step in this process once an initial objective metric is selected is to defining the scope of the data to be considered so it can be tuned as the investigation proceeds and causations can being re-hypothesized as better understanding is found. There are three aspects to defining the data needed: scope, reference dimensions, and level of activity.

$$Objective \approx Data(Scope, Reference, Activity) \quad (1)$$

The first area of interest is the scope required in the dataset being examined. This dictates what is being investigated e.g. a market, or is it a firm, account, or trader in a market. Even finer might it be a set of individual trades or individual orders that are placed into an exchange.

The next variable to consider is what reference points does the data so as to place as sense of dimensions to the data, and, thereby, give order to the data. Initially it might be useful to not have any and aggregate the data. However to effectively communicate the process and results of an investigation, usually a series of events must be explained in a chronologically manner subject to a set of conditions. This can especially necessary when multiple events occur in parallel and are difficult to explain using a linearly structured sentence. Some examples of reference points are: Clock Time, Trade Time, Order Time, Volume-weighted Time, Price-Time, and Price.

The third consideration is the level of activity that is encompassed by the investigation. The levels can be made finer or broader as the investigation process occurs, but it is essential to consider to what level your investigating should start. However, each underlying level of the exchange can later be examined for understanding as to its role in the investigation.

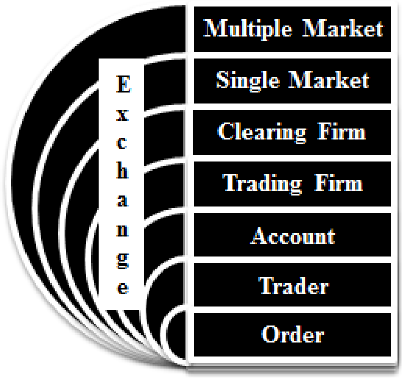

Fig. 12. Levels of Activity in Investigation

Once an understanding of how these three components fall into the objective of the investigation you can then make a selection of what visual is most appropriate for the task or metric the objective requires. Figure 13 gives a chart suggesting what visualization might best fit the task/metric using the scopes to help determine the selection.

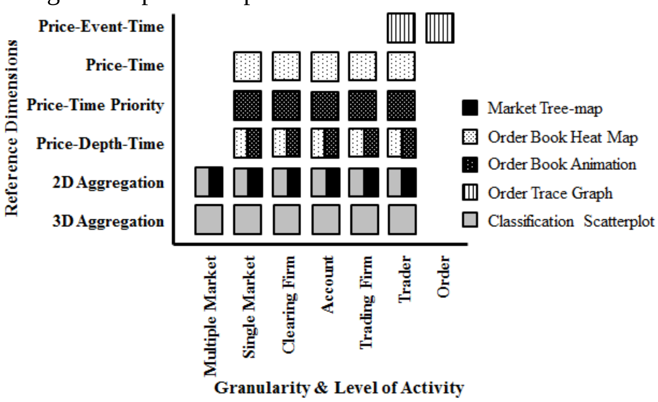

Fig. 13. Visualization Selection Chart

VI. Case Studies

The following example cases studies[8] are designed to demonstrate the benefit that visualizations can bring to regulatory tasks, including those regulators engage in on a daily basis. Each case example is prefaced with a storyline reflecting incidents that either have happened, or could happen, in current financial markets. Each of the cases described requires a subsequent regulatory review that may have to be done within a short response time window. By providing regulators with tools that aid efficient review, turnaround times required for answering causal questions can be greatly reduced. Therefore, within each case study, we cover how the integration of tools like visualizations can substantially increase productivity and the extraction of pertinent information from regulatory data.

A. Case 1: Market Oversight - Large Price Drop leads to Questions of Manipulative Trading

The price of ABC shares fell unexpectedly during afterhours trading by over 40% over a brief period of 10 seconds, sending investors into a panic. ABC's CEO assured investors the next day that the company had strong earnings and revenues and could not understand/explain the large price drop.

In financial markets it is not uncommon to see prices rise or fall very quickly after anticipated news announcements; this is often simply a price response reflecting the market's incorporation of new information. Less common are cases where prices in a given security change significantly without the obvious presence of new information. During these periods, it is hard for market participants, or observers, to point to reasons external to the order book as the ultimate cause of the increased volatility. Because a price discovery process matching of bids and offers must occur, one possible explanation may simply be unexpected changes in the order book itself. Given this, there could be related concerns about disruptive trading practices, either those done by mistake or done with the very purpose of disrupting price discovery. Answering questions of this type often requires regulatory review, and depending on the circumstances, enforcement review.

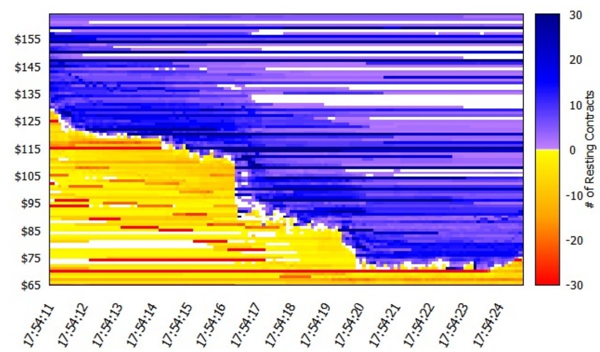

Fig. 14. Heat Map of Order Book Depth during the Price Shift of ABC

As with most investigations into market behavior, a blunt, yet helpful first analysis can be achieved through a relatively "naïve" depiction of actions published within the public market feed.[9] This feed provides the level of liquidity at each price point, along with the price and timestamp for executions. Using this feed, analysts can isolate the time period associated with the strongest, and most rapid, of the price movements.[10] In addition, regulators may wish to identify whether there were precursory events that could have given warning prior to the movement.[11] In the case above, as can be seen in the Figure 14, during the period at 17:54 between 16.30 and 16.40

[8] The example case studies are fabrication of the authors of this paper using data generated by a market simulator from the University of Virginia's Financial Decision Engineering Lab and are not based on completed or ongoing investigations at a financial regulator.

[9] Note that the public market feed, which indicates the current state of the limit order book, would not include hidden liquidity like iceberg orders.

[10] Regulatory data of this type often has timestamps with precision down to the millisecond, allowing for very granular event ordering.

[11] One metric which has been claimed to have these predictive characteristics is the VPIN (Volume Synchronized Probability of Informed Trading). More strictly order book related, large levels of stop orders placed at similar price points may indicate market vulnerability.

seconds the price of ABC shares drops 15 dollars, which through a quick viewing of the market feed can be determined as the most rapid price movement. Around this point, no large demand or supply buildup of orders can be seen in the market; perhaps more importantly, no there is also no indication of a large decrease in liquidity just prior to this point. With this, it appears clear that the bid side of the market did not anticipate such a violent after-hours move in ABC stock. This thought process would lead the analyst to consider the third of our three sources of liquidity change that cannot be seen within changes in the order book: marketable orders removing standing orders from the bid liquidity levels.

To get a better picture of this moment, and to dig further into the reasons for large liquidity demand, an analyst could move to the Order Book Animation tool (refer Figure 15). This visualization provides the state of orders that were resting in the order book (which we saw in a more static way within the Heat Map), but more importantly here an identification of those accounts which were responsible for the most new orders, cancellations, modifications, and trades in the market during the 15 dollar drop.

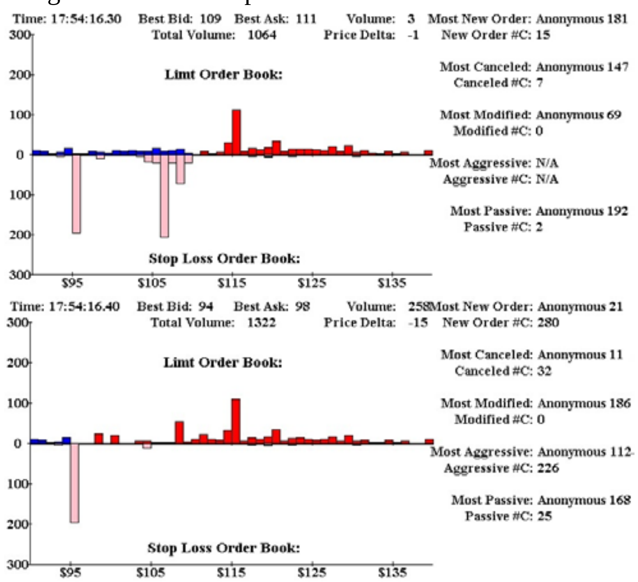

Fig. 15. Market Order Book Histogram Snapshot at 17:56:16.30 & .40 Price Movement

From the animation (from which the snapshots above have been taken), we can see in the before snapshot of the price shift a large number of stop-loss sell orders resting with trigger prices set between $109 and $104. This cluster of stop-loss orders, lined-up like dominos, when aggregated clearly show a volume far beyond that of the standing limit order buy depth. Because of this, in the after snapshot we can see the effects of this set of contingent orders getting triggered. The first stop-loss order, when triggered, overwhelms the standing depth at that price point and therefore triggers the stop orders just below, causing the dominos-like fall seen in the price feed. Because stop loss orders are automatically triggered and executed, within the matching engine, the speed of their impact can be extremely high. As they triggered each other, within a matter of 100 milliseconds, they consume a total of 15 ticks of the resting limit orders, a staggering sum.[12]

At this point, the first set of questions of investigation seem to have been answered: liquidity demand, in the form of very large stop orders, overwhelmed standing liquidity and forced prices to move several ticks prior to market stabilization. Information like this provides the regulator a means to understand the "why" of an event inside the matching engine.

However, if the interest is to determine the original intent of the orders, auxiliary information is important. Within the order book animation, the associated metrics provide information about the accounts that originally placed the stop orders. We can observe, as these metrics display through the event, that the majority of orders that changed from stop-loss to limit orders were placed by a single account (Anonymous 112) and that account made up nearly 90% of all the traded volume during that period of 100 milliseconds. It is clear that a sole individual, or desk, entered the full set of stop orders, at some point in the past, for some, as yet unidentified, reason. In continuing this investigation, an analyst could use the trace order graphs to identify when the stop-loss sell orders where placed to assess if trader Anonymous 112 might have tried to create this event maliciously versus simply trying to provide legitimate protection against adverse price movements.

One further indication helpful in assessing trader Anonymous 112's motive can be realized by observing the extent of other aggressive actions of this account just prior to the period of interest. It may be the case, assuming the account knows of the stop orders, that they actively worked to cause them to be triggered, perhaps by executing a few contracts at a close price point to set off the cascade. There may be other observable activity by them within the book that may also have made the price fall almost inevitable. That said, it should be noted that convincingly proving intent is often an extremely difficult task within an enforcement investigation.

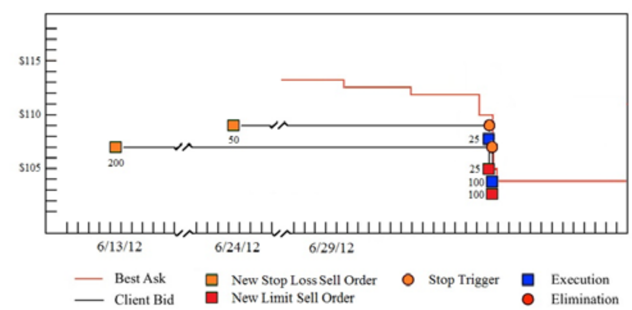

Fig. 16. Trace Order Graph of Stop-loss orders

In this case, the Trace Order Graph of their two stop-loss orders were placed far apart from one another and were not modified in any way that would be suggestive of trying to cause a price drop.

[12] This set of events, a set of contingent stop orders progressively triggering the next, is not of vanishingly small probability. Some of the futures exchanges have included functionality within the matching engine which will pause the market when this event is imminent (so called "stop-logic functionality"). This functionality was introduced to mitigate the effects of exactly this sequence of trades.

B. Case 2: Research - Examining the impact of an Exchange's New Policy

The Stock Trading Electronic Market (STEM), one of the largest stock exchanges in the US, has recently contemplated implementing a new rule to prevent its trading systems from being "clogged up" through excessive usage by inefficient high frequency trading (HFT) systems. To address this, STEM plans to implement an order-to-execution ratio program on its exchange in hope of lowering the volume of electronic messaging. The exchange has found that messaging levels have increased dramatically in certain products over the last few month to the point that they have slowed down its market matching engines. In many cases, these messages never result in executed trades, and seem to provide little information to price discovery. Most trading firms are proponents of this new rule since they believe it will help cut down on the amount of undesired HFT activity, which they feel has been increasing their latency and cutting into their profit margins.

Financial markets rely on efficient matching of interested buyers and sellers in a product, and rapid dissemination of this execution information to the more general market public. In perhaps rather basic terms, this set of activities is the primary purpose of the matching engine and its related service feeds. However, given that the matching engine is an automated system, the speed at which it can disseminate this information is proportional to the amount of information requiring processing; in addition, this processing comes at a cost to the exchange. Unneeded or inefficient processing imposes burdens on both the exchange and the exchange members. In particular, one growing class of potentially inefficient messaging at modern exchanges is messaging related to orders far from the current bid/ask which are modified frequently, though rarely executed. These messages require multiple updates within the matching engine, but rarely take part in the true price formation process. As a result, a growing number of exchanges have chosen to implement a "disincentive program" related to the ratio of orders to executions associated with a given account.

STEM is considering implementing a similar program, but is concerned that it may result in lowered liquidity levels, especially during periods of extreme volatility. The exchange would like to better quantify the costs and benefits associated with the implementation of such a program. The general perception on the impact of rules such as found in an order-to-execution ratio program is that they are targeted at the "HFT-subset" of the market; however, this conclusion may not be so straightforward. It is important that both regulators and the exchange consider the full list of stakeholders potentially impacted by the proposed rule.

Examining STEM's market more closely, it is simple to identify one set of participants who may be most affected by the new policy: the top hundred most active message submitters in the market; most of these are expected to utilize some level of high-frequency trading. In Figure 17 we depict the percent of executed orders versus the average percent of resting orders they offer to the market versus trading volume. In generating this rough classification (i.e. directly related to the policy under consideration) certain clusters of participants are clearly identifiable.

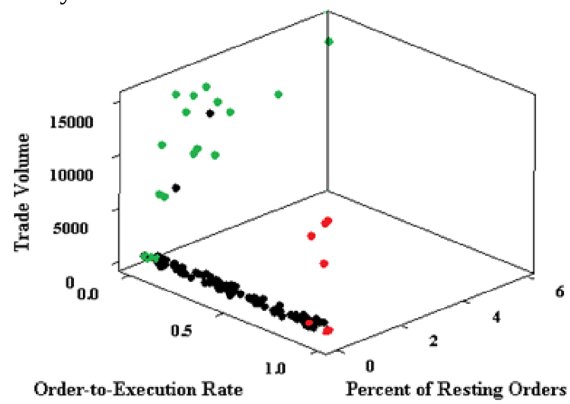

Fig. 17. Classification Scatterplot of HFTs in STEM Exchange (red = more than 95% aggressive, green: less than 5% aggressive)

First, it should be noted that not all market participants are included within the representation. Given the order-to-execution ratios currently in effect at various market venues and under consideration by STEM, the policy would only directly affect a very small number of participants (i.e. those considered "anomalous" in their behavior). Because of the quantity of messages required to hit the order policy limits, accounts must be submitting, modifying and/or cancelling orders at an elevated frequency. In the figure, only those accounts identified as high-frequency traders have been included.[13] This 'HFT' title, however, glosses over the fact that different categories of high frequency traders are likely to be affected in differing ways by an order charge. One subset of HFTs submits orders that are overwhelmingly aggressive (i.e. so almost every order is executed against liquidity sitting in the order book). Another set is primarily passive, but enters orders close to the BBO, with good chances for execution (similar to traditional "market makers"). A final set, the most passive, consists of accounts that primarily enter orders far away from the BBO and rarely experience executions; as a result, accounts within this set may be the most inefficient in order entry.

[13] Though it is often difficult to provide a formal definition for high-frequency traders, characteristics of this type of participant often include wide use of automated systems, low-latency connectivity solutions, co-location or proximity services, along with other technological trading methods.

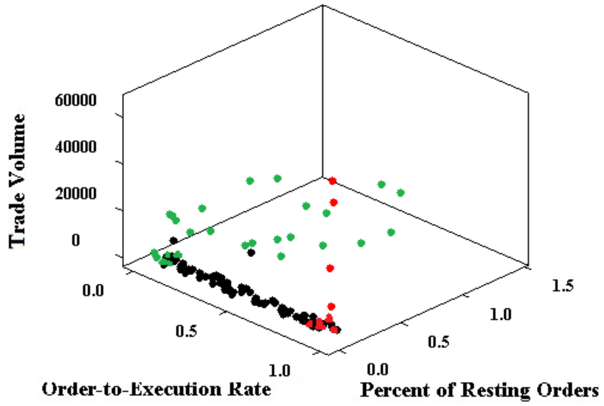

Fig. 18. Classification Scatterplot of HFTs in STEM under High Volatility Market (red = more than 95% aggressive, green: less than 5% aggressive)

As can be seen, the most passive of the traders (those who do not execute on a significant portion of their standing orders) provide a significant portion of the standing volume at any point in time. This volume does not necessarily imply an explicit provision of liquidity during traditional market movements, but may signal an implicit liquidity provision of orders ready to be filled if prices move beyond certain bounds. This liquidity could provide a "backstop" for price volatility during market instability. Hence, it is possible that these market participants are willing to provide liquidity at the very moment in time where liquidity is most necessary (i.e. that moment where large parts of the order book are experience cancellation or execution). In order to determine the validity of this hypothesis, it is useful to isolate the behavior of these accounts during a period of extreme price movements; by doing this, a regulator or exchange can differentiate between orders placed deep in the book, which express willingness for execution, and pure "phantom" liquidity. Figure 18 provides a liquidity provision depiction similar to that in the previous figure, but now isolates order percentage during high volatility periods.

Taking a closer look at liquidity provision at STEM during market periods of high volatility, we can see that, at the market level, liquidity offered by resting orders decreases; this is as expected, given the higher option value implied by resting limit orders. However, within this reduced order book depth, that group of accounts we have classified as the ultra-passive traders continue to make up the majority of the offered liquidity. It is also clear that the order-to-execution ratio of this group increases during these periods, as they experience execution at prices, which were recently deep in the order book. Increases in one-directional aggressive trading now face much of the "inefficient" liquidity identified earlier; disincentives against liquidity of this type could further exacerbate price movements during volatile times. In other words, any further decreases in liquidity, due to a revised incentive structure, could increase market price volatility at the extremes.

Examining such secondary features can easily be neglected by STEM, but clearly should be part of properly assessing the risk that the new policies may pose to market. The research teams at the regulatory agency may be the best resource to fully comprehend such market relationships and prevent unintended events from occurring, especially given the often cross-market transmission effects, for which the regulator is the sole data collector.

C. Case 3: Enforcement - Allegation of Spoofing

A Trader on the floor of the ABC Futures Pit in Chicago noticed that the concurrent electronic market order book had been flashing large sell demand orders throughout the day, though the trader rarely saw these orders executed. They reported the case to regulators as a suspected case of spoofing in the ABC market aimed to lower its price.

Spoofing has been identified as a market manipulative practice, wherein traders with a position in a financial instrument place an anonymous buy order (or in the opposite direction, a sell order) for a large quantity and soon after cancel, to avoid execution. The intention of the order is to provide the impression of large buying demand (without actual execution), resulting in an upward price movement. Often, the market participant will have a resting order sitting on the other side of the market that will get filled due to the price response. Once the market returns to its previous equilibrium level (due to the fact that the large demand was ephemeral), the participant can buy at the lower price, realizing a profit and flattening his risk. This act of placing orders with the intention of canceling them before execution was recently been made illegal under provisions of the Dodd-Frank Act.

In investigating the allegation of ABC future price manipulation, a regulator would need to determine the validity of the floor trader's complaint, including whether the identified set of orders came from a single individual. To do this, we first look at the window of time identified by the trader, and create a related tree-map signifying the quantity of order placement and cancellation in the ABC market.

In Figure 19, the area of an individual rectangle represents the relative size of order placed by a given participant (i.e. with the assumption that the orders of interest under investigation should be of an anomalous size). The color of a given rectangle is determined by the probability of the order being canceled by that participant. Red orders signify a high degree of "false" (or non-executed) liquidity provided by that account.

Fig. 19. Tree-map of Market Participant: Organized by Size according to Largest Observed Order and Highlighted by Percent of Contracts Canceled (smallest: white, largest: red)

Of course, these filters may not fully account for the "spoofer's" behavior. The orders associated with the spoofing event may only be representative of a relatively small fraction of an account's activity. As with all of the above applications, an iterative process in determining appropriate metrics will likely be necessary.

That said, in the above visualization, we notice that there are a handful accounts with metrics that may indicate potential spoofing activity: 23, 373, 415, and 424. All of the accounts in this set exhibit some measure of large orders during the day and have cancellation rates greater than 75%. At this point, with a manageable number of identified accounts, a regulator would be able to drill down further into the specific activity of the four participants. Through further investigation using the individual order book animation tool described above for each of the four accounts, we can see evidence of Anonymous Trader 23 taking actions that appear to match the definition of spoofing.

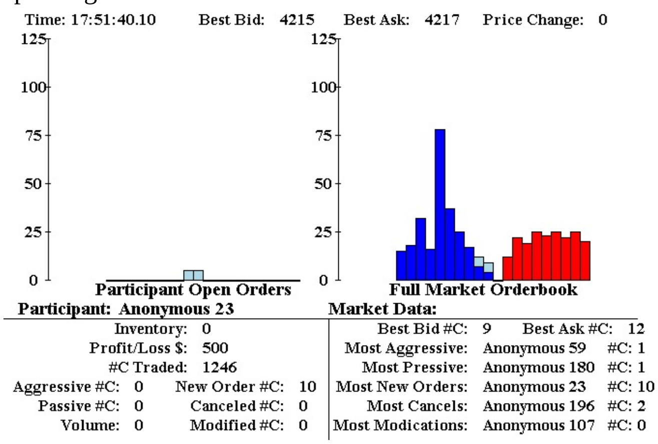

Fig. 20. Participant Anonymous 23's Order Book Histogram Snapshot Pre-Spoof

In Figure 20, we see that Trader 23 has placed two buy orders for five contracts at the level of the best bid and the bid level just below. These (relatively small) sitting orders appear to be in preparation for the subsequent "spoof" order to sell a hundred contracts at the best ask, seen in Figure 21, entered with the possible intention of driving down the price. This sell order is almost immediately cancelled (to avoid the possibility of execution), but results in a movement of price downward. As a result, this price movement causes both of his smaller, passive buy orders to fill at the short-term, lower equilibrium price.

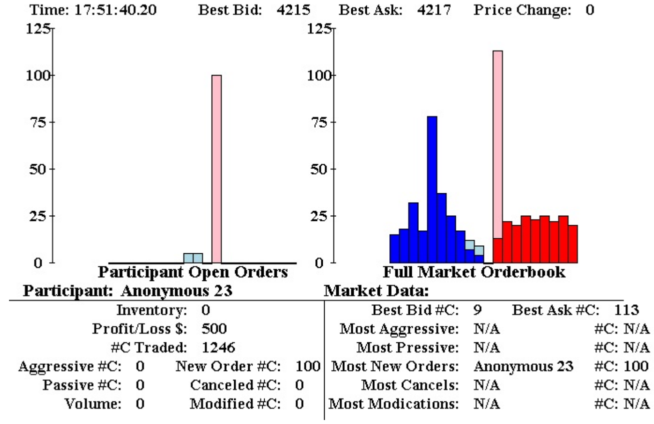

Fig. 21. Participant Anonymous 23's Order Book Histogram Snapshot during Spoof

We can see that in Figure 22, that the spoof is successful in its buy execution, with prices moving the sell interest lower, and the account finding itself now long ten contracts. Within a matter of a few additional seconds, the price has returned to its prior equilibrium, allowing the participant to flatten out risk at an elevated price.

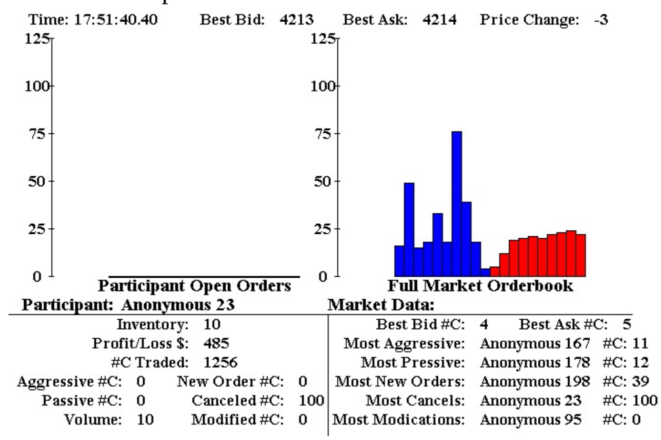

Fig. 22. Participant Anonymous 23's Order Book Histogram Snapshot Post-Spoof

VII. Conclusions

Analysis of behavior in financial markets has traditionally focused on consummated actions within a market environment, actions including completed trades and the resulting inventory held by a participant. Recent improvements in the regulatory audit trails available from financial markets now allow a far more complete and information rich reconstruction of the order book. This expansion in data quantity, however, comes with resulting difficulty. Processing and analyzing order flow data of this type, especially in the case of institutions not used to "big data" processing can present a major challenge.

With visualization techniques for retrieving, analyzing, and disseminating data, regulators can have access to the tools needed to tackle the cumbersome task of examining the large data sets relevant to their regulatory tasks. Such tools can facilitate a rapid analysis of changes in participant and market behavior and subsequent dissemination of this information to relevant parties (including the exchange, the clearing firm, or the participating firm itself). Visualizations are one of the most powerful means of analysis and subsequent information dissemination. By using appropriate visualization techniques, regulators can create a more detailed market picture not only including public market depth, but also detailed information on order status, history and priority.

REFERENCES

[1] Dow Jones & Company (2013). Map of the Market, NJ. [Online]. Available: http://online.wsj.com/mdc/public/page/2_3024markets_map_of_the_market.html

[2] T. Dwyer and P. Eades, "Visualizing a Fund Manager Flow Graph with Columns and Worms" in *IEEE IV*, 2002, pp. 147–153.

[3] CFTC and SEC. (2010). Preliminary Findings Regarding the Market Events of May 6, 2010. DC. [Online] Available: http://www.sec.gov/news/studies/2010/marketevents-report.pdf

[4] T. Hendershott, and C.M. Jones, (2005, Mar.). Island goes dark: transparency, fragmentation and regulation. *Review of Financial Studies*. 18(3), pp. 743–793.

[5] CME Group. (2013). CME MDP Message Statistics. IL. [Online] Available: http://beta.cmegroup.com/market-data/distributor/market-data-platform.html

[6] D. Keim, (2002, Jan.). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 7(1), pp. 100–107.

[7] D. Meadows, "The Basic" in *Thinking in systems: a primer*, VT., Chelsea Green Publishing, 2008, ch. 1, sec. 1, pp. 11–34.

[8] J. Mackinlay, (1986, Apr.). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*. 5(2), pp.110–141.

[9] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melancon, "Visual Analytics: Definition, Process, and Challenges" in *Information Visualization: Human Centered Issues and Perspectives*, Springer, 2008, vol. 4950, pp.154–175.

[10] M. Zhou and S. Feiner, "Visual task characterization for automated visual discourse synthesis," in *CHI Conference*, 1998, pp. 392–399.

[11] E.H. Chi, "A taxonomy of visualization techniques using the data state reference model," in *IEEE IV*, 2000, pp. 69–75.

[12] M. Chen and L. Floridi, *Synthase.* unpublished.

[13] H. Blume, "Behavior Identification in Markets using Visualization and Network Analysis," Ph.D. dissertation, Info. & Mkt. Engr. Dept., KIT, Karlsruhe, Germany, 2012.

[14] M. Paddrik, R. Hayes, A. Todd, S. Yang, W. Scherer, and P. Beling, "An Agent Based Model of the E-Mini S&P 500 and the Flash Crash," in *IEEE CIFER*, New York City, NY., 2012.

[15] M. Wattenberg, "Visualizing the stock market," in *CHI Conference*, 1999, pp.188–189.

[16] B. Shneiderman, (1992, Jan.). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*. 11, pp. 92–99.

[17] L. Harris, "Liquidity" in *Trading and exchanges: Market microstructure for practitioners*. NY. Oxford University Press, 2002, ch. 19, pp. 394–409.

[18] I. Spence and R. Garrison. (1993, Feb.). A remarkable scatterplot. *The American Statistician*. *47*(1), pp. 12–19.

[19] I. Spence, and S. Lewandowsky, "Graphical perception" in *Modern methods of data analysis.* CA., Sage, 1990, Ch. 1, pp. 13–57.

[20] E. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT., Graphics Press, 1983.

[21] A. Kirilenko, A. Kyle, M. Samadi, and T. Tuzun. "The flash crash: The impact of high frequency trading on an electronic market," unpublished, [Online] Available: *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686004*

[22] L. Wilkinson and M. Friendly, (2009, Jan.). The history of the cluster heat map. *The American Statistician*. 63(2), pp. 179–184.

[23] O.J. Blanchard, *The crisis: basic mechanisms and appropriate policies.* Vol. 9, International Monetary Fund. Cambridge, MA., 2009.

[24] J. Larkin and H. Simon, (1987, Nov.). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*. 11(1), pp. 65–99.

[25] B. Tversky, "Cognitive origins of conventions" in *Understanding Images*, NY. Springer-Verlag, 1995, pp. 29–53.

[26] IMF, *Global Financial Stability Report, October 2008: Financial Stress and Deleveraging Macrofinancial Implications and Policy*. International Monetary Fund, Oct., 2008.

[27] L. P. Rieber, (1989, Oct.). The effects of computer animated elaboration strategies and practice on factual and application learning in an elementary science lesson. *Journal of Educational Computing Research*. *5*(4), pp. 431–444.

[28] E.L. Ferguson and M. Hegarty, (1995, Jan.). Learning with real machines or diagrams: application of knowledge to real-world problems. *Cognition and Instruction*. 13(1), pp. 129–160.

[29] S. Srinivasan, D. Ponceleon, A. Amir, and D. Petkovic, (1999). "What is in that video anyway?": in *ICMSC Conference*, 1999, pp. 388–393.

[30] E. Perez and M. White, (1985). Student evaluation of motivational and learning attributes of microcomputer software. *Journal of Computer-Based Instruction*. 12, pp. 39–43.

[31] L.P. Rieber, (1991, Mar.). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*. 83(3), pp. 318–328.

[32] J. Morrison and B. Tversky, "The (in)effectiveness of animation in instruction," in *ACM CHI Conference*, 2001, pp. 377–378.