Visualizations for Financial Market Regulation

Mark Paddrik, Andrew Todd, William Scherer, Peter Beling
Systems & Information Engineering
University of Virginia
Charlottesville, VA, USA
{mp3ua, aet6cd, wts, pb3a} @virginia.edu

Richard Haynes
US Treasury
Washington, DC, USA
Richard.Haynes@treasury.gov

Abstract—Electronic markets and automated trading have resulted in a drastic increase in the quantity and complexity of regulatory data. Regulatory analysis now includes detailed analysis of all messaging and communications related to electronic limit order books. New order types, intra-market behavior and other exchange functionality further complicate analysis. 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 academic research.

Keywords- Visual Analysis, Financial Market Regulation, Financial Data, Electronic Order Book

I. INTRODUCTION

Financial markets have long been vital to efficient resource evaluation and allocation. Yet for markets to play their intended central role in price discovery, participants must trust that the mechanisms of the market are fair to all. Historically, a system of organized exchanges overseen by regulatory bodies has engendered the level of trust necessary for markets to function smoothly. The characteristics of financial markets have changed drastically in the last several decades, however. Exchanges have abandoned the open outcry format for electronic matching systems where automated trading is now the dominant mode of for negotiating and completing transactions. Market regulators who once wandered through the pits of commodity futures exchanges today are forced to rely on big data analysis and relevant statistics to ensure markets are free of fraud, manipulation and abusive practices.

The mission of regulators has been further complicated by market fragmentation and the rise of automated trading. Market participants employing high performance computing and low-latency market connections transact on millisecond time scales. The electronic order book has vastly reduced the ability of market participants to self-regulate themselves by obscuring traders’ identities and allowing manipulative behavior to go unrecognized. Events, such as the flash crash of May 6th, 2010 [1] have eroded market confidence and triggered calls for the regulation of what has come to be known as high frequency trading. Additional participant behaviors outside of the scope of current regulation have started to influence market quality. A key technical challenge faced by regulators today is to understand how to use available data to recognize participants who are manipulating markets or engaging in otherwise abusive practices.

While they do allow for a richer and more accurate reconstruction of the market, the sheer complexity and volume of today’s exchange data threatens to overwhelm regulatory analysis. The complexity of order types and intra-market capabilities offered by exchanges complicates the task of reconstructing market events such as trades and orders. 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 central limit order book [2] or the audit trail of a specific account. In addition the large volume of regulatory data, and the general need for rapid synthesis related to order activity, suggests that regulators have a need for strong and flexible tools for data analysis and visualization.

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. In order to digest a diverse collection of information, then, it is necessary to use a language that shares some of the same properties as the phenomena under observation [2]. Visualizations are an especially powerful method for summarizing multi-dimensional data, such as is seen in the evolution of financial markets. Visualizations are also useful because they provide a balanced portrait of the dynamic and static components of the market system.

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 monitoring, enforcement and general oversight.

1 By “regulator” we mean to 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, the term could include SRO’s such as FINRA or NFA within the US, and exchanges tasked with market oversight.

2 A trading mechanism used by most electronic exchanges globally for matching customer orders (e.g. bids and offers) on a 'price time priority' basis.
Visualizations have the potential to become a valuable part of regulatory analysis. Techniques, discussed in Section 2, which can help decompose the structure of financial data and further allow regulators to accomplish their mission are in demand. Visualizations offer a variety of tools that can further this cause and supplement data analysis. Section 3 develops a set of visualization techniques that can support both exploratory data analysis and core regulatory tasks. Section 4 goes through a case study in which we the set of visualizations are implemented as part of the analysis of the market event and finally Section 5 provides a summary of the conclusion.

II. BACKGROUND

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. Regulators must battle to stay current with these changing practices while still maintaining mandated protective roles in the market.

A. Regulatory

The challenge of managing and interpreting data on a large scale in order to meet the objectives set for regulators is far greater than they have had to deal with in the recent past. These objectives can be broken down into three primary roles: market oversight, enforcement, and research relevant to regulatory policy.

Market oversight must verify the integrity of market conditions and, conversely, identify those periods where integrity has been compromised. This requires the construction of accurate, reliable, and replicable information regarding market conditions around these periods. Although in substantially all cases the results of volatile events can be detected easily, determining the ultimate cause of many market failures requires methodical evaluation of the participants and actions taken on multiple markets up to the moment of an event.

In enforcement, it is often important to study the potentially illegal behavior of an individual entity, or set of entities that have chosen to collude. It is this behavior that sparks enforcement investigation and at times communication to an adjudicatory body like a court of law. Adjudication centers around two claims: that the identified individuals acted in an exceptional manner and that those exceptional actions should be considered illegal activity. Regulators face the difficult task of describing the behavior of market participants to adjudicators who are not experts in algorithmic trading. We hypothesize that visualization can be an important tool in this regard.

Finally, regulators engage in policy research when considering the impact of altered market structure. Recent structural changes of this type include such things as the introduction of dark pools [3], 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 evolutions likely have intended and unintended consequences for market liquidity, volatility or participation by various market groups.

B. Financial Data

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

\[\text{Figure 1. Data flow between Market Stakeholders}\]

As a result, it is necessary that exchanges have to capability to communicate quickly, simply and in a secure manner with market participants and regulators that are located in disparate geographical regions. 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. Much of this information eventually is passed on to market regulators. The following is a brief summary of these data channels.

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. These messages make it possible for the exchange to fairly execute trades using a matching engine, an order matching system following a set of publicly known trade prioritization rules.

Distinct from the private feeds that make up the individual parts of order flow data, there exist public, anonymized 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.

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. This data is cumbersome for regulators because of both the underlying structure of the data and the

\[3\text{ Errors and general status notifications from the matching engine are also part of order flow data.}\]
sheer size of records generated in a single day. Consider that at least 10 million contracts are traded per day within the CME Group exchanges alone, and overall activity is perhaps an order of magnitude larger still [4].

The fields of regulatory data provide the building blocks from which one can construct data structures that can effectively communicate information throughout the financial system. However the complexity inherent to logic and size can make interpretation very difficult for regulators.

C. Visualizations in Data Analysis

Visualizations often have acted as means of communicating results, by translating multidimensional data into a form 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 [5]. This efficiency can free the user’s memory to support further, necessary, cognitive operations or tasks. One example would be a city map for first time visitors, an object that stores unfamiliar information in an easy to retrieve format. A map allows visitors to spend time seeing attractions rather than learning all street names/intersections prior to travelling.

A design goal for a data analysis system is to ease the cognitive load on users, freeing time and energy for them to pursue insights or iteratively explore the data. In order for visualizations to be helpful in this regard, their design and use must be appropriate to the task at hand [6,7]. Some of these tasks, within a regulatory context, can be separated into three broad categories:

1) Information Retrieval is the operation of exploring the data space through overview, browsing, navigation, zooming, and observing derived quantities such as data ranges, distributions, errors, certainty and sensitivity of those values. For spatial and temporal data sets it involves inspecting features via viewing animated or sequential representations.

2) 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. Fig. 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.

3) Information Dissemination is the process whereby a visual acts as an aid for the presentation of information, allowing for easy data comprehension. In this case, the visualization should summarize, annotate, and illustrate analytics which support or reject a hypothesis.

III. VISUALS FOR ANALYSIS OF FINANCIAL MARKETS

The full transaction and order histories that are available from electronic markets contain information about the intentions of the participants and the breadth of market responses. Because of this, using data analysis is a promising approach to effectively combat unlawful activities and further the understanding of market anomalies. However, since the patterns used by abusive firms change when they are made aware of a detection mechanism, new patterns have to be reviewed and identified through adaptive means [9]. Therefore, while effective summary visualizations must provide storylines for market events, it is equally important that they stay flexible enough to adapt to changing market behavior.

In the following section, we cover three visualization techniques that can be used individually or in unison as part of integrated regulatory process. 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 [10]).

A. Order Book Heat Map

A heat map is constructed using a rectangular tiling of a data matrix; this tiling facilitates the inspection of three dimensional data using rows, columns, and a third attribute (often color) as it changes with respect to the first two. This allows large data matrices (several thousand rows/columns) to be displayed effectively on a high-resolution color images [11].

Heat maps, such as Fig. 3, depict the state of a financial market over many years, allowing a regulator to monitor the prices of sectors over time [12]. For this application, using a rectangular tiling of a data matrix facilitates spatial-temporal data which shows the price of a group of assets over time; the colors indicate a growing (green), stagnant (yellow), or declining (red) market. With this, the heat map applies a spatial relationship to price across multiple markets, typically not well represented [13,14].

![Figure 3. Example of Sector Heat Map [15]](image)

Another heat map, one for the order book, can be seen in Fig 4. This map allows a user to examine liquidity
expansion and contraction over a selected period of time. The colors are applied at specific price levels and colored using the resting limit order depth (for the 100 ms interval). The depth is color coded with buy orders shaded yellow, and sell orders shaded violet (the darkness of the color indicating size).

From Fig. 4 a user can extract simple information about the direction of price movement over time. Inferences can also be made about the changes in supply and demand of this market. A large selling depth relative to buying depth can be observed above, driving price downward over time.

B. 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 may be too much for a static image. Animation is the rapid display of a sequence of images to create an illusion of movement, helping to express such complex processes such as the behavior of market or the impact of single individuals over time.

In general, animations have been successful in conveying extra information and allowing for interactivity [16]. The ability for users to interact with an animation, combining technology together with a user interface, has been good in rapidly filtering and facilitating deeper comprehension of content [17,18,19,20].

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, Fig. 5 below gives an instantaneous snapshot of the limit and stop-loss order book along with a historical trading volume chart. The snapshot is put together sequentially in a video format that allows users to select a time interval between shots and then play the market at that speed.

Animations do not preclude 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. 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 limited, one can often just insert these values in text. By doing this, the user is able to move between the holistic view given by the animation and the targeted information in the associated text. This also simplifies a user’s task by avoiding the need to estimate values in the chart by eye. This 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 [21]. This hybrid approach attempts to incorporate multiple methods of information transfer to allow for multiple analytical responses.

Items of value for textual summary are dependent on the visualization. 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, 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

---

4 The Limit Order Book histogram depicts sell orders in red and buy orders in dark blue. The Stop-loss Order Book histogram depicts sell orders in pink and buy orders in light blue. The Trade Volume at Price histogram depicts executed trade volume at each price level, sell initiated volume red and buy initiated volume dark blue in the market.
lliquidity provision. The combination of the three indicates the net liquidity change. To summarize these levels, one can include information about the most important accounts, such as 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 occurred within a known period of time, the movement often can be 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 trader helps to break down the impact of that trader. For instance, the animation in Fig. 6 above shows the total number of limit orders provided by a specified individual (in this case a simulated participant, “Anonymous 23”) together with a reference market order book to compare the relative liquidity provided by the individual versus the rest of the market. Similar to the above, the information provided in the animation is supplemented by metrics in text.

C. Order Tracing Graph

The most granular element within the market system is the order and its evolution over time. However, 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 order book changes around it. An individual order can also alter 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 Fig. 7 below, depicts an order’s lifecycle.

![Order Trace Graph](image)

Figure 7. 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 Fig. 7, 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.

IV. APPLIED CASE STUDY

The following example case study is designed to demonstrate the benefits that visualizations can bring to regulatory tasks. The case is an example of a storyline reflecting an incident that has happened, or could happen, in current financial markets. It is followed by a subsequent regulatory review integrating the use of the new visualizations to demonstrate the substantially increased productivity and the extraction of pertinent information from regulatory data.

Case: Price Drop leads to Questions of Manipulation

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

---

5 The left limit order book histogram depicts sell orders, in pink, and buy orders, in light blue, for the individual participants resting orders. The right limit order book histogram depicts the sell orders in pink and buy orders in light blue for the individual participants resting orders, while the sell orders are in red and buy orders in dark blue for the rest of the market participants.

6 The following case study is fabricated by 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.
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.

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. 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. 8 In addition, regulators may wish to identify whether there were precursory events that could have given warning prior to the movement. In the case above, as can be seen in the Fig. 8, during the period at 17:54 between 16:30 and 16:40 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 Fig. 9). 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.

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. 9

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

---

9 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.
means to understand the “why” of an event inside the matching engine.

If the interest is to determine the original intent of the orders, then 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.

In this case, the Fig. 10, 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.

V. CONCLUSIONS

Recent improvements in the regulatory audit trails available from financial markets now allow a far more complete and information rich reconstruction of the state of the market. This expansion in data quantity, along with the complexity of market events, matching logic, and order types, has led to difficulty on the part of regulators to make effective use of this new data. Processing and analyzing order flow data of this type, especially in the case of institutions not used to big data can present a major challenge.

The incorporation of visualization techniques to retrieving, analyzing, and disseminating data, however, can offer regulators accessible tools to tackle the cumbersome task of examining large data sets relevant to their regulatory tasks. Such tools can facilitate the 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). These tools can further the mission of all the parts of these organizations; by allowing more detail to be explored, more effectively ensuring market integrity and fairness.

REFERENCES


