Assessing Financial Markets through System Complexity Management

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Mark Endel Paddrik

December

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

The dissertation has been read and approved by the examining committee:

W. T. Scherer

Advisor

P. A. Beling

P. J. Dennis

G. P. Learmonth

R. I. Webb

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

December

2013

ABSTRACT

Today's data-driven, electronic financial markets, wherein highly automated and, often, high frequency trading has become the norm for major participants, present significant technical challenges that complicate effective monitoring of their most fundamental functions, trade and the price discovery. Though regulators are implementing new surveillance systems capable of capturing all necessary market data to describe the details of market events, the fact that markets are fundamentally "complex systems" often undermines the use of conventional assessment techniques. The research work underlying this dissertation provides new approaches for analyzing financial markets that can help overcome these problems and demonstrates how each can be used to better understand the state of the market's trading and price discovery processes. These approaches utilize adaptations of methods employing data visualizations, Markov State modeling, and agent based simulations that have been more commonly used in other fields.

The occurrence of abnormal pricing anomalies in today's complex markets has particularly intensified the demand for better monitoring and assessment, as it is crucial for regulators to be able to work to protect orderliness in markets and, thus, build confidence for both the industry and the public. These anomalies are transient, unique events that suddenly and unexpectedly take place in markets which are fundamentally complex systems environments with large numbers of participants each having their own, individual, ever-adapting views and decision-making processes. Much of the research effort in the dissertation is aimed at improving the capabilities of financial regulators and exchanges to better understand the precursors of such abnormal events and approach their investigation using methods that help overcome the effects of systematic complexities inherent in today's data intensive markets.

Through examination of financial markets as complex systems designed to support efficient and fair trade and pricing, this work has focused on untangling behavioral complexities by demonstrating how to apply techniques used in engineering that are well-suited for the study of such systems. By breaking down the complexities based on the desired tasks and objectives, this work shows how issues arising from imaginary, periodic, and combinatorial complexities can be managed by using visual analytical techniques, Markov anomaly detection, and agent-based simulation. Through the use of such tools, regulators and exchanges can not only improve their analysis of markets post an event of interest, such as a "Flash Crash", but, also, analyze behaviors peri- and pre- events of interest.

ACKNOWLEDGMENTS

I feel very fortunate to have had the privilege for the past nine years of being part of Systems & Information Engineering Department at the University of Virginia with its so many talented and friendly faculty and colleagues.

First, I would like to thank my advisors, Prof. William Scherer and Prof. Peter Beling, for the opportunity to work on research that expanded my knowledge and for providing me essential insights as the work progressed. Thank you for your guidance and for allowing me to help build the Financial Decision Engineering Lab. Thank you, also, for your understanding and support during my difficult times. I have really learned so much from you both and hope we can continue to find ways to build on our work in the future.

Second, I would like to thank my other PhD advisory committee members Prof. Gerard Learmonth, Prof. Robert Webb, Prof. Patrick Dennis and Prof. Andrei Kirilenko. Thank you for your many valuable thoughts, suggestions, time, and efforts to improve and guide my dissertation work.

Third, I would like to thank my friends and colleagues, Roy Hayes, Andrew Todd, Mahbod Olfat, Matt Burkett, Kari Wald, Steve Yang, Richard Haynes, and Sheung Yin Mo, who worked closely with me during my PhD studies. Without your help, the completion of this dissertation would have been impossible.

I would also like to thank all the Professors with whom I have been so lucky to have worked as their TA or RA through the past years: Prof. Reid Baily, Prof. Barry Horowitz, Prof. Gerard Learmonth, Prof. Ginger Davis, Prof. Roman Krzysztofowicz, Prof. Greg Gerling, Prof. Stephanie Guerlain, and Dr. Kyle Enfield.

Finally, and most importantly, I would like to thank my family. My parents, Thomas and Viviana Paddrik, who, through their lifelong love and support, made it possible for me to follow my dreams in this educational journey.

TABLE OF CONTENTS

Chapter 1: Introduction **1**
1.1 Financial Market Evolution 2
1.2 Financial Regulatory Goals 7
1.3 Financial Markets as Complex Systems 10
1.4 Engineering with Complexity 12
1.5 Controlling for Complexity 16
1.5.1 Visualization and Imaginary Complexity 17
1.5.2 Monitoring Markets and Periodic Complexity 18
1.5.3 Agent-based Simulation and Combinatorial Complexity 19
1.6 Summary 21
Chapter 2: Overview of Market Data Structure **23**
2.1 Electronic Order Book Structure 23
2.2 Market Data Communication 26
Chapter 3: Data Visualizations and Imaginary Complexity **30**
3.1 Regulatory Data Objectives 31
3.2 Structure of Financial Regulatory Data 33
3.3 Visual Analysis Concepts 36
3.3.1 Features of Visual Communication 36
3.3.2 Cognitive Processes Served by Visuals Analysis 37
3.4 Applying Visualization to Financial Regulation 38
3.4.1 Visualizations for Market Regulatory Data 39
3.4.1.1 Market Tree Maps 39
3.4.1.2 Classification Scatterplot/Clustering 41
3.4.1.3 Order Book Heat Map 43
3.4.1.4 Order Book Animations 45
3.4.1.5 Order Trace Graph 50
3.4.2 Visualization Selection 51
3.5 Case Studies 53
3.5.1 Market Oversight: Price Drop Leads to Questions of Manipulation 54
3.5.2 Research: Examining the Impact of an Exchange's New Policy 59
3.5.3 Enforcement: Allegation of Spoofing 64
3.5 Summary 68

Chapter 4: Market Monitoring and Periodic Complexity 71
4.1 The Exchange System 74
4.1.1 Information Flow through Markets 74
4.1.2 Market Participants as Systems 76
4.1.2.1 Fundamental Buyers and Sellers 78
4.1.2.2 Opportunistic Trader 78
4.1.2.3 Intermediary Traders 79
4.1.3 The Exchange: System of Systems 81
4.2 Exchange as Complex Systems 82
4.2.1 Phase Transitions in Complex Systems 83
4.2.2 Detecting Phase Transitions 85
4.3 Monitoring for Exchange for Pricing Problems 86
4.3.1 The Markov State Model of the Exchange 87
4.3.2 Organizing Data for the Markov State Model 88
4.3.3 Application of the Markov State Model for Market Monitoring 90
4.3.3.1 Monitoring Capacity 92
4.3.3.2 Monitoring Trade Dynamics 93
4.3.3.3 Monitoring Intermediation 95
4.4 Results of the Markov State Model 97
4.4.1 Capacity Results 98
4.4.2 Trade Dynamics Results 98
4.4.3 Intermediation Results 100
4.4.4 Detecting Phase Shift Results 100
4.5 Case Study: Flash Crash 102
4.6 Summary 105
Chapter 5: Agent Based Simulation and Combinatorial Complexity......................107
5.1 The Agent-Based Simulation Framework 109
5.2 Validating an ABS of a Financial Market 113
5.2.1 Distribution of Price Returns 114
5.2.2 Volatility Clustering 115
5.2.3 Absence of Autocorrelation of Returns 116
5.2.4 Aggregation of Returns 116

5.3 Investigating with an ABS of a Financial Market 117
5.4 Forecasting with an ABS of a Financial Market 121
5.4.1 Forecast Variables 122
5.4.1.1 VPIN 124
5.4.1.2 Price Impact 124
5.4.1.3 Window Indicator 125
5.4.1.4 OBAW Spread 126
5.4.1.5 PH Spread 126
5.4.2 Testing Forecasts in Simulations 127
5.4.3 Vetting Forecast with Real Market Data 132
5.5 Rulemaking with an ABS of a Financial Market 135
5.6 Summary 137
Chapter 6: Conclusions 139
6.1 Visualization and Imaginary Complexity 140
6.2 Monitoring Markets and Periodic Complexity 141
6.3 Agent-based Simulation and Combinatorial Complexity 142
6.4 Future Work and its Limitation 144
6.5 Final Statements 145
Chapter 6: References 148

LIST OF FIGURES

Fig. 1.1 - The transition of markets over the last 25 years, from human to mostly computer driven auction matching systems and participants 3
Fig. 1.2 - Average Daily Trade Volume seen in the New York Stock Exchange and the Chicago Mercantile Exchange from 1993 to 2012 4
Fig. 1.3 - Frequency of Extreme and Rapid Price Events in U. S. Equities 6
Fig. 1.4 - General complexity concept in context of engineering a system 14
Fig. 1.5 - Combinatorial Complexity and Periodic Complexity in Design Range of a system 15
Fig. 1.6 - Levels of Investigation 16
Fig. 2.1 - The Layer of the Financial Market Data 22
Fig. 2.2 - Data Flow between Market Stakeholders 24
Fig. 3.1 - Model of Data Visualization Analysis 36
Fig. 3.2 - Tree Map of Market Participant Contribution to Order book Depth and Executions 38
Fig. 3.3 - Classification of Traders in E-Mini Futures Market 40
Fig. 3.4 - Example of Possible Classification of Traders in E-Mini Futures Market 41
Fig. 3.5 - Example of Sector Heat Map 42
Fig. 3.6 - Example of Heat Map of Order Book Depth 42
Fig. 3.7 - Example of Market Order Book Animation Snapshot 45
Fig. 3.8 - Example of Participant Order Book Animation Snapshot 47
Fig. 3.9 - Example of Participant Order Trace Graph 48
Fig. 3.10 - Levels of Scope and Activity in Investigation 50
Fig. 3.11 - Visualization Selection Chart 51
Fig. 3.12 - Heat Map of Order Book Depth during the Price Shift of ABC 53
Fig. 3.13 - Market Order Book Histogram Snap Shot at 17:56:16.30 & .40 Price Movement 55
Fig. 3.14 - Trace Order Graph of Stop-Loss Orders 57
Fig. 3.15 - Classification Scatterplot of HFTs in STEM Exchange 59
Fig. 3.16 - Classification Scatterplot of HFTs in STEM under High Volatility Market 60
Fig. 3.17 - Tree Map of Market Participant: Organized by Size according to Largest Observed Order and Highlighted by Percent of Contracts Canceled 63
Fig. 3.18 - Participant Anonymous 23’s Order Book Histogram Snapshot Pre-Spoof 64

Fig. 3.19 - Participant Anonymous 23's Order Book Histogram Snap Shot during Spoof 65
Fig. 3.20 - Participant Anonymous 23's Order Book Histogram Snapshot Post-Spoof 66
Fig. 4.1 - Financial exchanges and participant relationship 69
Fig. 4.2 - Market structure: decentralized market and centralized exchange market 72
Fig. 4.3 - A participant systems interaction with the Exchange System 74
Fig. 4.4 - Diagram of flow of assets through the exchange 79
Fig. 4.5 - Transformation of exchange system model to Markov model 84
Fig. 4.6 - Diagram of modified exchange model to fit changes required by threshold categorization method with descriptive statistics of each class of participants observed in the E-Mini S&P 500 futures market 86
Fig. 4.7 - Plot of transitional likelihood throughout the day 87
Fig. 4.8 - Plot of average I inventory of a contract thought out the month of February 89
Fig. 4.9 - Plot of l_1 for February 25th 91
Fig. 4.10 - Plot of intermediation path and number for the example calculation 93
Fig. 4.11 - Plot of average intermediation of a contract thought out the month of February 94
Fig. 4.12 - Average price change to n(I(i)) 95
Fig. 4.13 - Scatter plot of price change as it relates to l_1 96
Fig. 4.14 - Chart of price movement as it relates to l_1 96
Fig. 4.15 - Average price change to Intermediation 97
Fig. 4.16 - Histogram of I(i), l_1, and Intermediation with quintile extremes 98
Fig. 4.17 - Time series of I(i), l_1, and Intermediation with quintile extreme variables 98
Fig. 4.18 - Average price change depending on I(i), $l_1$and Intermediation in E-Mini and WTI Crude Oil 99
Fig. 4.19 - Example trading strategy in table and graph of resulting strategy implementation 99
Fig. 4.20 - The E-Mini May 6th 2010 results from the model 101
Fig. 5.1 - Components of an agent class 109
Fig. 5.2 - Normality test of real E-Mini price returns and simulated price returns 112
Fig. 5.3 - Volatility clustering of absolute price returns for S&P 500 E-Mini and simulation 112
Fig. 5.4 - Absence of autocorrelation for price returns for S&P 500 E-Mini and simulation 113

Fig. 5.5 - Aggregation of price returns for S&P 500 E-Mini and simulation 114
Fig. 5.6 - Real vs. simulated E-Mini: volume, moving average price and order book depth 116
Fig. 5.7 - Minimum price versus # of high frequency traders & rate of size of sell orders 117
Fig. 5.8 - Market state diagram 120
Fig. 5.9 - PDF of a Flash Crash in Simulated E-Mini market 125
Fig. 5.10 - Indicator PDF & CDF in Simulated E-Mini market 126
Fig. 5.11 - Indicator Advance Warning PDF & CDF 127
Fig. 5.12 - ROC Curves of indicators: 30 Second & scale modified of 30 Second 128
Fig. 5.13 - Book during lower resiliency 129
Fig. 5.14 - WTI Crude Oil & E-Mini S&P 500 Futures Price Events 131

LIST OF TABLES

Table 2.1 - Public Market Feed as seen by Market Participants 26
Table 4.1 - Participant categorization threshold criteria 85
Table 5.1 - Listing E-Mini S&P 500 market participation rates 110
Table 5.2 - Table of Events that occurred during the Flash Crash of May 6th 115

Chapter 1

Introduction

Today's data-driven, electronic financial markets, wherein highly automated and, often, high frequency trading has become the norm for major participants, present significant technical challenges that complicate effective monitoring of their most fundamental functions, trade and the price discovery. Though regulators are implementing new surveillance systems capable of capturing all necessary market data to describe the details of market events, the fact that markets are fundamentally "complex systems" often undermines the use of conventional assessment techniques. The research work underlying this dissertation provides new approaches for analyzing financial markets that can help overcome these problems and demonstrates how each can be used to better understand the state of the market's trading and price discovery processes. These approaches utilize adaptations of methods employing data visualizations, Markov State modeling, and agent based simulations that have been more commonly used in other fields.

The occurrence of abnormal pricing anomalies in today's complex markets has particularly intensified the demand for better monitoring and assessment, as it is crucial for regulators to be able to work to preserve orderliness in markets and, thus, build confidence for both business and the public. These anomalies are transient, unique events that suddenly, and unexpectedly, take place in markets which are fundamentally complex systems environments with large numbers of participants each having their own, individual, ever-adapting views and decision-making processes. Much of the research effort discussed in this dissertation is aimed at improving the capabilities of financial regulators and exchanges to better understand the precursors of such abnormal events and approach their investigation using methods that help overcome the effects of systematic complexities inherent in today's data intensive markets.

For example, on May 6^{th}, 2010, world financial markets suffered a pricing event that lead to a sudden drop in the price of thousands of stocks, bonds, and commodities over just a few minutes; at the bottom of this slide, assets averaged a 7% loss before recovering (CFTC, 2010). This event, known as 'The Flash Crash', was investigated by regulators, who eventually identified that it had been initiated by a collapse in price of E-Mini S&P 500 future, an asset viewed by many as an indicator of the overall health of the financial system. Though regulators, exchanges and market participants investigated The Flash Crash's causes and found the potential inceptor, it became clear that there was no straight forward way to effectively investigate and fully understand the cause of such a systemic event.

Some of the confusion that came as a result of this event helped highlight that the nature of effective methods needed for policing markets had changed. Today, a regulator can no longer simply walk the floors of the exchanges looking for misbehaviors in trading, or rely on personal requests to traders to help self-police. Immense amounts of electronic data and computational analysis are required to investigate a market event; the resources and capability for this are thought not to be adequately available (Bethel, 2011) and responsible for paralyzing regulators' best efforts. The Dodd-Frank Wall Street Reform and Consumer Protection Act has, in many ways, tried to address such issues by actively requiring that technical resources and computational capability be developed to collect, store, and analyze today's data. However, developing the analytical methodologies required for sorting through and translating available market data into reliable and actionable information has only begun to be addressed.

This dissertation's research, through examination of financial markets with recognition that they are "complex systems", suggests that there are opportunities to begin to untangle the individual events that occur in the markets, as well as, consider how participants' trading strategies may lead to systemic market and cross-market, events. By applying methods that are well-suited for the study complex systems, it is possible to not only learn how to analyze markets post an anomaly event, but, also, how they behave peri- and pre- such events.

1.1 FINANCIAL MARKET EVOLUTION

A key function of financial market exchanges is to provide a system that organizes the fluxes of information related to a financial asset's nature and condition, through which the asset's price can be discovered. They provide a structured and centralized system for trading in a generally efficient manner which serves to resolve the complex coordination problem of interactions between market participants. The structuring of this complex coordination serves to promote an orderly price discovery process which has typically been characterized in terms of the micro-structural elements of the market (e.g. price tick size, order types).

The mechanism for price discovery for most financial markets, until nearly the end of the 20th century, was known as the "open outcry" system. Traders, people trading for themselves and on the behalf of clients, were "physically present at a central location, publicly announcing their bids and offers. If a trader finds a bid or offer attractive, the trader simply hits the bid or takes the offer. The transaction price is then reported to the entire trading community." (Massimb, 1994)

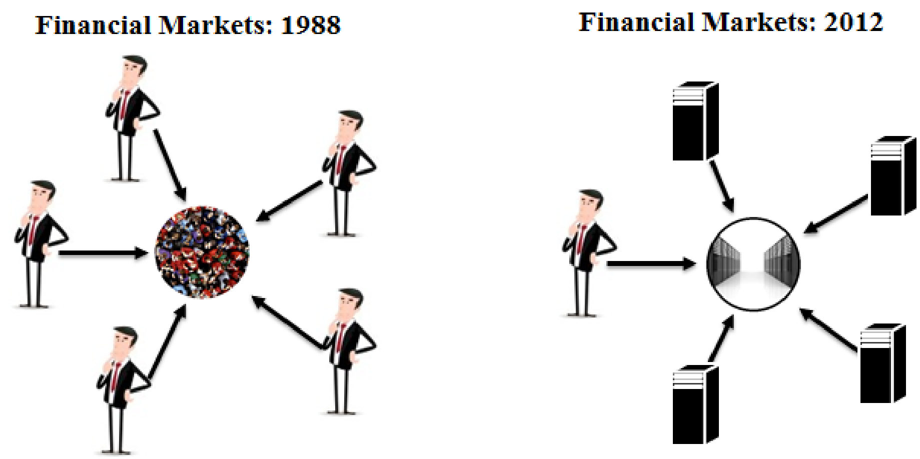


Figure 1.1: The transition of markets over the last 25 years, from human to mostly computer-driven auction matching systems and participants

As illustrated in Figure 1.1, over roughly the past quarter century, a major transition has occurred in the mechanism managing the interactions of market participants. Taking advantage of technological advancements, a new electronic, continuous auction system, known as the "electronic order book", has replace the manually managed open outcry system wherein floor traders directly made contact with one another in a trading pit. This change resulted in many improvements to basic functions of markets: the ability to match buyers and sellers became more rapid, the cost per transaction decreased, and the overall population and participation level of traders increased as the need to maintain physical presence was eliminated.

Though increases in trading volumes has been a trend for some time, the effects of broad market computerization has undoubtedly contributed to the nearly exponential rise in trading volume in the early 2000's, as illustrated in Figure 1.2 depicting trading volume in the major US equities and futures exchanges since 1993. These increases in volume seem to be far greater than simply what is needed to meet the demand of a growing population of traditional long term investors, who primarily trade to rebalance and hedge their portfolios risks (Grossman, 1991). As such the theory of rational markets fails to explain these increase in volume (Glaser, 2008) and questioning if this higher than expected volume levels of trade are really necessary to support effective price discovery.

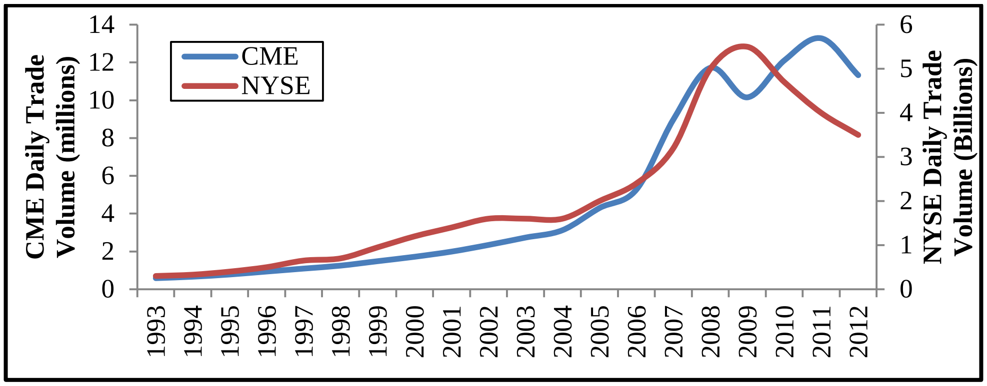


Figure 1.2: Average Daily Trade Volume seen in the New York Stock Exchange (NYSE) and the Chicago Mercantile Exchange (CME) from 1993 to 2012

The dramatic increases in the speed of transactions coupled with increases in the amount of instantaneous information regarding open interest (both produced and captured) are now widely available to most market participants; hence, much more trading can be done by more participants at higher frequencies during a trading day. Beyond this, more sophisticated subgroups of market participants have been able to build upon these trading technologies and develop associated strategies that attempt to maximize new opportunities that may arise purely from the market’s greater responsiveness and higher volumes.

Through the use of proprietary, automated trading systems (ATS), the large volumes of market data now created through the new electronic order book system can be quickly analyzed and interpreted utilizing the vastly increased computational power which is now widely available. Also, by using low latency time connectivity to markets, these interpretations can be applied very rapidly in making many trading decisions, almost immediately. These systems have been shown to be highly profitable, making millions of dollars a month (Baron, 2012). ATS's now make up a significant amount of total participation in the market, accounting for upwards of 70% of all trades (Brogaard, 2010) in equities and commodity markets.

Due to the developments and changes in both the environment (i.e. the rise of data-intensive, computerized exchanges) and the above related activities in trading, it is not surprising that the price discovery process has grown to be a far more complex. The fundamental ecology of markets to arrive at price (i.e. the manner in which market participants interact within the environment with which they are now presented by the markets) has shifted, and, as a result, new and surprising behaviors are now frequently being observed in the manner that market prices can unexpectedly fluctuate.

The NANEX group has noted that the number of extreme and rapid price anomalies appearing in US equities markets (i.e. similar to the large May 6th 2010 Flash Crash but on a smaller scale) has increased significantly (refer Figure 1.3). Though identifiable events have decreased in somewhat in the last few years, it has been estimated that about a dozen such anomalies happen on an average trading day in US equities (Farrell, 2013).

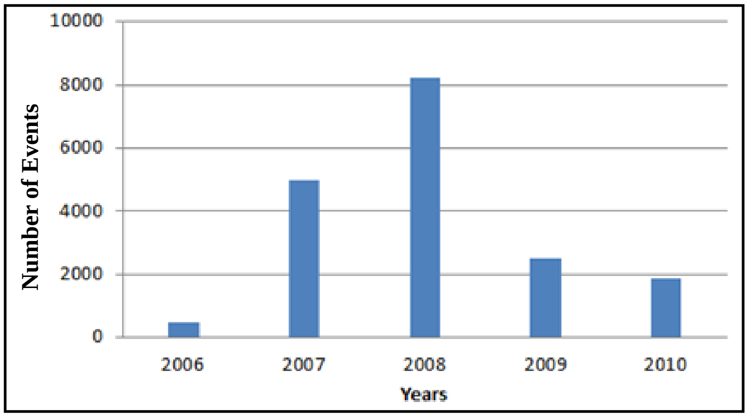


Figure 1.3: Frequency of Extreme and Rapid Price Events in U. S. Equities

Just in the last three years, there have been a number of more notable examples of major pricing anomalies that disrupted the market from a few minutes to a couple hours:

- May 6th 2010: The Flash Crash: Stock and commodities fell 7% in over 5 minutes.
- July 8th 2011: Natural Gas futures: Oscillating algorithm causes price to drop 8%
- August 2nd 2011: Earthlink: Market maker created fleeting orders to catch slower participant orders
- March 26th, 2012: BATS IPO: $15.25 down to a few pennies in 900 ms
- August 1st 2012: Knight Capital affects 148 equities in glitch incident

While the impact of such events is not always as noticeable to many individual long term investors, they have raised concerns for large investments firms who trade daily and continuously need to manage their risk exposures.

The capability to determine the causes of pricing anomaly events like these has been lost, in part, as a result of the automated matching mechanism's lack of complete transparency to market participants; they can no longer directly and openly just observe each other's actions on a physical trading floor. On the other hand, however, the use of electronic order books has allowed for better record keeping, making it possible to electronically capture all the details of the events that occur in markets with a higher degree of accuracy and fidelity than previously possible in the open outcry market. One consequence of this fully electronically documented system has been that free access to the full scope of all details of market transactions has been limited. Hence, the burden of analyzing and discovering the causes of price anomalies as they continue to occur has fallen primarily on exchanges and market regulators, who uniquely have the necessary level of access to these records.

1.2 FINANCIAL REGULATORY GOALS

Transforming available data into understandable and actionable information has become a focus for many different groups in both the private and public sectors. However, financial regulators, in particular, now face the biggest challenge, as they have responsibility for managing and interpreting the scope of all available data for oversight, enforcement, and market stability research purposes; the scale of effort that this demands today is, at least, several orders of magnitude greater than that of only a decade ago. With this in mind, this section provides a brief overview of some of the specific aspects of the goals of regulators as background perspective for discussions in later chapters of this dissertation.

Regulators tasked with market oversight must monitor the health of markets and work to identify those periods when it has been compromised. The inability to keep pace in real time with speedily evolving markets and interconnected financial systems has made oversight more focused on successful retrospective analysis, rather than trying to keep up with monitoring current conditions or looking ahead to the future problems being that might be taking root today.

The most volatile events, such as instances similar to the “Flash Crash,” are examples of timeframes when market integrity and orderly execution failed at a dramatic level and require extra attention. In cases such as these, the role of the market regulator would ideally be to help manage the event’s impact effects as they occur, quickly determine the proximate and ultimate causes of the resultant instability, and, finally, develop ways to help prevent similar occurrences in the future.

Markets have rules and regulation that are meant to protect both individuals and the integrity of the entire system as a whole. This requires enforcement through the regulatory actions that promote fairness for/among participants. Enforcement actions to this end can come in different forms, but all require reliable violation detection to enable enforcement steps to occur. To achieve this, it is often important to study the behavior of an individual entity, or set of entities, that have chosen to collude. While today’s technology allows for the collection of historical data for cases, it also provides more avenues for disguise illegal behavior in the complexity of market. Detecting this type of behavior by analyzing market data requires particularly effective investigatory tools followed by clear communication of findings to an adjudicatory body like a court of law.

A key outcome of a successful investigation is to clearly show that identified individuals acted in an exceptional manner, and that those exceptional actions should be considered illegal activity. Behavior, whether illegal or not, can be categorized as exceptional if it deviates from traditional market accepted practices; hence, it is useful 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 those of the rest of the market.

Demonstrating that exceptional behaviors are occurring, however, can be difficult because it may require detailed analysis of the dynamics of orders as they are entered, modified, and cancelled, in order to clearly portray how the intentions of a market participant(s) violate a regulation. The evidence must also show how an action had a negative impact consistent with the cause and effect relationship that the regulation was intended to prevent. The ability to demonstrate both of these is fundamental to successfully undertaking follow-on enforcement steps.

For regulators, the continually changing nature of the markets requires research and research expertise to continually update the fundamental understanding of these systems for the purpose maintaining meaningful policies and regulations. Recent developments include the introduction of dark pools (Hendershott, 2005), 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.

Beyond retrospective inquiry, regulatory experts and researchers are interested in the effects of such developments going forward in time. Achieving this goal demands the creation of analytical techniques that can be used to consider possible future outcomes. This includes the use of simulations capable of providing insight into the likely potential outcomes and future concerns to help shape policy directions and decisions.

In summary, regardless of the specific task at hand or timeframe of interest, regulators, to a greater or lesser degree, need to be able to gather and analyze of accurate, reliable, and replicable information regarding market conditions proximate to the periods of interest. Today, this means that regulators must have tools that enable sorting and analysis of meaningful data/data flows from many market data feeds that are uniquely available to regulators and pinpointing techniques that identify the true, expected, or potential root causes of past, present and future market problems /failures.

1.3 FINANCIAL MARKETS AS COMPLEX SYSTEMS

As noted by economist and Nobel laureate Friedrich Hayek (1952) noted, financial markets are an example of an evolutionary solution to human reasoning which enables people to solve problems, such as price, supply, and demand, which is otherwise impossible to do by direct rational calculation. The ability of a financial market to produce such attributes as price is the result of a system of:

i. agents (i.e. market participants), which are heterogeneous in nature
ii. interaction rules (i.e. the exchange) by which they trade with each other
iii. an environment having conditions that impact individual decisions of agents, but are outside of their control (e.g. regulations and cross market relationships).

Such systems are generally termed "complex systems" by physical, social, and engineering sciences to describe a system wherein the properties of the underlying components that dictate the outcome of the system's processes cannot be separately and individually examined to predict an outcome.

Complex systems typically include some, if not all, the following features that are also seen in financial systems:

i. *Non-stationarity*: It cannot be assumed that the dynamic or statistical properties observed in the system's past, will remain unchanged in system's future.

ii. *Agents*: The elements that make up a system which can interact and respond to information create a sufficiently complex system that has a "whole" response which is not just a reflection of the responses of its elements side by side.

iii. *Open system*: The system is coupled to the environment; hence it is hard to distinguish between exogenous and endogenous effects.

iv. *Self-organization:* The internal structure of the system evolves without external intervention.

v. *Feedback:* The use of information within the system that is remembered from the past can influence the similar phenomenon in the future, causing a chain of cause-and-effect that forms a circuit or loop.

Some of the earliest references to complex systems theory are found in the 18th century political economy writings of the Scottish Enlightenment. While the market system was first described as both a spontaneous or emergent tool and the result of human action, it was not seen as the execution of a human design. This early observation, well known through Adam Smith's metaphor of the "invisible hand", premises a disjunction between a system wide outcome and the design capabilities of individuals at a micro level, where there is absence of an external organizing force. (Hodgson, 1993)

More recently, as financial markets were formalized by exchanges and given structure via rules governing the manner in which trade occurred, the metaphor of an "ecological system" began to be used to describe the inter-relationships of financial agents with each other and their environment. "The interactions of financial agents are strangely mediated through a single variable, the price, which forms an important part of their environmental. Each trading strategy influences the price, and the price in turn influences each trading strategy." (Farmer, 2002)

In a chapter entitled "The Ecology of Markets", Niederhoffer (1997) likened financial markets to an ecosystem consisting of dealers as "herbivores", speculators as "carnivores", and floor traders and distressed investors as "decomposers". In his view of this complex set of "webs and multifarious habitats" of agents, his ideal vantage point would have been to use "a satellite … including X-ray, a time machine, and ESP." However, in today's markets, that are mostly electronic, the "ecology" no longer requires such technology to observe interactions, webs, and habitats as transactional data is available in real time and continuously recorded by exchanges such that it can be analyzed.

This process is, however, not as straight forward as it may sound. While information may be completely accessible, the ability to effectively select and utilize data in a manner that allows users to synthesize a clear picture of what is occurring in the markets is relatively weak and incomplete. Untangling the information that lies within financial market data requires understanding the nature and extent of the complex system properties inherent in the data, and their effect on being able to successfully complete an intended data analysis task. In Chapter 2, the structures and types of data produced as a result of the financial market systems' complex processes are broken down into their content and uses.

1.4 ENGINEERING WITH COMPLEXITY

In general, engineering methods support the design of systems that can logically achieve a desired objective by meeting a set of functional requirements. The system itself consists of a set of interconnected elements which function together in achieving this result. The system becomes ‘complex’ when the function of an individual element is not fixed and is dynamically dependent on how the other elements function.

The term ‘complex’ is often considered a synonym of ‘complicated’. However, there is a precise distinction that must be made between the two in terms of the interdependencies found in complex systems’ components. Complex systems have components whose individual behaviors are dependent on their interactions with other components in the system. On the other hand, components of complicated systems, behave independently of each other. In other words, the behavior of an individual component within a ‘complex system’ must also consider the influences all the other components in the system may have on it.

Understanding the causality, and, thus, being able to predict a complex system’s overall behavior (output) given initial conditions (input) is very difficult and requires the ability to explain how the overall behavior is produced from the collective behavior of its components. This can be done by making complexity, itself, a qualitative property described by several measurable or observable attributes and, ultimately, requires a systematic process of reduction so as to make it as controllable as possible. (Lee, 2003) In Figure 1.4, Lee demonstrates the process of breaking down a system into a set of measurable components that can be used use to determine the complexity of a system.

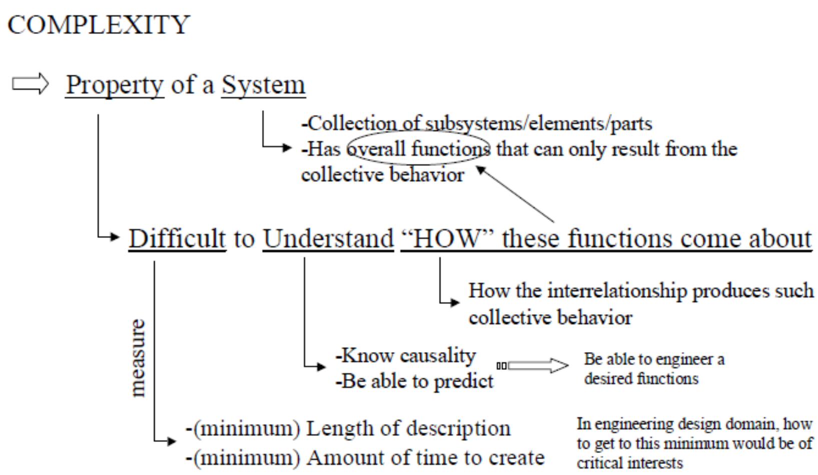


Figure 1.4: General complexity concept in context of engineering a system (Lee, 2003)

The process of determining the complexity of a system requires defining the uncertainties in the relationship of the functional requirements (FR) of the system and design parameters (DP) meant to achieve the requirements. In the case of financial markets, this means defining the relationship of effective price formation and the self-organizing behavior characteristics of market participants and their capability to control price formation as a collective system.

Complexity must also be viewed as a function of its relationship to time. It can be classified into: *time-independent complexity* and *time-dependent complexity*. *Time-independent complexity* further breaks down into two different kinds of sub-complexities:

i. *real complexity* - the uncertainties inherent in the system's design

ii. *imaginary complexity* - the uncertainties caused by lack of design knowledge, ignorance

These complexities maybe seen as the uncertainty that is a result of either a poor design of engineered systems or not understanding the system. When there are many FR that a system must satisfy at the same time, the quality of design in terms of the independence of FR affects the uncertainty of satisfying the FR (Suh, 2005). In trying to cope with meeting several FR's, market participants can get overly focused on satisfying a single FR under certain conditions and cause the entire system to malfunction as a result.

Time-dependent complexity can also be further broken down into two different kinds of real complexities:

i. *combinatorial complexity* - the uncertainty grows indefinitely more complicated as a function of time because the future events depend on the decisions made in the past

ii. *periodic complexity* - a periodicity in the systems that allows the system to regain its initial state

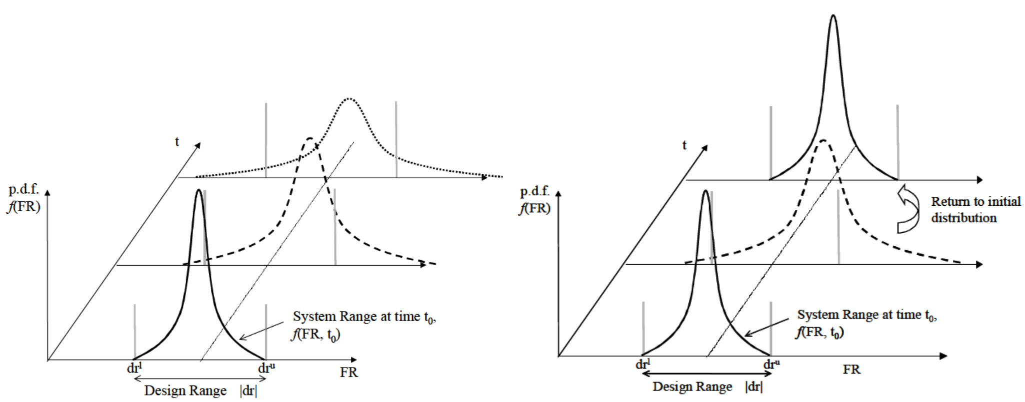


Figure 1.5: Combinatorial Complexity (left) and Periodic Complexity (right) in Design Range of a system (Lee, 2003)

Examples of how these two types of real complexities are reflected in how a system functions are depicted in Figure 1.5, where we can see how the designed range of a system changes over time. As combinatorial complexity grows, the design range will grow, as well, forcing the system's output to eventually go beyond its desired limits.

Periodic complexity will eventually return a system back to its original state, but this may not occur before the system's output goes beyond its desired limits.

In examining financial markets, it is critical to develop an understanding of how these types of complexity can each play a role in both the overall market's and the participants' underlying designs such that it enables us to identify where complexities lie and how they might be reduced. In principle, most well designed and engineered systems should have neither time-independent complexity nor combinatorial complexity. Markets systems differ in this aspect from most engineered systems, in that, by necessity, they do require complexity to self-organize to discover prices such that no single participant can monopolize the process.

Being able to handle complexity in the work process of analyzing data and behaviors is vital to the task of untangling financial markets. With proper consideration of complex systems engineering concepts, regulators can become better equipped to answer questions about the past, present and future aspects of events of interest that occur in the markets.

1.5 CONTROLLING FOR COMPLEXITY

In examining the behavior of financial markets as complex systems, it is necessary to consider what information is available to identify the presence of the types of complexity that could affect the specific question one is attempting to answer using available data. In answering, regulators need to first identify the levels of activity on which their investigations they must be focused. Figure 1.6 depicts the various levels that a regulatory question may address initially.

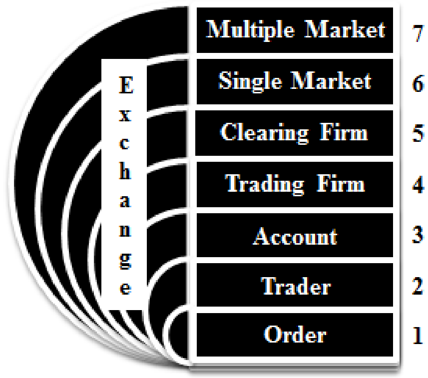


Figure 1.6: Levels of Investigation

As is the case with analyzing a complex data set, the level of investigation typically will likely need to grow or shrink, as possible answers or conclusions are proposed. This makes it very difficult for investigators, since investigation can easily end up going down a host of worm holes looking for answers. However, a regulator generally can identify the initial question and describe the event for which they need to develop answers. This is important since it can help dictate the type of analytical method to be used such that the presence of complexities does not overwhelm an investigation. As a result, a key objective of this research is to suggest methods that help investigators and researchers approach investigational challenges in ways that comprehend or work around the effects of complexity.

1.5.1 Visualization and Imaginary Complexity

Retrospective analysis of behavior in financial markets has traditionally focused on examining data related to a set of consummated actions within a market environment, such as completed trades and the resulting inventory held by a participant. Recent improvements in the regulatory audit trails available from exchanges now allow a far more complete and information-rich reconstruction of events related to key market elements such as the order book. Basic data inspection methods (such the number of trades in a day, average order size, or daily change in inventory), do not readily reveal the systemic effects and causal relationships that orders and trades have on markets. Hence, a more meaningful approach to analyzing the order flow dataset is needed, especially for institutions working with market data that are not experienced in to “big data” processing. This can be a major challenge simply due to the presence of the imaginary complexity that comes when facing, particularly for the first time, a large scale data set with numerous interconnected systems producing it.

In many fields, data visualizations have proven to be an invaluable tool for building intuition and enabling exploratory data analysis. Using data visualization techniques for retrieving, analyzing, and disseminating data, regulators can access the tools needed to tackle the cumbersome task of examining the complex structure of large data sets relevant to their regulatory tasks that review past events. 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). Chapter 3 provides a detailed discussion of how data visualizations can be incorporated into the workflow of multiple financial regulatory roles.

1.5.2 Market Monitoring and Periodic Complexity

As markets have gone electronic, the ability to monitor their function has deteriorated due to a lack of a floor perceptive of participants’ trading activities and a significant increase in trade due to automated trading. It has been especially difficult to gauge a markets functional ability to perform price discovery, the process by which changes in demand and supply cause price to change as new equilibriums are found. As it occurs, a periodic complexity (i.e. a set of market dynamics) usually allows the new price to be obtained in a stable fashion.

In Chapter 4, the specifics of a monitoring approach for tracking the passage of assets' transactions through the market are discussed. Such an approach can provide regulators with the ability to understand when a change is occurring within the current dynamics of a market. By being able to monitor for systemic changes in behavior, regulators can examine the system for factors that may be driving it into a non-periodic state of instability in pricing, wherein price discovery could become erratic and volatile.

1.5.3 Agent-based Simulation and Combinatorial Complexity

Through the use of agent-based simulation models (ABS) together with real market data, it is possible to investigate the behaviors that lead to market pricing events. An ABS has a structure, which is defined by a set of agents, a topology and an environment, that can readily provide a framework conducive to the creation of a simplified financial market simulation that can be used to approximate a real market. In the ABS of a market, the market participants are agents, the market mechanism is the topology, and the exogenous flow of information into the market is the environment. By having a replicable model of the market, the combinatorial complex behaviors that occur on a financial exchange can be explored.

Using an appropriately defined ABS for a given market, it is possible to use the characteristics of real past events of interest to study the potential for a future event and consider related sensitivity cases affecting the likelihood of such an event in that market. The effect of changes in the assumed characterization of agents, topology, and/or environment can also be used to facilitate policy debates or inform market stakeholders about the types of data they should be interested in collecting. Through simulation, the presence of combinatorial complexity can be assessed by preforming Monte Carlo testing to identify if a specific behaviors (inputs) are responsible for certain pricing characteristics or events (output).

Chapter 5 examines how an ABS can be built such that it replicates financial market functionality and behaviors. These tools can be used to replicate events in order to better understand their causes and, thus, also be used to understand if past circumstances and causes can be used as are reliably signal of potential future events. The ABS can provide in-depth, simulated datasets, whose availability would normally be restricted to exchanges and regulators, and the ability to reproduce numerous events for sensitivity analysis. Furthermore, it can be an effective means for regulators to assess and help explain the effects of new regulations in advance of their introduction in order to demonstrate their costs and benefits.

1.6 SUMMARY OF RESEARCH

Over the past two decades, the magnitude and nature of financial markets has changed quite significantly, as they and other systems (e.g. merchandise retailing) have transitioned from being physical venues where interpersonal communications and interactions occur in public to anonymous, computer-based interactions occurring on electronic platforms. For financial markets, these changes have resulted in significant challenges and opportunities, as well as many new considerations, for the regulators who now are tasked to solve problems whose solutions require the use big data analytics.

This research is intended as a first step at helping financial regulators better understand, and approach the investigation of events and decode the complexities inherent in electronic financial markets. It demonstrates that through the recognition of financial markets as fundamentally complex systems, there exist opportunities to use data visualization techniques, Markov State based monitoring, and agent-based simulation to more effectively assess and analyze these systems.

Chapter 2

Overview of Market Data Structure

This chapter is intended to summarize the structure, scope/magnitude, and communication of market data today for electronic order book markets. This, coupled with the increased number of market participants, the wider set of financial instruments traded, and the use of automated trading systems, drives the massive market data communications requirements needed for the operation of today's financial markets.

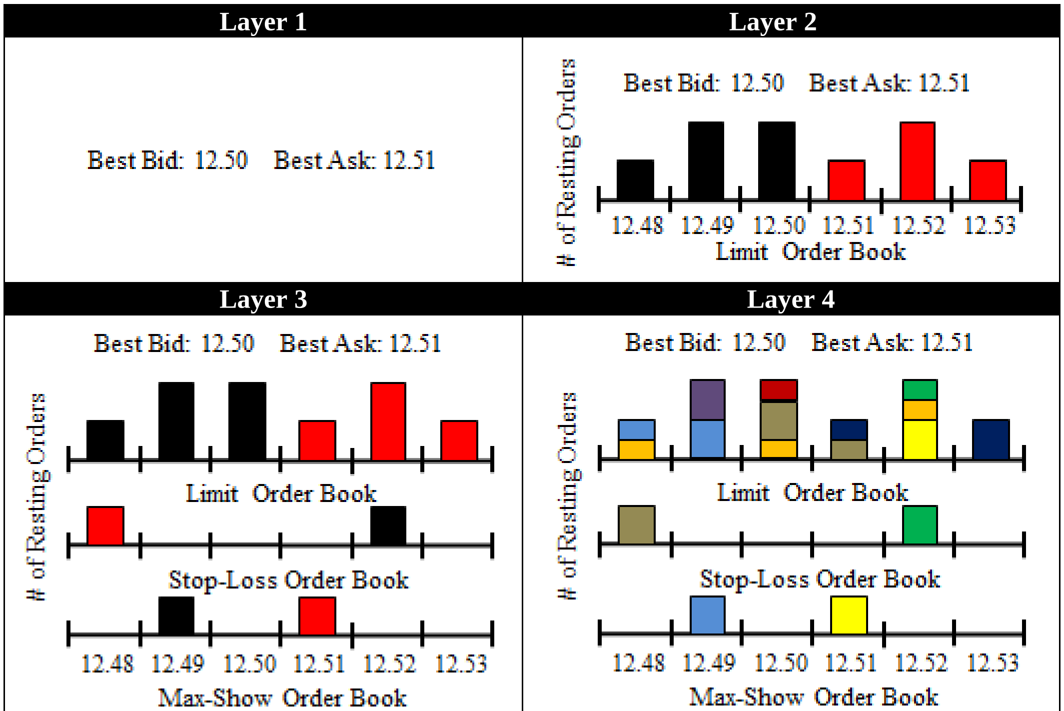
2.1 ELECTRONIC ORDER BOOK STRUCTURE

The electronic order book is modeled as a publicly visible screen providing bids and offers, each of which specifies a price and a quantity of an order (Glosten, 1994). A transaction occurs when an order is entered into the order book that removes an order from the book. This typically is done using a first in first out system based on a price-time priority rule such that orders are queued at price points and cleared from the order book based on being the highest bid price or lowest offer price. Generally order books are filled with mostly limit orders, however additional order types, like stop-loss and max-show[1] orders exist as well that follow additional rules for execution.

This generalized electronic order book structure is used by securities and financial exchanges has different levels of privacy and informational content. These levels are

[1] A stop-loss order is an order placed with an exchange designed to buy or sell a security once a defined price threshold is reached. A max-show order, also known as an "iceberg order", is a large single order that only shows a fraction of its total quantity to the market, for the purpose of hiding the actual order quantity.

often broken down into four distinct layers that in combination show the entire picture of the state of a market for a financial instrument at any moment of time, as illustrated in Figure 2.1.



Layer	Type	Data
1	Price Level	Price: Best Bid/Ask
2	Public Market Structure	Ten Deep Limit Order Book Depth Executed Trade Order Flow
3	Private Market Structure	New, Modification, & Cancelation Order Flow Stop-Loss Order Book Max-Show Order Book
4	Account Level	Account Level Order Flow Account Level Order Book

Figure 2.1: The Layer of the Financial Market Data

Layers 1 and 2, also known as 'level 1 and 2' data, typically are available on public market feeds that allow participants to examine price, trade and market depth data which are the result of aggregated order flow communications between an exchange and all individual participants of a market. The distinction between the layers 1 and 2 is that the second layer provides the participant some underlying knowledge as to the state of the market when making a decision to trade, rather than simply utilizing current price as a single reference point.

Layers 3 and 4 are private market data available to exchanges and regulators which can be used to verify if rules and processes are being complied with by both the market participants and the exchanges themselves. Layer 3 provides a full picture of a market's aggregate state by including all order flow and order book data regardless of order type (i.e. marketable, limit, stop-loss, max-show, etc.); this makes it possible to comprehensively track and account for all events that occur in a market. Finally, layer 4 provides full details of orders by individual accounts allowing users to group together and study the actions of specific individual traders.

While the data layers that build-up the picture of the market are relatively simple, they can be complicated to follow as a whole. There are at least 10 million contracts traded per day within the CME Group exchanges alone which result from an order of magnitude larger incidence of overall order book activity (CME Group, 2013). With such quantities of data, users to different degrees including participants, exchanges, and regulators need to be able to pinpoint and summarize the activity of a market or a select a subset(s) of participants easily and quickly when markets begin to behave oddly. However, there is little experience and capability in the industry for analyzing and creating good summary indicators that can provide a useful understanding of the situation before and during market price discovery anomalies. Beyond this, monitoring the activity of all individual participants is a particularly monumental task.

2.2 MARKET DATA COMMUNICATION

Exchanges need to be able to communicate data quickly, simply, and securely with market participants and regulators located in disparate geographical regions. Given the sensitivity of the data, these communication channels are often divided into public and private data feeds. Order flow data between individual participants and exchanges is transmitted through single, private information channels, whereas public feeds such as updates to data to the order book, are sent equally to all connected market participants. The following discussion describes this diverse set of market data systems.

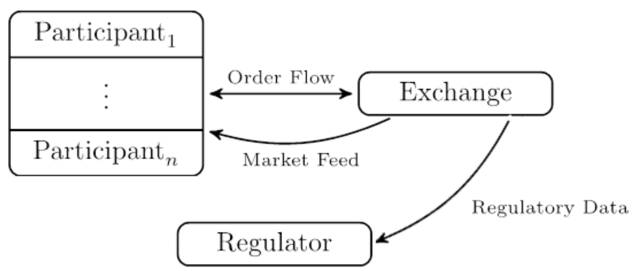


Figure 2.2: Data Flow between Market Stakeholders

Order flow data is the aggregation of bidirectional private communications between individual participants and a financial exchange and is originated by the exchange. Order flow data consists of all messages between individual participants to/from exchanges including 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.[2] These messages make it possible for the exchange to fairly execute trades using its matching engine within the electronic order book; the matching engine follows a set of publicly known prioritization rules with which it identifies appropriate trade counterparties.

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

In order to efficiently transmit the high-volume of information required, roughly 80-100 million messages per day (CME Group, 2013), the exchange is aided by a messaging format designed specifically for financial data, the Financial Information Exchange, or FIX Protocol. First introduced in 1992, it has come to be the standard communication language between exchanges and market participants. Through its flexible structure and continual updates and extensions, FIX Protocol, has been able to successfully build in the necessary fields required for communicating both complicated order types and new financial instruments such that the flow of orders and the software required to process them has only gotten faster.

Distinct from the private feeds that make up the individual parts of order flow data, there exist public aggregations of these private communications. These order aggregations are transformed into market feeds and broadcast to market participants. These feeds traditionally include the price, trade, and market depth data that result from net order flow and trade executions. Table 2.1 provides a sample of Layer 2 data that might be provided to all market participants in such a data feed (i.e. in this example, the top ten best bid and ask prices and quantities); it conveys this in a more easily digestible, and commonly used, format. Between private and public feeds, market participants must manage their individual private communication with the exchange with respect to the public aggregated exchange data available such that they consider their notional contribution to the market.

Table 2.1: 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, to enable regulators to verify that rules and processes are being followed by both market participants and exchanges regulatory data feeds with Layer 4 content 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 consisting of all the individual private data feeds. With such a super-sized 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 as broad and as granular, as possible, and contains unique identifiers that allow regulators to associate each action with its source entity.

As a result, the data set provided to regulators is extremely large and not easily parsed, stored, and managed. With over 10 million contracts traded per day in the CME Group exchanges alone, the overall order book messaging activity for all exchanges is likely to be well in excess of one billion order related messages per day. Because of the sheer size of the data set, regulators are often not able to fully analyze the 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.

Chapter 3

Data Visualizations and Imaginary Complexity

Recent improvements in the regulatory audit trails available from financial markets now allow a far more complete and accurate reconstruction of the order book; yet, regulators now face the task of creating meaningful detailed pictures of the market that not only provide traditional information, such as market depth, but, also, can depict the systemic influence of individual participants on the overall market. Processing and analyzing order flow data for order book reconstruction given the variety of order types and the intra-market capabilities offered by exchanges complicates this task. In addition, the nature of the data structures can be quite complex, and the volume of regulatory data is quite large (100 GB+ for a single day on one exchange).

As a result, the pictures of the markets ultimately portrayed based on data analysis are often very simplified depictions of reality in which important information can be lost or easily misrepresented. These simplifications come as a result of what has been termed as imaginary complexity (Lee, 2003), where a lack of comprehension how the underlying structures functions or poor depiction of a system can make the system appear to have complexity features where in fact, there are none.

The structure and rules required to reconstruct events exist, as does the computing power to accomplish that task in a timely manner. The challenge, then, is really to do technically valid and rapid syntheses of data from market events in a form that users can readily comprehend while minimizing any effects of imaginary complexity in describing the system. This requires technically well-founded tools that facilitate this task, as well as the subsequent dissemination of the clearly understandable results to relevant parties, such that an investigator can interpret them in a self-evident manner.

In this chapter, we propose to minimize the effects of imaginary complexity on data analysis tasks through data visualizations, which can be incorporated into the workflow of regulators engaged in a variety of activities. Visualizations can support a range of functions within the regulatory sphere, including: real-time and day-after market monitoring, reviews for regulatory enforcement, and, more abstractly, academic research supporting regulatory policy. Typically, good quality market data visualizations require careful and effective analysis of extremely large quantities of data (often in the order of millions of trades). Additionally, for those tasks that are focused on imminent risks to the market system, the results of the analysis might be required within constrained time periods. The visualizations must communicate the meaning of data accurately and in ways that are both information rich (e.g., from which non-trivial market connections may be inferred) and easily digestible (e.g., understandable within a required time frame or without overly burdensome prior knowledge).

This chapter begins by examining, in Sections 3.1 and 3.2, regulatory objectives and the details about structure of their financial data. Section 3.3 discusses the concepts underlying visualizations and their application to market data. Section 3.4 develops criteria and a set of visualization techniques that can support both exploratory data analysis and core regulatory tasks. Section 3.5 goes through several example cases in which market data visualizations have been developed to accomplish specific tasks. Finally, Section 3.6 provides a summary for this chapter.

3.1 REGULATORY DATA OBJECTIVES

With the explosion of market activity in recent years, financial exchange operators needed to adapt to a new environment in which the storage, management, and transmission of data are at the very forefront of their key activities. Technological advances have enabled exchanges to disseminate market information over high-speed networks to automated trading systems within milliseconds, with equally quick responses. 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. In particular, Financial regulators, who have access to the fullest possible scope of all available data, face the greatest challenge of managing and interpreting data at a scale much greater than ever before for oversight, enforcement, and research purposes.

Regulators tasked with market oversight must verify the integrity of market conditions and be able to identify those occasions where integrity has been compromised. Volatile events, such as instances similar to what is now known as the Flash Crash of May 6, 2010, are extreme examples of time frames where market integrity and orderly execution in markets have failed dramatically. In such cases, it is the role of market regulators to help avoid or prevent similar occurrences in the future, often by determining the proximate and ultimate causes of the underlying instability. Each of these tasks, to a greater or lesser degree, requires the construction of accurate, reliable, and replicable information regarding market conditions surrounding these periods. While market anomalies and other volatile events can be detected in multiple—and public—market feeds, determining the ultimate cause of such market failures often requires data that is only included within sets of private or limited availability feeds.

The initial steps in developing these causal links typically involve examining 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 this limited subset of information can reveal quite complex, “emergent” market behavior.

One demonstration of this utility can be seen through an analysis of the Flash Crash of May 6, 2010 (CFTC, 2010), where market depth sitting on the ask side was far greater than that of the bid side[1], nearly ten to one. Charts providing information regarding market depth can isolate the initiation of the occurrence and potential reasons for a subsequent price fall. A discussion of a flash crash–type event is given in Section 3.5, “Case Studies.”

Studying unexpected and anomalous behavior is of importance to market oversight, even when such behavior occurs within the context of practices generally accepted by the market. However, in other cases, traders may deliberately initiate such events, hoping to trigger a market reaction; often behavior of this type is considered illegal either by the market venue or its regulatory body. In a subsequent enforcement investigation, it is often important to study the behavior of the individual entity or set of entities that initiated the illegal behavior. To successfully conduct this type of investigation, regulators must establish two facts: that the identified individuals acted in an exceptional

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

manner and that those exceptional actions should be considered illegal activity. Behavior becomes exceptional if it deviates from traditional market-accepted practices, determined by comparing participants' behavioral characteristics with other market participants. In addition, this analysis may use a dynamic portrait of orders as they are entered and canceled, providing a window into the possible intentions and goals, in real time, of a given market actor.

In addition to the challenges above, market research is becoming increasingly important for regulation and policy makers. New types of market structures are rapidly developing within the financial environment. Recent changes include the introduction of dark pools (Hendershott, 2005), 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, involving millions of data points. We propose that the most effective means of comprehending this vast array of information need to involve visualizations that can provide succinct and unifying summary of the data. Such visualizations would also have to be responsive to highly sophisticated analytical techniques designed by experts and academic researchers.

3.2 STRUCTURE OF FINANCIAL REGULATORY DATA

In examining financial regulatory data, one first needs to understand its underlying structure. Simple features included within the data might be the time a new order was 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 to (or subtracts from) the price level in the queue, and how other market participants respond to that order entry.

An order flow message, a data type often used as a building block, typically arrives at a exchange's matching engine or market participant with little advance processing and minimal structure, allowing for quick, automated processing. 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 to perform simple aggregation analysis, such as identifying the most active users and markets within a given day or specified time period, or isolating accounts with aberrant modification or cancellation levels.

Order flow datasets, however, also provide a building block to construct higher dimensional structures; for example, these could include re-constructing an order book or determining the relative risk metrics associated with given traders. These blocks must be arranged in a manner such that higher order information can be obtained from the resultant structures. While the knowledge required to build these structures may appear complex at first, it only requires layering the structural rules of the system to effectively to realize the complexity is imaginary.

For a regulator, the first useful additions to this atomic data (i.e. the lowest level data structure) are often temporal, where an outside time reference point is imposed on the data stream, most commonly wall-clock time. Such an approach makes it possible to weight orders by trade time, by volume, or relative to the timing of another series of events. By incorporating a temporal dimension, a number of visualizations can 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. A “spatial” dimension can be given to the data, where orders are given coordinates associated with the logic, such as the price-time priority within the order book. This spatial dimension allows a user to track a given order as an element of the order book, thus, lowering the cognitive demand placed on an investigator’s memory such that they may focus on higher order questions about the relevance of an order to the markets pricing. Regulators can 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?
- Is this order visible in the public market feed data, and how might it impact the decisions of other firms/accounts?

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

3.3 VISUAL ANALYSIS CONCEPTS

3.3.1 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, producing imaginary complexity. This task is made more difficult when the data being examined is itself part of a much larger system (as with order flow data in financial markets), because verbal and statistical descriptions tend to be linear and one-dimensional, whereas large systems are, by their very nature, multidimensional. That is, individual activity and the system within which that activity occurs develop continuously; both the activity and the system evolve not just in one dimension, but in multiple dimensions simultaneously (Keim, 2002).

As an example, consider a simple trade execution on a market platform. Here a trade denotes a market participant account actively choosing to execute against a standing order sitting at the top of the order book (originally entered by another market participant). 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 served to move new orders to the top of the order queue. If price did in fact change during execution, the order might have also have triggered additional market liquidity contingent on the prevailing price point, such as stop orders. 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 the information set need to be updated simultaneously at the time of the trade. To depict all these types of information, it is necessary to use a language that shares some of the same properties as the phenomena under observation (Meadows, 2008). A visualization should embed important information while effectively communicating with, and capturing the attention of, the human visual system (Mackinlay, 1986).

3.3.2 Cognitive Processes Served by Visuals Analysis

Visualizations, by definition, are 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 both to help externalize the memory associated with the data, and to more closely represent a user's mental model of the data (Keim, 2008). The objective is to eliminate overloading the cognitive memory of a user who is trying to understand how a system functions. This efficiency frees the user's memory to support further cognitive operations or tasks. A simple example is a map for first-time visitors to a city which organizes and stores unfamiliar information in an easy-to-retrieve format; this allows visitors to spend time seeing attractions, rather than learning all of the city's street names, or even worse, getting lost.

For visualizations achieve this, their design and use must be appropriate to the task at hand. (Zhou, 1998; Chi, 2000). The objectives and tasks typically fall into one of three broad categories:

i. *Information Retrieval* explores the data space through overview, browsing, navigation, zooming, and observing derived quantities such as data ranges, distributions, and the errors, certainty, and sensitivity of those values. For spatial and temporal data sets, it involves inspecting features by viewing animated or sequential representations.

ii. *Information Analysis* serves as a method for gaining further insight, by fostering the constructive evaluation, correction, and rapid improvement of model and hypothesis, which in turn enhances decision-making. Figure 3.1 shows how the analytical process makes use of visualization and model construction to deliberate and build upon current knowledge. This includes a range of analytical tasks, such as identifying anomalies or correlations, and evaluating hypotheses.

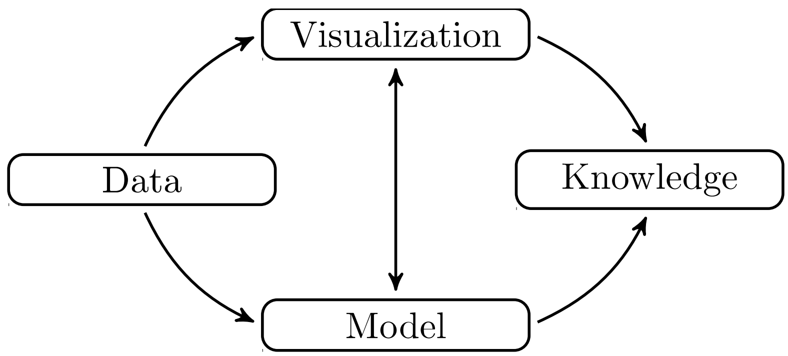


Figure 3.1: Model of Data Visualization Analysis (Chen, 2011)

iii. *Information Dissemination* involves presenting information to others, in the form of visual aids, so as to allow for easy data comprehension. In this case, the visualization should summarize, annotate, and illustrate analytics that support or reject a hypothesis.

3.4 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 about market responses. Analyzing this data thus offers regulators a promising basis from which to effectively combat unlawful activities and understand market anomalies. However, because firms engaged in abusive practices change their trading patterns as they learn about a new detection mechanism, regulators must identify and review new patterns through adaptive means (Blume, 2012). Therefore, while summary visualizations must be designed so that they provide story lines for market events, what makes them especially useful is their ability to stay flexible enough to adapt to changing market behavior and structure in ever-changing market environments.

3.4.1 Visualizations for Market Regulatory Data

In the following section, we cover five visualization techniques that can be used individually or collectively as part of an integrated process for satisfying regulatory needs. Each technique is meant to capture different aspects of structure, hierarchy, and information that exist in regulatory data (in this case using market-simulated data) (Paddrik, 2012).

3.4.1.1 Market Tree Maps

Tree maps are a recent design for organizing complex tree-structured data; in contrast to the traditional treeing approach, they use a two-dimensional space-filling approach where rectangles represent data "atoms" like individuals. The area of a given rectangle is proportional to a chosen attribute (Wattenberg, 1999). 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.

Traditional tree diagrams use root nodes, often at the top of the page and children nodes below the parent node (selection criteria), each with connecting lines. 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 ordering and, in the case of numerical data, appropriate group cutoffs. Nonfinancial 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 (Shneiderman, 1992). Even in these cases, only the structural relationship is shown; additional information, such as the size or importance of each node, is often ignored or included in a summary external to the visualization. Tree maps satisfy similar requirements within a more compact, and information-rich, space.

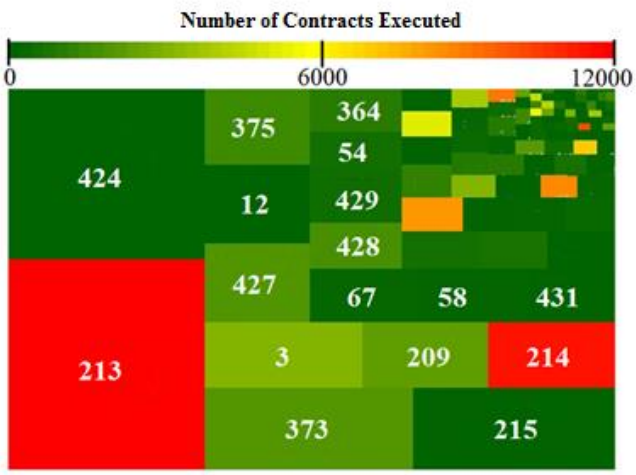


Figure 3.2: Tree Map of Market Participant Contribution to Order book Depth and Executions

Figure 3.2, above, is an example of a tree map that depicts the contribution individual accounts make to overall liquidity for the period of an hour in a fictional market. The size of a node (individual rectangles in the diagram) denotes the average relative quantity of contracts that each trader offers to buy and sell in this market. This order depth, which can be viewed as a proxy for the level liquidity an individual brings to the market, is an important variable in measuring the effects a given participant has on the market (Harris, 2002).

The importance of liquidity to an individual participant can also be measured using the number of trades the participant makes, denoted by the color scale at the top of the figure. This representation gives us a better picture as to who in the market is participating in liquidity provision most often (and so may cause important changes to market quality if their behavior changes). 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 inability to organize data points, or individuals, into related groups.

3.4.1.2 Classification Scatterplot/Clustering

Regulators are tasked with the objective of trying to create policy to help ensure markets are fair and transparent, which, in part, requires a basic understanding of their stakeholders (market participants). Often this means that regulators must develop an understanding of how regulations may affect actions and behaviors of certain classes of market participants, with potential advantages for some, and disadvantages for others.

Whenever organization and classification is the objective of a chart, the most common and effective graphic is the scatterplot. It is the predominate graphing tool used in the physical, biological, and social sciences, making up an estimated 75% of the graphs used in science (Tufte, 1983). A part of its usefulness lies in the fact that it may stimulate the production of more variations of plotted variables or guide the choice of a more rigorous model (Spence, 1993). Linear or nonlinear relations are easy to discern from groups of data points; moreover, the human eye is keenly aware of outliers and other aberrations in the data (Spence, 1990).

This chart type can support financial analysis by classifying market participants according to objectives and trading behavior. A notable example includes the investigation of the May 6, 2010, Flash Crash conducted by the Commodities Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC); this investigation focused largely on understanding causal chains and responsibility (Kirilenko, 2011). As shown in Figure 3.3, trading volume and end- of-day trading positions were used to classify traders into five groups based on clustering. This process allowed investigators to determine what cut of data points they might use to investigate traders considered “high frequency.”

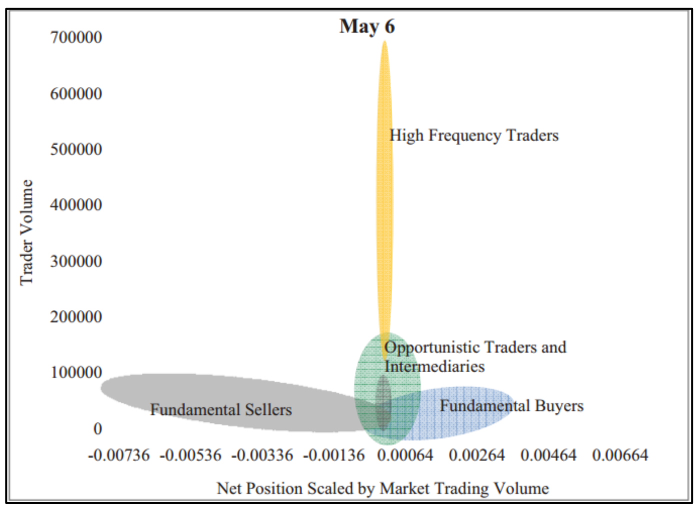


Figure 3.3: Classification of Traders in E-Mini Futures Market (Kirilenko, 2011)

Clustering techniques, however, often have disadvantages, largely due to the fact that they tend to rely on two-dimensional representations. As a result, depending on the data set, there may be a significant amount of overlap between groups, making it difficult to effectively classify them. At times, this problem can be solved by adding additional characteristics or dimensions. 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. The added 3rd dimension of Figure 3.4 illustrates the layered clustering of opportunistic and intermediary traders are no longer mixed together as they were in the two-dimensional example shown in Figure 3.3.

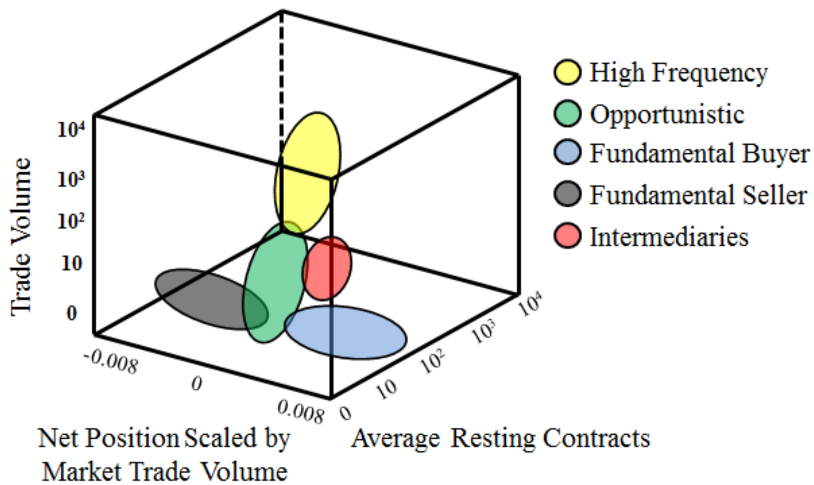


Figure 3.4: Example of Possible Classification of Traders in E-Mini Futures Market

3.4.1.3 Order Book Heat Map

A heat map is constructed using rectangular tiling of a data matrix; by using a row-and-column structure, the heat map facilitates inspection of three-dimensional data, showing the transition of a third attribute as it changes in respect to the other two. This allows large data matrices (several thousand rows and columns) to be displayed effectively on high-resolution color images (Wilkinson, 2009).

Heat maps such as the one shown in Figure 3.5 are constructed to depict the state of financial markets over many years as a method of monitoring the prices of sectors over time (Blanchard, 2009). For this application, the rectangular tiling of a data matrix facilitates the expression of spatial-temporal data by using the price of a group of assets and time to structurally describe the transition of a set of a markets using the colors to indicate a growing (green), stagnant (yellow), or declining (red) market period. This allows the heat map to apply the spatial relationship to price such that it organizes information, facilitate memory, and empower spatial inference (Larkin, 1987: Tversky, 1995).

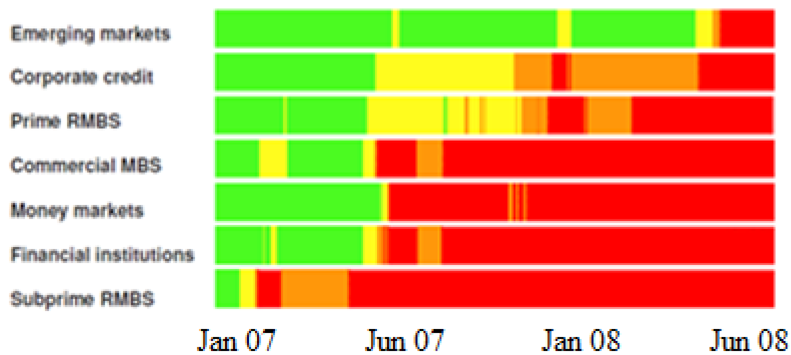


Figure 3.5: Example of Sector Heat Map (IMF, 2008)

The order book heat map, shown in Figure 3.6, allows for the visual examination of liquidity expansion and contraction in the market. The colors are applied at specific price levels and colored by resting limit order contract depth (for the 100 ms interval), to show the overall depth of the order book. The contract depth is color coded to represent buy (violet to dark blue) and sell (yellow to red) orders based on the number of resting contracts.

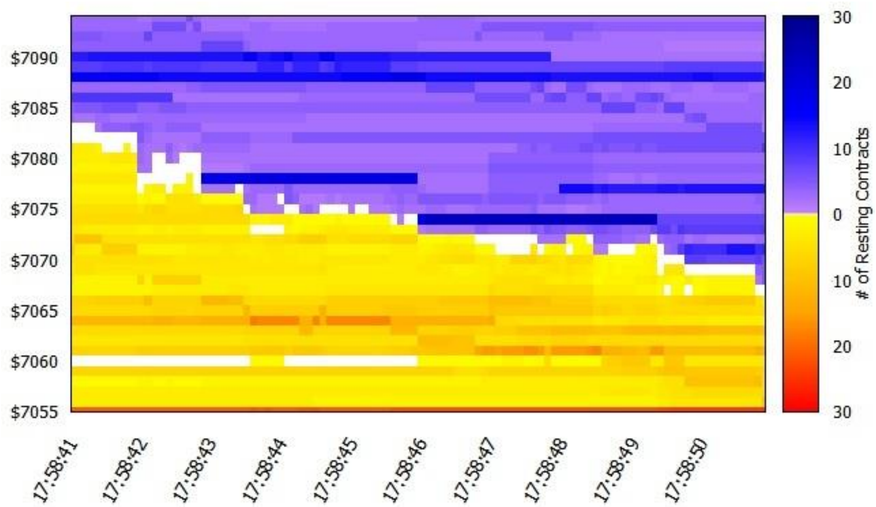


Figure 3.6: Example of Heat Map of Order Book Depth

From Figure 3.6 we can extract simple information about the direction of price over this time. Moreover, we can also make inferences about the supply and demand in this market by examining market depth over time. In this example, a large selling depth relative to buying depth can be observed above the current best bid/ask spread of the market, which drives price downward over time.

3.4.1.4 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. Individual graphics can depict only a limited number of structural dimensions, which in turn limits their ability to provide useful information. Animation offers a robust alternative; it consists of the rapid display of a sequence of images to create an illusion of movement. By portraying changes over time, it can be used to better express such complicated processes such as the behavior of market or the impact of single individuals.

The effectiveness of animation is still a matter of debate. Animation does overcome some of the difficulties associated with static graphics; in particular, it addresses concerns raised in recent research regarding both the design and the limited ability of static graphics to convey complex systems. Other research has shown, though, that animated graphics are not more effective than static graphics in testing hypothesis or learning interdependencies (Rieber, 1989).

In fact, the success of animation lies primarily in its capacity to convey extra information through freeing the temporal space and producing interactivity (Ferguson, 1995). The ability for user to interact with the animations, combining technology together with a user interface, can provide a mechanism for rapidly filtering and facilitate deeper comprehension of content through interactivity (Srinivasan, 1999; Perez, 1985; Rieber, 1991).

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 3.7 gives an instantaneous snapshot of the limit and stop-loss order book along with a historical trading volume chart.[2] The snapshots are put together sequentially in a video format that allows users to select a time interval between shots; users also have the ability to play the video at different speeds or directions.

[2] 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.

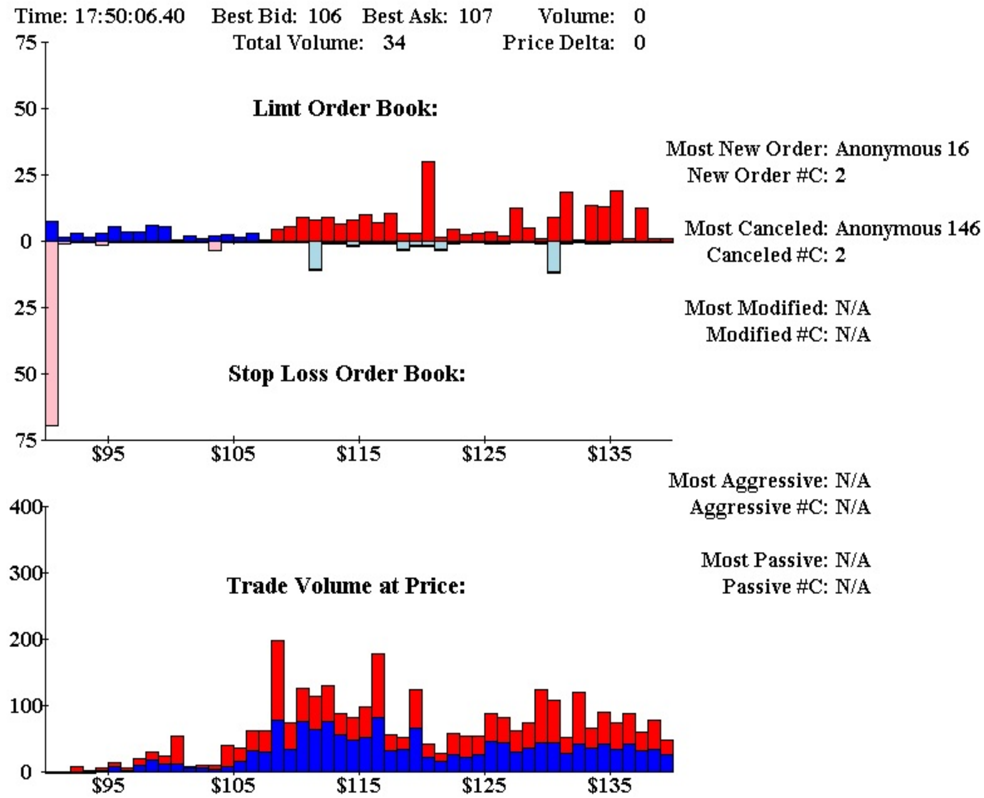


Figure 3.7: Example of Market Order Book Animation Snapshot

Animations do not preclude the inclusion of more standard methods of information summation. As noted above, one of the strengths of an animation is its ability to convey large amounts of information within a compact space.[3] 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

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

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 (Morrison, 2001). 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 how the visualization is used. 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 placed 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.

There is a 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 built for examining the practices of traders, helps to break down the impact a single trader can have on a market's behavior. The animation in Figure 3.8 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 liquidity importance (i.e., resting orders) that the individual plays relative to the rest of the market.[4] As discussed earlier, 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.

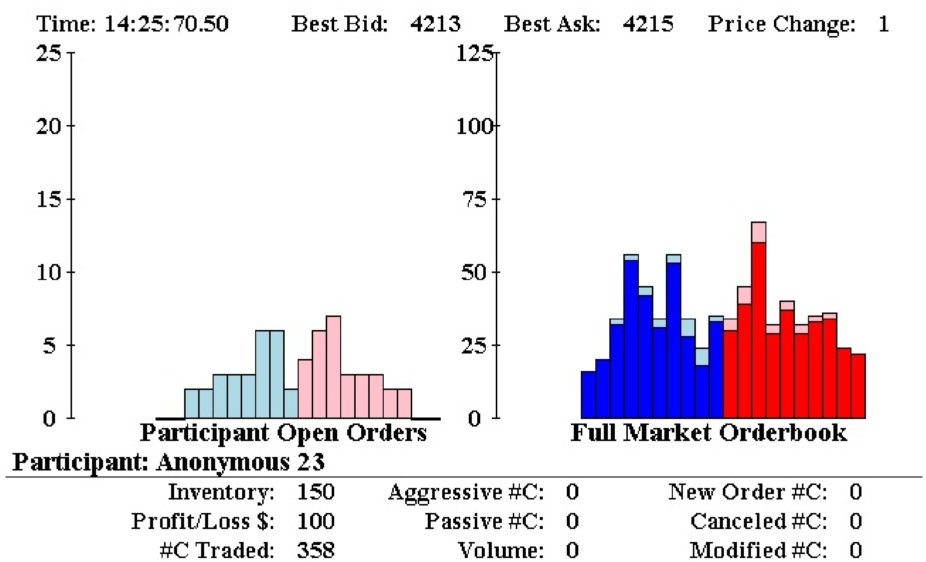


Figure 3.8: Example of Participant Order Book Animation Snapshot

[4] 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.

3.4.1.5 Order Trace 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, 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 canceled (in part or in full). The following visualization in Figure 3.9 presents an order trace graph depicting an order's lifecycle.

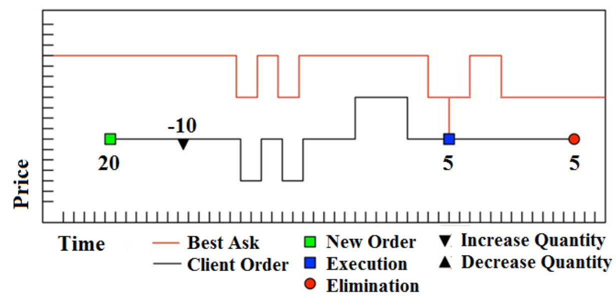


Figure 3.9: Example of Participant Order Trace Graph

In Figure 3.9, the order trace graph depicts an individual order over time and to visually process the events during its lifecycle through a spatial-temporal framework of price level and clock time. 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 lifecycle, both in price, as seen in the line's vertical movement six times, and in the size of the order, illustrated with shapes specifying the altered levels. Using this graphic, an analyst can construct, within a single visualization, the entire "storyline" of an order and relate it to the order book's activity.

3.4.2 Visualization Selection

One notable problem with visualizations is their scalability. After reaching a certain size or level of complexity, they can become too large for efficient examination. When a graphic contains too many elements, it may become visually cluttered, which may impair the ability of a user to interpret when more data is included. The critical threshold—that is, the point of diminishing returns—depends on the graphical density of the data. The graphic must fully illustrate the dimensions of structure and hierarchy, while at the same time effectively achieving the goal of the visual. Thus, the proper selection of a visual framework is critical to effectively answering a question that might be posed by an investigator.

In order to take advantage of the visual frameworks introduced in this chapter, a proper understanding of the question an investigator is trying to answer must be considered so that both data and analysis are effectively selected. This requires selecting an objective for the investigation that can be properly defined, such that a metric can be constructed that proves or demonstrates the objective. For example, if a regulator was investigating "quote stuffing" in a market (the practice of slowing down a market's matching system with large quantities of messaging traffic) one such objective metric might be the modification and cancellation rates per minute.

Once an initial metric is selected, the next step is to define the dataset to be considered, such that it can be fine-tuned as the investigation proceeds, and more importantly, so that the underlying causes of an event can be re-hypothesized as a better understanding of the event becomes established. There are three aspects to defining the data needed: scope, reference dimensions, and level of activity.

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

The first area of interest is the scope of investigation. Here, scope refers to the scale of the entity that is being investigated: it might refer to an entire market, but it could also apply to a single firm, account, or trader within a market. On an even finer level, scope might refer to a set of individual trades or individual orders that are placed within an exchange.

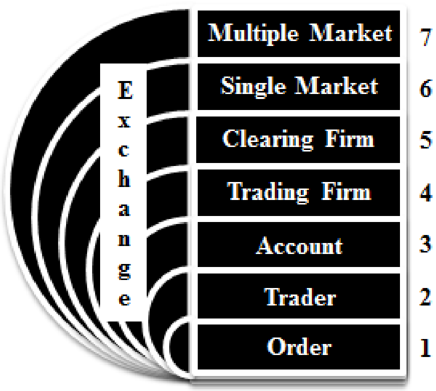


Figure 3.10: Levels of Scope and Activity in Investigation

The next step is to establish appropriate reference points—typically temporal or other, quantifiable factors—that add dimension (and therefore order) to the data. Initially, it might seem more efficient to bypass this step and simply to aggregate the data. However, to effectively communicate an investigative process, a series of events must usually be explained in a chronological manner, subject to a set of conditions. This is especially necessary when multiple events occur in parallel, making them difficult to explain using a linear structure. 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 the data needs to have in it to properly assess the events under investigation. This is similar to scope in that it requires a level of granularity to investigate; however, the activity level can be further refined or aggregated as the investigation proceeds. This is an iterative process such that the investigator can take advantage of the visualization to determine what level best explains and express the sets of complex interacts of concern.

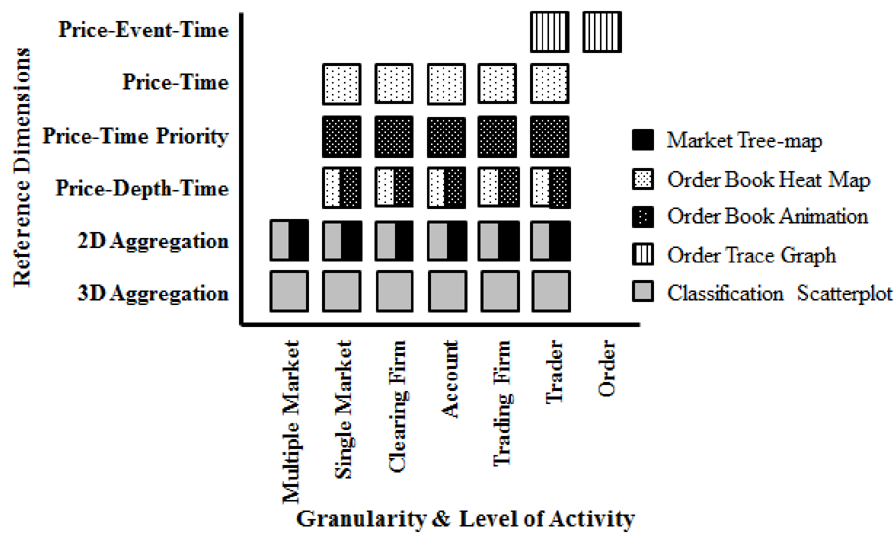


Figure 3.11: Visualization Selection Chart

Once these three variables are established, one can then select the most appropriate visual for the task, as well as the metric the objective requires. The chart depicted in Figure 3.11 suggests how different visualizations might best be matched with a range of tasks/metrics, based on scope, reference point, and level of activity

3.5 CASE STUDIES

The following example cases studies[9] are designed to demonstrate the benefit that visualizations can bring to regulatory tasks, including those in which regulators engage on a daily basis. Within each case study, we explore how the integration of tools like visualizations can substantially increase productivity and the extraction of pertinent information from regulatory data without appearing to become an overwhelming task that is, perhaps, too complex to perform.

3.5.1 Case 1: Market Oversight - Price Drop Leads to Questions of Manipulation

The price of ABC shares fell unexpectedly during after-hours 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 that he 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 the price discovery process involves the matching of bids and offers, 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.

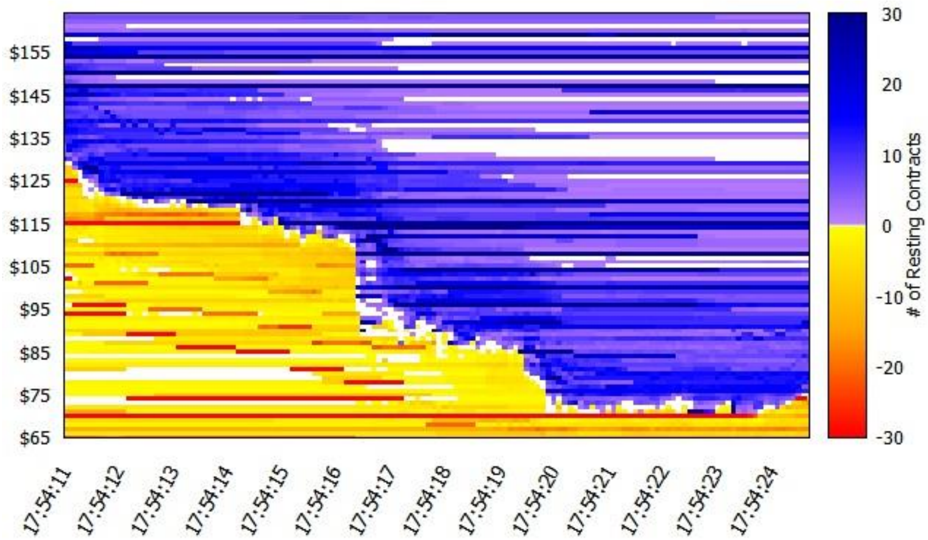


Figure 3.12: Heat Map of Order Book Depth during the Price Shift of ABC

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.[5] This feed provides the level of liquidity at each price point, along with the price and time stamp for executions.[6] Using this feed, analysts can isolate the time period associated with the strongest, and most rapid, of the price movements.[7] In addition, regulators may wish to identify whether there were precursory events that could have given warning prior to the movement.[8] In the case above, as can be seen in the Figure 3.12, during the period at 17:54 between 16.30 and 16.40 seconds the price of ABC shares dropped $15, 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

[5] 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.

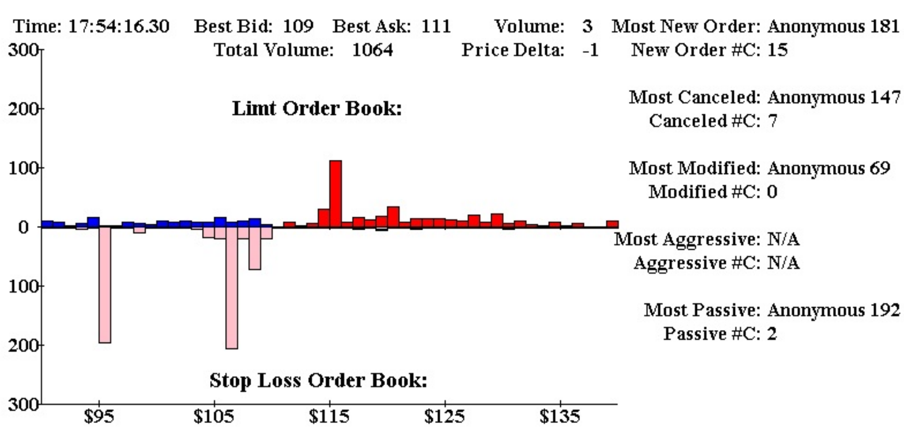
[6] Note that the public market feed, which indicates the current state of the limit order book, would not include hidden liquidity like iceberg orders.

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

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

of orders can be seen in the market; perhaps more importantly, 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 of the order book: marketable orders removing standing orders from the bid side of the book.

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 to Figure 3.13). 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 it identifies those accounts that were responsible for the most new orders, cancellations, modifications, and trades in the market during the $15 drop.



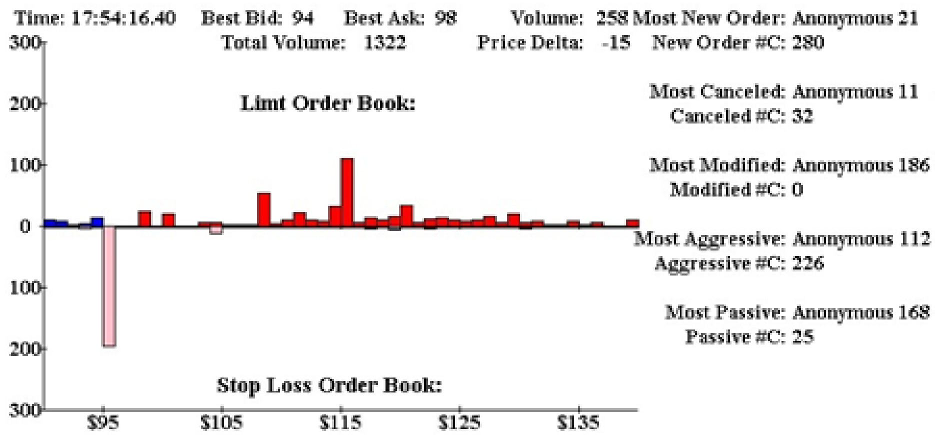


Figure 3.13: Market Order Book Histogram Snap Shot at 17:56:16.30 & .40 Price Movement

From the animation (from which the above snapshots have been taken), we can see in the before snap shot 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, they clearly show a volume far beyond that of the standing limit order buy depth. Because of this, in the after snap shot 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 domino-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 trigger each other, within a matter of 100 milliseconds, they consumed a total of 15 ticks of the resting limit orders, a staggering sum.[9]

At this point, the first set of investigative questions seems to have been answered: liquidity demand, in the form of very large stop orders, overwhelmed standing liquidity

[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 that 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.

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 single entity, 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 were 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 this trader actively worked to cause these actions 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 the trader within the book that may also have made the price fall almost inevitable. That said, it should be noted that convincingly proving intent may be an extremely difficult task within the context of a formal enforcement investigation.

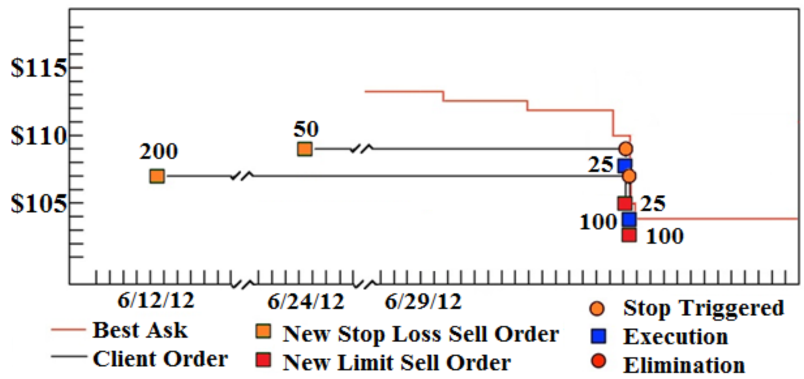


Figure 3.14: Trace Order Graph of Stop-Loss Orders

In this case, per Figure 3.14, the trace order graph of the trader's 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.

3.5.2 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 United States, 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 with the aim of lowering the volume of electronic messaging. The exchange has found that messaging levels have increased dramatically in certain products over the last few months, 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 aim to provide efficient matching of interested buyers and sellers in a security, and, then, act to rapidly disseminate of this execution information to the general market. In simple terms, these activities represent 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 that 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. Figure 3.15 depicts the percent of executed orders versus the average percent of resting orders these traders 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.

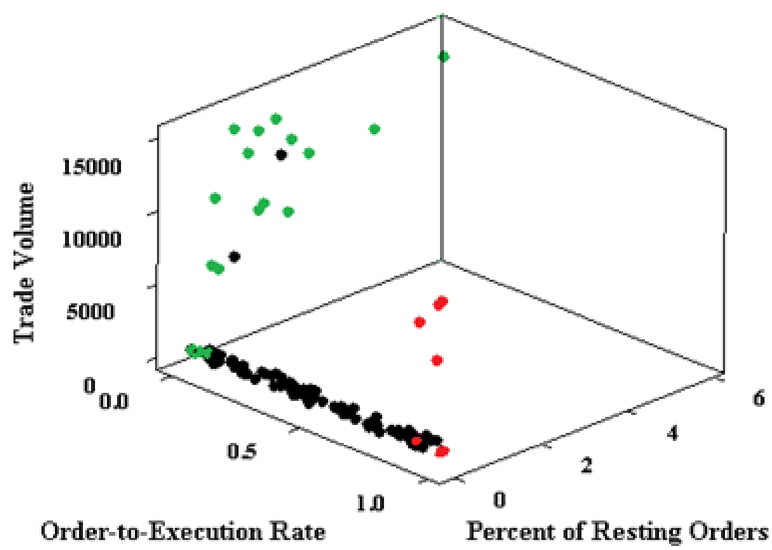


Figure 3.15: Classification Scatterplot of HFTs in STEM Exchange (red = more than 95% aggressive, green = less than 5% aggressive)

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 canceling orders at an elevated frequency. In the figure, only those accounts identified as high-frequency traders have been included.[10] 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 best bid offer 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 best bid offer and rarely experience executions; as a result, accounts within this set may be the most inefficient in order entry.

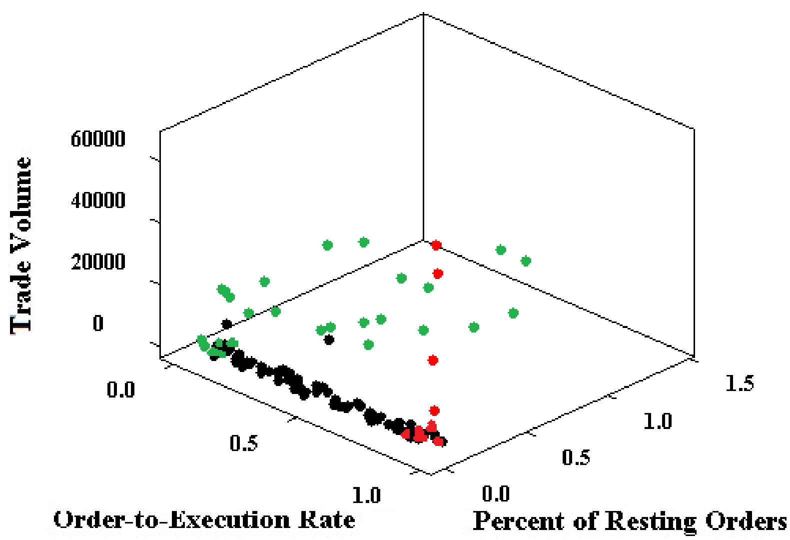


Figure 3.16: Classification Scatterplot of HFTs in STEM under High Volatility Market (red = more than 95% aggressive, green: less than 5% aggressive)

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

[10] 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.

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 when liquidity is most necessary (i.e., that moment when large parts of the order book 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 in the book, which express willingness for execution, and pure "phantom" liquidity. Figure 3.16 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 that were recently deep in the order book. Increases in unidirectional, aggressive trading now face much of the "inefficient" liquidity identified earlier; hence, disincentives to provide 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, 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 may be the sole data collector.

3.5.3 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. The trader reported the case to regulators as a suspected case of spoofing in the ABC market aimed at lowering 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 has 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, the regulator would first need to look at the window of time identified by the trader and, then, create a related tree map depicting the quantity of order placement and cancellation in the ABC market.

In the tree map shown in Figure 3.17, 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 nonexecuted) liquidity provided by that account.

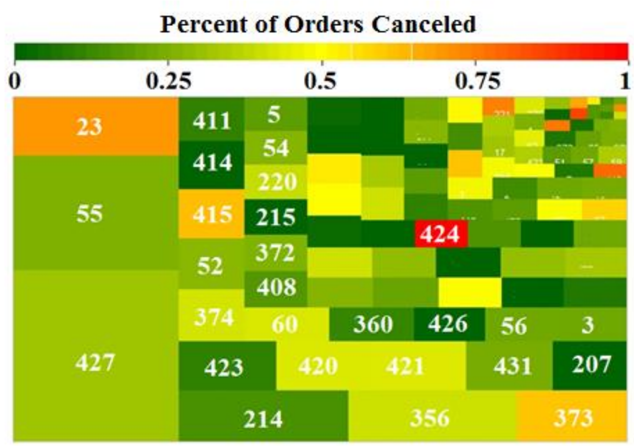


Figure 3.17: Tree Map of Market Participant: Organized by Size according to Largest Observed Order and Highlighted by Percent of Contracts Canceled (smallest: white; largest: red)

Of course, this selective variable filtering may not fully account for the spoofer's behavior. The orders associated with the spoofing event may only represent a relatively small fraction of an account's activity. As with all of the above applications, an iterative process in determining appropriate metrics is likely to 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 (seen in red and orange rectangles). 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 (refer Figure 3.18), we can see evidence of Anonymous Trader 23 taking actions that appear to match the definition of spoofing.

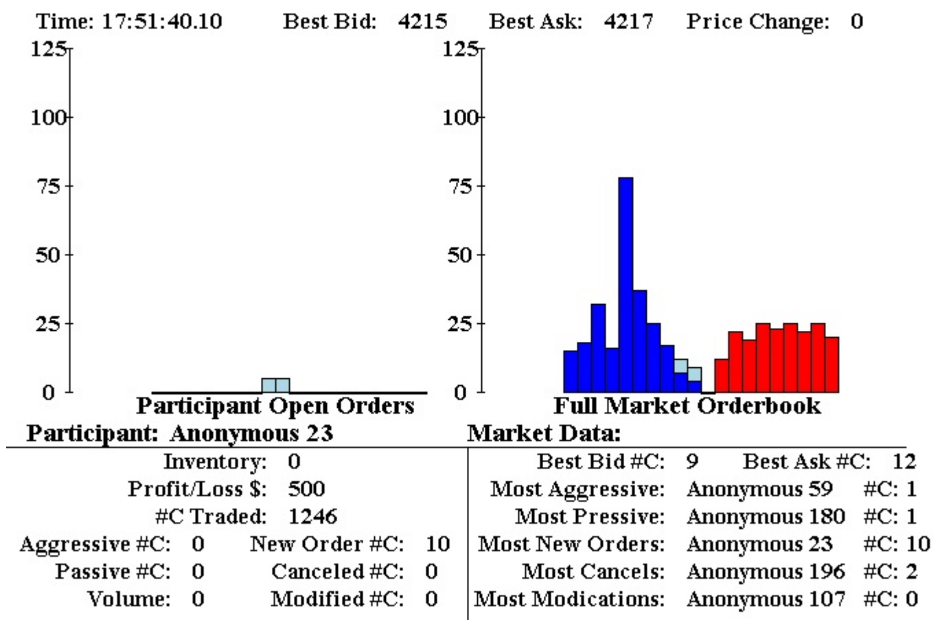


Figure 3.18: Participant Anonymous 23's Order Book Histogram Snapshot Pre-Spoof

In Figure 3.18, 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 100 contracts at the best ask, seen in Figure 3.19, entered with the possible intention of driving down the price. This sell order is almost immediately canceled (to avoid the possibility of execution), but results in a movement of price downward. As a result, this price movement causes both of the trader's smaller, passive buy orders to fill at the short-term, lower-equilibrium price.

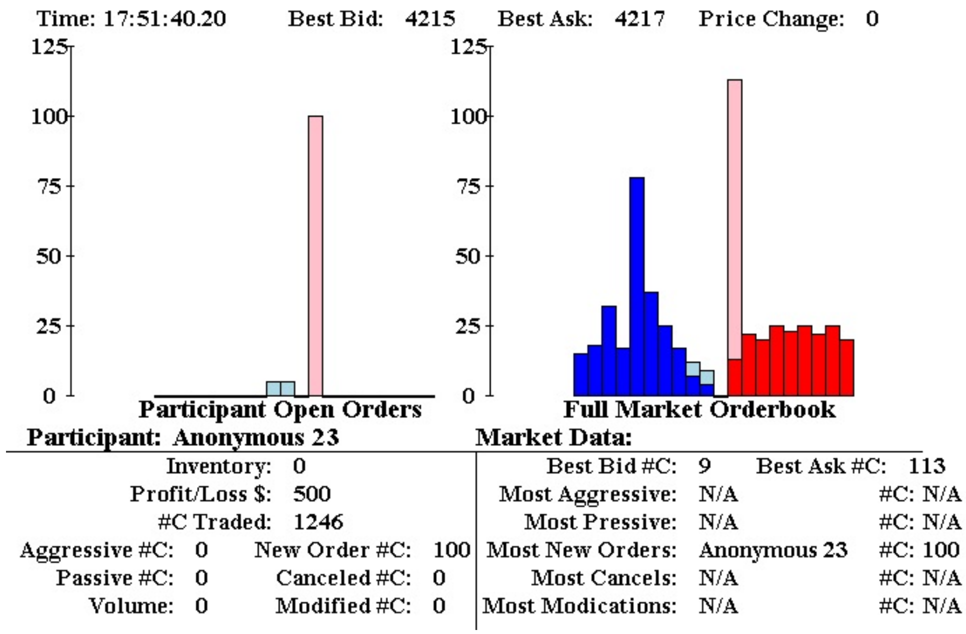


Figure 3.19: Participant Anonymous 23's Order Book Histogram Snap Shot during Spoof

In Figure 3.20, we can see that the spoof is successful in its buy execution, with prices moving the sell interest lower, and the account finding itself now long 10 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.

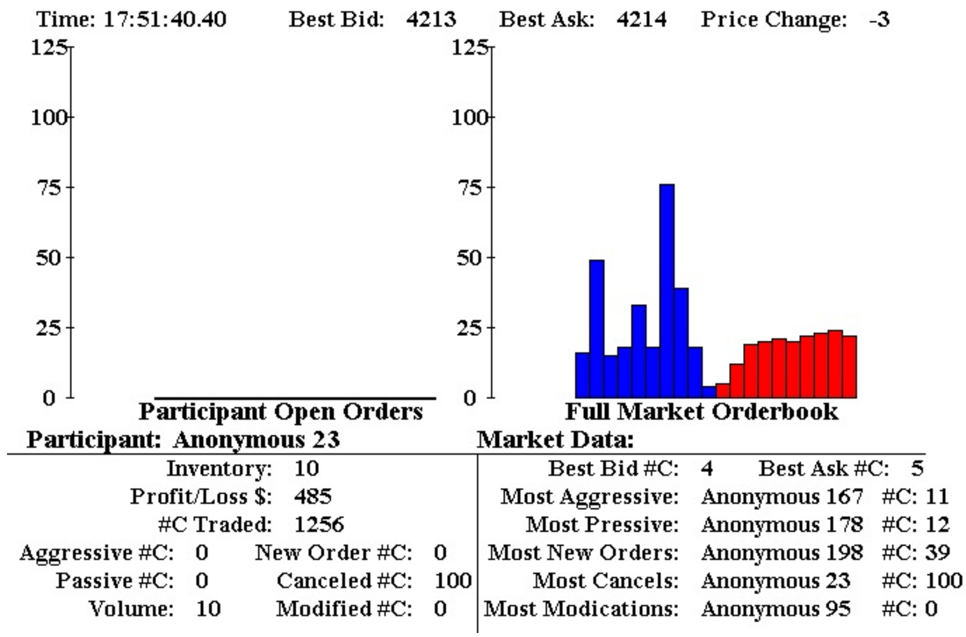


Figure 3.20: Participant Anonymous 23's Order Book Histogram Snapshot Post-Spoof

3.6 SUMMARY

Like many "big data" processes, sorting through billions of orders and trades can feel like "searching for a needle in a haystack", especially if one is not exactly sure what "the needle" looks like that causes large price shocks or manipulates trading. Though these tasks may appear as if the complexity of the financial system makes interpretation overwhelmingly difficult, many times the complexity is just an imaginary one due to either, or both, a lack of knowledge of the system the data of which the data is a product or/and the manner in which it is presented. Simply removing the imaginary complexity component of data is the first step in the untangling of the complexity that drives most questions in a typical investigation. One way this can be done is by selecting the right data visualization technique.

As discussed and demonstrated in this chapter, by using appropriate visualization techniques, regulators can create a more detailed market picture not only of the markets but the layers of data that build up to create them. With data visualization tools based on financial exchange system knowledge, regulators can have the ability to retrieve and analyze data which otherwise might not be accessible. Such tools can facilitate the rapid analysis of changes in participant and market behavior and subsequent disseminate this information to relevant parties (including the exchange, the clearing firm, or the participating firm itself).

Chapter 4

Market Monitoring and Periodic Complexity

The financial exchange is a system that provides individual market participants with a means of interacting with other participants to facilitate trade and price discovery through a centralized system. Such an over-arching system, known as an “organization” or “system of systems”, serves to combine and support the needs of many individual sub-systems (i.e. those of the market participants). They focus on meeting the holistic needs of the entire group of individual systems by fostering efficient trade and pricing, rather than any one individual’s needs (Ackoff, 1971). Exchanges manage the total-system’s performance by realizing a set of consistent operational objectives meant to aid and ensure fair trade and pricing while the underlying component systems (i.e. the participants) continually change.

Financial exchanges achieve this through a dynamic, whereby an individual participant can impact the demand and supply of a market by making changes to its own inventories, by executing against open orders, and making adjustments to their own individual open orders. Conversely, as seen in Figure 4.1, the market can influence an individual’s demand and supply by doing the exact same thing. The aggregate change in individual inventories (and orders) is what leads to changes in price through a process known as price discovery. This process results in trade at agreed prices that is responsive to the demand and supply dynamic present in the system at that moment in time.

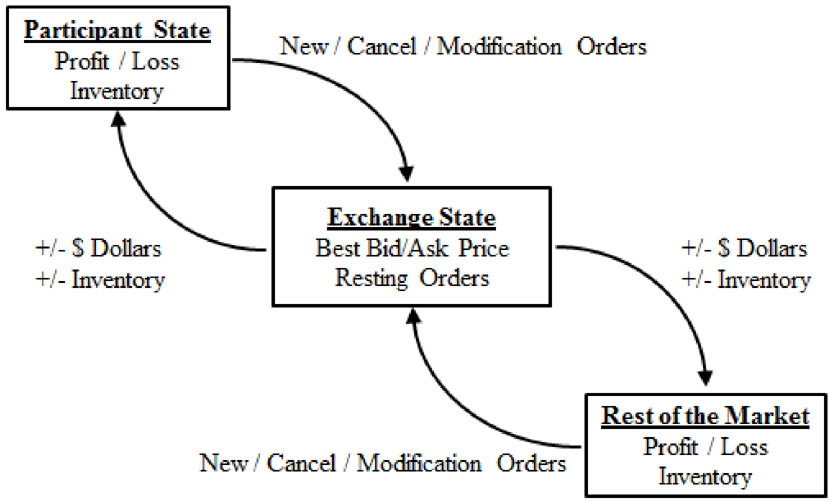


Figure 4.1: Financial exchanges and participant relationship

The value of exchanges comes not only in facilitating trading but, also acting as a pricing tool; this is viewed by many as the most important product of a market and public good to society because of its ability to 'fairly' determine price (Hasbrouck, 1995). As participants are continually adjusting their individual demand and supply needs, the exchange system must support price discovery in manner that can collectively comprehend all these dynamics while remaining stable and transparent. To achieve the points of equilibrium at which buyers and seller can agree on price, this system has a complex property in its organizational behavior that allows it to regain equilibrium. This organization behavior is what Lee (2003) terms the periodic complexity of a complex system. This results from the interaction of the mechanisms and rules (i.e. structure) of the exchange and the individual participants' objectives; however it is continually evolving dynamic reflective of the stream of many underlying individuals' decisions.

Regulators have been tasked with monitoring the functional health of these systems to offer 'fair' and effective pricing, a process that in the past relied heavily on monitoring participant behavior on the floors of an exchange. As exchanges have moved to being entirely electronic, regulators have lost this floor perspective, but have gained an immense amount/flow of data in its place.

In this chapter, we review an exploratory analysis of developing a Markov State model meant to capture the behaviors of the market as an aid to understanding the complex dynamics that create stable trade and pricing. Using the Markov modeling approach, which lends itself to system control and analysis, a set of variables was developed that characteriz**es** the market's ability to function. Using historical trade data of the E-Mini S&P 500 contracts on the Chicago Mercantile Exchange, we examine how this model can potentially be used by regulators as a monitoring system for identifying conditions under which a market**'s** organizational behavior may compromise its ability to discover price and efficiently support trade. Section 4.1 explains the general framework of the exchange system and how it serves several types of participants in trading and pricing. Section 4.2 explores the importance that the complexity plays in market pricing systems and how it can be monitored. Section 4.3 introduces a Markov State model approach to capture the dynamics of the system, and how to set descriptive variables for the model such that they can provide information to regulators about the trade and pricing conditions of the market. Section 4.4 reviews the results of applying the model to a historical data set and demonstrates its ability to predict price change. Section 4.5 provides an example case study of how such a tool works during an extreme market event, the Flash Crash of May 6th 2010. Finally, Section 4.6 provides a summary of what this tool maybe used to indicate about how well a market is functioning.

4.1 THE EXCHANGE SYSTEM

A well-known story recounts how three centuries ago, traders and speculators would gather to trade under a buttonwood tree in old New Amsterdam, a short distance from what would become Wall Street, the present center of financial trade. Such gatherings eventually lead to the formation of formal financial market exchanges providing a recognized venue to support the exchange of publicly traded financial assets. The formation of these formal systems for trade allowed market forces to be consolidated at venues where defined structures for pricing and transactions could be established. The exchanges benefited their users by increasing the expediency of trade, decreasing adverse selection from open pricing, and giving stability to pricing; all of which are emergent results of consolidating many independent market participant trade decisions.

An exchange system, as shown in Figure 4.1, links all participants to a centralized mechanism of trade. Though an exchange is vital to defining how pricing and trade occurs, it is important to recognize that it serves as a system of systems dedicated to interfacing market participants' resources and capabilities together to create a more functional and better performing overall system for efficient market pricing. Within this context, market participants are systems in and of themselves individually working to achieve their own objectives through the exchange.

4.1.1 Information Flow through Markets

Hayek (1945) was one of the first to look at the pricing system as a mechanism for communicating information. He viewed markets as information distribution systems designed to process information from different sources and distribute it to create new understanding (Huber, 1991). This new understanding is what allows price to be determined for an asset, such that price can be agree upon by participants to complete a trade. However, the manner in which a market accomplishes pricing is, in part, dictated by the structures used for distributing information to and between participants. The quality of information flow is an area of concern to financial economists and regulators since it is considered by many as the most important product of a security or futures market and a public good in society (Hasbrouck, 1995).

In Figure 4.2, we have two examples of markets with different information flow structures where the nodes represent participants (T#) and the links represent the ability to trade or witness trade in the market. In a decentralized system, information is disseminated as the participants and connections are structured, which lead to asymmetries in market knowledge derived from non-uniform distribution of information and several price formulations. A centralized exchange-based system, collects and uniformly distribute all the information; thus, all participants are equally informed, allowing for no one participant to have a structural advantage such that one price is determined among all the participants.

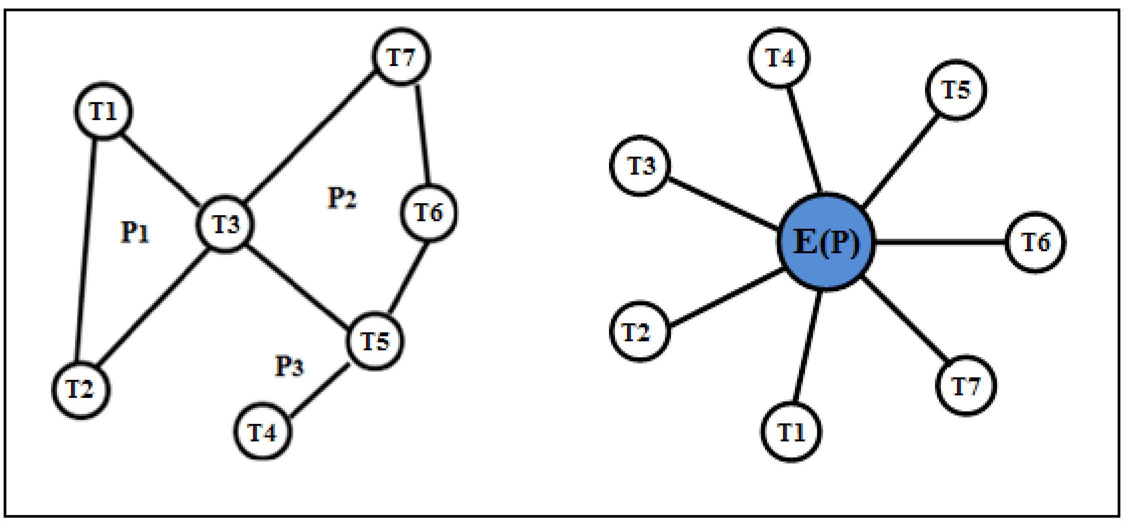


Figure 4.2: Market structure: decentralized market (left) and centralized exchange market (right)

In its simplest form, the flow of information in exchanges using an order book structure consists of executed trade data[1]. This data about the flow of traded assets through the market can be thought of as the information flow that is used to determine price. However, as the price discovery process is not instantaneous or exact; the information flow can reflect prices that swing over a range of values until participants can come to an agreeable and, at least momentarily, stable price. The process is best characterized as more of a 'deliberation' between the buyers and sellers through the actions of placing, modifying, and canceling orders and trades until an agreeable price can be found.

4.1.2 Market Participants as Systems

Market participants have individual objectives for participating in the market and follow a specific execution strategy that allows them to achieve these objectives. Their strategies, regardless of formality or strictness, are systems in their own right that dictate their behavior within the market while also acting as constraints on their decisions.

Figure 4.3 below provides a diagram intended to represent a generalized framework (Haimes, 1998) of the relationship that each participant's own system has with the overall exchange system. The exchange state, described by price and participation, are a construct of outside variables (exogenous, random, input, and output). While the participant's state, described by profit/loss and inventory, are based on the decision variable to either buy or sell, or do both or neither and the inputs and outputs of the exchange

---

[1] Other forms of data are available, like order book depth that may also provide insightful information but are less understood.

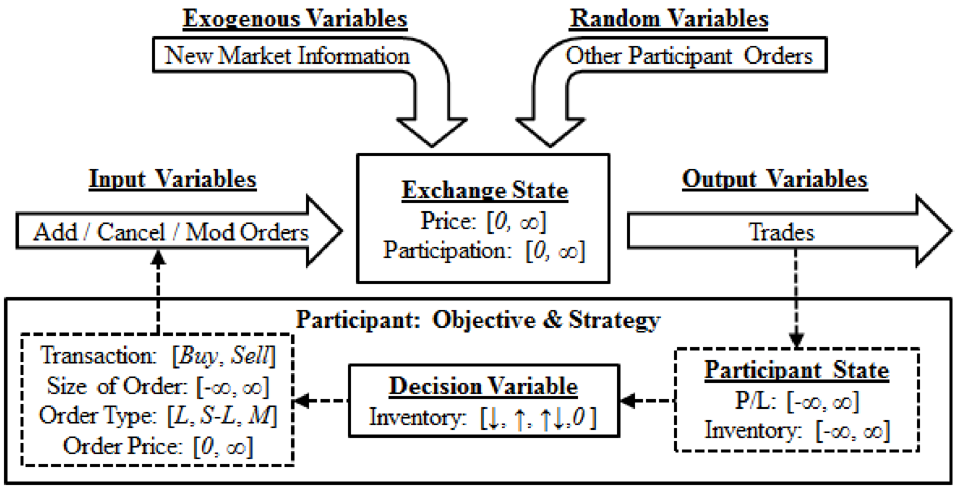


Figure 4.3: A participant systems interaction with the Exchange System

The objectives of participants can differ greatly, such as individuals, who trade on their forecast of prices, versus market makers, who earn a fee per transaction, versus those that simply place orders to lower the risk exposures to an inventory they are holding. Regardless of the benefit or strategy implemented, observing the orders and trades that a participant places throughout a day provides insight and understanding of the flow of information they provide. This is to say that, even though their participation clearly has its own motives, their activity creates flows of information to the exchange in the form of trades that can help classify their contribution to the dynamic of the exchange.

The following four types of participant classes, Fundamental Buyers, Fundamental Sellers, Opportunistic traders and Intermediaries, were selected such that they created a simple but formal model of the exchange system by definition each class by how they managed their inventory decisions. More complicated models of the system have been built to show further granularity in participant classes but for purposes of demonstration the model was kept generic. The following sections give a description of each of these participant classes.

4.1.2.1 Fundamental Buyers and Sellers

The behaviors of Fundamental Buyers and Sellers on the exchange are simple: Fundamental Sellers (FS) create the flow (sell) and Fundamental Buyers (FB) absorb it (buy). They act as long term “value” investors, individuals looking to hold on to an asset for longer than a day and use exchanges as a vehicle for get into and out of value investment positions safely and at a good price. They benefit from the exchange’s ability to price efficiently via the property of liquidity that the exchange creates (i.e. that markets make it possible to get in or out of a position with little slippage, or movement in price.) (White, 1981).

This property is attractive to FB and FS since it makes the exchange and movement of capital both cheaper (Atje, 1993) and safer, through the ability to easily diversify (Allen, 1997). Thus, this property is protected and encouraged by sovereign countries looking to attract investment capital since it attracts and reassures FB and FS looking for safe and established venues for executing trade.

4.1.2.2 Opportunistic Trader

The Opportunistic traders (O), or strategy traders, are individuals who aren’t restricted to the direction of their flow of trade since the direction depending on their strategy. They have the shared belief that they can profit from inefficiencies in a market’s ability to correctly assess the price of an asset or that there is predictability to the price of an asset.

These inefficiencies are a result of the tendency of individual participants to overreact to information as it becomes available, a popular view held by most economists and demonstrated by De Bondt and Thaler (1985, 1987). They show that there may be greater than normal returns to be had by leveraging opportunistic corrections of these inefficiencies in pricing.

The manner in which opportunistic traders identify these inefficiencies, and how they execute on them, however, can vary greatly. These strategies maybe mathematical based in time series calculations like moving averages, Fibonacci, or correlation functions or based simply on looking at charts for patterns in candle sticks and relative strength indices. The activity of their trades can vary greatly; it can be vary from long term, monthly or weekly trades, to short term trend followers that only hold for a few hours down to only a few seconds.

All opportunistic traders share a common feature in that their trading strategies implement signal processing elements that use external information and past prices as inputs to project future prices. From the view point of the exchange system, though these participants may use different tools and have varying levels participation, they all serve to achieve the same purpose of increasing the efficiency of pricing.

4.1.2.3 Intermediary Traders

Intermediaries, much like retail stores and wholesale dealers, help manufacturers to sell assets and final end users to buy. In financial exchanges, they are known as specialists or market-makers who help create trades by acting at times as sellers for buyers entering the market and vice versa for sellers entering the market. In this manner, they have the ability to create and absorb flow, like opportunistic traders, with plans to temporarily hold the assets. However, they differ in the manner in which they control the flow of trade.

Intermediary traders achieve their profits much like retailers and wholesalers, by marking up the price of goods. They accomplish this by being willing to buy or sell and making their profit only through the ask-bid spread[2] or fees. This compensation they receive is for managing orders and assuming risk by standing ready to carry out trades on their own account on either side of the order book (Demsetz, 1968). By controlling their own inventories such that they stay near zero, intermediaries remove risk; this is what differentiates their information flow characteristics from those of opportunistic traders.

This process of market intermediation is analogous to the job performed by catalysts in chemical reactions. Catalysts are able to increase the rate of chemical reaction due to their participation, but, unlike other reagents that participate in the chemical reactions, they may participate in multiple chemical transformations and are never consumed by the reaction itself. Market-makers act in this very same manner by helping to induce trade by creating liquidity on the buy and sell side of the market.

Intermediation activities, as a whole, play a significant role contributing to the U.S. economy. Conservative estimates indicate they have accounted for one-quarter of gross domestic product (Spulber, 1996). In securities and futures markets, it is more difficult to estimate a figure, but intermediaries do play a significant role in the price discovery process by helping to smooth trade on exchanges by creating market liquidity through their willingness to sell and buy inventories of financial securities. This process helps

[2] The bid-ask spread is the difference in price between the highest price that a buyer is willing to pay for an asset (the best bid price) and the lowest price for which a seller is willing to sell it (the best ask price).

increase the ease of trade and promote optimal selection of portfolios diversity (Allen, 1997).

4.1.3 The Exchange: System of Systems

Markets act as unstructured systems aimed at obtaining cooperation among a collection of individuals or units who share congruent objectives of selling at the highest price while buying at a lower one (Ouchi, 1979). As a pure mechanism for organization, a market is a very efficient mechanism of control, where price exactly represents the value of a good or service and decision-makers need no other information (Arrow, 1964). The financial exchange is an organizational system meant to take hold of a market and funnel its demand and supply needs by using specified criteria and logic based transactional rules to create trade and determine price. They also incorporate features to deal with the real world imperfections of markets like transactional costs, information asymmetry, and counter party risk.

Today's exchanges have typically adopted a price-time based system for matching and an order book as their mechanism for structuring transactions and simplifying the queuing of orders to be executed at set price points. As new orders are entered, trades, modifications, or cancelations may occur, and the volume of indicated interest at a given price level often changes. This dynamic of order behavior occurs throughout the trading day and represents how these mechanisms create a unifying market system for pricing.

This dynamic can be observed through the exchange by grouping the participant types according to their direction of flow of trade and if they have inventory limits (i). The result is Figure 4.4 where the Fundamental Buyers (FB) and Sellers (FS), trade to either increase or decrease a position throughout the day, since they are long term holders.

Opportunistic (O) traders, who are taking advantage of perceived inefficacies throughout the day, take positions in either direction. Intermediary (I) traders, trade in either direction as well by providing liquidity to the market while managing their inventory risk.

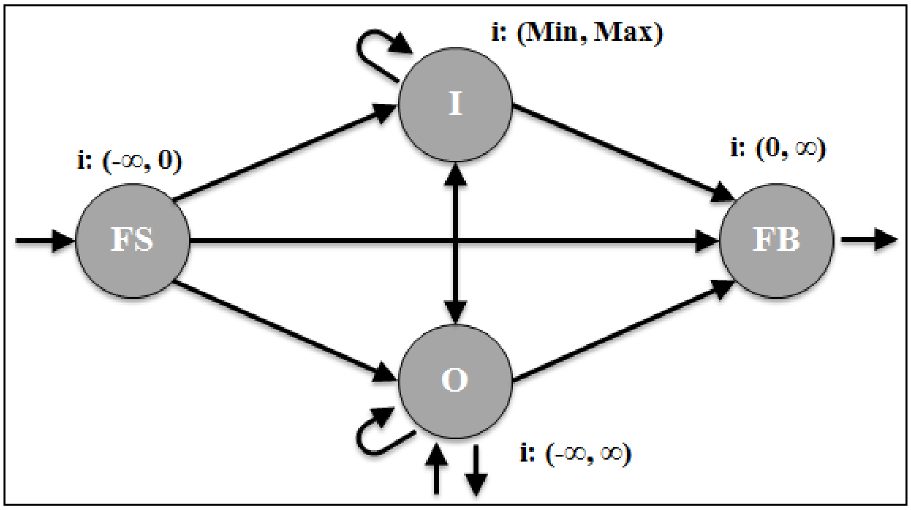


Figure 4.4: Diagram of flow of assets through the exchange

4.2 EXCHANGES AS COMPLEX SYSTEMS

Financial exchanges, as stated earlier, are complex organizational systems, systems of the systems of individual participants, designed to coordinate the demands of many market participants by providing venues for trade with efficient pricing. Though we can identify the components that make these systems, the manner in which these systems function and behavior are still an area of much research.

The paradigms that describe system behavior have historically mimicked the prevailing paradigm of that era's scientific theories (Ackoff, 1972; Hayles, 1991). Most of that science leading up the 20th century was heavily influenced by the scientific principles of Newton, LaPlace, and Descartes (Capra, 1982). Their paradigms were used to create principles in science, such as the natural state of a system was equilibrium and that departures from equilibrium would be damped out (Dooley, 1997). Their approaches reflected reductionist and deterministic mindsets that described how systems function from a mechanistic view. By taking the component elements of a system and the manner in which they interact, they believed that the future states of the system could theoretically be predicted (Bohm, 1957).

However, this mechanistic view of systems hasn't been effective in describing the complex behaviors of systems that are continually evolving, such as financial exchange markets and natural ecological environments containing numerous sub-systems. Such systems work toward equilibriums, as previous principles suggest, but follow a periodic damping process that can never fully dampen out. In these systems the periodic dampening of the larger system itself causes changes to the behavior of component systems, and, thus, never allows the system to reach a stable equilibrium, or low entropy.

In order to better understand how the components and their dynamic relate in complex systems, models using agent based simulation have been provided experimental laboratories for examining the exchanges and what may cause sudden disruptions to their functionality leading to market crashes (Farmer, 2005; Paddrik, 2012). Other similar work has looked at the outcomes of market crashes, observing pricing data to understand if there are ways to detect changes in pricing behavior due to critical relationship changes between participants within the exchange, known as "phase transition" (Sornette, 2003).

4.2.1 Phase Transitions in Complex Systems

In complex organizational systems, a phase transition is characterized by a change in the pattern or relationship of the component systems to one another which alters the system's behavior. Such events can often signal functional weakness and, typically, can be identified by observing increased correlation, much like a portfolio of assets has increased in risk as the cross-correlation between assets increases (Onnela, 2003) or when biological systems become correlated prior to fatal infections (Fairchild, 2012)[3].

Exchanges, which are a subsystems of systems, are functionality reliant on a complex periodic function that a diverse set of participants (subsystems) provide as they seek to determine price, which results from their gathering of information and sharing it through orders/trade and transforming price (Gault, 1996). This functionality, though periodic in nature, requires the exchange system be able to deal with a continually changing set of order flows, information and knowledge, such that the system is able to steadily arrive at a price by maintaining a both buyers and sellers engaged and willing to trade.

In exchange systems, phase transitions occur when there is a change to the system that makes them unable to maintain the complex periodicity necessary for price discovery as a result of the loss of either buyers or sellers willing to participate; this can cause the multi-objective behaviors related to trade and pricing to no longer be met. An example of such a transition is the Black Monday stock market crash of 1987; an event believed to have been caused by what was then the recent introduction of Portfolio Insurance, a hedging techniques, that lead to many of the market participants of the time to behaving in the same manner (Shiller, 1988) and selling formulaically during sharp down turns in the market. This behavior that was later detected by Kiyono et al. (2006), who saw an abrupt transition of the probability density functions of price behavior of numerous stocks

[3] The heart rate characteristics monitor system, known as HeRO, was developed with this purpose in mind. It was designed for monitoring several body life signs for the detection of sepsis and other infections pruned for preterm infants. The system assesses a score based on two heart rate components: decreased variability and decelerations, which normally act independently of one another and don’t act during a variety of pathologic conditions.

during the 1987 crash; this deviation from normal functioning is suggestive of a problem with pricing behavior in those markets.

Regulators interested in protecting the market from such phase transitions in the future must first determine under what conditions can the complex periodicity of these systems be lost and be able to pinpoint their root causes. However, in order to detect such behaviors, a means of successfully monitoring these ever changing market environments must be constructed such that the dynamic relationships, or interaction patterns, of the components of the system, as a group, can be monitored; this is much more useful than focusing on a mechanistic view that requires that every subsystem be individually examined.

4.2.2 Detecting Phase Transitions

In detection systems used for complex biological systems, there is a reliance on signature-based systems that are designed to detect known conditions, such as blood markers used in screening for bacteria or parasites. Such systems require frequent updates, and are not capable of detecting new conditions (or behaviors) that maybe dangerous (Patcha, 2007). Unlike the evolution of most biological systems (e.g. where the yearly cold virus's evolution closely followed so appropriate vaccines can be adapted), financial markets evolve constantly as a result of new participants entering and old ones leaving. This constant evolution makes it difficult for exchanges and regulators, mandated to keep markets functioning safely, to keep up with the latest techniques of market manipulation or dangers trading algorithms.

A more applicable approach to financial markets ia anomaly detection systems, used in computer security to combat new virus and hacking threats, model the normal state of evolving systems behavior offers (Patcha, 2007). They offer the ability to monitor the relationships of continually growing computing networks by looking for unique and new behaviors between subsystems as a manner of early warning. Many of these relationship monitoring systems use Markov chains models (simply referred to as a Markov models) at their core.

Markov models offer a tool for computationally analyzing the relationship such systems for stability and functionality. By using a probability matrix of the likelihood of different components of a system communicating with one another, a probability distribution of a stationary Markov chain can be learned from the observations of the system over time (i.e. a patterns) that describes how the relationship of parts normally function relative to one another.

Markov models also offer the ability to used tools and concepts from system control and analysis to recognize new conditions as possible threats to the functionality of a system, even if the pattern has not been seen before. This allows detection of conditions relationships wherein states may be absorbing or non-communicative. Such conditions are typical early warning signs of unstable behavior if persistent for long periods of time.

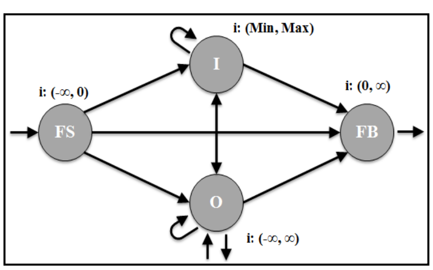
4.3 MONITORING FOR EXCHANGE PRICING PROBLEMS

Identifying when exchanges are not able to properly organize the participants such that the periodic function of deliberating prices is not working efficiently can signal the beginnings of market price disruptions that can send prices on unnecessary rollercoasters. This section lays out how a Markov model framework can be used to look at the exchanges' systems functionality while maintaining a multisystem level perspective on how each participant class type can influence the greater exchange through examining individual inventories and possible price constraints. By doing so, it makes it possible to help determine which parts or sub-functions may be causing systemic pricing issues.

4.3.1 The Markov State Model of the Exchange

Using the structure of a Markov model, a platform for examining the flow of assets between classes of participants through their buying and selling decisions can be created. By looking at flow during increments of time, we can examine how assets are dispersed through the system as they enter and ultimately exit the system, and how this dynamic changes with respect to pricing and inventory.

Using the diagram of the exchange system, described at the end of Section 4.1, we have a simple, but useful, trade network that can be fit to the Markov framework which can, then, be used to search for relational changes in the networks behavior. Figure 4.5 builds on Figure 4.4 showing the relational matrix taken from the diagram of flows where, $\chi_{i,j}$ is the percent of assets sold from class r to class s. This is then linked to a vector, Q, that has the probability distribution of daily inventory of each group. The combination of transition probability matrix of traded assets and the inventory state of the classes create the Markov State model from observations of the system.



$$\begin{bmatrix} \chi_{FS,FS} & \chi_{FS,I} & \chi_{FS,O} & \chi_{FS,FB} \\ \chi_{I,FS} & \chi_{I,I} & \chi_{I,O} & \chi_{I,FB} \\ \chi_{O,FS} & \chi_{O,I} & \chi_{O,O} & \chi_{O,FB} \\ \chi_{FB,FS} & \chi_{FB,I} & \chi_{FB,O} & \chi_{FB,FB} \end{bmatrix}$$

$$\boldsymbol{Q} = (\boldsymbol{i_{FS}} \quad \boldsymbol{i_I} \quad \boldsymbol{i_O} \quad \boldsymbol{i_{FB}})$$

$$where,\ \ \chi_{r,s} \geq 0,\ \sum_{I,FS,FB,O}^{s} \chi_{i,j} = 1,$$

$$and\ \chi_{FS,FS}, \chi_{I,FS}, \chi_{O,FS}, \chi_{FB,FS}, \chi_{FB,I}, \chi_{FB,O}, \chi_{FB,FB} = 0$$

Figure 4.5: Transformation of exchange system model to Markov State model

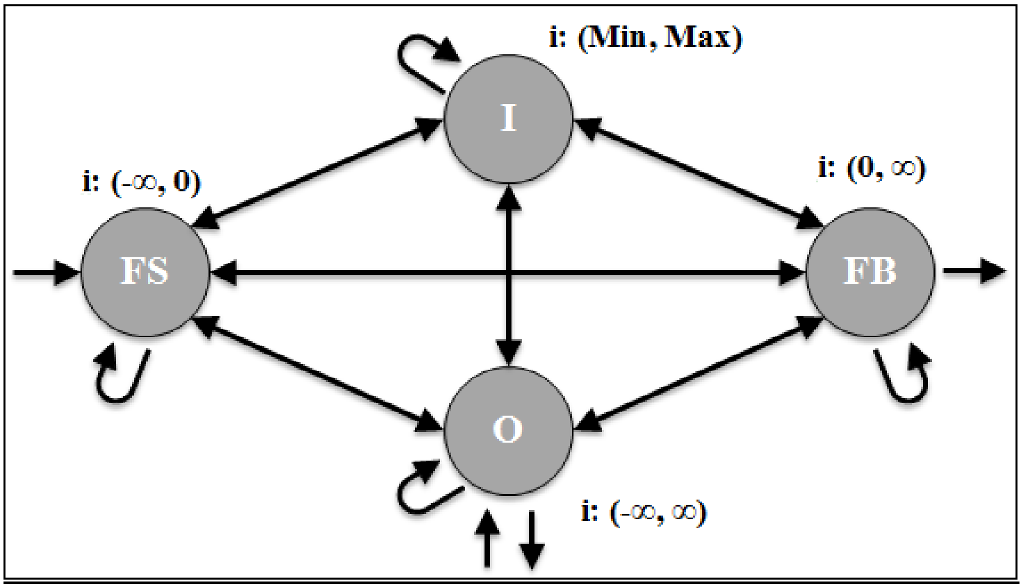
4.3.2 Organizing Data for the Markov State Model

In order to implement the exchange model in the Markov State framework, the market participants must first be assigned to one of the four participant classes using a set of threshold criteria defined by constraints on the direction of flow and inventory magnitude in relation to trade, as shown in Table 4.1. Using the thresholds as defined, those that meet the given single constraints are simply classified as either: I, FB, or FS, and the remainder are categorized as O. As a result of the approach, characteristics of market participants within each category can vary considerably relative to one another; however, from observations of trade data by Kirilenko et al. (2011) these threshold points divide the percent spectrum of participants most clearly.

Table 4.1: Participant categorization threshold criteria

Class	Flow Constraint	Inventory Constraint: Max(ABS(i)) / Total Contracts Traded
I	-	> 20
FB	80% < Buy	-
FS	80% < Sell	-
O	-	-

A practical consequence of applying the thresholds specified is that the Markov framework constraints placed on the direction of flow of FB and FS participants have to be ignored to permit both buying and selling by the classes (a departure from the original model). With this adjustment and the criteria, it is possible to apply the Markov model to capture trade data. As an example of the results of applying this process Figure 4.6, demonstrates what the average participant looked like for the active month contract of E-Mini S&P 500 futures in February 2011, using account level transaction data from the CME. The breakdown of the average dynamic of these groups and their contribution to the exchange system can be observed; it reflects our expectation of how these groups typically act.



Class	Size of Class	Avg(# Contracts Traded)	Avg End of Day (i)	Min(i)	Max(i)
I	160	13867	-3	-158	158
FB	453	426	401	-9	402
FS	1018	192	-188	-189	0
O	9547	545	-7	-103	80

Figure 4.6: Diagram of modified exchange model to fit changes required by threshold categorization method with descriptive statistics of each class of participants observed in the E-Mini S&P 500 futures market

In the Figure 4.7, a series of plots show the average price transition probabilities seen throughout the day between classes of traders. These plots reflect our expectation that the majority of contracts transitions are between I and O and that this behavior is constant throughout the day.

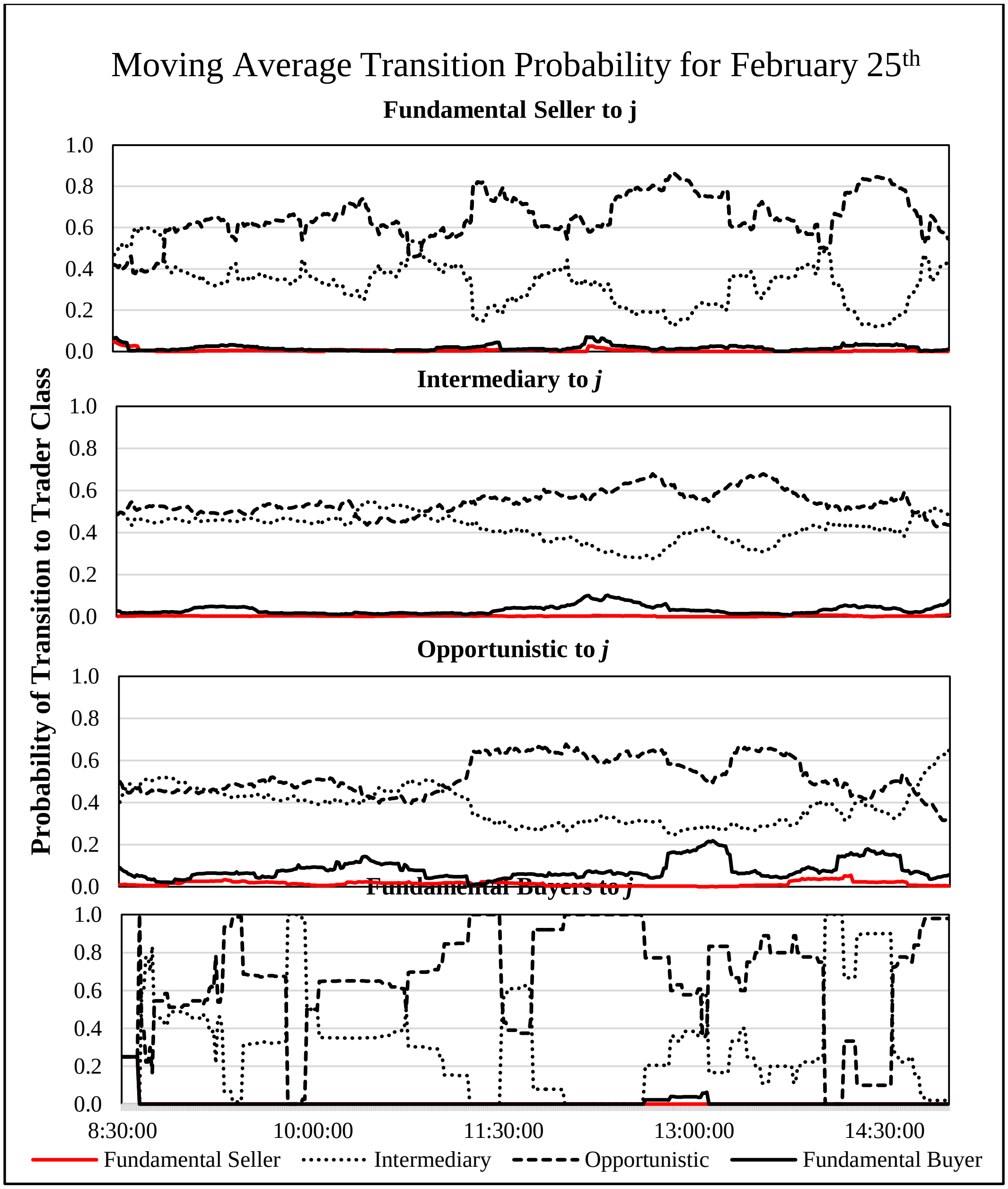


Figure 4.7: Plot of transitional likelihood throughout the day

4.3.3 Application of the Markov State Model for Market Monitoring

Today, financial market regulators must typically work post hoc to determine if a market's performance appears to be "normal"; however, prior to electronic markets, market regulators had the ability to walk the trading floors and directly observe participant behaviors for suspicious actions, or how pricing discrepancies were being managed. The Markov State model approach proposed in this research may offer regulators a means to, once again, examine the system from a vantage point closer to "the floor level". The Markov State model can provide a means for limited real-time observation of trading behavior to allow regulators to observe how a participant's (or participant class's) trading behavior maybe impacting the overall exchanges system's ability to function towards its objectives.

Markov models are a typical tool used in system control and analysis. This tool can provide the capability to monitor the capacity of the system to see how/if a limit is reached, the dynamics of the relationship of its parts, and changes in the flows between them. (Flemming, 2006). Through the application of a Markov State model on a market exchange system, we can**:**

i. Monitor the **capacity** of the system to buy or sell assets, which is essential to understanding its flexibility to deal with demand and supply shocks as pricing occur. Knowing these limit is allows regulators to understand where tipping points may exist in exchange's ability to price efficiently.

ii. Monitor the **trade dynamics** of the Markov State model using the probability of trade between participant classes permits regulators to observe the behavior of the system as demand and supply change, or the periodic complexity of system.

iii. Monitor the passage of an asset through the exchange, known as its **intermediation**, can give regulators insight into the lifecycle of an asset traded through the system. By examining the length of the path and the influence that a participant (or participant class) has on the flow of an asset through the market, regulators can learn look for unusual trading behaviors.

4.3.3.1 Monitoring Capacity

The capacity of any system is the limit of system such that it no longer can successfully accommodate changes to its inputs while still maintaining the objectives set forth; in the case of a financial exchange, the limit is reached when the exchange can no longer successfully trade and price due to demand and/or supply changes. In the Markov State model applied to an, this capacity is related to the inventory, *i*, of the participant class.

Table 4.2: Participant inventory limits

Class	i(min)	i(max)
I	min	max
FB	0	-
FS	-	0
O	-	-

Of these limits, the I class participants have both a min and max capacities that are observable by looking at their inventory over time. When such limits are reached, the typical resiliency of the system is lost (i.e. resiliency being the capability to function in the face of swings in demand and/or supply, or a phase transitions).

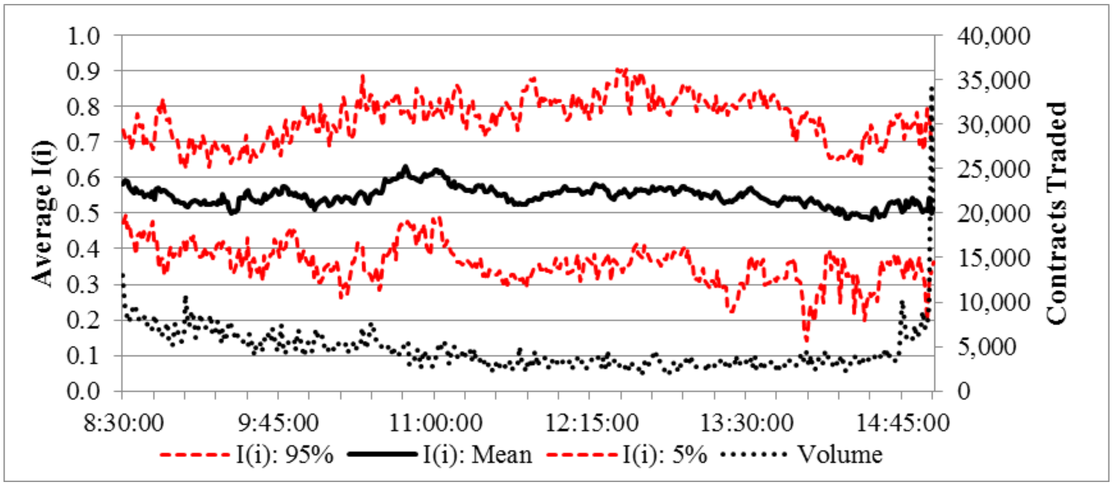


Figure 4.8: Plot of average I inventory of a contract thought out the month of February

In Figure 4.8, the plot shows the average inventory of the I class participants observed using the E-Mini S&P 500 Futures data series where the I(i) has been normalized between 0 and 1, and has an inventory level of zero at 0.54. Of particular not**e,** while I participants usually keep their inventories near zero, they preferred to keep their inventories positive rather than negative over this time period. Also, the average inventory level is very steady such that there is no pattern to how they control their inventories throughout the day.

4.3.3.2 Monitoring Trade Dynamics

In examining the state of trade flow across the exchange system, the Markov chain can be used as a representation of the temporal behavior of the pricing process. As the exchange system continuously has to deal with participants entering and leaving the system, the flow of trade between participant classes must change in order to adapt. By monitoring the level of stationarity of the Markov matrix of the system, it is possible to measure the periodic complexity of the system as it adapts to shocks in demand and supply, a requirement to achieve smooth pricing.

To do this, typically, a long term norm profile of temporal behavior is built**,** which**,** then**,** permits comparison with the current/recent temporal behavior for the purpose of detecting a significant difference. A moving average profile is used in monitoring the adaptive changes in the markets trade flows. Using the formulation in Figure 4.5, stationary Markov State models of both long-term norm profiles and a moving average profile can be built and trained for the temporal behavior.

To calculate the mathematical Euclidian norm, a comparison test must be done giving a quantifiable difference between the current transaction matrix and the long term and moving average matrices. These comparisons take into account the relative inventory level of the I class participants as they, unlike the other classes, have known inventory constraints.

$$l_1 : \sum |P_5 - P_{60}|$$

where,

$$P_5 : 5\ minute\ Markov\ matrix$$
$$P_{60} : 60\ minute\ Markov\ matrix$$

The most widely used comparison test is the l_1-norm which is used in almost every field of engineering and science. By examining these values for their relationship to price and volatility, there exists an opportunity to monitor the market participants' ability to adapt in their decision making as new assets enter the system.

.

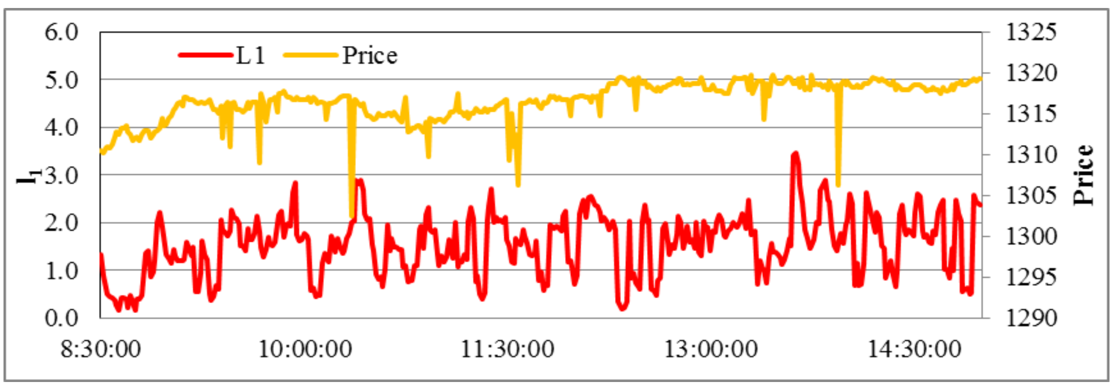


Figure 4.9: Plot of l_1 for February 25th

Figure 4.9 is an example of what l_1-norm looks like on an average day with a few large price changes throughout the day. Both the moving average and long term average l_1-norm vary greatly throughout the day which indicates that trade between classes is non-stationary. This result implies that participants' decisions are continually changing causing their trade dynamics to change, as well.

4.3.3.3 Monitoring Intermediation

It is important to consider the path an asset 'travels' as it is traded through the exchange; as the path is a record of the price deliberation process of that asset. Intermediation is a measure of the number of trades it takes for an asset entering the system to leave the system. This measure offers regulators an understanding of the amount of pressure the entire market is under to trade an asset out of its inventory. The Markov State model transition matrix formulation provides a mechanism for estimating the number of trades an asset will experience prior to leaving the system**,** while not directly having to trace the details of its trading.

$$P = \begin{bmatrix} 0.14 & 0.33 & 0.43 & 0.1 \\ 0.23 & 0.29 & 0.36 & 0.12 \\ 0.12 & 0.43 & 0.42 & 0.03 \\ 0.12 & 0.3 & 0.45 & 0.13 \end{bmatrix} P^2 = \begin{bmatrix} 0.159 & 0.357 & 0.405 & 0.08 \\ 0.157 & 0.351 & 0.409 & 0.084 \\ 0.17 & 0.354 & 0.396 & 0.08 \\ 0.155 & 0.359 & 0.407 & 0.078 \end{bmatrix}$$

$$P^\infty = \begin{bmatrix} 0.162 & 0.354 & 0.403 & 0.081 \\ 0.162 & 0.354 & 0.403 & 0.081 \\ 0.162 & 0.354 & 0.403 & 0.081 \\ 0.162 & 0.354 & 0.403 & 0.081 \end{bmatrix}$$

$$\begin{array}{ccccc} K = & (0.162 & 0.354 & 0.403 & 0.081) \\ 0 & FS & I & O & FB \end{array}$$

Regular stochastic matrices, such as the Markov model of assets flow described in the above matrices ("P"), have the property that when raised in power, all rows tend to converge to a unique vector, K, which represents the average percentage of times an asset is sold to each participant class during a asset's duration in the system. By using this formulation, we can estimate the importance a class has in the overall evaluation of a asset. Based on the example equation above, we would estimate that I and O classes were responsible for 75% of the trades of this asset.

In order to calculate intermediation, we must account for the assets flow into and out of the system as it relates to the Markov matrix. By simulating the passage of contracts through the Markov transitions and taking into account the entrance and exit points for an asset from the system, the correct rate of intermediation can be determined. Given that in the above matrix *P* that 90% and 10% of assets arrived from FS and O, and 90% and 10% of assets exited FB and O, we would expect that the intermediation rate would average twelve.

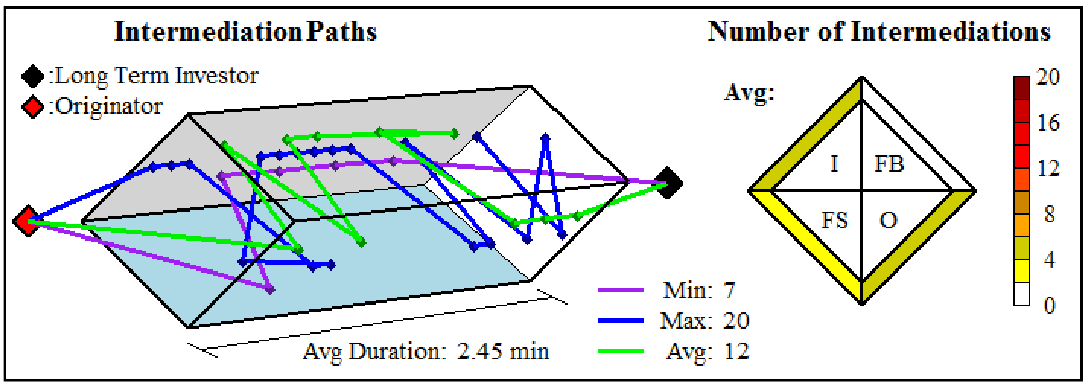


Figure 4.10: Plot of intermediation path and number for the example calculation

Figure 4.10 illustrates the example path of an assets intermediation as a three dimensional path where the four sidewalls of the cube represent ownership class, and lines showing the trading path of a contract. This is helpful since it can give a glimpse into how each participant class is part of the development of the pricing of an asset. Through examining this journey, we can more visually comprehend the impact a participant class has on the length of a journey (i.e. intermediation), as well as quantify the significance of its role on the pricing process by specifically monitoring for inventory changes as price changes.

In Figure 4.11, the plot shows the average intermediation observed using the E-Mini S&P 500 Futures data series. Of particular note is the visible drop in intermediations during the last four hours of trading, as well as, the final decreased intermediation that occurs as normal markets close in the US when trade volume usually increases dramatically as individuals try to zero out their risks.

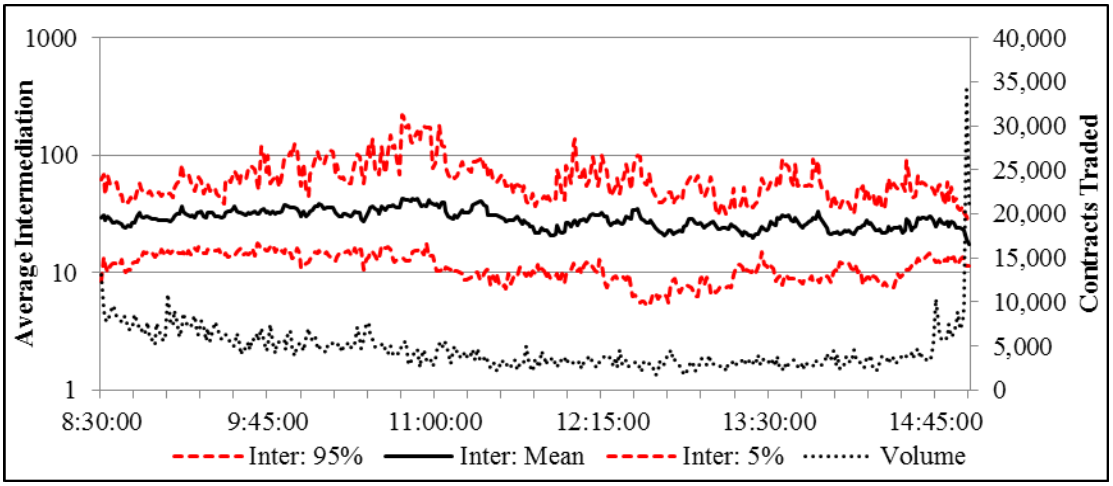


Figure 4.11: Plot of average intermediation of a contract thought out the month of February

When the intermediation path for an asset lengthens, there may be cause for concern that the capacity of the system is being reached. This type of trading is known as “hot potato” trading, where unwanted positions continue to be traded from one dealer to another following an initial customer trade (Burnham, 1991) and can lead to violent price swings. Such events in the Markov State model will look like absorbing or non-communicating states, which, if left for long periods of time (measurable by l_1-norm), are a possible warning sign of a phase shift in the system’s ability to price stability.

4.4 RESULTS OF THE MARKOV STATE MODEL

Using the active weekday trading hours of 8:30 AM to 3 PM CT of the E-Mini S&P 500 futures data from February 2011, we formulated our Markov State model to the market data provided by the CFTC. The following sections discuss the results for each of the variables that describe aspects of the Markov State model (i.e. the capacity (n(I(i)), the change in dynamic of trade (l_1-norm), and the intermediation rate) for the purpose of detecting a phase transition in pricing.

4.4.1 Capacity Results

In examining the capacity of the exchange, we found that the inventory limits of the I class participants showed a strong negative correlation to price change, as seen in Figure 4.12. Though this does not show direct causation, it does show that price tends to be negatively impacted if I class participants are unable to control their inventories effectively and quickly.

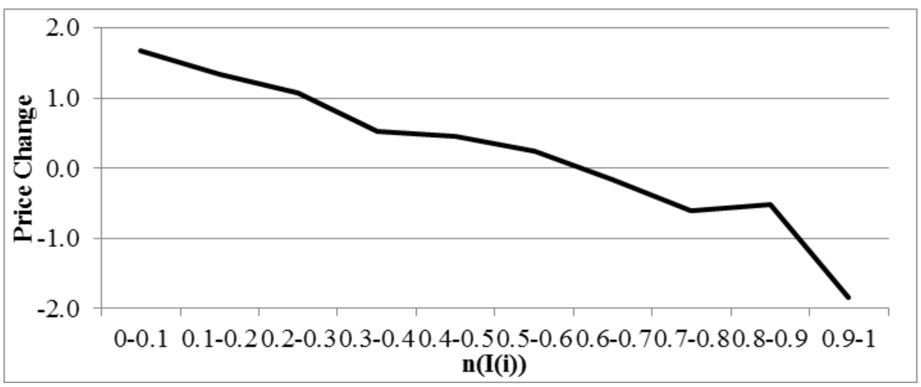


Figure 4.12: Average price change to n(I(i))

4.4.2 Trade Dynamics Results

In examining the trade dynamic of the exchange using the l_1-norm, we were interested in seeing if there were any similar correlations with price change as was seen above in capacity I(i). When looking at price changes at one minute increments, we found that the distribution of price changes was roughly normal with a mean of 2 for the l_1-norm and that the largest price changes occurred roughly at the means, as well. Hence, we found no significance in the relationship of l_1 to a large price change (see Figure 4.13).

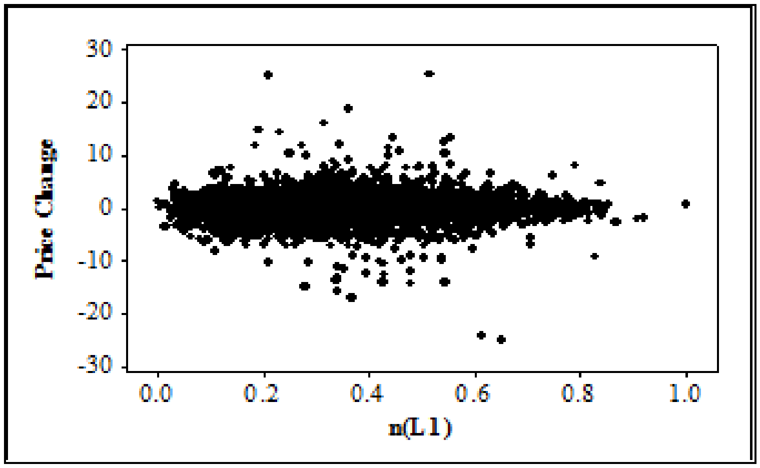


Figure 4.13: Scatter plot of price change as it relates to l_1

It was found, however, that the number of price changes (what we have termed "price movements" in Figure 4.14) seen over a minute increased as the l_1-norm value approached zero. This suggests that, when market participants trade in the same manner with one another for an extensive period of time, such that their trade decision drivers are not changing, the ability of the exchange system to support stable pricing decreases and prices can change rapidly.

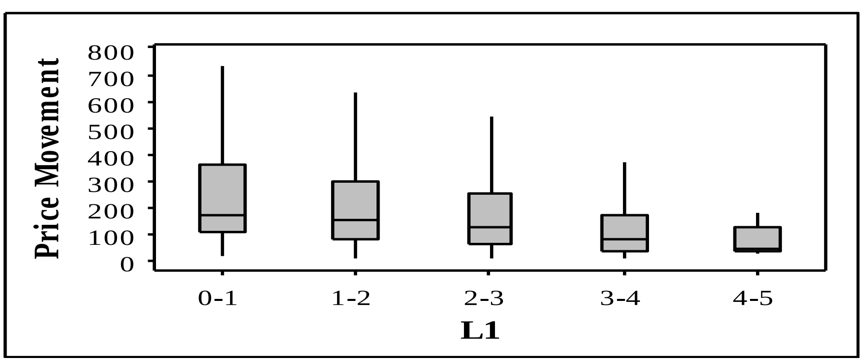


Figure 4.14: Chart of price movement as it relates to l_1

4.4.3 Intermediation Results

In examining the level of intermediation occurring in the exchange we found that it showed a strong negative correlation to price change in the contract. Figure 4.15 shows that as intermediation increases the likelihood of a price change being negative increases.

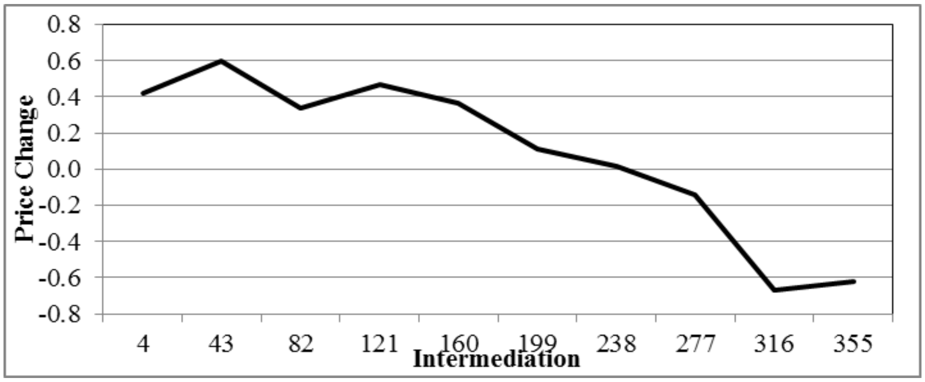


Figure 4.15: Average price change to Intermediation

4.4.4 Detecting Phase Shift Results

Given the above results for the separate measures provided by the Markov State model, we also explored the February data set of the E-Mini for phase shifts in the markets pricing by looking at combinations of the above three variables. We examined both the top and bottom quintiles of I(i), where capacity of the I participants inventory limits would be reached that had been correlated with price change. We also examined the bottom quintile of l_1-norm when the state of the system showed less flexibility and was correlated increased price movement, and the top quintile of Intermediation, representing a "hot potato" trading situation; both of these are potential warning signals of large demand or supply shock pricing. Figure 4.16 shows the distributions of these three variables and the cut off points selected based on the quintiles.

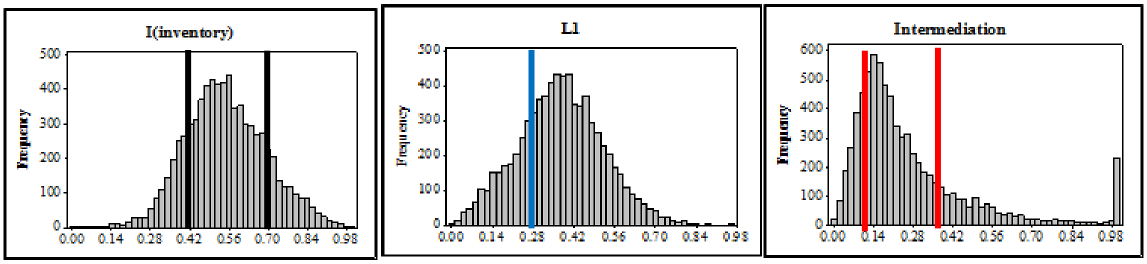


Figure 4.16: Histogram of I(i), l_1, and Intermediation with quintile extremes

By combining these three variables, we created a simple binary counting system that measured the count of how many of the variables had crossed the quintile thresholds. By using these thresholds, we wanted to determine if there were any early detection capabilities to permit a large price change or lower price stability (larger number of price movements) situation to be detected. Our expectation was that if I(i) was in lower quintile, l_1 and Intermediation where both in their upper, price would be significantly higher. However if I(i), l_1 and Intermediation where all in their upper quintiles, price would be significantly lower. Figure 4.17 shows the full minute by minute time series of the three variables implement on the E-Mini data, along with the quintile thresholds.

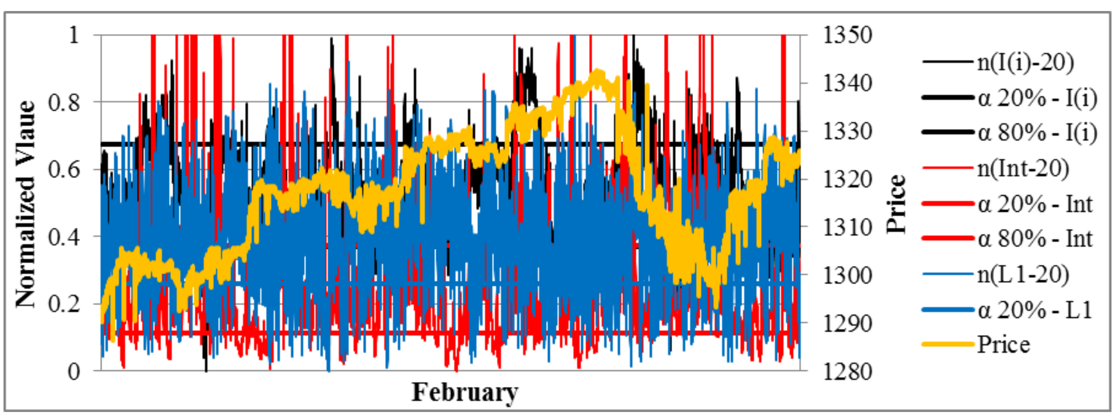


Figure 4.17: Time series of I(i), l_1, and Intermediation with quintile extreme variables

The results of this analysis (seen in Figure 4.18) found that, on average, price was higher when I(i), l_1 and Intermediation were in the first quintile; however, these results were not found to be statistically significant. The results of test of l_1 in the first quintile, I(i) and Intermediation in the fifth quintile, also found price was lower, but this result was also not statistically significant. As a test of the strength of the relationship seen in the E-Mini data set, the analysis was rerun on one month of WTI Crude Oil market data; where a similar relationship was found.

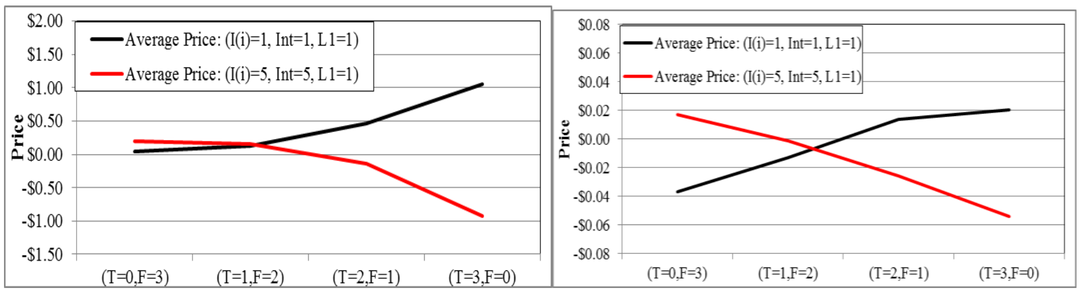
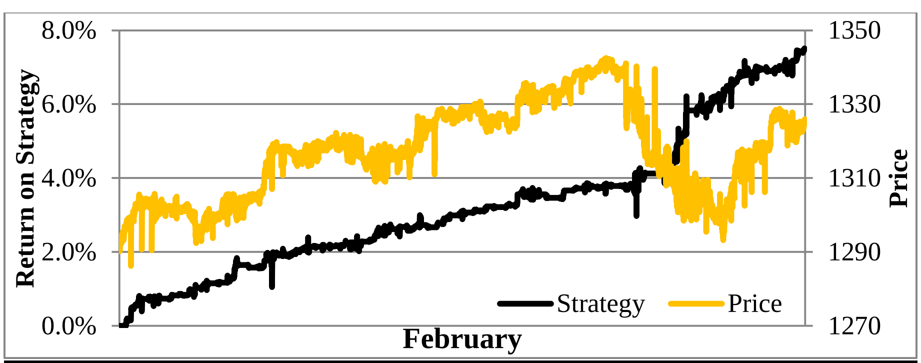


Figure 4.18: Average price change depending on I(i), l_1 and Intermediation in E-Mini (left) and WTI Crude Oil (right)

To demonstrate the predictive capabilities of this model, a simple trading strategy was constructed using the three variables from the model, where one would buy or sell-short at a rate of 33%, 66%, and 100% of available investment capital based on which quintile each variable was in (see table in Figure 4.19 details). The outcome of this simple strategy, which simply used the model, was able to produce an 8% return over the month.



Trade	%	Q(I(i))	Q(Int)	Q(l_1)
Buy	33	1	-	-
Buy	67	1	1	-
Buy	100	1	1	1
Short	33	5	-	-
Short	67	5	1	-
Short	100	5	1	5

Figure 4.19: Example trading strategy in table and graph of resulting strategy implementation

4.5 CASE STUDY: Flash Crashes Price Events

There is good potential for regulators and exchanges to use this approach to accurately identify stressed markets in advance of large price shocks to enable abnormal market price discovery states to be controlled. To understand if this model could prove effective, we examined how it would have performed in interpreting the E-Mini S&P 500 Futures price decline on May 6th 2010, a precursory event to the “Flash Crash” [CFTC & SEC. 2010] when the price dropped $100 over a 4 minute period at 1:45PM CST. At that time, it reached its low and triggered the CME circuit breaker; afterwards the contract’s price was able to recover.

In Figure 4.20, the three variables of the model are normalized to the E-Mini sample month data to permit a comparison with the normal along with the 80% and 20% alpha values, such that abnormally high and low values can be identified. In this figure, we can see that the I’s inventory had been abnormally high and stressed throughout the day and, just prior to the crash, was stressed to nearly double its normal limit while the intermediation of the contract increased dramatically, as well.

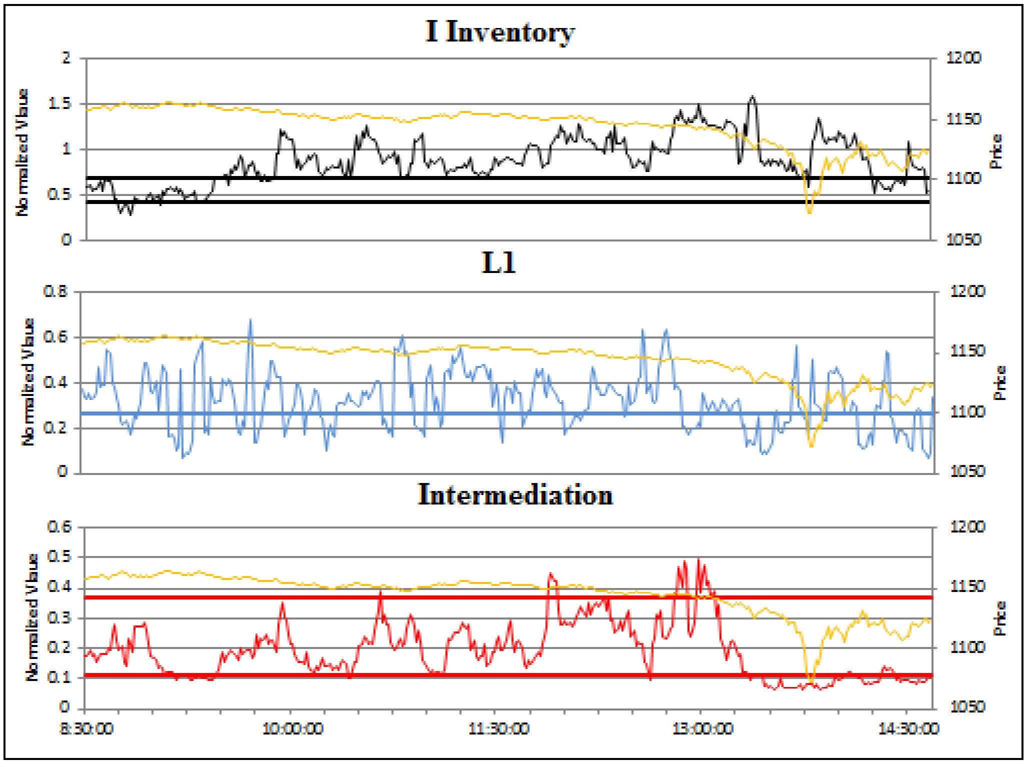


Figure 4.20: The E-Mini May 6th 2010 Results from the model

A “hot potato” like trading situation was created, where the I class participants found it difficult to sell off their inventory causing them to change their trading tactics. As a result, as a group, they began to decrease their buying and only sold-off their inventory; thus, they individually lowered their risk by removing themselves from the market. This action to collectively decrease inventory was persistent, as L1 shows, which resulted in lowered intermediation and a sudden decrease in price.

The actual formal investigation into this event, which took 5 months to complete, determined that three large forces were at play. First, it was a predominantly seller’s market, as is visible by the higher inventories being held by the I class participants. The second was that a large selling algorithm had gone awry, which could promptly be seen at 1 PM by the inventory levels of I. Lastly, this selling algorithm caused a “hot potato” like trading situation among high frequency traders trading contracts between themselves as prices began to fall, which could be seen in the intermediation of contracts around 1:15 PM.

Though these results have the benefit of hindsight, the model approach helps to highlight the importance of having a mix of divergent trading views in a market system, which if not maintained, can lead the system to a phase shift in its objectives in which all participants suddenly end-up prioritizing their trading actions to remove risks rather than seek the price. Though a market system in this state can naturally right itself; participants and regulators should consider if the manner in which trading is conducted during such a transient period of turmoil serves the public's interest.

4.6 SUMMARY

As markets have gone electronic, the ability to monitor their function has deteriorated due to a loss of perceptions from the physical market floor interaction of participants and the use of automated trading systems to handle a very significant portion of all trading actions. The Markov State model developed in this chapter takes a complex systems perspective of markets systems by reconstructing the exchange as a system of systems wherein the participants are subsystems that the infrastructure provided by the exchange system combines to solve for a demand and supply equilibrium while achieving the basic objectives of trade and stable pricing.

Through the application this model, we are able to identify important features of the periodic complex behavior of supply and demand equilibrium while price discovery takes place in the markets. By recognizing characteristic features using the model, it is possible to better understand the state of a market. As was illustrated, it is possible predict the likely direction of price movements for periods when elevated stress in the market participants' inventories and actions can be observed via the model. Using such a model, regulators can regain some of the floor perspective of the markets that is hidden in the transactional data they currently receive; providing them a means of monitoring the stress being felt by a market.

Chapter 5

Agent Based Simulation and Combinatorial Complexity

Markets are fundamentally complex systems having numerous participants that all have their own individual ever-adapting views and decision-making processes; this makes the markets' granular movements very unpredictable. Although a market's daily outcomes can reflect many complexity effects, the manner in which a market behaves is still governed by three relatively well defined interacting parts: participants, trading rules, and external environment. As such, when an extreme event occurs in the price discovery process of a financial instrument, one of these component areas is very likely to be closely connected to the root cause of the event.

As systems becomes more complex, the ability to connect macro-level behavior with the changes in the underlying micro-level behavior. This is due to the combinatorial complexity effect where uncertainty grows more complicated with time because the future events depend on the string of past decisions; making it difficult to pinpoint the cause of any small change in price. Typically when larger moves in the price of an instrument occur, external environmental variables can be identified, such as a news announcement that impacts price directly due to a fundamental relationship between the news and the instruments evaluation. However when a large price change occurs that don't have an obvious culprit to connect to the move, concerns of manipulative trades by the participants or the functionality of the markets governing certain trades are pointed to as possible the root causes, which are difficult but important to identify.

Since market events typically occur as a result of specific conditions that can be quite unique and short-lived, and, thus, it is difficult to perform analysis and draw comparisons with other events to easily identify root causes. Secondly, the events' datasets, though detailed, are just stamps of the past and, thus, do not easily lend themselves for use in an experimental framework such that tests, beyond statistical data comparisons, can be performed to validate a hypothesis. Agent based simulations (ABS) offer a means of re-creating an event in a simplified, interactive model. An appropriately designed ABS can provide a means to adjust market variables in a controlled system such that experimentation can be performed to reveal the impact on a system's behavior of changes to specific variables of interest. Through multiple Monte Carlo simulation runs of systems, the effects of adjusting variables can be examined in depth. Such active exploration may help identify the reason for an irregular price event, or what the impact of implementing a new rule change may have on the market's behavior going forward.

In this chapter, ABS's of financial markets are explored as possible means to work around the combinatorial complexity problems encountered when trying to work with only one-off real market data sets that are created by a unique set of market condition. In particular, this chapter examines how an ABS can be used as an experimental platform to identify what information about a market's conditions maybe significant in providing indication of poor stability and resiliency that has a high risk of precipitating a price flash event. Using datasets from the ABS's, resiliency measures are examined for their predictive power in identifying situations that are suggestive of a market problem. The development of such measures and others can help to improve the understanding of the stability and resiliency of a market.

This chapter focuses on how ABS's can serve as a tool to develop a better understanding of cause and effect relationships in the markets. By using this tool, regulators can study at past events to in an effort to predict problematic situations and create prevention techniques for the future. Section 5.1 the benefits and frame work of a financial market ABS for the evaluation of markets. Section 5.2 how to validate the market such that discusses the combinatorial complexity that creates the price discovery process demonstrate the macro stylized facts of pricing data. Section 5.3 how to use the ABS to investigate market anomalies. Section 5.4 looks at how ABS can be used to help learn further knowledge about the market such that forecasts can be developed. Section 5.5 discusses how these tools can finally help to test potential solutions and other rule changes and what they may look like if implemented. Section 5.6 is a summary of the role agent based simulations can play in dealing with combinatorial complexity.

5.1 THE AGENT-BASED SIMULATION FRAMEWORK

Regulators, facing a quickly evolving and complex environment, require new and flexible tools and decision aides to evaluate policy proposals (Hayes, 2012). Agent-based modeling, which is capable of capturing the organization of exchanges, intricacies of the trading process, and the heterogeneity of market participants, is a powerful method for analyzing financial market (Bookstaber, 2012). Agent based simulations (ABS) are simplified models of complex systems such that include a set of individual agents, a topology and an environment (Farmer, 2005).

In a typical ABS of a financial market, the market participants are agents, the market mechanism is the topology and the exogenous flow of information that is relevant to market is the environment. As a system's complexity increases, the ability to directly correlate its macro-level behavior with the changes in the underlying micro-level behavior and parameters diminishes. This requires that the micro-level behavior and parameters be more fully examined and defined to understand their relationship to the macro outcomes. ABS can provide regulators with an experimental environment that can comprehend complex system outcomes.

Designers of ABS's for this purpose face a trade-off between complexity and explanatory power. In the context of financial markets, for example, agents are endowed with varying degrees of sophistication with regard to how they adapt, predict and optimize within their environment. In creating an ABS of a financial market, it is necessary to determine what features are necessary for creating a proper representative market without overly complicating the framework. As a market is a system constructed of the participants and the rules governing their interaction (e.g. trade and pricing), it is necessary to focus on these two parts and set aside, for simplification purposes, the environment aspect which is an uncertain and has indefinite set of conditions. This selection to focus on the agents and the topology is important to note since it presumes that the market participants in this simulation will focus on their individual conditions and the single market in which they operate and not on any other market or economic condition that could be affecting their decisions.

Of these two components, the agents are more difficult to specify as their design needs to mimic that of true market participants. Current literature suggests that the markets are divided into subcategories of participants and the combinations of trading styles that are responsible for emergent market events. These combinatorial aspects led to the design of multiple categories of trading agents in the simulation. From work done by Kirilenko et al. (2011), it was possible to characterize market participants in to one of six categories of agent types:

i. & ii. ***Fundamental Buyers and Sellers****:* take long or short positions on the asset during the entire duration the markets exists and trade with a low frequency.

iii. ***Market Makers:*** take the position of straddling both sides of the market by taking long and short positions on an asset. These intermediaries' trades are meant to give the market liquidity.

iv. ***Opportunistic:*** take long or short positions on the asset during the duration of the market day like a fundamental participant. However, they implement trading strategies that make them resemble Intermediaries because they do not take a large position.

v. ***High Frequency****:* take long or short positions on an asset for short periods and trading with high frequency near the best-ask and best-bid sides of the book. HFTs in the simulation use a simple momentum strategy. As the bid/ask queues becomes imbalanced, HFTs will tend to trade in the same direction as the imbalance. In other words, if there are excess bids in the order book, HFTs have a higher probability of placing a buy order; such a strategy was indicated in real market data provided by the CFTC. In general, HFT's will allow themselves to take large positions for short periods of time, but will try to be neutral by the end of day.

vi. ***Small:*** take either long or short positions on the asset during the entire duration of the markets exist and trade with a very low frequency.

The foregoing descriptions of participants in the market require that they be represented through a set of variables that can be used to design the autonomous agents in the simulation. As depicted in in Figure 5.1, four key quantitative components were extracted through decomposing the agent descriptions: frequency of action taken, inventory limits, a distribution to the order sizes they placed, and a distribution of orders prices they placed in reference to best bid/ask. Taking this approach, it is possible to extract the necessary information from real data sets such that a specific market situation can be rebuilt with agents defined by these features and placed with the same relative population proportion into the simulation.

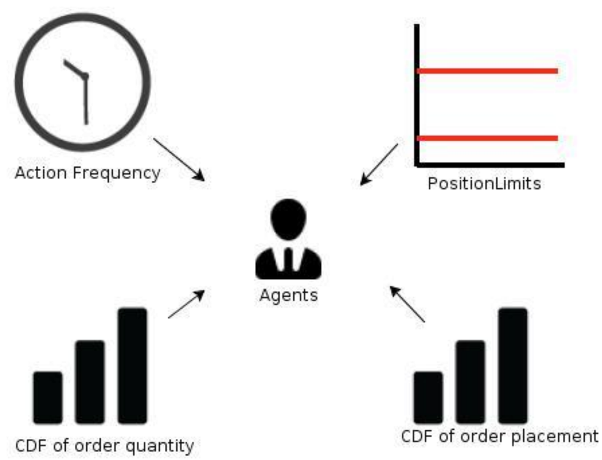


Figure 5.1: Components of an agent class

The second component of the simulation is topology. Although it is rather complicated to recreate the design and rules used in an order book system, they are known and set. This certainty in topology makes the explanatory power of these ABS's stronger since they do not require interpretation due to the transparent structure upon which they are built. This allows for asset price creation to emerge from individual actions (e.g. order, cancel) of market participants and the market matching engine connecting them to create trade.

5.2 VALIDATING AN ABS OF A FINANCIAL MARKET

In order to implement an ABS as an "experimental laboratory" that will allow one to perform in-depth analysis, the laboratory itself must be validated. In creating an ideal laboratory, it is necessary to first demonstrate that one can effectively replicate the features specific to the market scenario one wishes to use to perform experiments later. To reach this level of quality in an ABS, it is necessary to achieve some basic characteristic and behaviors of a financial market scenario of interest.

As an initial step, gathering all the necessary parts to build the ABS is required, namely the markets rules and data of the time period in question. The market rule and structure allow the ABS topology to build such that is resembles the exchange. Taking data of the market period under investigation, specifically order level data, agents can be created. Using the classification system of agent types, based on Kirilenko *et al*, each participant in the market can be categorized and finally added to the artificial market parameterized to the statistics of each participant. An example of a set of participants' statistics is provided in Table 5.1, which describe what was observed in aggregate of the E-Mini S&P 500 Futures market.

Table 5.1: Listing E-Mini S&P 500 market participation rates

Participant Type	# of Participants	Arrival Speed	Position Limits	Market Volume
Small	6880	2 hr	-30 – 30	1%
Fundamental Buyers	1268	1 min	-∞ - ∞	9%
Fundamental Sellers	1276	1 min	-∞ - ∞	9%
Market Makers	176	20 sec	-120 – 120	10%
Opportunistic	5808	2 min	-120 – 120	33%
High Frequency	16	0.35 sec	-3000-3000	38%

With the agents and the correct market topology in place, the model must be validated to see that that it resembles the manner in which that market typically functions by assessing the "stylized facts" (Kaldor, 1961) of the pricing series associated with that market. Stylized facts are a set of statistical characteristics found in the price time series data which, as stated in chapter 4, represents the macro dynamic of the underlying demand and supply of individual participants in the market. These features include the distribution of price returns, volatility clustering, absence of autocorrelation of price returns, and aggregation of price returns.

The following subsections describe the stylized facts requiring comparison between the simulation and the real market data. Each validation includes an example of how the E-Mini S&P 500 Futures ABS result compared to the real market's data.

5.2.1 Distribution of Price Returns

It has been widely observed that the empirical distributions of financial returns and log returns are fat-tailed. Mandelbrot (1963) observed that the tails of a distribution of prices changes are extraordinarily long and the sample second moment of price typically varies in an erratic fashion. This has caused various suggestions regarding the form of the distribution, ranging from the Student-t, hyperbolic, normal inverse Gaussian, and others, but no general consensus exists for the form of the tails for all markets. In the observed data from the S&P 500 E-mini market and its ABS recreation, the normality of the distribution of price returns, as seen in Figures 5.2, illustrates that both the real and simulated data diverge from normality at the tails.

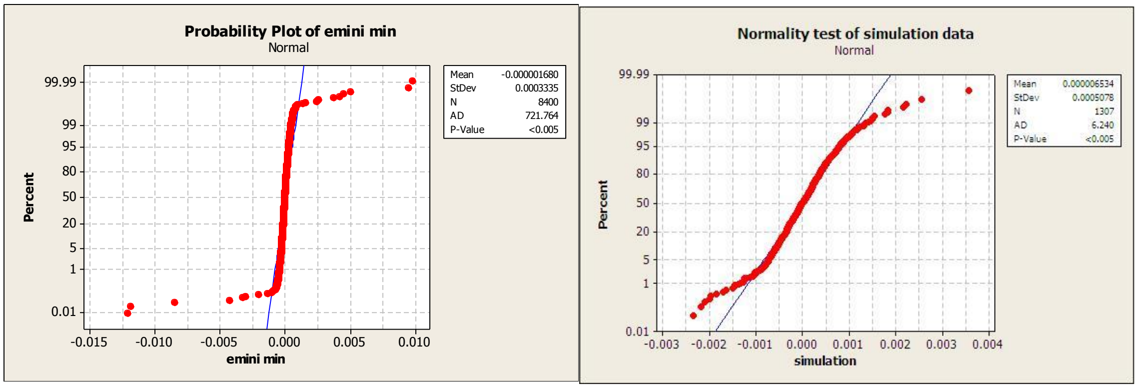


Figure 5.2: Normality test of real E-mini price returns (left) and simulated price returns (right)

5.2.2 Volatility Clustering

The characteristic of volatility clustering is seen in the absolute price returns for securities that have slowly decaying autocorrelation in variance (i.e. price changes tend to follow other price changes of the same size). This was first noted by Mandelbrot (1963) and was finally translated into agent based models by Kirman and Teyssiere (2002), when they discovered a model would exhibit autocorrelation patterns in the absolute returns if a variable was herded by the positive or negative opinion of an asset:

$$Herd_{t+1} = Herd_t + UniformDistribution(-N, N)$$

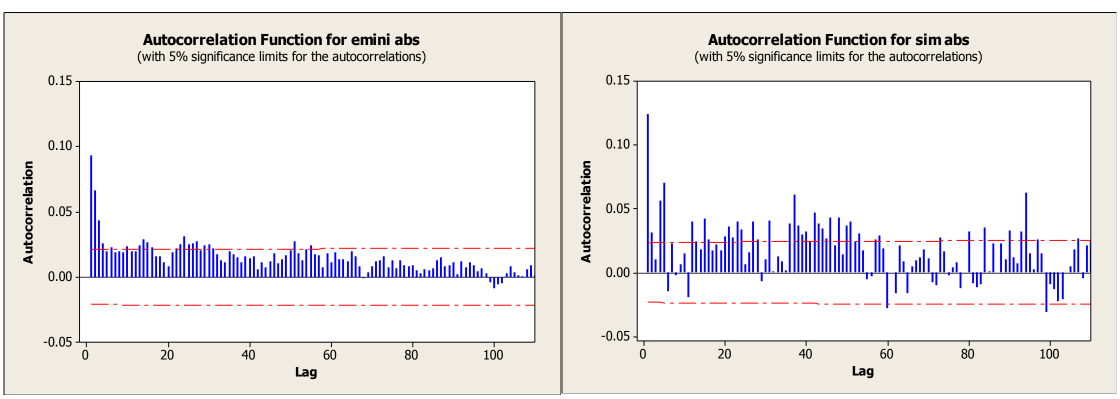


Figure 5.3: Volatility clustering of absolute price returns for S&P 500 E-mini (left) and simulation (right)

The simulation implements the same herding variable to influence the decision of opportunistic participants in the model (see Figure 5.3). This creates a similar autocorrelation pattern in the absolute returns that is seen in the S&P 500 E-Mini Futures contract's real minute price returns.

5.2.3 Absence of Autocorrelation of Returns

In validating that markets are efficient, it has been common practice to show that there is no predictability in the price returns of assets. In demonstrating this, the autocorrelation of returns to show that there is no predictability of markets. Figure 5.4 illustrate that this property exists in both the real and simulated market data.

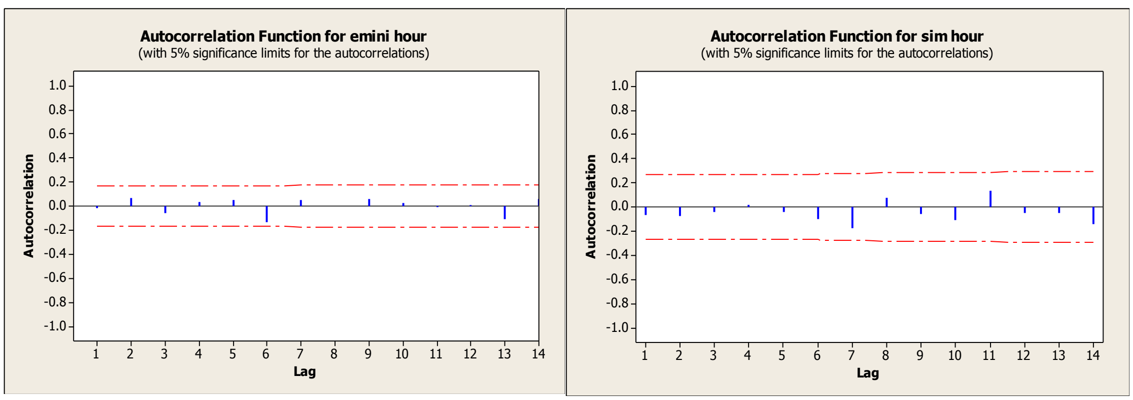


Figure 5.4: Absence of autocorrelation for price returns for S&P 500 E-mini (left) and simulation (right)

5.2.4 Aggregation of Returns

The final stylized fact, aggregation of returns, shows that as one increases the time scale over that one measures between price returns, the distribution approaches the Gaussian form. This cross-over phenomenon was noted first by Kullmann *et al* (1999), where the evolution of the Pareto exponent of the distribution with the time scale was studied. Kyle and Obizhaeva (2013) shows that this distributional change is a result of the aggregation of informational units (orders) such that they will only look the same if the velocity (the rate of orders are submitted to a market) is equivalent.

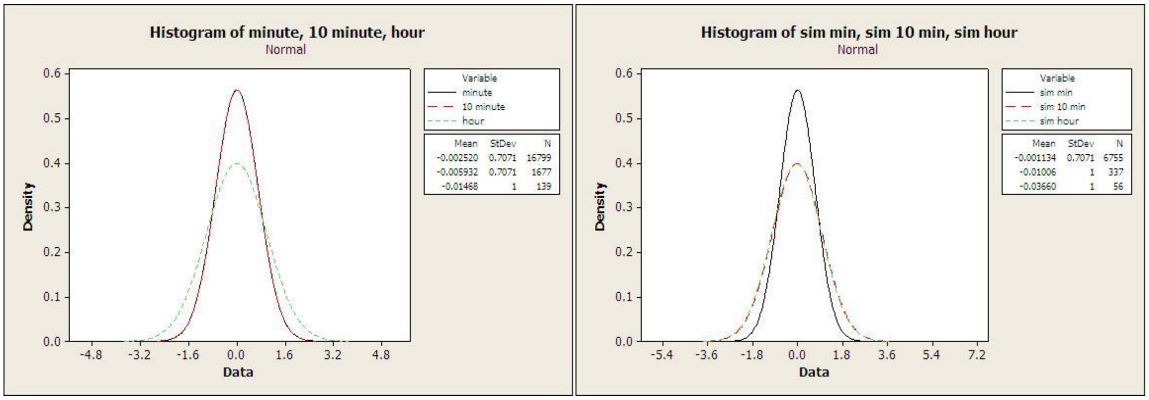


Figure 5.5: Aggregation of price returns for S&P 500 E-mini (left) and simulation (right)

Figure 5.5 illustrates the standardized distributions of returns for S&P 500 index and the simulation, for one minute, ten minute, and sixty minutes. As hoped the longer the time scale, the more Gaussian both sets of distributions become and that observed distributions have the same moments reflecting that the velocity of orders flow in the two markets match informational value.

5.3 INVESTIGATING WITH AN ABS OF A FINANCIAL MARKET

Once an ABS laboratory is established and validated, regulators have the ability to examine specific market events of interest. Events can be examined for better understandings of root cause and variable dependencies, such as the impact of participants. The ABS framework can allow regulators to systemically appreciate the impact of certain participants or behaviors on the market; for example, the impact a participant/behavior can have on the market will typically be non-linear. Through the testing of the sensitivity of variables in a simulation regulators can learn whether a variable or value is insignificant and significant to the outcome of a complex system, typically a difficult, but important aspect to proper assessing a cause.

With the foregoing in mind, an ABS was created using the standard trading rules for limit orders and was calibrated to the stylized price and participant characteristics of a E-Mini S&P 500 futures market at the time of the May 6^{th} 2010 Flash Crash. During this event, the E-Mini market fell nearly 6% in 5 minutes, after which it recovered nearly as quickly (i.e. indicative of a price flash).

Price flashes are events characterized by the price of an asset drastically changing for very short periods of time until the market can recover naturally or until a 'flash crash' occurs that requires the exchange's intervention via price circuit breakers. Over 9000 such events were reported between 2007 and 2010 wherein an asset's price change exceeded 0.8% within 1.5 seconds (Nanex, 2011). However, it was not until 2:45 PM EST on May 6^{th} 2010 that their presence made headlines, when most major equities, futures, and commodities markets plunged in unison causing over a trillion dollars of asset value to be lost over a few minutes before a recovering.

The CFTC-SEC Staff Report (2010) on the events of May 6^{th}, 2010 Flash Crash identified an automated execution algorithm that sold a large number of contracts as one catalyst for the crash. The algorithm, which was being run on the E-mini S&P 500 futures market, kept pace with the market selling about 9% of the previous minute's volume. However, this was only the trigger for a cascade of selling by ATS's, particularly high-frequency trading firms, which Kirilenko et al (2011) described as "hot potato trading" where these firms rapidly acquired and, then, liquidated positions among themselves, resulting in steadily declining prices. These events unfolded as described in Table 5.2 below.

Table 5.2: Events that occurred during the Flash Crash of May 6^{th}

Event	Model Price Points	Effect
Large Sell Execution Algorithm Initialized	-	Large aggressive sell orders enter the market
Market makers begin to retreat from the market as the price falls	24 ticks below the moving average	Market depth beings to disappear
Fundamental participants withdraw from the market as stop-loss orders are triggered	70 ticks from price at start of day	Market depth disappears and more sell orders are executed
A market pause is initiated by the exchange	Price $\downarrow \geq 1.3\%$ per second	Gives time for slower market participants to enter trades into the order book

Considering these events, an agent was created to represent the large sell algorithm during that day and the other participants were added in the same proportions as observed during those days. Figure 5.6 illustrates the impact on the simulated price and the moving average volume at the moment in the simulation when the large sell algorithm participant is started. The graph of the actual E-mini S&P 500 flash crash is shown for comparison.

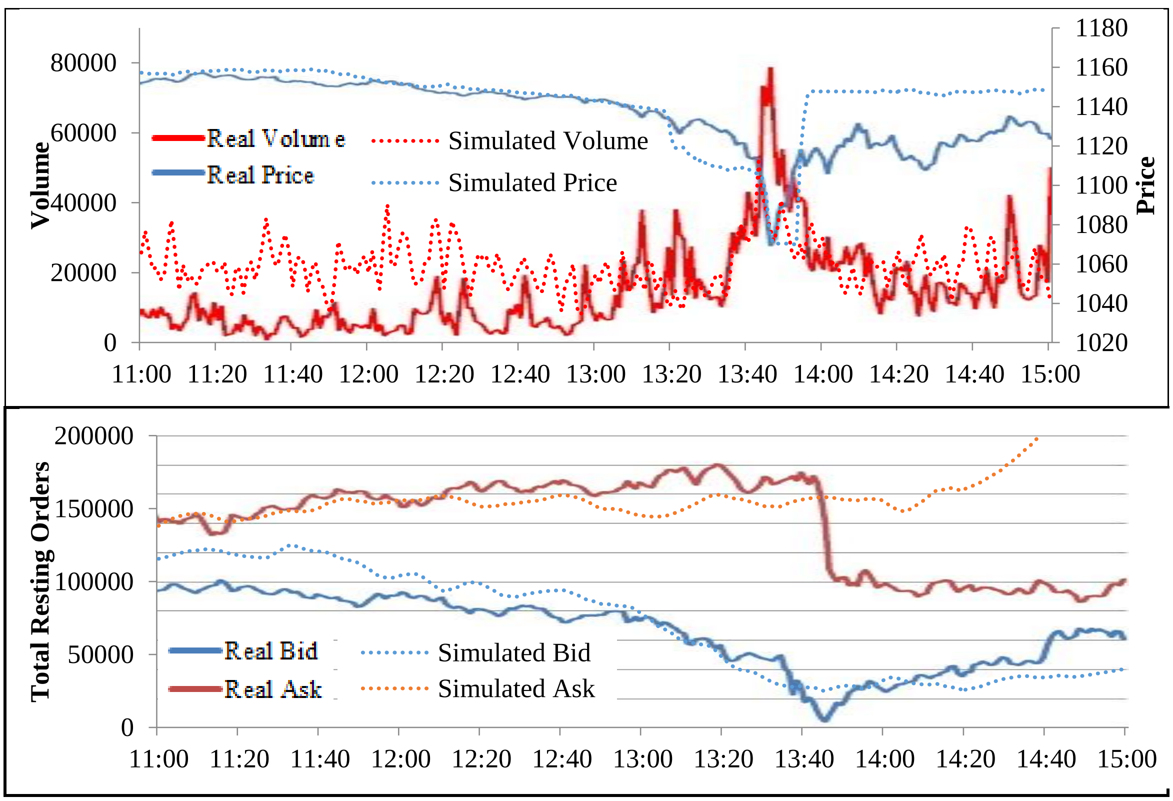


Figure 5.6: Real vs. Simulated E-Mini: Volume, Moving Average Price and Order book Depth

Using the ABS laboratory, created to mimic the event of May 6th, the two culprits that the CFTC & SEC report named as causes of the Flash Crash, the large selling algorithm's impact and high frequency participants' impact, were examined further for their possible roles in causing the event. Two variables associated with these were varied over a series of simulation experiments using the ABS to see how these affected the price flash in the simulation (i.e., the lowest price was observed during these experiments):

i. the percentage of previous minute's volume that the large sell algorithm used to sell orders, and

ii. the number of high frequency participants that were present in the market.

A total number of 40 simulations were run for each variable pair and the minimum price in each simulation was recorded. The median price for each variable pair is illustrated in the Figure 5.7.

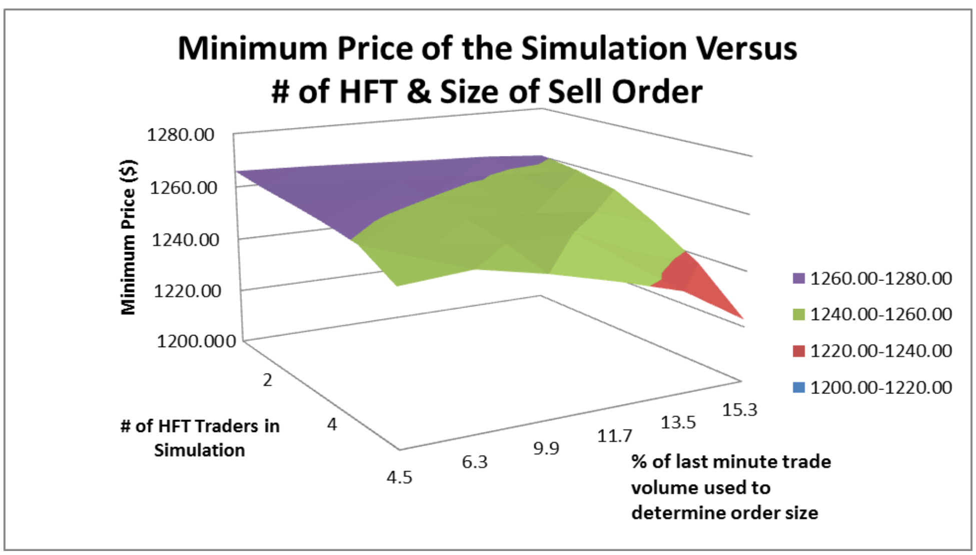


Figure 5.7: Minimum Price versus # of High Frequency Traders (HFT) & Rate of Size of Sell Orders

As the number of high frequency participants was reduced to zero, the minimum price result of the simulation increased. This illustrates that high frequency participants were necessary for the events on May 6th to occur. It was hypothesized in the CFTC & SEC report that the high frequency participants had to have traded amongst themselves, as liquidity dried up, to create the rapid trading resulting in what was termed the "Hot Potato" effect, which was concluded to be the cause of the flash crash.

In examining the large sell algorithm's impact, the sell volume percentage executed by the simulated agent was decreased, which lead to the severity of the price drop also decreasing. Although, the large algorithm selling contracts causes the price of the simulated contract to drop, because of the large quantity it want to sell, the rate at which it sells the contracts has a larger impact on creating the price flash. The rate at which contracts being sold has a non-linear in relationship to the drop of the market.

The final conclusion is that the flash crash would not have occurred without both the HFT's and large sell orders execution algorithm. The foregoing is a simple example of how an ABS can be used to test the soundness of conclusions from data analysis in the CFTC & SEC Report.

5.4 FORCASTING WITH AN ABS OF A FINANCIAL MARKET

As the previous section demonstrates ABS's, are excellent tools for constructing market experiments 'in silico' and provide a means to study of the causes of disruptions in the market in the past. ABS's, however, also offer opportunities for forecasting by enabling the study of market disruption pre-conditions to detect the presence of risks in advance. As ABS's enable replications of and variations to the complex sets of interactions that occur in a market, they can generate a variety of data sets of interest to support development of robust means of detection and forecasting.

With the large increase in price flash events in markets today, there is a need for the development of a tool capable of signaling an increased risk for the occurrence of a flash event. As an extension of the investigation of the May 6th 2010 Flash Crash's causes, work was done using the ABS laboratory to see if basic precursors could be detected in advance. Testing for precursors could be useful in the development of risk avoidance protocols for market participants and, perhaps, help further inform exchanges and regulators interested in preventing such events from occurring.

5.4.1.1 Forecasting Variables

In the case of forecasting price flashes, an effective indicator must reliably be able to signal the risk of a large price change by detecting changes in the market's price discovery process based on monitoring a real time stream of available data, such that it can determine the pricing process is not functioning properly. Unfortunately, there is no unique and widely accepted definition of what is required to define a well-functioning and healthy market. However, it can be said that if a market allows an "asset to be transformed into another form of asset in a short period without losing its value, i.e. change in price" (Benic, 2008), the market is considered to be liquid and working efficiently.

It is difficult to fully encompass the liquidity concept in a single measure, as it is a multi-dimensional phenomenon. Harris (2003) defined four dimensions of financial market liquidity:

i. *Trading time:* defined as the ability to execute the transaction immediately at the current price. The waiting time between trades is the measure for trading time.

ii. *Tightness:* the ability to buy and to sell an asset at about the same price at the same time. Hasbrouck [2004] argues that tightness shows the cost associated with transacting or the cost of immediacy. Measures for tightness are the different versions of spread.

iii. *Depth:* the ability to buy or to sell a certain amount of an asset without influence on the quoted price. A sign of illiquidity would be an adverse market impact on price when trading occurs.

iv. *Resiliency:* the ability to buy or to sell a certain amount of an asset with little influence on the quoted price. While the aspect of market depth regards only the volume of best bid and best ask prices, resiliency takes the elasticity of supply and demand into account.

Of these variables, resiliency of a market may be the most important dimension to measure because flashes occur when liquidity is 'low' and prices can change quickly (i.e. low resiliency). Measuring resiliency requires an understanding of the current state of the market, as well as, the inventory of resting orders indicating strength in demand and supply. It also requires an understanding of the impact actions can have on the market since orders flowing into the market will change its state.

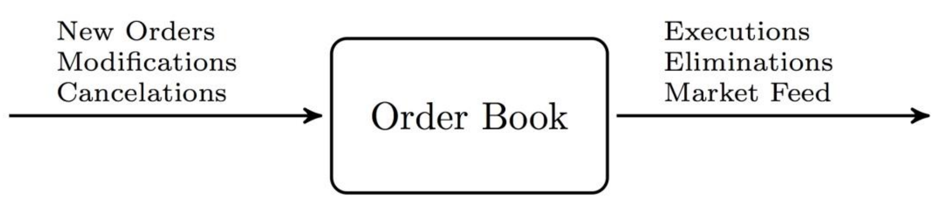


Figure 5.8: Market State Diagram

The diagram in Figure 5.8 depicts the market state. It is characterized by order IN flows consisting of new orders and orders that are modified to increase the quantity which increase the inventory of resting orders held in the order book. Order OUT flows are completed orders that have been executed or canceled, and orders that are modified to decrease the quantity which decrease the inventory of resting orders held in the order book.

To-date, most of the study work done on examining a market's liquidity has been focused on the impact order out flows of the market have on price (Brandt, 2004; Love, 2008; Evans, 2002; Easley, 2010), but very little has been done examining the impact of order in-flows or the state of limit orders in the order book. This study examined five different indicators that measure the order flow and order book state to different degrees, and tested them as predictors of weakness in resiliency and, hence, the possibility of a market price flash.

5.4.1.1 VPIN

Volume-Synchronized Probability of Informed Trading (VPIN) is an existing indicator developed in Easley, López de Prado and O'Hara (2010); it is the ratio of average unbalanced volume to total volume in a period and requires Layer 2 data. Heuristically, the VPIN metric "measures the fraction of volume-weighted trade that arises from informed participants as the informed tend to trade on one-side of the market and so their activity leads to unbalanced volume" (Easley, 2010). This metric is parameterized using the volume bucket size (V) and the number of buckets needed to get an accurate measure (n). In the simulation V = 50, and n = 10.

$$VPIN = \frac{\sum_{i=1}^{n} \left| V_i^S - V_i^B \right|}{nV}$$

$\mathbf{V}$: *the volume in every bucket*
$\mathbf{V_i^B}$: *volume from buyer-initiated trades*
$\mathbf{V_i^S}$: *volume from seller-initiated trades*

n: *the number of buckets used to approximate the expected trade imbalance*

5.4.1.2 Price Impact

Price Impact in limit order book markets is an existing indicator that is based on a measure of the effect a trade has on an asset's value. For this study, the adverse selection component found by decomposing the price spread along the lines of Glosten (1987) and used by Hendershott et al. (2011) in examining the price impact of algorithmic trading on financial markets was selected. This is based on empirical evidence supporting the belief that order book liquidity and adverse selection are inversely related (Frey, 2006).

The Price Impact indicator is the cumulative response of the quoted price to a one-time unit shock in the order flow as a measure of adverse selection; this accounts for persistence in order flow which can be translated as an indication of sharp order flow imbalance (Hasbrouck, 1991). This metric requires using aggression (i.e. the side of the order book that initiated the trade) as found in Layer 2 data. It is then parameterized for a time in the future ($t+x$) to which the current trade price (at time, t) is compared (note: In this study, x was set to 30 seconds.).

$$Price\ Impact = \frac{Q_t(M_{t+x} - M_t)}{m_t}$$

Q_t: *an indicator variable (+1: for buyer-initiated trades, -1: for seller-initiated trades)*
M_t: *midpoint price is the average of the best bid /ask at the time of the trade (t)*
x: *time in the future*

5.4.1.3 Window Indicator

The window spread is new indicator developed as part of this work and is a measure of the price movement over a period of time. It is meant to demonstrate the range in current pricing due to market order flow without regard to quantity traded; hence it only requires Layer 1 data. This metric is parameterized to the length of time that the window cares minimum and maximum values. In the simulation this is set to 5 seconds.

$$Window\ Spread = \max_{t=0,\ldots,i}(P_t) - \min_{t=0,\ldots,i}(P_t)$$

$\mathbf{P_t}$: *prevailing midpoint price at the time of the trade (t)*

5.4.1.4 OBAW Spread

The Order Book Average Weighted Spread (OBAW) is new indicator developed as part of this work and is a weighted average bid price and ask price spread of the order book using the first ten best bid and best ask ticks' prices. The spread gives an aggregated measure of where the majority of resting orders in the order book. The smaller the value indicates, the closer in agreement exists on price between the majority of participants. OBAW requires Layer 2 data.

$$OBAW\ Spread = \frac{\sum_{i=1}^{n} P_i S_i}{\sum_{i=1}^{n} S_i} - \frac{\sum_{j=1}^{m} P_j S_j}{\sum_{j=1}^{m} S_j}$$

S: *order size*
P: *trade price*
i: *ask order* , **j**: *bid order*
n: *number of bid limit orders,* ***m****: number of ask limit orders*

5.4.1.5 PH Spread

The PH Indicator, also developed as part of this work, follows the same formulation as the OBWA indicator but it uses an altered order book in its calculation. Orders made by ATS's are not included in the average to remove any false indications of market resiliency that might be introduced (i.e. ATS orders may not provide a true indication of price spread compared to traditional orders). Through their removal, it was expected a better indication of the strength of the market to take 'acidic' order flow (i.e. orders that consume the depth of the book). The key caveat that comes with this indicator is that you must be able to categorize participants and classify their orders, requiring use of layer 4 data. ATS accounts were identified by the quantity of trading they did solely, 2000 or more actions (new, canceled, or modified orders) in a single market. A quantity viewed to beyond a human's ability to manage.

5.4.2 Testing Forecasts in Simulations

Market ABS's can be useful in understanding the features and characteristics of a market that underlie price discovery anomaly events and serve as a controlled means to create market data to test indicators. This was particularly important for this study as very little real Layer 3 and 4 electronic order book data is accessible for academic efforts.

The five indicators discussed in Section 5.4.1 were implemented in the market simulation and examined under the conditions of a normal market having relatively equal demand and supply and compared to a market with conditions of a heavy supply side sell off versus demand in which a flash crashes have occurred. This was done using marketable orders with increasing volume based on the previous minute's volume (i.e. the precipitating algorithm of the price flash that occurred on the E-Mini S&P 500 Futures market on May 6th 2010). Admittedly, this experiment focused on how one price flash occurred, but it is believed that it might help in demonstrating some stylized facts of what occurs to the underlying microstructure of the market during any such flash price event.

To test how well the indicators discussed above respond to the market conditions that arise due to both random chance, as well as, the direct result of the impactful actions of an individual or group, the simulation was run 2000 times keeping the initial conditions with respect number of participants, their behavior, and market rules constant while varying solely the selling behavior of a single large selling ATS. The simulated ATS had a preset selling volume percentage ranging from 0 to 9 percent of volume per minute, which resulted in a rapidly increasing probability of a flash crash above about 4%, as seen in Figure 5.9.

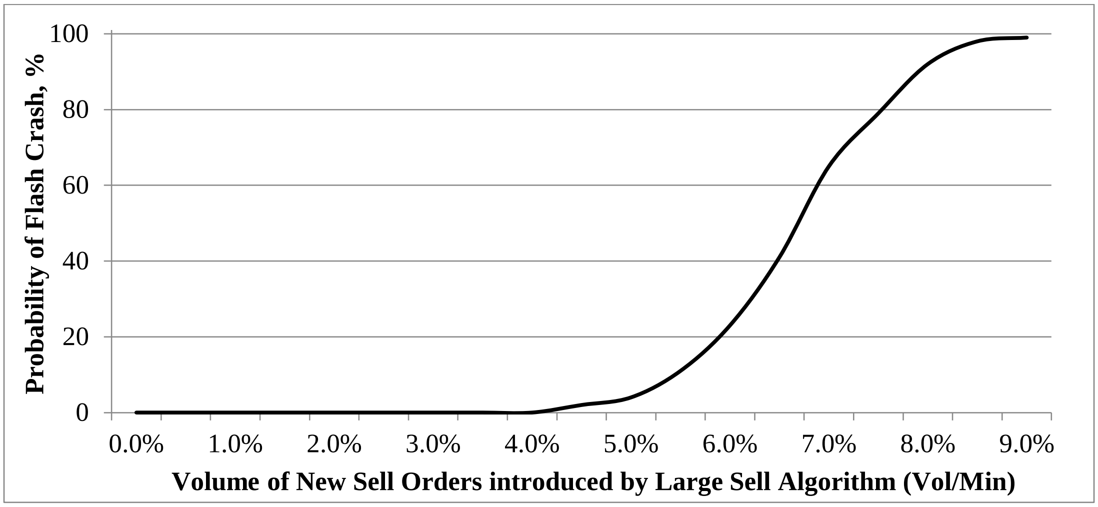


Figure 5.9: PDF of a Flash Crash in Simulated E-Mini Market

The indicator results observed from each of these simulations runs yielded the following set of distributions for the indicators under the two conditions of a normal price decline (i.e. non-market pause) and flash crash resulting runs as seen in PDF and CDF curves in Figure 5.10. The results showed the center of the all the indicator distributions shifted to the right (most notably the OBWA distribution), demonstrating that on average all the indicators have elevated values prior to the flash crash run versus the normal run.

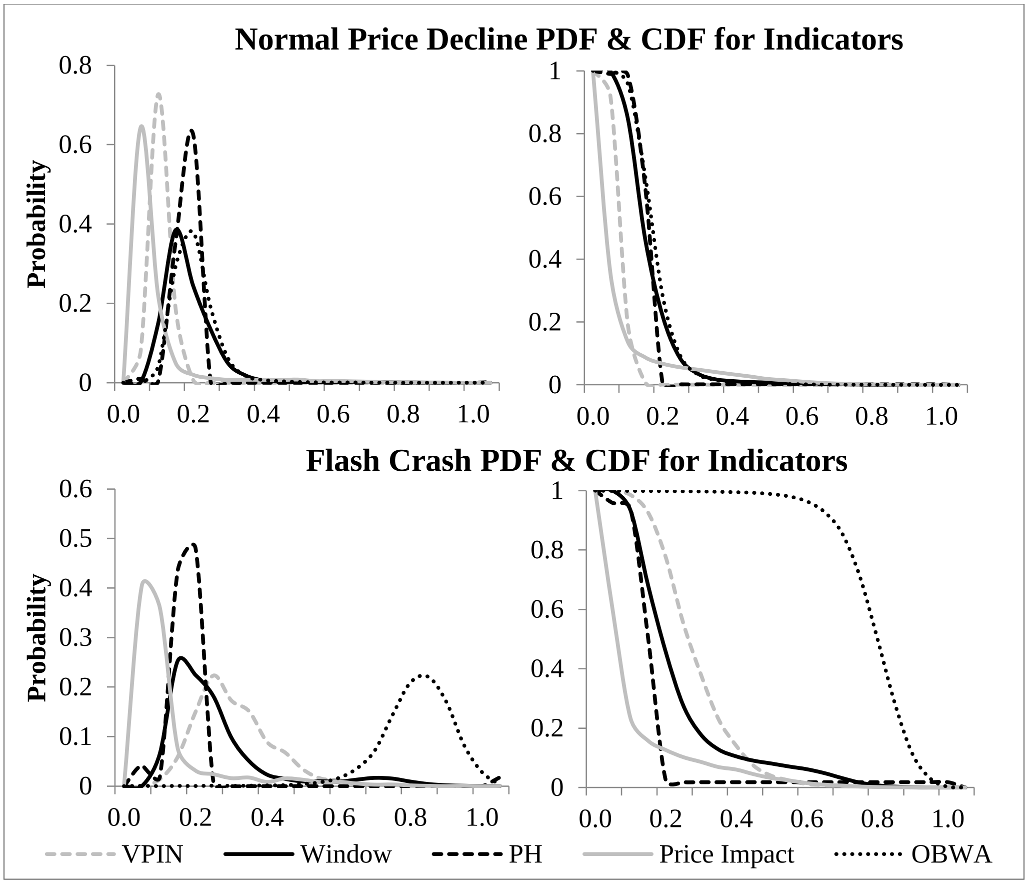


Figure 5.10: Indicator PDF & CDF in Simulated E-Mini Market

It is important consider the extent of advance warning an indicators can provide prior to a market event. In Figure 5.10, the PDF and CDF of the maximum value observed prior to the price flash showed that the highest values were seen within 30 seconds prior of the event for all the indicators; the Window and Price Impact indicators seemed to be slightly better as the early warning predictors.

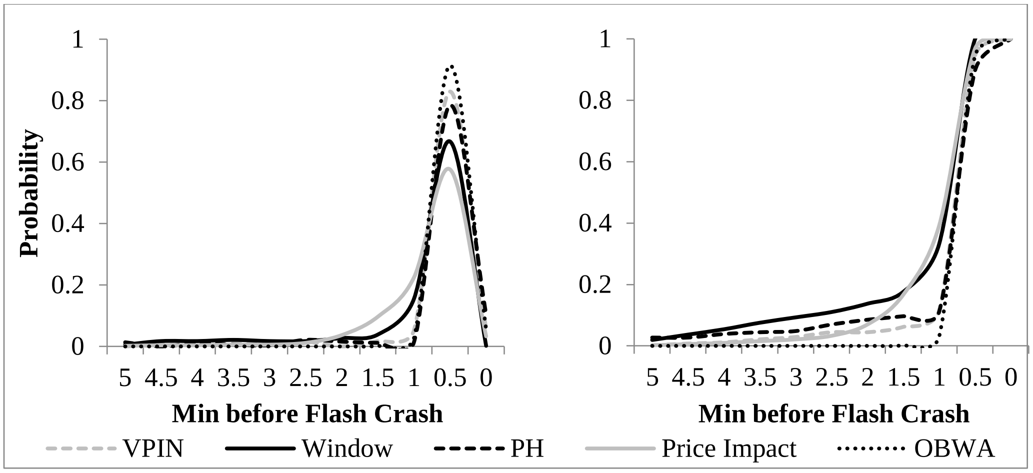


Figure 5.11: Indicator Advance Warning PDF & CDF

Finally, the reliability of the predictive capabilities of the five indicators were compared using receiver operating characteristic (ROC) curves. The plot Figure 5.12 illustrates the performance of the indicators in relation to their true-positive (a market pause was predicted and occurred) and false-positive (a market pause was predicted but did not occur) as their discrimination thresholds were varied. Measures of the maximum value were observed for the indicators at 30 second intervals from 5 minutes to 30 second prior to the event of a market pause. Figure 5.11, shows the strongest predictive power at 30 seconds, as expected given the distributions of the indicators seen in Figure 5.10.

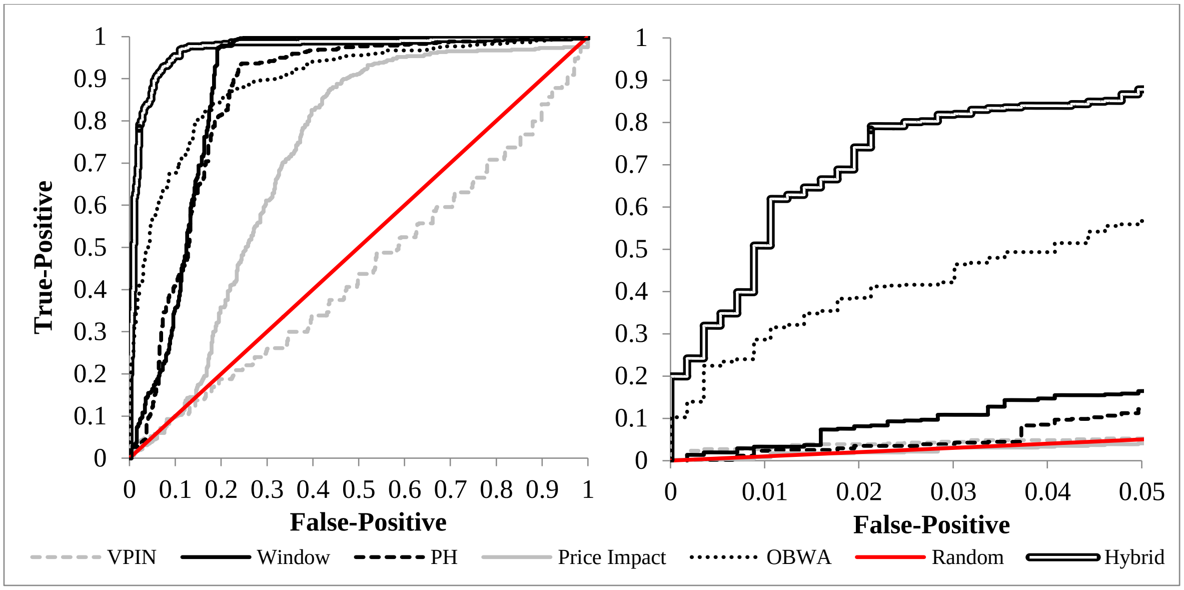


Figure 5.12: ROC Curves of Indicators: 30 Second (Left) & Scale Modified of 30 Second (Right)

As the results show, there is no simple or perfect indicator of a markets resiliency and susceptibility to a price flash. The indicators focusing on measuring the market microstructure state of the order book are able to predict the price flash with more accuracy; however, they do not capture the impact of order flow. This suggests that a combined hybrid indictor might be able to provide stronger predictive value and earlier warning. Using logistic regression, the predictive power of Window, PH, and OBWA indicators when combined was found to be significant and resulted in a more trustworthy, early warning indicator that functions both when a market near equilibrium and when high quantities of marketable order being placed.

It is important to note that the hybrid indicator is able achieve a perfect detection for 20% of all the price flashes, as seen in the zoomed in portion of the ROC curse in Figure 5.12. This significance of this rate is an important consideration in evaluating if an indicator can be a reliable instrument in triggering potential new market protocols to avoid a potential price anomaly (e.g. a pausing of trading).

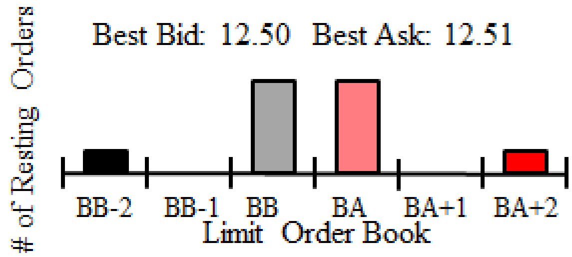


Figure 5.13: Book during Lower Resiliency Event, ATS = light shade, Non-ATS = dark shade

A second interesting result of the simulation work is the difference in predictive power observed in the OBWA and PH indicators. The predictive power of the OBWA indicator is typically stronger, showing that ATS are not overly skewing the market liquidity during the most extreme of points. Yet, during tipping points, the resting orders that the ATS's place can hide the true spread in demand/supply between non-ATS bid/ask resting orders; this would otherwise be demonstrated by a growing gap of ticks between the best bid/ask. Figure 5.13 illustrates this point by showing an example of an order book with low resiliency in price in which ATS orders keep the spread tighter than the rest of the market believes it should be. This makes it appear that demand and supply are tight which serves as a catalyst for trades to occur potentially causing the "hot potato" effect during large one sided demand/supply periods.

5.4.3 Vetting Forecast with Real Market Data

Taking what was learned from the ABS exploration, the performance of indicators were vetted using two sets of real market data samples. Each sample required tracking every limit order as it entered, modified, and exited from the limit order book; as well as the execution of other order types so as to have perfect reconstruction of events. In addition, each order was given an identifier that classified it as belonging to an ATS or not. Finally, the asset data had to be over a period in which the markets experienced a price flash or flash crash event, such that a test could be performed.

Two unique data sets were obtained for investigating the predictive power of the indicators:

- The first of these the September 19th, 2012 drop in the front month WTI Crude Oil Futures market when the price dropped $4 over 10 minutes (much of which occurred during 30 seconds within that period) and reached its low at 12:55PM CST. This event is classified as being a price flash event since the CME's circuit breakers were not triggered.
- The second event being the "Flash Crash" of the May 6th, 2010 that originated in the front month E-Mini S&P 500 Futures market [CFTC & SEC. 2010] where the price dropped $100 over 20 minutes at. At 1:45PM CST, it reached its low when the CME circuit breaker was triggered before the price was able to recover in the contract.

Using 4 days of data, 3 full days selected at random from 2011 & 2012, and the full day of the price events' days mentioned above, the indictors' behavior was examined. By comparing the values seen during normal days and those seen during the price event days, a "likelihood ratio" was constructed for each indicator to represent how it changes from of the market acting normally without a flash event to one experiencing a flash event.

$$Price\ Flash\ Likelihood\ Ratio = \frac{1 - CDF(F)}{1 - CDF(N)}$$

N: *normal day data*
F: *flash day* data

The results for real market data are shown in Figure 5.14 and were very similar to those seen in the simulated data sets. The window indictor was found to be the most confident indicator of the group prior to the lowest price observed in the set. However, the PH and OBWA indicators were able to show earlier warning signs of the markets becoming abnormal. By taking the average of the three as a hybrid indicator, a likelihood ratio of a price flash of above 100 was signaled over a minute in advance of the worst phase of these two price events.

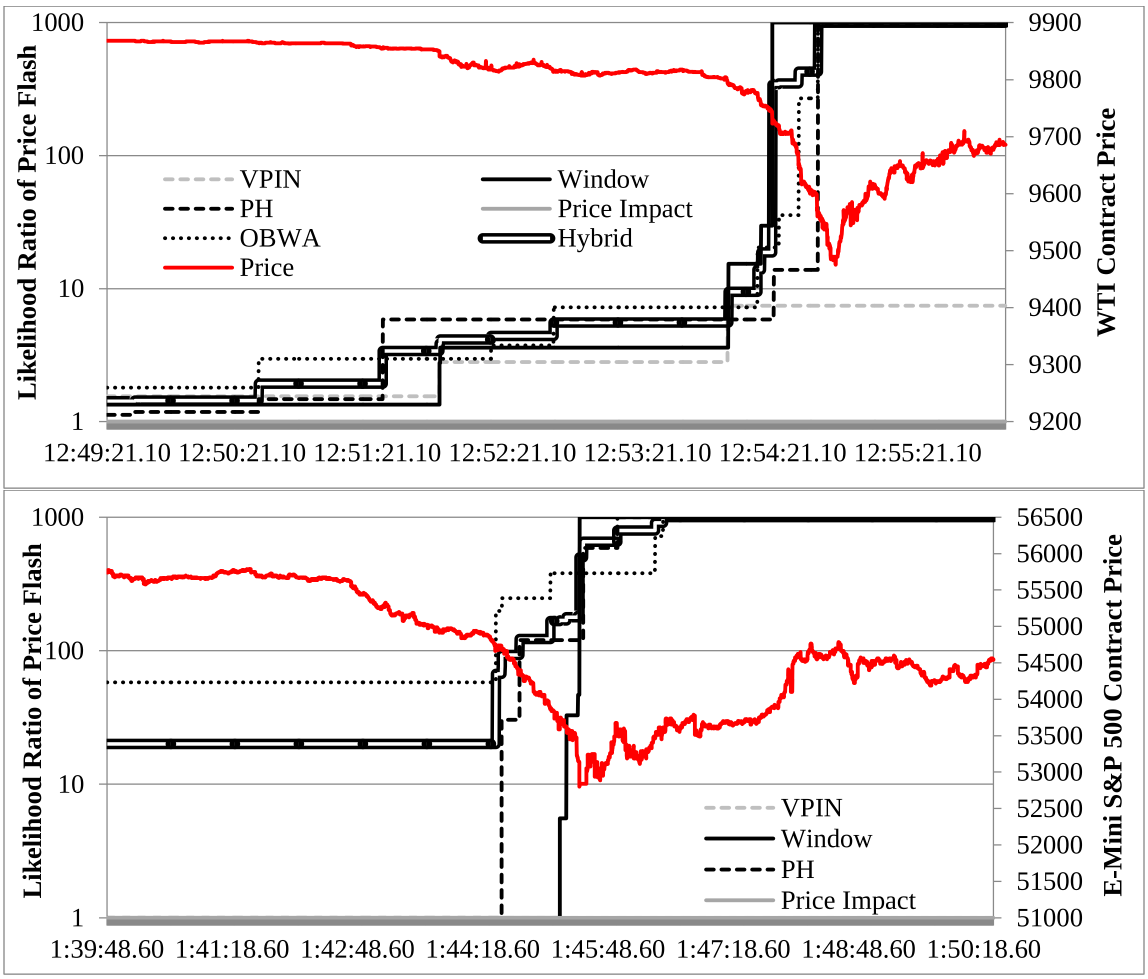


Figure 5.14: WTI Crude Oil (top) & E-Mini S&P 500 (bottom) Futures Price Events

Through examination of the relationship that order flow and order book state data have with one another and the price discovery process, this chapter has identified a possible set of data driven indicators that might be implemented in tools to provide a level of predictive capability to potentially help avoid short term irrational pricing events. Applying the indicators developed based on market simulation data, it was possible to realize a one minute advanced warning signal for two well-known real market price flash events. The results demonstrate that, given data availability, the underlying market microstructure state of the order book can be a more accurate predictive indicator of market's resiliency than order flow, even though order flow can also play an important role.

5.5 RULEMAKING WITH AN ABS OF A FINANCIAL MARKET

As illustrated in the foregoing study of indicators for forecasting price discovery problems, ABS's permit extensive re-creation of market data surrounding specific events and behaviors seen in the real market which can be useful in forecasting or detecting conditions for anomalous events. Thus, an ABS laboratory environment can be used by regulators to support analysis in shaping of policies or rules that will need to be considered in order to prevent negative consequences in a market.

When new law is passed that changes the policy of financial markets, detailed regulations to implement the policy must be developed through a process known as "rulemaking". Rulemaking is designed to insure the quality of the policy change has a net positive influence on the market by following a regimented set of steps that allow stakeholders to have a say in a policy's final creation.

A traditional first step requires a regulatory body to publish an advance notices on the proposed rule that provide preliminary information on the subject area. After this initial step, a proposed rule is published with an analysis and justification for its implementation. Then, the proposed rule is opened to the public for comments. Written issues and concerns about the rule are sent to the regulator during this period and the regulator must address the comments. After the public comment period is closed the final rule is written and published. The final release of the rule lays out how the new regulatory policy will be implemented, addresses all public comments, and contains the regulatory agency's final analysis and justification.

A key feature of the rulemaking process is that the analysis must demonstrate that the benefits of the rule outweigh the costs; otherwise, the rule maybe further contested through litigation. The use of ABS's can help regulators examine the cost and benefits of proposed rules by allowing them to consider/explain how the regulation can affect:

i. the protection of market participants and the public

ii. efficiency, competitiveness, and financial integrity of markets

iii. price discovery

iv. risk management practices

An example of such use of ABS's was done by Hayes *et al*. (2012) where they tested a rule the SEC was contemplating (Securities and Exchange Commission, 2010), which would create a minimum time before an order can be cancelled. The minimum quote life rule was designed to dampen volatility (Securities and Exchange Commission, 2010) by throttling the speed of order revisions by high frequency traders', a group of traders who have been theorized to be increasing volatility (Zhang, 2010).

An experiment was run where the minimum time that an order must remain untouched was varied from 0-1.25 seconds. As the minimum quote life time was increased the absolute value of minute price returns decreased suggesting a decline in volatility and, thereby, that high frequency trading is closely related to its root cause. Those in opposition to this rule suggested that it would decrease liquidity. However, the simulation developed for this work also addressed those comments showing that unless other participants in the market changed their behavior, nothing in the market would change from a liquidity standpoint.

In summary, the foregoing example shows that ABS's can give regulators a means of analyzing and justifying why new rules should be implemented. Although an ABS's cannot replace fundamental economic analysis, they can be used to facilitate a decision maker's understanding of underlying complexities in the response of the financial market.

5.6 SUMMARY

An understanding of cause and effect in complex systems, like financial markets, is dependent on being able to correctly link the parts of a system together, such that responsibility can correctly be assigned. However, as markets are continually adapting and changing, these linkages are non-constant and, thus, can create an overwhelming level of combinatorial complexity which maybe too difficult to overcome to permit effective use of traditional data analysis as a means to gain a true understanding.

ABS's can offer a constructive manner to deal with the combinatorial complexity of the market by enabling experimentation aimed at identifying the effects of the relational and behavioral parts of a financial market on the overall function and behavior of the system. With an ABS experimental framework, fairly simple to rather complicated scenario changes can be made to robustly designed models of a market to gain an understanding of their possible consequences on the market's overall outcomes. This permits both retrospective analysis of market to determine the cause and effect relationships that have led to past market events and research into the development of forecasts/warning systems to prevent negative market events in the future. Through ABS, repetitive and controlled experiments can be run such that the combinatorial complexity that usually interferes with the market analysis, given that a market has thousands (if not millions) of moving parts potentially driving outcomes, can be obviated.

The opportunities that well-constructed ABS offers are wide-ranging - from allowing traders to test their new trading algorithms to helping exchanges to design controls for averting market malfunctions to helping regulators test new regulations. In essence, ABS's offer a platform to run experimentally driven question of the "what if" variety.

Chapter 6

Conclusions

Today's electronic financial markets, wherein highly automated, and often high frequency trading, is the norm for major participants, presents problems that complicate effective monitoring of their most fundamental functions of trade and the price discovery. Though regulators have proposed new surveillance systems for capturing all data for U.S. financial markets, due to the fact that markets are fundamentally "complex systems" the assessment techniques to leverage this data is still far from developed. The research underlying this dissertation provides new approaches for analyzing financial markets that can help overcome these problems and demonstrates how each can be used to better understand the market's trading. These approaches utilize adaptations of methods employing data visualizations, Markov State modeling, and agent based simulations that have been more commonly used in other fields.

Abnormal pricing anomalies in today's complex markets have intensified the demand for such monitoring and assessment, as it is crucial for regulators to effectively protect markets and build confidence for both the industry and the public. These anomalies are transient, unique events that suddenly take place in markets which are fundamentally complex systems environments with large numbers of different types of participants that all have their own individual ever-adapting views and decision-making processes. Much of this research effort is aimed at improving the capabilities of financial regulators and exchanges to better understand the precursors of abnormal events and approach their investigation using methods that help decode the systematic complexities inherent in today's data intensive electronic financial markets.

Through examination of financial markets as systems designed to support efficient and fair trade and pricing, this work has focused on untangling behavioral complexities by demonstrating how to apply techniques used in engineering that are well-suited for the engineering study of complex systems. Through their use, regulators and exchanges can not only improve their analysis of markets post an event of interest, such as a Flash Crash, but, also, analyze behaviors peri- and pre- events of interest. The following sections summarize our findings, recommendations, and highlight the foundational achievements this work has been able to demonstrate.

6.1 VISUALIZATION AND IMAGINARY COMPLEXITY

Retrospective analysis of behavior in financial markets has traditionally focused on examining data related to a set of consummated actions within a market environment, such as completed trades and the resulting inventory held by a participant. Recent improvements in the regulatory audit trails available from exchanges now allow a far more complete and information-rich reconstruction of events related to key market elements such as the order book. Basic data inspection methods (such the number of trades in a day, average order size, or daily change in inventory), do not readily reveal the systemic effects and causal relationships that orders and trades have on markets. Hence, a more meaningful approach to analyzing the order flow dataset is needed, especially for institutions working with market data that are not experienced in to "big data" processing. This can be a major challenge simply due to the presence of the imaginary complexity that comes when facing, particularly for the first time, a large scale data set with numerous interconnected systems producing it.

In many fields, data visualizations have proven to be an invaluable tool for building intuition and enabling exploratory data analysis. Using data visualization techniques for retrieving, analyzing, and disseminating data, regulators can access the tools needed to tackle the cumbersome task of examining the complex structure of large data sets relevant to their regulatory tasks that review past events. 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).

Chapter 3 of this paper has provided a detailed discussion of how data visualizations can be incorporated into the workflow of multiple financial regulatory roles. As discussed and demonstrated in this chapter, by using appropriate visualization techniques, regulators can create a more detailed market picture not only of the markets and the layers of data that build up to create it. With the types of data visualization tools based on financial exchange system knowledge discussed, regulators can have the ability to retrieve and analyze data which otherwise might not be accessible. Such tools facilitate the rapid analysis of changes in participant and market behavior and subsequent dissemination of information to relevant parties (including the exchange, the clearing firm, or the participating firm itself).

6.2 MARKET MONITORING AND PERIODIC COMPLEXITY

As markets have gone electronic, the ability to monitor their function has deteriorated due to a lack of a floor perceptive of participants' trading activities and a significant increase in trade due to automated trading. The Markov State model methodology discussed in Chapter 4 offers a viable means for gaining new perspectives on the functioning of the markets; it views the systems of market participants as an infrastructure for developing trade and pricing.

Using this modeling approach that captures the flow of assets between classes of participants with different trading objectives, we examined how these interactions can affect the outcome of the price discovery process. From the vantage point of examining the market as a construct of individual participants, sub-systems, we are able learn about the periodic complexity (i.e. market organizational dynamic) of the entire market such a useful picture of the state of demand and supply stress in the system can be observed. Using this knowledge, we were able to show that when the system is under stress, there are identifiable features in its behavioral state that follow the natural rules of supply and demand which allow for a degree of prediction of the direction of price.

With the implementation of such a model, regulators can once again have an "on the floor" perspective that provides a sense of a market's state. Permitting regulators the ability identify stressful conditions that may not be good for market participants.

6.3 AGENT-BASED SIMULATION AND COMBINATORAL COMPLEXITY

Through the use of agent-based simulation models (ABS) together with real market data, it is possible to investigate the behaviors that lead to market pricing events. An ABS has a structure, which is defined by a set of agents, a topology and an environment, that can readily provide a framework conducive to the creation of a simplified financial market simulation that can be used to approximate a real market. In the ABS of a market, the market participants are agents, the market mechanism is the topology, and the exogenous flow of information into the market is the environment. By having a replicable model of the market, the combinatorial complex behaviors that occur on a financial exchange can be explored.

Using an appropriately defined ABS for a given market, it is possible to use the characteristics of real past events of interest to study the potential for a future event and consider related sensitivity cases affecting the likelihood of such an event in that market. The effect of changes in the assumed characterization of agents, topology, and/or environment can also be used to facilitate policy debates or inform market stakeholders about the types of data they should be interested in collecting. Through simulation, the presence of combinatorial complexity can be assessed by performing Monte Carlo testing to identify if a specific behaviors (inputs) are responsible for certain pricing characteristics or events (output).

Chapter 5 demonstrated how an ABS can be built such that it replicates financial market functionality and behaviors. These tools can be used to replicate events in order to better understand their causes and, thus, also be used to understand if past circumstances and causes can be used as are reliably signal of potential future events. The ABS can, then, provide in-depth, simulated datasets, whose availability would normally be restricted to exchanges and regulators, and the ability to reproduce numerous events for sensitivity analysis. Furthermore, it can be an effective means for regulators to assess and help explain the effects of new regulations in advance of their introduction in order to demonstrate their costs and benefits.

Also, as discussed in Chapter 5, ABS's offer a constructive manner to address the effects of combinatorial complexity in the market system by enabling experimentation aimed at identifying the effects of the relational and behavioral parts of a financial market on the overall function and behavior of the system. With an ABS experimental framework, fairly simple to rather complicated scenario changes can be made to robustly designed models of a market to gain an understanding of their possible consequences on the market's overall outcomes. This permits both retrospective analysis of market to determine the cause and effect relationships that have led to past market events and research into the development of forecasts/warning systems to prevent negative market events in the future. Through ABS, repetitive and controlled experiments can be run such that the combinatorial complexity effects that usually interfere with the market analysis, given that a market has thousands (if not millions) of moving parts potentially driving outcomes, can be obviated.

6.4 DIRECTION OF FUTURE WORK

The underlying work done for this dissertation offered many glimpses into the potential tools and examination methods that financial market regulators have at their disposal as they continue to work to stay abreast with the latest changes in technology and participants' potentially illegal (but, perhaps, creative) trading strategies.

The opportunities to build specific visualizations for targeted analysis objectives are endless, no matter what the field. The challenge for regulators is identifying the key gaps in their ability to uncover, assess, and communicate accurate conclusions about the expansive data they have at their fingertips. As their data access continues to get richer, uncovering the hidden meanings and knowledge in it will become an area of growing and greater importance. The value of future work will depend on specific, hand-in-hand development opportunities and solutions working closely with regulators to determine how to best incorporate visualizations in the daily activities of regulators to further the fundamental objectives of their work.

The Markov State model offers a new perspective to the function of the markets and sees the systems of participants as an infrastructure for developing trade and pricing. This model demonstrated markets can be seen and treated as systems for meeting demand and supply, but, just like other systems, have limits to their function when stressed.

The model gives regulators a tool to start asking important questions about the fairness of a market and if their function is meeting a public good or only favoring a few. Future application opportunities include specifically looking at individual participants impact in the overall dynamic of trade in order to identify issues in the overall function of the system. This could aid in the identification of speculative limits, a limit meant to protect markets from excessive speculation, which can result in unreasonable or unwarranted price fluctuations, by helping to assess the correct limit levels such that there is no danger in unintentionally setting a limit too large or too small.

Markets are inherently complex and thus are difficult to predict given that they involve so many different individuals with a wide range of reasons for participating and manners of doing so. ABS offers a manner to final test markets and begins to understand them from a scientific level rather than relaying purely on perceptual/theoretical ideas of how they should function.

With the proper access to data to infuse ABS in future, laboratory like environments could be built for economists that would allow for scientific examination and justification of theory to be performed. As this work has shown, there are many purposes for which such laboratories could be used that would further the understanding of market events, thru their deconstruction, and the impact of future market rules and circuit breakers changes, through forecasting and policy testing.

6.5 FINAL STATEMENTS

A key goal of this work has been to present financial markets in a new, engineering light, as complex systems, and develop alternative guided paths for financial regulators to use in market data analysis that help to tackle the problems created by system complexity. By breaking down the effects of complexity in data related to trade and pricing, this research has demonstrated how different tools (though they may not be perfect) can be used to help overcome these effects to improve the quality of analysis.

Complex systems come in numerous forms and can be seen all around us. All living things are complex systems; they are built of biological systems which are, in turn, built by and from chemical and physical subsystems. Sciences have developed to explain fundamentals and provide structure to understanding of the role each part plays in the function of the systems around us.

For their part, engineers have traditionally started with the fundamental building blocks of knowledge provided by these sciences and built new systems intended to meet sets of objectives and demands useful to humans. What has typically been key to the success of any man-made system has centered around achieving control and predictability in the way the engineered system functions, such that the system can be relied upon to operate predictably. However, this feature is often considered lost when a system's nature becomes complex. As this dissertation demonstrates for financial market systems, complex systems present significant additional, often unique, difficulties in monitoring and assessment to those tasked with recognizing problems and developing effective solutions; they require their own "scientists" to continually evaluate them.

As such, it is important to consider, as more and more systems naturally evolve from simple systems, oriented with a single objectives in mind, to systems meant to manage many objectives simultaneously (so that they may capture the benefits of lower costs, higher capacity, and greater speed), how do we ensure the ability to continue to effectively assess and control them. It is important to note that though newer systems may offer large quantities of performance data for analyzes, a gap often exists in the skills needed to translate the data into practical information.

One of the key lessons learned in developing the visualization, state-based monitoring systems, and ABS for exploring markets is that, although these tools may serve to allow us to examine complex systems with higher fidelity than statistics may offer, they definitely have limits in their capability for untangling the information hidden in large and complex datasets. As we continue to build more complicated systems leveraging technological advances, it is important to consider how we will simultaneously build capabilities to assess their ability to function reliably and successfully; the warning signs of malfunctions maybe too difficult to perceive till it is too late. For the financial markets in particular, we must not let our ability to monitor and assess to fall to the way side to ensure that these critically important systems, which directly affect the lives of almost everyone, always function fairly and for the greater public good.

References

Ackoff, R. (1971) Towards a system of systems concepts. *Management Science*, 17(11): 661–671.

Ackoff, R. and Emery, F. (1972) *On purposeful systems*. Chicago: Aldine-Atherton.

Allen, F. and Gale, D. (1997) Financial markets, intermediaries, and intertemporal smoothing. *Journal of Political Economy*, 105(3): 523–546.

Arrow, K. (1964) Control in large organizations. *Management Science*, 10(3): 397–408.

Atje, R. and Jovanovic, B. (1993) Stock markets and development. *European Economic Review*, 37(2): 632–640.

Baron, M., Brogaard, J. and Kirilenko, A. (2012) The trading profits of high frequency traders. Working Paper.

Benic, V. and Franic, I. (2008) Stock Market Liquidity: Comparative Analysis of Croatian and Regional Markets. *Financial Theory and Practice*, 32(4): 477–498.

Bethel, E., Leinweber, D., Rübel, O. and Wu, K. (2011) Federal market information technology in the post flash crash era: roles for supercomputing. In *Proceedings of the fourth workshop on High performance computational finance*.

Blanchard, O. (2009) *The crisis: basic mechanisms and appropriate policies*. Vol. 9, Cambridge, MA.: International Monetary Fund.

Blume, H. (2012) *Behavior Identification in Markets using Visualization and Network Analysis*. (Doctoral Dissertation, Karlsruhe Institute of Technology).

Bohm, D. (1957) *Causality and chance in modern physics*. Philadelphia: University of Pennsylvania Press.

Bookstaber, R. (2012) Using Agent-Based Models for Analyzing Threats to Financial Stability. *Working Paper*. Office of Financial Research.

Brandt, M. and Kavajecz, K. (2004) Price discovery in the U.S. Treasury market: The impact of order flow and liquidity on the yield curve. *Journal of Finance*, 59(6): 2623–2654.

Brogaard, J. (2010) High frequency trading and its impact on market quality. *Working Paper*. Northwestern University.

Burnham, J. (1991) Current structure and recent developments in foreign exchange markets. In *Recent Developments in International Banking and Finance*. 123–153.

Capra, F. (1982) *The turning point*. Toronto: Bantam.

CFTC and SEC. (2010) Findings regarding the market events of May 6, 2010. September 30, 2010.

Chen, M. and Floridi, L. (2013) *Synthase.* Unpublished.

Chi, E. (2000) A taxonomy of visualization techniques using the data state reference model. In *Proceedings of the IEEE Information Visualization Conference.*

CME Group. (2013) CME MDP Message Statistics. http://beta.cmegroup.com/ market-data/distributor/market-data-platform.html. [Accessed April 16, 2013]

De Bondt, W. and Thaler, R. (1985) Does the stock market overreact?. *The Journal of Finance*, 40(3): 793–805.

De Bondt, W. and Thaler, R. (1987) Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance*, 42(3): 557–581.

Demsetz, H. (1968) The cost of transacting. *The Quarterly Journal of Economics*, 82(1): 33–53.

Dooley, K. (1997) A complex adaptive systems model of organizational change. *Non-linear Dynamics, Psychology and the Life Sciences*, 1(1): 69–97.

Easley, D., de Prado, M. and O'Hara, M. (2010) The microstructure of the 'flash crash': Flow toxicity, liquidity crashes and the probability of informed trading. *Working Paper*. Cornell University.

Evans, M. and Lyons, R. (2002) Order flow and exchange rate dynamics. *Journal of Political Economy*, 110(1): 170–180.

Fairchild, K. D., & Aschner, J. L. (2012) HeRO monitoring to reduce mortality in NICU patients. *Research and Reports in Neonatology*, 2: 65–76.

Farmer, J. (2002) Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5): 895–953.

Farmer, J., Patelli, P. and Zovko, I. (2005) The predictive power of zero intelligence in financial markets. In *Proceedings of the National Academy of Sciences of the United States of America*, 102(6): 2254–2259.

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

Fleming, W., and Soner, H. (2006) *Controlled Markov processes and viscosity solutions*. New York, NY: Springer.

Frey, S. and Grammig, J. (2006) Liquidity supply and adverse selection in a pure limit order book market. *Empirical Economics*, 30(4): 1007–1033.

Gault, S. and Jaccaci, A. (1996) Complexity meets periodicity. *The Learning Organization*, 3(2): 33-39.

Glaser, M., Nöth, M. and Weber, M. (2008) Behavioral Finance. In *Blackwell Handbook of Judgment and Decision Making*. Malden, MA: Blackwell Publishing Ltd. 525–546.

Glosten, L. (1987) Components of the bid ask spread and the statistical properties of transaction prices. *Journal of Finance*, 42(5): 1293–1307.

Glosten, L. R. (1994) Is the electronic open limit order book inevitable?. *Journal of Finance*, 49(4): 1127–1161.

Grossman, G. and Helpman, E. (1991) Trade, knowledge spillovers, and growth. *European Economic Review*. 35(2): 517-526.

Haimes, Y. (1998) *Risk modeling, assessment, and management*. New Jersey: John Wiley & Sons.

Harris, L. (2003) Trading and Exchanges: Market Microstructure for Practitioners. New York: Oxford University Press.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46(1): 179–207.

Hasbrouck, J. (1995) One security, many markets: Determining the contributions to price discovery. *Journal of Finance,* 50(4): 1175–1199.

Hayek, F. A. (1945) The Use of Knowledge in Society. *The American Economic Review*, 35(4): 519–530.

Hayek, F. (1952) *The Counter-Revolution of Science*. Glencoe: Free Press.

Hayes, R., Paddrik, M., Todd, A., Yang, S., Beling, P. and Scherer, W. (2012) Agent Based Model of the E-Mini S&P 500 Future: Application for Policy Making. In *Proceedings of the 2012 Winter Simulation Conference*.

Hayles, N. (1991) Introduction: Complex dynamics in science and literature. In *Chaos and order: Complex dynamics in literature and science*. Chicago: University of Chicago Press. 1–36.

Hendershott, T. and Jones, C. (2005) Island goes dark: Transparency, fragmentation, and regulation. *Review of Financial Studies*, 18(3): 743–793.

Hendershott, T., Jones, C. and Menkveld, A. (2011) Does algorithmic trading improve liquidity?. *Journal of Finance,* 66(1): 1–33.

Hodgson, G. (1993) *Economics and Evolution: Bringing Life Back Into Economica*. Michigan: University of Michigan Press.

Huber, G. (1991) Organizational learning: The contributing processes and the literatures. *Organization Science*, 2(1): 88 –115.

International Monetary Fund. (2008) *Global Financial Stability Report, October 2008: Financial Stress and Deleveraging Macrofinancial Implications and Policy*.

Kaldor, N. (1961) Capital Accumulation and Economic Growth. *The Theory of Capital: Proceedings of a Conference Held by the International Economic Association*. London: Macmillan. 177–222.

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

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J. and Melancon, G. (2008) Visual Analytics: Definition, Process, and Challenges. In *Information Visualization: Human Centered Issues and Perspectives*, Springer. 154–175.

Kirilenko, A., Kyle, A., Samadi, M. and Tuzun, T. (2011) The Flash Crash: The impact of high frequency trading on an electronic market. *Working Papers*. University of Maryland.

Kirman, A. and Teyssiere, G. (2002) Microeconomics model for long-memory in the volatility of financial time series. *Studies in Nonlinear Dynamics and Econometric,* 5(4): 281–302.

Kiyono, K., Struzik, Z. and Yamamoto, Y. (2006) Criticality and phase transition in stock-price fluctuations. *Physical review letters*, 96(6):

Kullmann, L., Toyli, J., Kertesz, J., Kanto, A. and Kaski, K. (1999) Characteristic times in stock market indices. *Physica A: Statistical Mechanics and its Applications*, 269(1): 98–110.

Kyle, A. and Obizhaeva, A. (2013) Market Microstructure Invariance: Theory and Empirical Tests. Revise and Resubmit, *Econometrica*.

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

Lee, T. (2003) *Complexity Theory in Axiomatic Design* (Doctoral dissertation, Massachusetts Institute of Technology).

Love, R. and Payne, R. (2008) Macroeconomic news, order flows, and exchange rates. *Journal of Financial and Quantitative Analysis*, 43(2): 467–488.

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

Mandelbrot, B. (1963) The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4): 394–419.

Massimb, M. and Phelps, B. (1994) Electronic trading, market structure and liquidity. *Financial Analysts Journal*, 50(1): 39–50.

Meadows, D. (2008) *Thinking in Systems: a primer*. VT.: Chelsea Green Publishing.

Morrison, J. and Tversky, B. (2001) The (in)effectiveness of animation in instruction. In *Proceeding of the 2001 ACM Conference of Human Factors in Computing Systems*.

Nanex. (2011) Flash Crash Analysis Continuing Developments: Flash Equity Failure for 2006, 2007, 2008, 2009, 2010, and 2011. http://www.nanex.net/FlashCrashEquities /FlashCrashAnalysis_Equities.html [Accessed April 16, 2013]

Niederhoer, V. (1997) *Education of a Speculator*. New York: John Wiley & Sons.

Onnela, J. P., Chakraborti, A., Kaski, K., Kertesz, J., & Kanto, A. (2003) Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5): 056110.

Ouchi, W. (1979) A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9): 833–848.

Paddrik, M., Hayes, R., Todd, A., Yang, S., Scherer, W., and Beling, P. (2012) An Agent Based Model of the E-Mini S&P 500 and the Flash Crash. In *Proceedings of the IEEE Computational Intelligence and Financial Engineering Conference.*

Patcha, A. and Park, J. (2007) An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12): 3448–3470.

Perez, E. and M. White, M. (1985) Student evaluation of motivational and learning attributes of microcomputer software. *Journal of Computer-Based Instruction*, 12(2): 39–43.

Rieber, L. (1989) 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): 431–444.

Rieber, L. (1991) Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, 83(3): 318–328.

Securities and Exchange Commission. (2010) Concept Release on Equity Market Structure. http://www.sec.gov/rules/concept/2010/34-61358.pdf [Accessed April 16, 2013]

Shiller, R. (1988) Portfolio insurance and other investor fashions as factors in the 1987 stock market crash. *NBER Macroeconomics Annual*, 3: 287–297.

Shneiderman, B. (1992) Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1): 92–99.

Sornette, D. (2003) Critical market crashes. *Physics Reports*, 378(1): 1–98.

Spence, I. and S. Lewandowsky, S. (1990) Graphical perception. In *Modern methods of data analysis.* CA.: Sage. 13–57.

Spence, I. and Garrison, R. (1993) A remarkable scatterplot. *The American Statistician*. 47(1): 12–19.

Spulber, D. (1996) Market microstructure and intermediation. *The Journal of Economic Perspectives*, 10(3): 135–152.

Srinivasan, S., Ponceleon, D., Amir, A. and Petkovic, D. (1999) What is in that video anyway?. In *Proceedings of the International Conference on Mechanics, Simulation and Control*.

Suh, N. (2005) *Complexity: Theory and Applications*. Oxford: Oxford University Press.

Tufte, E. (1983) *The Visual Display of Quantitative Information*. Cheshire, CT.: Graphics Press.

Tversky, B. (1995) Cognitive origins of conventions. In *Understanding Images*. New York: Springer-Verlag. 29–53.

Wattenberg, M. (1999) Visualizing the stock market. In *Proceeding of the 1999 ACM Conference of Human Factors in Computing Systems*.

White, H. (1981) Where do markets come from?. *American Journal of Sociology*, 87(3): 517–547.

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

Zhang, F. (2010) High-Frequency Trading, Stock Volatility, and Price Discovery. *Working Paper*. Yale University.

Zhou M. and Feiner S. (1998) Visual task characterization for automated visual discourse synthesis. In *Proceeding of the 1998 ACM Conference of Human Factors in Computing Systems*.