Financial Market Monitoring and Surveillance Systems Framework:
A Service Systems and Business Intelligence Approach

A Thesis submitted to the University of Manchester for the degree of
PhD in Business Administration
In the Faculty of Humanities

2011

David Diaz

Faculty of Humanities

Contents

LIST OF TABLES V

LIST OF FIGURES VI

ABSTRACT X

DECLARATION AND COPYRIGHT STATEMENT XI

ACKNOWLEDGEMENTS XIII

PREFACE XV

PUBLICATIONS RELATED TO THIS WORK XVI

1 INTRODUCTION 17

1.1 BACKGROUND AND MOTIVATION 17

1.1.1 WHAT ARE THE DRIVERS BEHIND THE MARKET STRUCTURAL CHANGES? 19

1.1.2 WHAT ARE THE CONSEQUENCES OF MARKET STRUCTURAL CHANGES? 21

1.1.3 WHAT ACTIONS ARE BEING CONSIDERED TO RESTORE TRUST IN THE MARKETS? 22

1.2 AIMS AND OBJECTIVES 25

1.3 RESEARCH QUESTIONS 26

1.4 STRUCTURE OF THE THESIS 27

2 LITERATURE REVIEW 29

2.1 REVIEW OF FINANCIAL ECONOMICS LITERATURE 29

2.2 MARKET MONITORING AND SURVEILLANCE: CONTEXT, CAPABILITIES AND PERFORMANCE 34

2.2.1 REVIEW OF MODERN FINANCIAL MARKETS ORGANISATION 34

2.2.2 MARKET MANIPULATION DEFINITIONS 38

2.2.3 TYPES OF MARKET MANIPULATIONS AND TAXONOMY 40

2.2.4 CROSS-BORDER, CROSS-MARKET, CROSS-PRODUCT AND CROSS-JURISDICTION MANIPULATIONS 49
2.2.5 REGULATORY FRAMEWORKS RELATED TO CROSS-BORDER TRADING 52
2.2.6 DEFINITION OF MARKET SURVEILLANCE AND MONITORING 55
2.2.7 MARKET MONITORING AND SURVEILLANCE: ALLOCATION OF RESPONSIBILITIES 56
2.2.8 ALLOCATION OF MARKET MONITORING AND SURVEILLANCE ACTIVITIES 59
2.2.9 MONITORING AND SURVEILLANCE: INTERNAL ORGANISATION AND PROCESS 61
2.2.10 MONITORING AND SURVEILLANCE: SCOPE, PERFORMANCE AND CROSS-MARKET CAPABILITIES 64
2.3 MARKET MONITORING AND SURVEILLANCE: TECHNOLOGIES AND SYSTEMS 67
2.3.1 FRAUD AND DATA MINING 67
2.3.2 EXISTING SYSTEMS AND FRAMEWORKS 69
2.3.3 CHARACTERISTICS AND REQUIREMENTS 73
2.4 GAPS IN THE LITERATURE AND REQUIREMENTS FOR A CONCEPTUAL MMSS 75

3 RESEARCH DESIGN AND METHODOLOGY 82

3.1 DESCRIPTION OF THE METHODOLOGY 82
3.1.1 STEPS AND OUTPUTS OF DESIGN RESEARCH METHODOLOGY 84
3.2 JUSTIFICATION OF THE SELECTION OF THE METHODOLOGY 88
3.3 METHODOLOGY APPLICATION 92

4 TAXONOMIES OF MARKET MANIPULATION AND MARKET MONITORING AND SURVEILLANCE SERVICE SYSTEM CONCEPTS 96

4.1 A TAXONOMY OF ELEMENTS IN MANIPULATION SCENARIOS 97
4.1.1 EXAMPLES 102
4.2 A TAXONOMY OF STAKEHOLDERS OF FINANCIAL MARKETS 104
4.3 A TAXONOMY OF MARKET MONITORING AND SURVEILLANCE DETECTION' ENGINES AND COMPONENTS 111

5 THE PROPOSED MARKETS MONITORING AND SURVEILLANCE SERVICE SYSTEMS FRAMEWORK 115

5.1 CONTEXTUAL THEORIES, PRINCIPLES AND THEIR APPLICATION TO THE FRAMEWORK ..115
5.1.1 SERVICE DOMINANT LOGIC ..115
5.1.2 SERVICE SCIENCE ..121
5.1.3 SYSTEMS THINKING ..123
5.1.4 THINKING *ABOUT* MARKET MONITORING AND SURVEILLANCE SERVICE SYSTEMS ..124
5.1.5 THINKING *FROM* MARKET MONITORING AND SURVEILLANCE SERVICE SYSTEMS ..127
5.1.6 'LIMITS TO GROW' IN FINANCIAL MARKETS ..129
5.1.7 'SHIFTING THE BURDEN' IN FINANCIAL MARKETS ..132
5.1.8 SUMMARY ..135
5.2 SERVICE VALUE NETWORKS AND CUSTOMER DRIVEN VALUE CO-CREATION ..136
5.2.1 BASIC CONCEPTS ..136
5.2.2 VARIATIONS OF VALUE NETWORKS AND VALUE PROPOSITIONS ..140
5.3 '2M-3S' FRAMEWORK' WORLDVIEW ..143
5.4 '2M-3S' SERVICE VALUE NETWORKS AND VALUE PROPOSITIONS ..149
5.5 '2M-3S' FRAMEWORK' DETECTION ENGINE ..153
5.6 CHAPTER SUMMARY ..157

6 CASE STUDIES ..159

6.1 CASE STUDY ONE. 'TRADE-BASED' MANIPULATIONS ..160
6.1.1 BACKGROUND ..160
6.1.2 DATA MODELLING AND ANALYSIS ..161
6.1.3 ANALYSIS AND RESULTS ..169
6.1.4 EVALUATION AND INTERPRETATION OF THE RESULTS ..185
6.2 CASE STUDY TWO: 'INSIDER TRADING' MANIPULATIONS ..193
6.2.1 BACKGROUND ..193
6.2.2 DATA MODELLING AND ANALYSIS ..195
6.2.3 ANALYSIS AND RESULTS ..197
6.2.4 EVALUATION AND INTERPRETATION OF RESULTS ..207
6.3 CASE STUDY THREE. USING SOCIAL NETWORKING TOOLS FOR THE DETECTION OF 'CROSS-JURISDICTION' 'WASH-SALE' AND 'MATCH-TRADE' MANIPULATIONS ..211
6.3.1 BACKGROUND ..211
6.3.2 DATA MODELLING AND ANALYSIS ..213
6.3.3 ANALYSIS AND RESULTS ..214

6.3.1 Evaluation and Interpretation of Results 233
6.4 Case Study 4: High frequency trading Quote Stuffing 239
6.4.1 Background 239
6.4.2 Methodology and data used 243
6.4.3 Data Modelling and Analysis 243
6.4.4 Analysis and Results 250
6.4.5 Evaluation and Interpretation of Results 257
6.5 Chapter summary 270

7 CONCLUSIONS 276

7.1 Overview of the work 276
7.2 Contributions and limitations of the work 278
7.3 Implications and Future Work 280

8 REFERENCES 283

APPENDIX 1: WATCH-LIST FOR SUSPICIOUS BLOCK HOURS AND HIT RATIO 290

APPENDIX 2: KNOWLEDGE DISCOVERY PROCESS FLOW 291

APPENDIX 3: ABNORMAL RETURNS CALCULATION 292

APPENDIX 4: EXAMPLE OF A NEWS ITEM 294

APPENDIX 5: DIFFERENCE OF MEANS 295

Word Count: 70,380

List of Tables

Table 2-1 Market Manipulations, Categories, Definitions and Types. Adapted from (Cumming and Johan, 2008). 45
Table 2-2 Extract on the implementation of MAD across EU countries 53
Table 2-3 Allocation of Market Monitoring and Surveillance Detection Responsibilities 60
Table 2-4 Descriptive Statistics of the ASX Surveillance Process. Adapted from (Comerton-Forde and Rydge, 2006) 64
Table 2-5 Summary of Requirements of MMSS 79
Table 3-1 Characteristics of the Addressed Problem 88
Table 3-2 Philosophical Viewpoints of Three Research Perspectives. Taken from (Vaishnavi and Kuechler, 2004) 92
Table 5-1 Principles of SD Logic and Application to the 2M-3S Framework 118
Table 5-2 Four Fingerprints of Complex Systems in the 2M-3S Framework 125
Table 6-1 Design Research in Case Studies 159
Table 6-2 Case Study 1: Summary Statistics 164
Table 6-3 Case Study 1: Distribution of the Type and Number of News Items and Events per Sample 164
Table 6-4 Case Study 1: Return, Abnormal Returns, Liquidity, and Volatility of Manipulated Stocks 178
Table 6-5 Case Study 1: Frequency Analysis of the Outliers per Sample 179
Table 6-6 Case Study 1: Unsupervised Modelling Results 180
Table 6-7 Case Study 1: Supervised Modelling Results (Level 1) 182
Table 6-8 Case Study 1: Supervised Modelling Results (Level 2) 182
Table 6-9 Case Study 3: HYHY All-Day Nodes Network Metrics 229
Table 6-10 Case Study 3: Graphs Metrics 231
Table 6-11 Case Study 4: Exchanges for AGQ 243
Table 6-12 Case Study 4: Descriptive Statistics AGQ Dataset 243
Table 6-13 Case Study 4: Simulation Parameters 247
Table 6-14 Case Study 4: GapLife Example 248

Table 6-15 Case Study 4: Gap and Gap Life Summary Statistics.....251
Table 6-16 Case Study 4: Gap and Gap Life Count of Events.....254
Table 6-17 Case Study 4: Correlation of Gaps vs Ratios.....255
Table 6-18 Case Study 4: Travelling Times.....260
Table 6-19 Summary of Case Studies.....271

List of Figures

Figure 2-1 Taxonomy of Market Manipulations and Abuses. Adapted from (Putnins, 2009) and (Cumming and Johan, 2008). * Indicates original taxonomy has been expanded by the author including a new category or element. 43
Figure 2-2 Surveillance Process ASX-SOMA (Comerton-Forde and Rydge, 2006) 62
Figure 2-3 Surveillance Workflow of the ThaiBMA (Mongkolnavin and Tirapat, 2009).....71
Figure 2-4 The Combination of Economic and Behavioural Analysis Approaches to Trading Analysis (Mongkolnavin and Tirapat, 2009).....71
Figure 2-5 Conceptual Architecture of the SONAR System (Goldberg, Kirkland et al. 2003).....72
Figure 3-1 Reasoning in the Design Cycle. Taken from (Takeda et al., 1990) as in (Vaishnavi and Kuechler, 2004).....83
Figure 3-2 General Methodology of Design Research. Taken from (Vaishnavi and Kuechler, 2004).....85
Figure 3-3 Information Systems Design Research Framework (Hevner et al., 2004).....87
Figure 3-4 Adopted Information Systems Research Framework. Adapted from (Hevner et al., 2004). Examples in parenthesis ( ) were added by the author. 93
Figure 3-5 Research Mapping. Adapted from (Vaishnavi and Kuechler, 2004). Thesis Mapping Chapters column was added by the author. 95
Figure 4-1 Elements Involved in a Manipulation 98

Figure 4-2 Taxonomy of Agents 98
Figure 4-3 Taxonomy of Actions 99
Figure 4-4 Taxonomy of Venues 99
Figure 4-5 Taxonomy of Assets. Red indicates branch was added by the author 100
Figure 4-6 Taxonomy of Time Concepts 101
Figure 4-7 Taxonomy of Effects 102
Figure 4-8 Financial Markets' Stakeholders 105
Figure 4-9 Taxonomy of Investors - Intermediation 106
Figure 4-10 Taxonomy of Regulators 108
Figure 4-11 Taxonomy of Support Organisations 110
Figure 4-12 Taxonomy of Detection Engines 113
Figure 4-13 Taxonomy of Sample Analysis Components 114
Figure 5-1 Named Resources as building Blocks of Service Systems 121
Figure 5-2 Value Propositions and Multi-Stakeholder Measure of Value 123
Figure 5-3 Limits to Grow Archetype. Taken from (Senge, 1994) 129
Figure 5-4 Limits to Grow in Financial Markets 132
Figure 5-5 'Shifting the Burden' Archetype. Taken from (Senge, 1994) 134
Figure 5-6 'Shifting the Burden' in Financial Markets 135
Figure 5-7 A Service System and its Entities, Stakeholders and Worldview. Taken from (Kwan and Min, 2008) 136
Figure 5-8 Service Supply Chain 138
Figure 5-9 SVN Dichotomy and Determinants of Value Co-creation. Taken from (Kwan and Yuan, 2011) 139
Figure 5-10 Value Co-creation Variation Scenarios and Variants of SVNs. Adapted from (Kwan and Yuan, 2011) 141
Figure 5-11 '2M-3S' Framework' Worldview 143
Figure 5-12 Financial Services Systems Worldview 145
Figure 5-13 Financial Markets Organisation – Single Jurisdiction 146
Figure 5-14 Financial Markets Organisation – Multiple Jurisdiction 148
Figure 5-15 Market Monitoring and Surveillance SVN. 149
Figure 5-16 'Closed' Market Monitoring and Surveillance SVN. 150

Figure 5-17 'Closed' Service System Worldview of MMSS. 151
Figure 5-18 'Open Stakeholder' Market Monitoring and Surveillance SVN 152
Figure 5-19 'Open Stakeholder' Service System Worldview of MMSS. 152
Figure 5-20 Market Monitoring and Surveillance Engine Model. 154
Figure 5-21: 2M-3S Framework Scope Summary 158
Figure 6-1 Case Study 1: Indicators for BIF, Dec. 2003 172
Figure 6-2 Case Study 1: Hourly Average Returns Indicator (ZO1 Indicator) 173
Figure 6-3 Case Study 1: Hourly Average Liquidity Indicator (ZS1 Indicator) 174
Figure 6-4 Case Study 1: Hourly Average Volatility Indicator (RV3 Indicator) 175
Figure 6-5 Case Study 1: C5.0 Tree – Suspicious Block Hour Flag, Level 1 184
Figure 6-6 Case Study 1: CR&T Tree – Suspicious Trade Flag, Level 2 185
Figure 6-7 - Case Study 1: Instantiation of Detection Engine 189
Figure 6-8 Case Study 1: Market Monitoring and Surveillance SVN. 190
Figure 6-9 Case Study 1: Service System Worldview 191
Figure 6-10 Case Study 1: Service System Alternative Worldview 191
Figure 6-11: Case Study 1 Taxonomy Instantiation 192
Figure 6-12 Case Study 2: Data Mining Outlier Detection Engine Visualisation ..199
Figure 6-13 Case Study 2: Time Series Visualization of Key Indicators 200
Figure 6-14 Case Study 2: OLAP Analysis IFDF. 202
Figure 6-15 Case Study 2: Visualization Analysis of Insiders Activity (BM) 203
Figure 6-16 Case Study 2: Visualization Analysis of Insiders Activity (JA) 204
Figure 6-17 Case Study 2: Forms 4 and 144 are Used to Explore Associative Relationships 206
Figure 6-18 Case Study 2: Scenario Instantiation of the '2M-3S' Framework 208
Figure 6-19 Case Study 2: Service System World view 209
Figure 6-20 Case Study 2: Taxonomy Instantiation 210
Figure 6-21 Case Study 3: Data Mining Outlier Detection Engine Visualisation ..216
Figure 6-22 Case Study 3: Time Series Visualization of Key Indicators 217
Figure 6-23 Case Study 3: Social Network of Manipulators and Accounts 220
Figure 6-24 Case Study 3: Benchmark Network 223
Figure 6-25 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 1 225

Figure 6-26 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 2 226
Figure 6-27 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 3 227
Figure 6-28 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 4 227
Figure 6-29 Case Study 3: Buyer-Seller Network of HYHY. All Day 228
Figure 6-30 Case Study 3: Buyer-Seller Network HYHY Case. All Day Flags 232
Figure 6-31 Case Study 3: Scenario Instantiation of the '2M-3S' Framework 234
Figure 6-32 Case Study 3: Service System Worldview 236
Figure 6-33 Case Study 3: Service System SVNs View 236
Figure 6-34 Case Study 3: Service System Alternative Worldview 237
Figure 6-35 Case Study 3: Taxonomy Instantiation 238
Figure 6-36 Case Study 4: Processing Timelines 246
Figure 6-37 Case Study 4: Total Gap 252
Figure 6-38 Case Study 4: Total Gap Life 253
Figure 6-39 Case Study 4: Gap vs Ratios in Simulation 4 256
Figure 6-40 Case Study 4: Markets' Travelling Times 260
Figure 6-41 Case Study 4: Order Protection Indicators 265
Figure 6-42 Case Study 4: Trade Protection Indicators 266
Figure 6-43 Ideal Market Monitoring and Surveillance Engine Model 268
Figure 6-44 Case Study 4: Taxonomy Instantiation 269

Abstract

The thesis introduces a framework for analysing market monitoring and surveillance systems in order to provide a common foundation for researchers and practitioners to specify, design, implement, compare and evaluate such systems. The proposed framework serves as a reference map for researchers and practitioners to position their work in the context of market monitoring and surveillance, resulting in a useful instrument for the analysis, testing and management of such systems. More specifically, the thesis examines the new requirements for the operation of financial markets, the role of technologies, the recent consultations on the structure and governance of EU and US markets, as well as, future usage scenarios and emerging technologies. It examines the context in which market monitoring and market surveillance systems are currently been used. It reports on their processes, performance, and on the organisational and regulatory environments in which they exist. Furthermore, it develops a set of taxonomies which cover the majority of the concepts of market manipulation, market monitoring, market surveillance, entities, technologies and actors that are relevant for the work in this thesis. Building on the gaps and limitations of the current systems, it proposes a new framework following the Design Science methodology.

The usefulness of the framework is evaluated through four critical case studies, which not only help to understand with practical exercises the way how markets monitoring and surveillance systems work, but also to investigate their weaknesses, potential evolution and ways to improve them. For each case study, the thesis develops a fully working prototype tested using a sample prosecution case and evaluated in terms of the appropriateness and suitability of the proposed framework. Finally, implications relating to policies, procedures and future market structures are discussed followed by suggestions for future research.

Declaration and Copyright Statement

Declaration

I hereby swear that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

The following three notes on copyright and the ownership of intellectual property rights:

i. The Author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of School Manchester Business School (or the Vice-President) and the Dean of the faculty of Humanities, for faculty of Humanities' candidates.

To my family, my parents Ricardo and Susana, my aunts Raquel and Carmen Gloria, and my great-grandmother Raquel. They taught me everything that is essential in life. They gave me values, inspiration and support and raised me to be the person that I am. For all their love and sacrifice I thank God almighty.

Acknowledgements

First and foremost, I thank my academic supervisor Dr Babis Theodoulidis for his immense wisdom, constant support and guidance since the day I first arrived in Manchester. I also thank my academic co-supervisor Dr Pedro Sampaio, who was always accessible and offered me insightful advice every time I needed it. Babis and Pedro are not only great supervisors, but also amazing people. It has been a pleasure and an incredible experience to work with and learn from them, and without doubt I will remember the time under their supervision as the most challenging, but also the most enjoyable years of my life.

I thank my colleague PhD (c) Mohamed Zaki for his advice and friendship during the long hours we dedicated to our research. Also to my colleagues, Dr Ilias Petrounias, Dr Stefanos Strickland, Azar Shahgholian, fellow research students, and friends at the Manchester Business School.

I gratefully acknowledge all the support from the people and colleagues in the University of Chile (Universidad de Chile): My interest in research started while I was still an undergraduate student and I was motivated and encouraged at a very early stage by Dr Antonino Parisi, whom together with his brother Dr Franco Parisi, became my mentors and MSc supervisors. They not only guided and supported me to pursue a research career, but in fact convinced me to quit my job (at a prestigious bank) and follow my dreams when doing a PhD seemed impossible. I am most grateful for all the knowledge, patience and friendship they gave me during the time I was part of their research team.

I owe thanks to the Head of the Departamento de Administración MSc Pedro Hidalgo; former Dean of the Facultad de Economía y Negocios, Dr Felipe Morandé; and current Dean Dr Manuel Agosin, for giving me the opportunity to travel abroad and fulfil my PhD studies. Special thanks go to MSc Pedro Hidalgo for his kind guidance. I greatly appreciate the time and encouragement given.

This research would have not been possible without the support provided by several funding bodies. I gratefully acknowledge research funding from: the Facultad de Economía y Negocios, Universidad de Chile; the Chilean Ministry of

Planning and Development (Ministerio de Planificación y Desarrollo, MIDEPLAN), Beca Presidente de la República; and the Chilean Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Becas Chile.

Finally, I thank my friends who were always there to share joys and sorrows and were a source of encouragement throughout my work on this thesis: Many and special thanks to my girlfriend Mateja; to my flatmates Alejandro and Fabiola; to the Scoobies, Carolina, Amalia, Paula, Rodrigo and Ron; to the Chilean community in Manchester; and to all my friends around the world.

Preface

Some of the work in this thesis has been presented as joint work with my supervisors and colleagues at the Manchester Business School. Case Study 1 has been published in the *Expert Systems with Applications Journal* under the title "Analysis of Stock Market Manipulations Using Knowledge Discovery Techniques Applied to Intraday Trade Prices". A previous version of Case Study 1 was presented as a working paper at various academic conferences, including the IV European Conference on Intelligent Management Systems in Operation, Salford, UK, 2009; and the Doctoral Conference at Manchester Business School, England, June 2008.

A version of Case Study 2, together with parts of Chapter 2, Chapter 4 and Chapter 5, was presented as a working paper under the title of 'A Systematic Framework for the Analysis and Development of Financial Market Monitoring Systems' at the SRII 2011 Global Conference, March 30-April 2, San Jose, California, USA, and it is expected that proceedings papers will be published by IEEE in a forthcoming special SRII/IEEE journal.

The researcher also co-authored several papers on the application of text mining for the detection of manipulations in the stock markets that were not included here, as these were related, but not a essential part of the main topic of this thesis. A paper was published in the *Journal of Manufacturing Technology Management* Special Issue on Intelligent Management Systems and Operations. Previous version of this paper were also presented at the 16th International Conference on Information and Software Technologies, IT 2010, Kaunas, Lithuania, 2010. Another related paper entitled 'Using Text Mining to Analyze Quality Aspects of Unstructured Data: A Case Study For ‘Stock-Touting’ Spam Emails' was presented at the 16th Americas Conference on Information Systems (AMCIS), Lima, Peru, 2010.

Publications related to this work

Journal Paper:

- Díaz, D., Theodoulidis, B., and Sampaio, P. "Analysis of Stock Market Manipulations Using Knowledge Discovery Techniques Applied to Intraday Trade Prices," Expert Systems with Application Journal (38:10) 2011, pp 12757-12771.

- Zaki, M., Theodoulidis, B., and Díaz, D. "Stock-Touting Through Spam Emails: A Data Mining Case Study. Special Issue on Intelligent Management Systems and Operations" Journal of Manufacturing Technology Management (22:6) 2011, pp 770-787.

Conferences Proceedings:

- Díaz, D., Theodoulidis, B., and Sampaio, P. "Analysing Fraudulent Stock Market Manipulations. A Data Mining Case Study," IV European Conference on Intelligent Management Systems in Operation, Salford, UK, 2009, pp. 22-34.

- Díaz, D., Zaki, M., Theodoulidis, B., and Sampaio, P. "A Systematic Framework for the Analysis and Development of Financial Market Monitoring Systems," in: SRII 2011 Global Conference, San José, California, USA, 2011.

- Zaki, M., Díaz, D., and B.Theodoulidis "Using Text Mining To Analyze Quality Aspects Of Unstructured Data: A case Study For 'Stock-Touting' Spam Emails," 16th Americas Conference on Information Systems (AMCIS 2010), Lima, Peru, 2010.

- Zaki, M., Theodoulidis, B., and Díaz, D. "A Data Mining Approach for the Analysis of 'Stock-Touting' Spam Emails," in: 16th International Conference on Information and Software Technologies, IT 2010, Kaunas, Lithuania, 2010, pp. 70-79.

1 Introduction

1.1 Background and motivation

Since the early days of trading in Antwerp and Amsterdam in the 1600s, financial markets have arguably represented the most efficient and effective way to grow savings and help society to achieve economic prosperity. In fact, financial markets have a very important role to play in the creation of economic growth by enabling the flow of funds from investors (savings) to businesses (capital/loans) and then back through dividends and capital gains. However, for this virtuous circle to prosper, participants rely intensively on the quality of the markets and on fair access to them.

Indeed, it is important for society as a whole, not only for business and financial communities that financial markets operate in transparent and efficient ways while ensuring fair access for everyone. This is crucial to guarantee that investors are able to identify the best investment opportunities and consequently direct their funds to the companies that would make them grow. These are recurring topics in the economic literature, but they are also the essence behind numerous laws and regulations that protect investors globally. Generally, they are referred to as *market efficiency* and *market integrity* principles, the former describing the condition where all traders have the ability to transact easily and at low costs; whereas the latter describes the ability to transact in a fair and informed market where prices reflect all publicly available information (Comerton-Forde and Rydge, 2006).

In order to guarantee that the efficiency and integrity principles are maintained while trading is carried out, society should subject markets to continuous monitoring and surveillance following strict and rigorous standards of record keeping and audit trail. Moreover, progress should not be achieved at the expense of sacrificing or compromising the advancement of one principle to the detriment of the other. Markets need to be monitored not only with the objective of reducing market manipulations and fraud, but also with the purpose of levelling the playing field for all participants, assuring normal operation of trading systems. At the same time, good quality records should be kept and made available in a transparent and

timely manner, to completely satisfy accountability requirements when needed. If the basic principles are safeguarded, financial markets will continue to play a positive role in our society. In contrast, if they are hindered, trust quickly deteriorates and markets move away from this role, generating negative feedback which destroys prosperity and creates unfair advantages for privileged participants.

Over the years, and especially since the 1970s, the structure of the financial markets has changed substantially and regrettably the balance of principles has suffered a great deal during the process. The increased sophistication of financial markets and products and the explosive growth of complexity have not been matched by appropriate reporting requirements and monitoring structures to allow regulators and market participants to collect and audit information on market activities as well as to analyse the effect of trading behaviour on efficiency and integrity.

The changes in financial market structures have been dramatic; they include changes in the number and type of trading venues, the number and type of financial products, in addition to changes in operating rules and procedures. Very often, these changes are driven by increased competition or by the need to bypass regulatory constraints. The biggest change, however, relates to the increased use of technology and the replacement of manual transactions and human traders with automated systems. As a result, the basic principles have been put to the test and we are increasingly witnessing high-profile failures that have contributed to the erosion of trust and have put the very existence of the financial markets in jeopardy.

During the last year, both the European Commission and the US Securities and Exchange Commission (SEC) have called for public consultation on some of these issues. In the US consultation — the first in over a decade since the last comprehensive review — less than fifty percent of the buy-side respondents expressed confidence in the current market structure (Schapiro, 2010). In The EU consultation, more than four thousand two hundred comments were received (European Commission, 2010). Although market crashes and financial recessions are not new, we have witnessed an unprecedented breakdown of trust, which

demonstrates that finance and economic markets have reached a turning point, and not necessarily a good one.

1.1.1 What are the drivers behind the market structural changes?

For most people, the way markets operate has become a puzzle, and even sophisticated investors have been faced with disruptive technological innovations that have forced them to rethink and reinvent the way they conduct their business. Thanks to information technologies, an impressive amount of transactions today occur at close to the speed of light, and transactions costs, including bid-ask spreads, have gone down dramatically. Furthermore, a significant decrease in connectivity costs and a mind-boggling decrease of latency have made an almost instantaneous and seamless trading experience possible, thereby giving birth to the long-awaited realisation of truly unified and globalised markets. However, improved connectivity between markets also implies increased complexity, faster replication and contagious effects when things go wrong.

Costs, speed and connectivity are not the only things that have changed. The very essence of human-market interaction has also been transformed in the process. Human traders are increasingly replaced by computers that decide what trading strategy to deploy, execute it, and adapt it in case market conditions change; all of these actions happen faster than the blink of an eye. Because of this, some market participants argue that unequal access to technology has left them in an unfair position compared with investors who make intense use of sophisticated technology, and that they cannot follow the speed at which other firms are trading.

The public complaints appear reasonable in light of the evidence. Recent statistics show that as much as 70% of total share trading made in the US (U.S. Securities and Exchange Commission, 2010a) and 40% of the Euro markets (European Commission, 2010), was carried out using some type of algorithmic trading technique, and that the average speed of execution for small, immediately executable orders in the US markets dropped from 10.1 seconds in 2005 to 0.7 seconds in 2009 (U.S. Securities and Exchange Commission, 2010a). Furthermore, the evidence shows that in the US a typical algorithmic trading or

High Frequency Trading (HFT) firm could submit as much as 90 million orders in a single day, and cancel around 90% of those before any human eye can even glance at them (U.S. Securities and Exchange Commission, 2010a). As a consequence, a typical US stock is now only held for an average of 22 seconds (Chlistalla, 2011), which confirms that trading is decreasingly driven by fundamental reasons.

Claims of unfairness are also augmented by other forms of asymmetrical access to technology. Orders compete to use the same networking facilities, and thus orders originating from faster servers or from close-proximity servers can gain an advantage over servers at remote locations. After all, an advantage of a few milliseconds could effectively mean the difference between millions of dollars worth of losses or gains. Commissioner Barner, member of the European Commission Internal Markets and Services Directorate General, addressed these matters at a public hearing on the 'Review of the Markets in Financial Instruments Directive (MiFID)' (Barnier, 2010). He stated that:

'Technical complexity or technological developments cannot be used as an excuse to derogate from the fundamental principles of transparency, responsibility and appropriate regulation.'

Increased competition due to new regulations[1] is a powerful driver that has fostered important changes in market settings. In most countries, exchanges have ceased to hold the exclusive rights to trade the products that are listed on them, and now stocks can be traded in other markets across countries and jurisdictions. A typical world exchange has on average almost 600 listed companies, which compete to attract funding from investors from all around the globe (Cumming and Johan, 2008). For instance, at least 5% of the companies listed in a typical exchange are foreign, and least 2% are cross-listed on US exchanges (Cumming

[1] See for example the EU 'Framework Directive' (Directive 2004/39/EC), 'Implementing Directive' (Directive 2006/73/EC) and 'Implementing Regulation' (Regulation 1287/2006), as well as the, 'Regulation NMS' (SEC Act Release 51808/2005), and the 'Order Handling Rules' (SEC Act Release 37619A/1996) of the US.

and Johan, 2008). In addition, the sheer number of trading venues has also grown extensively.

Following the trend of deregulation, investors in the European zone now have the chance to access a single venue to trade across the whole European Economic Area (EEA) rather than doing it on a per venue basis. Moreover, the new regulations broke a wide range of national monopolies, not only at the trading venue level, but also in a wide variety of services, such as investment services and other specialised services (European Commission, 2010). As a result, the number of financial products offered by a particular exchange has risen from only a few to a median of eight products (Cumming and Johan, 2008).

Developing countries are also following the trend of integration and de-monopolisation. For example, Chile, Colombia and Peru merged their stock markets in May 2011, and replaced their old trading platforms with new ones capable of high-frequency trading. The new Integrated Latin American Market (MILA) allows cross-border electronic trading of shares of 565 companies listed in the three exchanges (Iturrieta and Boadle, 2011). Deregulation has also reached the derivative markets. In 1999, the US Commodities Futures Trading Commission allowed banks to hold large positions in commodities securities, and in 2004 the US SEC further expanded the banks' options by allowing them to hold positions for as much as forty times the capital they held in collateral (Knaup et al., 2011).

1.1.2 What are the consequences of market structural changes?

The structural changes in financial markets have two key consequences. Firstly, although most self-regulated exchanges do consolidate their trading information, an increasing portion of trading is done outside these exchanges. Therefore, the consolidated data represents an ever smaller view of what is happening in the markets. There is also the risk that whatever data is available gets distributed unevenly, either in terms of speed, quality, or completeness. In addition, very little consolidation is done for non-equity markets, such as derivatives, commodities or bonds. This very important information gap highlights not only the need for better and more up-to-date consolidation and record-keeping systems, but also the

significant need for ubiquitous cross-product and cross-border audit trailing, monitoring and surveillance systems.

Secondly, most of the old questionable practices and weak points of the trading system still exist, or worse still have increased. Little is known about the possible consequences if, for example, market manipulations, hacks or attacks were to be carried out with the help of sophisticated computing technologies. Furthermore, even in the absence of bad intentions, current trading and monitoring systems could behave in ways that were not foreseen, as these were not designed to cope with extremely distributed and fragmented markets, high order cancellation ratios, or a gigantic volumes of trading taking place within milliseconds. This is also true for other traditional safeguard mechanisms adopted by markets, such as circuit breakers and volatility stoppers.

1.1.3 What actions are being considered to restore trust in the markets?

The European Commission and SEC have recently commented on the extent to which these issues are perceived by market participants as real problems, and changes in regulation are foreseen. The US authorities have taken steps in the right direction by eliminating the exception of 'flash orders'[2] from quoting requirements, and by prohibiting market maker 'stub quotes'. In addition, they have launched a pilot programme in which market makers must enter quotes that are not more than 30% away from the National Best Bid and Offer (NBBO) for listed stocks among other corrections[3] (U.S. Securities and Exchange Commission, 2010b). In

[2] 'Flash quoting' is a technology made available by NASDAQ, BATS, and DirectEdge to market participants, giving them the option of showing or 'flashing' their quotes inside their markets for 20-30 milliseconds, with the objective of allowing other participants — usually HFT firms — to quote a matching offer or bettering the national best offer. If no market participant provides a matching or bettering offer, the order would be then routed to another exchange. For more information see (Brogaard, 2010)

[3] The 'Quotation Standard for Market Makers' states that quotes cannot be more than a certain percentage away from the NBBO. If the stock belongs in the S&P or Russell 1000 indices and a quote is placed during market hours, the quote cannot deviate more than 8% from the NBBO. If the

addition, a new circuit breaker pilot programme, the 'Inter-Market Volatility Trading Pauses', has been approved. A revision exercise has also started to approve new rules requiring the exchanges to clarify up-front how and when trades would be halted, and to customise circuit breakers by incorporating limit-up/limit-down types of mechanisms on a security-by-security basis (Schapiro, 2010).

SEC is also considering imposing affirmative obligations on HFT firms to provide liquidity and price continuity, as well as negative obligations to deter trading activities that would exacerbate price moves in periods of increased volatility. For example, one step would be to require a minimum 'time-in-force' for quotations, or maximum order cancellation ratios. Another step would be to require firms that use semi-market-making strategies to support market liquidity in reasonable ways during episodes of unusual volatility (Schapiro, 2010).

Additionally, both the SEC and the European Commission are trying to limit or avoid 'naked' or unfiltered access to markets for firms that have not been registered as investment firms. The authorities' proposal is to subject all HFT firms over a certain size to obligatory registration, and therefore make them subject to full regulatory oversight (European Commission, 2010). The authorities have also proposed additional obligatory requisites, such as risk management obligations and minimum capital requirements. Such rules would effectively prohibit broker-dealers, as well as HFT firms, from providing customers with unfiltered access to exchanges, and would assure they operate with appropriate risk management controls (Chlistalla, 2011).

In order to increase transparency and potentially reduce any harmful effects of dark liquidity, regulators have proposed extensive requirements for the development of new consolidated audit trailing systems. This would enhance the regulators' ability

---

quote is placed during pre-market hours it should not be more than 20% away from the NBBO. Finally, if a quote is made for non-S&P or non-Russell 1000 stocks, it should be within 30% up or down from the NBBO. In the absence of an NBBO, the same percentages apply but the consolidated last sale should be used as a reference point.

to identify significant market participants, collect information on their activity, and analyse how their trading behaviour affects the market (Schapiro, 2010).

Despite the great efforts that regulators in developed countries have made, the advances have not necessarily covered all aspects or regions in the world. Developing countries, for example, are only starting the cycle of adoption of HFT technologies, and their regulations are lagging behind. This could be seen as an irrelevant fact to developed market participants, but the reality is that the lack of consolidated data and real-time information, especially when originating from cross-market and cross-border trading, represents a weak link in the actual configuration of the markets. For instance, global surveillance providers, such as the SMARTS group, supply services of cross-market monitoring, but such systems are considerably more expensive and more difficult to coordinate, due to both cultural and human factors, as well as legal and regulatory issues (Cumming and Johan, 2008).

This weak link should make us wonder what could happen if Exchange Trade Funds (ETFs), American Depositary Receipts (ADRs), or any other cross-listed security were attacked with the intention to manipulate correlating securities in other countries. Moreover, it raises the question of what can be done to ensure that front-running practices and other typical market manipulations are correctly monitored across products, markets and borders. Perhaps more critically, how can we ensure that we not only have the ability to detect market failures but also to recover from them quickly?

In the researcher's opinion, the first barrier to tackling these important questions is the great secrecy surrounding the real capabilities and performance of market monitoring and market surveillance systems. Possibly, the reason behind so much confidentiality is the fact that monitoring and surveillance systems are far less understood and much less efficient than one could expect. According to (Cumming and Johan, 2008), for example, self-regulating exchanges carry out surveillance of on average fourteen or fifteen types of manipulations on a single-market basis, and only on two or three cross-market type of manipulations. (Cumming and Johan, 2008) also reveals that exchanges rank their surveillance efforts very poorly, with

an average score of 2.15 out 5 for effectiveness, but little is known about the reasons why these self-performance indicators are so poor.

1.2 Aims and Objectives

The aim of the thesis is the development of a framework for analysing market monitoring and surveillance systems in order to provide a common foundation for researchers and practitioners to specify, design, implement, compare and evaluate such systems. The proposed framework could serve as a reference map for researchers and practitioners to position their work in the context of market monitoring and surveillance, resulting in a useful instrument for the analysis, testing and management of such systems.

In order to achieve this aim the thesis examines the new requirements for the operation of financial markets, the role of technologies, the recent consultations on the structure and governance of EU and US markets, as well as, future usage scenarios and emerging technologies. The usefulness of the framework is evaluated through four critical case studies, which not only help to understand with practical exercises the way how markets monitoring and surveillance systems work, but also to investigate their weaknesses, potential evolution and ways to improve them. More specifically, the objectives of the thesis are:

i. Examine the context in which market monitoring and market surveillance systems are currently being used or the context in which they are expected to operate e.g. to operate in single markets and single products, as well as in cross-border or high-frequency trading scenarios.

ii. Analyse their processes, performance, and the organisational and regulatory environments in which they exist and operate. On the basis of this, identify gaps and limitations, and develop a set of requirements for market monitoring and surveillance systems.

iii. Develop a taxonomy of concepts and their relationships which form a comprehensive proposal for a body of knowledge for the different types of manipulative actions and their consequences, the agents that perform these actions, the venues where these actions takes place, the types of assets

that are manipulated and the temporal dimension of the manipulative actions.

iv. Propose and design a new framework for market monitoring and surveillance. The design of the framework will be based on a service systems approach and will explicitly address service networks.

v. Examine the proposed framework using systems thinking concepts, analyse its properties and elaborate on the use of system archetypes as a way to identify and analyse behaviour patterns of market monitoring and surveillance systems.

vi. Evaluate the proposed framework through its instantiation with four case studies that are based on real cases that have been either prosecuted by the authorities or highlighted by market participants for their importance. These four cases are critical cases as they represent a selection of scenarios in which it is possible to observe the characteristics of the proposed framework that permits logical deduction of the type 'if this (not) works/is (not) valid for this case, then it can be applied to all (no) cases' (Flyvbjerg, 2006).

vii. Discuss the implications of the proposed framework with respect to the existing market monitoring and surveillance systems, processes, policies and identify priority areas and requirements for regulatory bodies, market operators and researchers. Finally, identify corresponding business models for the operation of markets in the future adopting a culture of 'many-to-many', open, transparent and global monitoring and surveillance services.

1.3 Research Questions

In order to achieve these objectives, the thesis addresses the following research questions:

1. What are the gaps and limitations with existing market monitoring and surveillance systems in relation to process, performance, organizational and regulatory issues?

2. What are the underlying ontological foundations for the body of knowledge in market monitoring and surveillance systems?
3. How could existing market monitoring and surveillance systems be examined under a socio-technical perspective that considers markets as service systems?
4. Is it possible to design a framework that would address the gaps and limitations? How could this proposed design be able to be instantiated and evaluated under different case study scenarios?
5. What are the implications of potential changes identified in relation to processes, policies, market stakeholders, and the future of the markets?

In order to answer the above research questions the thesis develops a number of prototype detection engines that use data mining, text mining, and other data processing components and furthermore, the thesis develops a market simulator system that is used to study the behaviour of markets in a high-frequency trading context.

1.4 Structure of the thesis

The thesis is organized as follows. Chapter 2 reviews the literature and discusses existing market monitoring and surveillance systems. In addition, it reviews relevant finance topics, and it gives background information on data mining tools used in the detection of financial fraud. Chapter 3 presents the research methodology, discussing the chosen 'Design Science and Information Systems' method (Vaishnavi and Kuechler, 2004) and the 'Information Systems Design Research Framework' (Hevner et al., 2004) that are used in this work. Chapter 4 introduces the different taxonomies that have been developed, including a taxonomy of basic elements and entities present in different manipulation scenarios. It continues with the introduction of a taxonomy of the interacting agents in financial market systems, and finishes with a taxonomy of the sub-systems, components and information management concepts of a model detection engine. Chapter 5 presents a complete definition of the proposed Market Monitoring and

Surveillance Service Systems ('2M-3S') framework and examines the proposed framework using systems thinking concepts elaborating on the use of system archetypes.

Chapter 6 presents individual tools and components for the detection of information-based and trade-based manipulations and provides details of the case studies that give them context. In Case Study 1, the challenge of building market monitoring and surveillance components for the detection of traded-based manipulations is addressed. These are built using real trading and textual data, and are based on real cases of manipulation. Case Study 2, deals with the challenge of detection of insider trading manipulations. This case study highlights the ability of the systems to discriminate between this and other types of market manipulations, in particular trade-based manipulations, as these tend to have similar patterns, but represent a different family of manipulation, i.e. information-based manipulations. Case Study 3 addresses the challenge of using social network analysis components for the detection of any kind of manipulation in which the coordination of different traders is taking place. Although the proposed prototype analysis architecture is very similar to that presented in Case Study 2, the difference lies in two aspects: firstly in the availability of fully functional social networking analysis components that are introduced in this case using the NodeXL component; and secondly in the instantiation of the framework as a special case in which the customer and provider of the detection engine are both part of a single regulatory entity, in this case US the regulator. In Case Study 4 the consequences of HFT practices is studied. In order to do so, a market simulator is built and used to model trading systems and their behaviour when confronted with HFT quote-stuffing practices. The case presents evidence on how HFT quote stuffing can increase the gap of best bid and ask prices between markets, contrary to the evidence that HFT is helping the alignment of prices across markets.

Chapter 7 summarises this thesis and its contributions and discuss the implications of the proposed framework with respect to the existing market monitoring and surveillance systems and corresponding business models. The chapter also

identifies priority areas for regulatory bodies, market operators and discuss avenues for future research.

2 Literature review

This chapter reviews the literature on Market Monitoring and Surveillance Systems ('MMSS'), describing the context in which they exist as wells as their capabilities and performance. It begins with a revision of financial topics regarding market manipulations and continues by reviewing existing systems. The MMSS are discussed in terms of the types of market manipulations they detect, the number of products for which they are deployed, the countries that use them, the performance of their components, and other typical processes and policies that characterise them. Next, this chapter briefly reviews the technologies that facilitate the tasks performed by the MMSS, and continues by presenting examples of systems that are reported in the literature. The chapter concludes with a summary and discussion of the requirements, limitations and gaps in the existing MMSS.

2.1 Review of financial economics literature

Irresistible impulses to predict the stock market have existed since trading first began. This matter has intrigued generations of traders, and it is a major point of controversy among the academic community. In this controversy, both academics and market traders have taken an active interest and contributed to the development of the field.

On one hand, the Efficient Market Hypothesis (EMH) (Fama, 1970) implies that the investor can only obtain a normal return when investing in the stock market. As a corollary, if the EMH is correct in its strong efficient form, the market will react instantaneously to the appearance of new information, adjusting the prices in such a way that investors will always be paid a fair price. Thus, since the appearance of new information is of random character, the prices would likewise share the same condition. On the other hand, the financial literature has uncovered and analysed many peculiarities in equity returns that do not hold under the EMH. If the prices are random, there will be no room for regularities, yet there is evidence of regularity

in certain kinds of assets under certain kinds of conditions. For instance, closing prices should reflect all the information that has been made available during trading hours but, according to (Harris, 1989), transaction prices systematically rise at the close without any relation to the appearance of new information.

The possibility that these regularities bear some relation to artificial manipulation of the price has been an important issue for everybody who is involved in trading stocks. For example, as trader performance is benchmarked against the closing prices, it is possible that traders try to influence the closing price, in particular if they have acquired a large net position in a stock during the trading day (Felixson and Pelli, 1999). Among the financial community, (Vila, 1989) was one of the first authors to present market manipulation as an example of game theory. Previous research considered perfect competition to be a standard assumption in financial economics. Stock markets and futures markets were often given as textbook examples of a Walrasian paradigm. Following the work from (Vila, 1989), (Benabou and Laroque, 1992), (Kumar and Seppi, 1992), and (Bagnoli and Lipman, 1996) developed special models to demonstrate the possibility of manipulation. (Allen and Gorton, 1992) introduced an asymmetry to the models and showed that profitable manipulation is possible.

(Allen and Gale, 1992) introduced a theoretical framework for a special kind of manipulation called trade-based manipulation. This is regarded as the manipulation implemented by buying and selling shares. Previously, it had been thought that trade-based manipulation could not be profitable. (Allen and Gale, 1992) showed that in a rational expectations framework, where all agents maximise expected utility, it is possible for an uninformed manipulator to make a profit.

In a dynamic model of asset markets, (Jarrow, 1992) investigated market manipulation trading strategies by large traders. Market manipulation trading strategies are shown to exist under reasonable hypotheses on the equilibrium price process. Profitable speculation is possible if there is price momentum, so that an increase in price caused by the speculator’s trade at one date tends to increase prices at future dates.

(Kumar and Seppi, 1992) investigated Salomon Brother's market corner of a Treasury note auction in May 1991. (Gerard and Nanda, 1993) examined the potential that security markets have for market manipulation. (Jarrow, 1994) studied a new theory for pricing options in a large trader economy that relies on the derivative security markets' impacts on market manipulation. (Zweig, 1997) demonstrated year-end seasonality in equity funds due to the manipulation of year-end valuations by some fund managers to improve their fund's return, given some insights in marking-the-close activities.

(Felixson and Pelli, 1999) tested for closing price manipulation in the Finnish stock market and found evidence that this occurs. Block trades and spread trades explained a part, but not all of the observed manipulation. This work also discusses some empirical studies on the effect of expiration-days on the underlying stock prices ((Chamberlain et al., 1989), (Stoll and Whaley, 1991)). The general results of these studies were that effects of manipulation could be found in at least the last hour before expiration. (Mahoney, 1999) examined stock price manipulation leading up to the Securities Exchange Act of 1934. (Vitale, 2000) examined manipulation in the foreign exchange market. (Leinweber, 2001) presented a time-line of market manipulation from the 1600s to internet-based scams of the 2000s.

(Carhart et al., 2002) directly syndicated market manipulations, especially 'marking-the-close', as one of the main causes of the phenomena, on both a day-to-day and intraday basis. They present evidence that fund managers inflate quarter-end portfolio prices with last-minute purchases of stocks already held, and that inflation is greatest for the stocks held by funds with the most incentive to inflate. They also propose two hypotheses about the reasons for marking-up: 'leaning for the tape' and 'beating the S&P500 benchmark', of which they accept the first reason to be valid.

(Chunsheng et al., 2003) constructed a theoretical example in which 'smart money' strategically takes advantage of investors' behavioural biases and manipulates the price process to make a profit. As an empirical test, their paper presents some evidence from the US SEC prosecution of pump-and-dump manipulation cases. The findings from these cases are consistent with their model.

(Khwaja and Mian, 2005) analysed a dataset containing daily firm-level trades for every broker trading on the stock exchange in Pakistan. They found that brokers earn at least 8% higher returns on their own trades. While neither market timing nor liquidity provision offer sufficient explanations for these results, they found compelling evidence for a specific trade-based pump-and-dump price manipulation scheme. (Merrick et al., 2005) provided empirical evidence on learning in the market place and on the strategic behaviour of market participants by studying an attempted delivery squeeze in the March 1998 long-term UK government bond futures contracts traded on the London International Financial Futures and Options Exchange (LIFFE).

Extending the framework of (Allen and Gale, 1992), (Aggarwal and Wu, 2006) presented a theory and some empirical evidence on stock price manipulation in the United States. At the theoretical level, they considered what happens when a manipulator can trade in the presence of other traders who seek out information about the stock's true value. They found that in a market without manipulators, these information seekers unambiguously improve market efficiency by pushing prices up to the level indicated by the informed party's information. In a market with manipulators, however, the information seekers play a more ambiguous role. More information seekers imply greater competition for shares in a market with manipulators, making it easier for a manipulator to enter the market and potentially worsen market efficiency.

At the empirical evidence level they found some interesting patterns. Using a dataset from SEC actions in cases of stock manipulation, they found that more illiquid stocks are more likely to be manipulated and manipulations increase stock volatility; daily stock prices rise throughout the manipulation period and then fall in the post-manipulation period; prices and liquidity are higher when the manipulator sells than when the manipulator buys; and at the time when the manipulator sells, prices are higher when liquidity is greater and when volatility is greater.

Other works also looked at the role of new information in the behaviour of prices when traders disseminate rumours in the market about their trades; one of the most influential of these studies is that of (Van Bommel, 2003). Continuing this line

of enquiry, based on a large sample of touted stocks listed on the Pink Sheets quotation system and a large sample of spam emails touting stocks, (Frieder and Zittrain, 2007) found that stock experienced a significantly positive return on days prior to heavy touting and negative returns in the following days.

In one of the most recent and complete reviews on the topic, (Putnins, 2009) examined the prevalence, effects and determinants of closing price manipulations. He founds that approximately 1.1% of closing prices are manipulated. Furthermore, for every prosecuted closing price manipulation, there are approximately three hundred instances of manipulation that remain undetected or are not prosecuted.

In summary, this economic literature review presented a brief overview of how the problem of market manipulation has been historically defined and investigated by the financial and economic community, and has highlighted that manipulations do indeed exist, and that they can be very profitable for manipulators but detrimental to other market participants.

2.2 Market monitoring and surveillance: context, capabilities and performance

2.2.1 Review of modern financial markets organisation

Cheaper, better and faster servers, as well as the networking technology that supports them, have enabled the integration of traditional markets, but also the creation of new, private restricted-access venues, also known as 'undisplayed' or 'dark' liquidity venues. Typical users of these venues are now institutional investors who want to execute big trading orders without substantial alteration of prices. Their transactions become 'dark' because prices and quantities traded in private venues are normally not under the same reporting standards demanded for traditional exchanges, and for example, the best bid-ask orders are not required to be included in national or regional consolidated trading data tapes. In the US, approximately 7.9% of the share volume in equity trading is executed in these 'dark pools' (U.S. Securities and Exchange Commission, 2010a).

Dark liquidity is also present in broker-dealers' internalisation processes in which client orders are internally matched before routing them into a trading venue. Broker-dealers rely on 'crossing-systems'; automated systems which use algorithms to break big client 'parent' orders into 'child' orders and automatically search for matches between them. The broker-dealers' internalisation processes accounted for 17.5% of share volume in the US during the 3rd quarter of 2009 and it was routed using different over-the-counter (OTC) market maker networks[4].

[4] Broker-dealers are broadly grouped into two categories: Over-the-Counter (OTC) market makers and block positioners. Rule 600(b)(52) of the US Regulation of National Market System (NMS) defines an OTC market maker as 'any dealer that holds itself out being as willing to buy and sell to its customers, or others, in the United States, an NMS stock for its own account on a regular or continuous basis otherwise than on a national securities exchange in amounts of less than block size'. 'Block size' is defined in Rule 600(b)(9) as an order of at least 10,000 shares or for a quantity of stock having a market value of at least $200,000. A block positioner, then, is any broker-dealer that executes block-sized trades, either for himself or herself or on behalf of others.

The combined effects of the integration and fragmentation of markets are reflected in a new distribution of trading volume. A decade ago, the New York Stock Exchange was responsible for around 80% of the volume traded in the US. Today that figure has gone down to around 26% (U.S. Securities and Exchange Commission, 2010a). Trading volume is now spread among four types of semi- or fully-automated trading venues, namely: *traditional registered exchanges*, such as NYSE or AMEX; *alternative trading networks* (mainly electronic communicating networks such as Direct Edge or BATS); more than thirty *private* or *'dark liquidity' venues*; and more than two hundred *broker-dealer internalisation facilities* (U.S. Securities and Exchange Commission, 2010a).

In this new configuration of markets, exchanges have come up with new and inventive ways to compete, offering a full range of new products and services. For instance, it is now common for exchanges to offer co-location services for trading firms interested in hosting their servers as close as possible to the exchange servers in order to gain a few milliseconds' speed advantage. It is also increasingly common for these trading firms to provide 'naked' or 'sponsored access services' to third parties.

The fact that many exchanges have changed their traditional revenue models in an effort to attract trading volume is perhaps the most significant consequence of the integration-fragmentation-competition trio. Exchanges have adopted a 'maker-taker' pricing model, in which makers or providers of liquidity are encouraged to trade in the venue by giving them a rebate. In some cases, venues have adopted an even more aggressive revenue strategy, offering rebates that are higher than the access fees, and thus effectively paying makers of liquidity more than the venue receives from takers (U.S. Securities and Exchange Commission, 2010a).

This pricing model has encouraged certain HFT firms to adopt a semi-market-making role, in which they continuously provide liquidity through slightly better non-marketable orders than the ones available on the market. This is done by sending large amounts of resting orders at high speed that are very briefly available to the market with the purpose of being the first to meet the demand for a particular security, thereby profiting from the rebates with minimum risk exposure.

HFT signifies as much as 70% of the volume traded in the US, but as much as 90% of those orders are cancelled within milliseconds. This practice of sending thousands of orders in a very small time window has been termed 'stuff-quoting'. Most high frequency trading firms are proprietary firms, and consequently risk their capital to provide this type of liquidity. Hence, HFT firms typically avoid holding positions in a particular security for long periods and, accordingly, quickly abandon the semi-market-making role when facing volatile markets. The new revenue model similarly enables them to move quickly from the buy to the sell side, or vice versa, and usually HFT firms complete their strategies by closing the day holding flat or near flat positions, i.e. holding either only cash, or only stocks and securities that were purchased before the trading day started.

In practice HFT firms have become unofficial market markers with none of the traditional market-maker obligations. Moreover, HFT firms are not required to quote good quality two-sided orders and could stop providing them at any moment. HFT orders are typically of reduced size, and thus their contribution is marginal to the depth-of-the-book (Brogaard, 2010). In simple terms this means that even though HFT firms are willing to offer a buy or sell side at a given price, the amount of shares offered at this price will be typically small and available only for a few milliseconds.

2.2.1.1 The 'flash crash'

Increased competition and new technology drivers have created the setting for big investors to move towards the dark pools. However, market integration and de-monopolisation allow big investors to stop trading in dark pools at any moment and return to trading in traditional markets if they consider it necessary. This means that traditional markets have to be ready not only for light investors' trading, but also have to be prepared for large trading volumes originating from dark pools without prior warning.

Unfortunately, in reality most trading venues are under- or unprepared for this, and such episodes could carry terrible consequences. In what is now known as the 'flash crash', that took place during the afternoon of 6th May 2010, equity markets fell sharply in a matter of minutes, only to recover almost as fast as they dropped.

The Dow Jones Industrial Average (DJI) fell 600 points, and blue-chip companies such as Accenture and Procter and Gamble saw their stocks plummeting as much as 30% before recovering. In that brief period over 20,000 trades for around 300 securities were executed at prices that differed more than 60% from their previous values, the majority of them executed at a penny or less (U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, 2010a).

In words of the US SEC Chairman:

'*... markets had experienced the worst price decline and reversal since 1929. At 2:40 on the afternoon of May 6, the broad market indexes dropped more than 5 percent in five minutes, only to rebound almost entirely in the next 90 seconds*' (Schapiro, 2010).

Yet, in the world's most modern and sophisticated trading venues, the exact origin of such a major fault is still not clear. Once again, despite the severity of the problem, the huge complexity of the market interactions left the authorities clueless about what had just happened. Alarmingly, it took six days to preliminarily discard the possibility that these events has been triggered by terrorist attacks, hackers or human errors. However, even then such causes could not be completely discarded (U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, 2010b). What went wrong? When the official explanations finally came out — six months later — the authorities reported that a large fundamental trader sold 4.1 billion dollars worth of 'E-Mini S&P 500' future contracts in a small amount of time. A computerised 'sell' algorithm, acting on behalf of the trader, sold 75,000 E-Mini contracts in less than 20 minutes (approximately 62 contracts per second), which started one of the biggest and quickest snowball effects ever seen in financial history. According to the joint US Commodity Futures Trading Commission (CFTC) and SEC investigation, an incredible amount of two billion dollars of stop loss orders were activated in less than half an hour between 2:30 and 3:00 p.m. (U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, 2010a). Why did the exchange systems not stop the rogue orders from being executed before the market collapsed? Because the systems were simply not prepared to do so.

What really triggered the crisis, however, were the unforeseen interactions of rogue trading algorithms and 'stuff-quoting' practices. Initially, the selling pressure was absorbed by the HFT semi-market-making behaviour, but eventually they ceased to provide liquidity when their positions started to accumulate at one side of the market. The crisis was then further exacerbated by the unforeseen interactions of 'stub quoting[5]' and 'stop loss orders[6]'. When the prices plummeted, HFTs withdrew from the markets, and market makers' stub quotes interacted with stop loss orders, reinforcing the downward trading spiral. Automated systems continued to execute trades at more and more absurdly reduced prices. Finally, the circuit breakers that were there to stop this kind of episode failed to activate. They had been designed to halt trading if the DJI index dropped more than 10% in less than five minutes. This limit, however, was never reached (U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, 2010a).

2.2.2 Market manipulation definitions

Market manipulations matter because they are detrimental to financial markets and their participants (Putnins, 2009). In particular, manipulations can encourage investors to trade in alternative or 'dark markets', thus decreasing liquidity and increasing trading costs. Moreover, manipulations can have a direct effect on price formation, making it increasingly difficult for savers to analyse and detect the best investments. If this happens, they will demand a higher return for their investments, thus making it more expensive for firms to raise funds for their projects. As a result,

---

[5] A 'stub quote' differs from 'stuff quoting' because the former are offers made by an official market maker, while the latter are made by HFTs that are not obliged to do so. According to SEC, a 'stub quote' is an offer to buy or sell a stock at a price so far away from the prevailing market that it is not intended to be executed, such as an order to buy at a penny or an offer to sell at $100,000. A market maker may enter stub quotes to nominally comply with its obligation to maintain a two-sided quotation at those times when it does not wish to actively provide liquidity

[6] Stop loss orders are put in place, usually by individual investors, to reduce the potential losses in the presence of a volatile market by selling the stocks when market prices reach certain levels.

firms will postpone projects that would have been implemented otherwise, eventually reducing the creation of jobs and slowing economic activity in general.

After the Great Crash of 1929, the US authorities created one of the first and most influential laws to enact anti-fraud provisions for financial markets. The US Securities Exchange Act of 1934 outlawed several types of market manipulation and established a dedicated Commission to tackle these issues.

(Allen and Gale, 1992) classify manipulations into three categories. The first category comprises *action-based manipulations*, which are attempts to alter the price of the stocks by taking actual actions that modify the future cash flows of the companies, for e.g. by selling a branch of the company without informing stockholders; the second is *information-based manipulations*, which are attempts to alter the price by releasing false information or rumours. The third category, *trade-based manipulations*, occurs when a trader attempts to manipulate a stock by simply buying and selling stocks, without taking any other publicly observable action or releasing false information to change the price.

The first category has been tackled by, among other things, making it illegal for directors and officers to sell the securities of their own firm short, or more generally speaking making it illegal to trade using any kind of information that has not been previously announced to the public, such as the selling of a strategic asset, or the merging of the company. The second category has been tackled by requiring firms to issue information to the public on a regular basis, and making it illegal for anybody to attempt to raise or depress the price by making statements which they know to be false (Allen and Gale, 1992). The third category is harder to eradicate, and implies that suspicious transactional behaviour could be prosecuted after fraud transactions occurs, leaving a trace that could be extremely difficult to detect. Because action-based manipulations are mostly related to activities that occur outside the market place, this thesis mainly focuses on information- and trade-based manipulations.

Legal definitions of trade-based and information-based manipulations, by contrast, are often very vague, and usually the courts bear the responsibility of judging what constitutes a manipulation and what does not (Putnins, 2009). For instance, in

Section 10(b) of the US Securities and Exchange Act 1934, it is considered against the law *'to use or employ, in connexion with the purchase or sale of any security... any manipulative or deceptive device or contrivance'*.

The Commodity Futures Trading Commission (CFTC) considers market manipulation as an illegal way of *'inhibiting market trades and describes market manipulation as activities with the capability of forcing a change in price thereby, causing artificial price*'. Market manipulation also includes unusual high-volume of sales and purchases, and the release of false information with the aim of causing the distortion of prices.

In the EU, statutory law Section 1041A, Corporations Act 2001, declares that *'market manipulation shall mean transactions or orders to trade which give, or are likely to give, false or misleading signals as to the supply of, demand for or price of financial instruments, or which secure ... the price of one or several financial instruments at an abnormal or artificial level'*.

The UK's Financial Services Authority Handbook defines trade-based manipulations as *'the behaviour which consists of effecting transactions or orders of trade which: give, or are likely to give a false or misleading impression as to the supply of, or demand for, or as to the price of one or more qualifying investments or secure the price of one or more such investments at an abnormal or artificial level'* (Financial Services Authority, 2007). Examples of trade-based manipulation given in the Handbook include matched-order transactions, wash-sales, runs, collusion, market stabilisation, and marking the close.

2.2.3 Types of market manipulations and taxonomy

As can be seen, the broad definitions utilised by the authorities make it very difficult to classify and distinguish specific types of manipulation. In order to try as much as possible to follow a standard definition of market manipulations, this thesis adapts and extends the taxonomy proposed by (Putnins, 2009). In order to make Putnins' taxonomy more complete, additional techniques and types of market manipulation are included, using mainly the collection of definitions suggested in (Cumming and Johan, 2008). This was then further expanded and adapted with inputs from the

author, but preserving where possible the original hierarchies. According to (Putnins, 2009), at the highest level manipulations can be divided into three categories: 'abuse of market power', 'contract-based' and 'runs and raids'. At the second level, the categories are divided following the (Allen and Gale, 1992) definitions of 'information-based', 'trade-based' and 'action-based' manipulations. At the base level, the taxonomy was expanded to consider a fourth category, namely, 'breach of fiduciary duty', and was simplified by excluding 'action-based' manipulations from the second level. These changes were made to accommodate more types of information-based manipulation, such as insider trading, and to cater for new types of manipulation, such as stuff quoting practices (see Figure 2-1).

The types of manipulation listed above have the following properties. In a ‘run’, a manipulator attempts to inflate or deflate the price of a security by either buying or short selling the targeted security, and then tries to attract liquidity to it. The manipulator later sells to, or buys from, unaware investors and makes a profit by reversing the original position. The most common types of run are known as 'pump and dump' schemes. In a bear raid, the manipulator firstly borrows or 'short sells' a security, and then profits from manipulating the price towards a depressed position. The manipulator then buys back the security at the depressed price, and returns the security to the original borrower, making a profit from the difference between the prices she obtained in the first instance by selling the security that was borrowed and the price at which it was repurchased. It is also common to find runs and raids which include information-based manipulations. In a 'boiler room', for example, registered and unregistered salespersons engage in fraudulent, high-pressure sales tactics in the offer and sale of securities and stocks to retail customers. The pressured selling is usually done over the phone, in person, or by mailing/e-mailing campaigns, disseminating false or misleading information through these means.

It is important to notice that in runs and raids, the manipulator profits directly from the market in which the manipulation occurred. In contrast, in contract-based manipulations the manipulator makes the profits outside the market of the security. For instance, this is the case for marking-the-close manipulations, in which portfolio

managers try to manipulate the prices of stocks or securities which are included in the funds they manage. If the prices of the securities included in the portfolio are artificially high, then they can charge their customers extra fees and commission. ‘Abuses of market power’ manipulations are closely related to the runs, but tend to be associated mainly with manipulations that take place in markets which trade in physical goods. In a 'squeeze', for example, the manipulator could profit by controlling or altering the supply and demand of a security or commodity.

There are also manipulations which are very difficult to classify as only one type. For instance, a derivative contract holder could profit by manipulating the price of an underlying asset, holding the entire physical inventory of the commodity (or at least by giving the impression of doing so). This is typically the case in cross-product and cross-market manipulations such as 'corners'. In a corner, a manipulator may take a position in both the derivative and the commodity markets and attempt to control one, or both, markets. For instance, this could be done by artificially increasing the price of a commodity by placing thousands of orders to buy that are later cancelled. If the increased number of orders is perceived by the other participants as evidence of resting market buying power, they could be deceived into actually placing and executing trading orders, which would in turn move the derivative price in the other market.

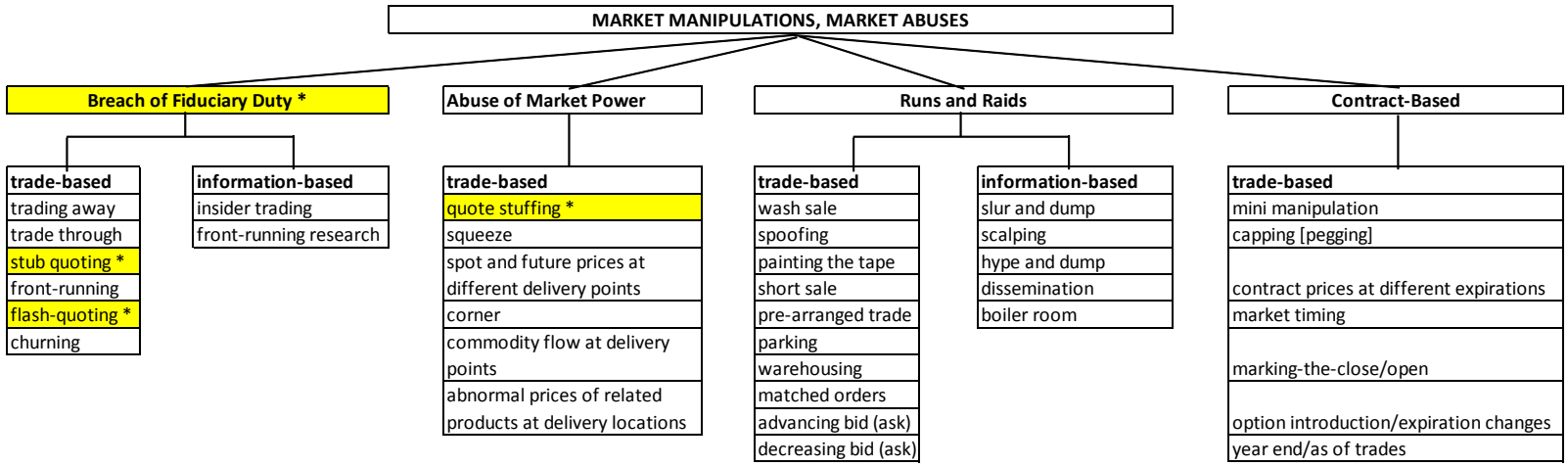


Figure 2-1 Taxonomy of Market Manipulations and Abuses. Adapted from (Putnins, 2009) and (Cumming and Johan, 2008). * Indicates original taxonomy has been expanded by the author including a new category or element.

As mentioned, a fourth group of manipulations has been added to the taxonomy, namely, 'breach of fiduciary duty'. This refers to market abuse practices or manipulations in which a market participant disrespects a regulatory obligation or contractual duty and is usually related to the way trading should be conducted in an exchange. This could be the case of 'front running orders', in which a broker-dealer executes orders to trade ahead of the orders coming from customers, to the customers' detriment. For instance, if a customer places a large order to buy a security, the increased demand is likely to push the price up. A broker-dealer could profit from this information by placing buying orders that immediately precede the customer's order.

Given the almost endless possibilities and combinations of manipulation, and the continuous adaptation and evolution of the manipulators' strategies, it is impossible to present a full and definitive list of manipulative behaviours. Table 2-1, however, summarises a list of market manipulations and their definitions that could be use as a working reference.

Table 2-1 Market Manipulations, Categories, Definitions and Types. Adapted from (Cumming and Johan, 2008).

Manipulation	Known as	Category	Type	Definition
1. Abnormal prices of related products at delivery locations		Abuse of market power	Trade-based	The expiring futures price and the spot price at the delivery market are abnormally high relative to prices at other, non-deliverable locations; the prices of related products; and prices of non-deliverable grades of the same commodity.
2. Advancing [decreasing] the bid [ask]		Runs and raids	Trade-based	Increasing [decreasing] the bid [ask] for a security or derivative to increase [decrease] its price.
3. Boiler room	pumping	Runs and raids	Information-based	A tactic in which registered and unregistered salespersons engage in fraudulent, high-pressure sales tactics in the offer and sale of securities and stocks to retail customers. The pressured selling is usually done over the phone, in person, or by mailing/e-mailing campaigns.
4. Capping [pegging]		Contract-based	Trade-based	Effecting transactions of the instrument underlying an option shortly before the option's expiration date to prevent a rise/decline in price of the instrument so previously written call/put options will expire worthlessly, protecting premiums previously received.
5. Churning		Breach of fiduciary duty	Trade-based	Frequent and excessive trading of a client's account.
6. Commodity flows to delivery points (1)		Abuse of market power	Trade-based	Large shipments of the commodity flow to the delivery point immediately prior to and during the delivery period. Moreover, shipments from the delivery point are abnormally small during the delivery period as traders amass stocks to make delivery.
7. Commodity flows to delivery points (2)		Abuse of market power	Trade-based	Delivery point receipts are abnormally small after the delivery period because of the glut of the commodity at the delivery point that results from the artificially large receipts during the delivery period. Shipments from the delivery point increase after the end of a corner as some of the excess shipments are returned to their original sources and delayed shipments are released.
8. Contract prices at different expirations		Contract-based	Trade-based	The price of the manipulated contract is abnormally high relative to the price of the contracts expiring later (that is, the price of the 'front month' contract is artificially high relative to the deferred or 'back month' contracts).
9. Corner		Abuse of market power	Trade-based	Securing control of the bid/demand-side of both the derivative and the underlying asset. The dominant position can be exploited to manipulate the price of the derivative and/or the asset.
10. Dissemination	touting, spamming, e-mail blasting	Runs and raids	Information-based	Dissemination of false or misleading market information.

11. Flash quoting*	flashing	Breach of fiduciary duty	Trade-based	Flash quoting' is a technology that some venues make available to market participants, giving them the option of showing or 'flashing' their quotes inside their markets for 20-30 milliseconds, with the objective of allowing other participants — usually High Frequency Trading firms — to quote a matching offer or better the national best offer. If no market participant provides a matching or bettering offer, the order would be then routed to another exchange.
12. Front running client orders	intervention	Breach of fiduciary duty	Trade-based	A transaction to the detriment of the order giver on the basis of and ahead of an order which he is to carry out for another.
13. Front running research		Breach of fiduciary duty	Information-based	Misuse of price or volume confidential and sensitive information as a result of research activities.
14. Hype and dump	pump and dump; pumping, touting	Runs and raids	Information-based	Buying at increasingly higher prices. Securities are sold in the market (often to retail customers) at the higher prices.
15. Insider trading	pump and dump; pumping, touting	Breach of fiduciary duty	Information-based	When a trade has been influenced by the privileged possession of corporate information or price-sensitive market order that has not yet been made public.
16. Market timing	late trading	Contract-based	Trade-based	Market timing includes (a) frequent buying and selling of shares of the same mutual fund or (b) buying or selling mutual fund shares in order to exploit inefficiencies in mutual fund pricing. Market timing, while not illegal per se, can harm other mutual fund shareholders because it can dilute the value of their shares if the market timer is exploiting pricing inefficiencies, or disrupt the management of the mutual fund's investment portfolio and can cause the targeted mutual fund to incur costs borne by other shareholders to accommodate frequent buying and selling of shares by the market timer.
17. Marking the close	'painting the tape'; 'stabilisation'	Contract-based	Trade-based	Buying or selling securities or derivatives contracts at the close of the market in an effort to alter the closing price of the security or derivatives contract.
18. Marking the open	'painting the tape'; 'stabilisation'	Contract-based	Trade-based	The placing of purchase orders at slightly higher prices/sale orders at lower prices to drive up/suppress the price of the securities when the market opens.
19. Matched orders	'pools'; 'collusion'	Runs and raids	Trade-based	Transactions where both buy and sell orders are entered at the same time with the same price and quantity by different but colluding parties.
20. Mini manipulation		Contract-based	Trade-based	Trading in the underlying security of an option in order to manipulate its price so that the options will become in-the-money.

21. Option expiration date stock price or volume changes		Contract-based	Trade-based	Unusual changes in the stock price and/or trading volume around the date of expiration of the option.
22. Option introduction date stock price or volume changes		Contract-based	Trade-based	Unusual changes in the stock price and/or trading volume around the date of introduction of the option.
23. Parking or warehousing		Runs and raids	Trade-based	(a) Hiding the true ownership of securities/underlying by creating a set of fictitious transactions and trades. (b) the sale of securities subject to an agreement or understanding that the securities will be repurchased by the seller at a later time and at a price which places the economic risk on the seller.
24. Pre-arranged trade	"pools'; 'collusion'	Runs and raids	Trade-based	Transactions in which the price, terms, or contra-side have been prearranged.
25. Scalping		Runs and raids	Information-based	A practice in which a manipulator, usually a blogger, TV presenter, or newsletter reporter, writes/disseminates a false or misleading report about a company whose stock he/she already owns.
26. Short sales	bear raid; free riding	Runs and raids	Trade-based	A market transaction in which an investor sells stock he does not have or he has borrowed in anticipation of a price decline. This is not per se manipulative but is considered manipulative in some jurisdictions in conjunction with other types of actions; for example, in Canada, under UMIR Rule 6.2(viii)(ix), a short sale cannot be at a price that is less than the last sale price.
27. Slur and dump	pump and dump; pumping, touting	Runs and raids	Information-based	Buying at increasingly higher prices. Securities are sold in the market (often to retail customers) at the higher prices.
28. Spoofing/painting the tape		Runs and raids	Trade-based	Engaging in a series of transactions reported on a public display facility to give the impression of activity or price movement in a security (e.g., misleading trading, switches, giving up priority, layering bid/asks, fictitious orders for the case of spoofing, etc.).
29. Spot and futures prices at different delivery points		Abuse of market power	Trade-based	The spot price in the delivery market declines both absolutely and relative to deferred month futures prices and spot prices at other locations around the end of futures trading or the delivery period.

30. Squeeze		Abuse of market power	Trade-based	Taking advantage of a shortage in an asset by controlling the demand side and exploiting market congestion during such shortages in such a way as to create artificial prices.
31. Stub quoting*		Breach of fiduciary duty	Trade-based	According to SEC, a 'stub quote' is an offer to buy or sell a stock at a price so far away from the prevailing market that it is not intended to be executed, such as an order to buy at a penny or an offer to sell at $100,000. A market maker may enter stub quotes to nominally comply with its obligation to maintain a two-sided quotation at those times when it does not wish to actively provide liquidity. A 'stub quote' differs from 'stuff quoting' because the formers are offers made by an official market maker, while the latter are made by HFTs that are not obliged to do so.
32. Quote stuffing*		Abuse of market power	Trade-based	Sending large amounts of orders at high speed usually cancelling the majority within milliseconds.
33. Trade through		Breach of fiduciary duty	Trade-based	The completion of a client's order at a price inferior to the best posted bid or ask. This is not per se considered manipulative, but many commentators (and the surveillance authorities themselves) consider it manipulative because the market maker who receives the order is unable or unwilling to fill it at the best posted bid or ask price, and hence the trade is instead executed at the market maker's price.
34. Trading away		Breach of fiduciary duty	Trade-based	Refers to brokers trading for their personal account through a brokerage other than their employer. It can also involve private placement transactions structured between a broker and a client without the knowledge of the employer.
35. Wash sale	'pools'	Runs and raids	Trade-based	Improper transaction in which there is no genuine change in actual ownership of the security or derivative contract.
36. Year end /as of trades		Contract-based	Trade-based	Transactions executed at a particular date to establish gains or losses or conceal portfolio losses or true positions.

Shaded box indicates definition was updated by the author. * Indicates definition was added by the author.

2.2.4 Cross-border, cross-market, cross-product and cross-jurisdiction manipulations

Touching on the previous topics are the concepts of cross-border and cross-market manipulations. According to the World Trade Organisation, under the General Agreement on Trade Services (GATS), cross-border trade is defined *as the provision of financial services by a financial firm located in one country to a costumer residing in another country without the establishment of a commercial presence, such as a branch or subsidiary, in the country of the customer (the 'host country')* (Christiansen, 2000). Such financial services include both wholesale and retail market segments and include services such as, wholesale commercial banking, investment banking, retail commercial banking, wealth management, insurance services, and financial information and data processing. In relation to securities markets, the governance and regulatory structures of cross-border trade with reference to monitoring and surveillance vary greatly and they depend on the characteristics of national and international regulatory frameworks and agreements.

The relationship between the financial markets and the traded commodities and the possibility of profits to be earned by manipulating cross-markets was studied by (Pirrong, 1995, Pirrong, 2004). Manipulation can also be carried out among cross-markets. As manipulation produces adverse effects on single markets, so it does it have detrimental effects on cross-border markets. Price discovery is meant to be natural and fair in individual markets as well as cross-border markets, but cross-border manipulation can have negative effects on the price discovery between these markets.

On this basis, the challenges with respect to cross-border monitoring activities are manifold and include not only the lack of global regulatory frameworks, but also business and technological issues. The first challenge is that data quality is at the heart of the concerns of regulators, including pre and post-trade transparency, data consolidation and access costs (European Commission, 2010). Furthermore, the cross-border challenges include the development of processes and the use of ICT

technologies that support the detection and investigation of potential trading rules violations with special emphasis on market place rules, and the codes of conduct of exchanges and trading venues (Polansky et al., 2004). Indeed, it is not clear how the traditional view of exchanges as the custodians of the enforcement functions –their capacity to monitor markets and bring cases to the attention of regulators- can be effective without the necessary ICT technology and business incentives especially considering the transformation of exchanges into listed companies amid a shrinking revenue base caused by increased competition (Christiansen and Koldertsova, 2009).

The ownership and control of stock exchanges by overseas entities can raise concerns with respect to their role as custodians of enforcement functions. To address these concerns, some governments such as the UK one, have been explicit in protecting their regulatory frameworks by passing bills that ensure that their regulatory approach is not threatened by any takeover of UK exchanges or clearing houses (Balls, 2007).

The challenges of cross-border monitoring and surveillance are reflected in the number of market manipulation cases that have been reported by the regulating authorities. For example, in the US, only seven cases of cross-border manipulation have been prosecuted during the last ten years as opposed to the thousands of cases of insider trading that have been investigated during the same period. In the UK, where the regulating authority FSA operates a risk-based approach to enforcement where only selected cases are released to the public (Conceicao, 2006), only one case has been reported since 2009. The small number of cases that have been reported raises concerns of how effective are the regulatory frameworks and supporting ICT technologies in detecting and tackling cross-border market manipulations.

The various types of manipulation that occur in single-securities markets also exist in cross-border markets. That is, any manipulation that exists in single markets also exists in cross-border markets. In this thesis the following definitions are proposed: *Jurisdictions* are entities that have legal frameworks for monitoring and

regulatory authorities responsible for carrying out the monitoring; these are normally countries but may also be regions, such as the EU.

A *cross-border market manipulation*, which is in the general case a *cross-jurisdiction manipulation*, can be defined as a deliberate attempt to interfere with the fairness of market operations where:

i) a security traded in two or more different jurisdictions is a target of manipulation, either by manipulating one to alter the others; or by manipulating two or more at the same time to enhance the effect of the manipulation; for example, when a stock of the same company is listed in both a European and a US market.

ii) two or more, different, but co-related securities are traded in different jurisdictions and one is a target of manipulation with the intention to alter the other(s); or where two or more securities are manipulated at the same time to enhance the effect of the manipulation; for example, when a stock is manipulated with the intention of altering its American Depositary Receipt (ADR).

iii) one security is traded in only one jurisdiction but is the target of manipulative trades originating from manipulators based in other jurisdictions; for example, when manipulators in China and Brazil buy stocks of a German company listed in the DAX with the intention of decreasing its price.

Similarly, *Cross-market manipulations* refer to deliberate attempts to interfere with the fairness of market operations that involve any type of 'breach of fiduciary duty', 'runs and raids', 'abuse of market power' and 'contract based' manipulations where:

i) a security traded in two or more markets in the same jurisdiction is a target of manipulation, either by manipulating one to alter the others; or by manipulating the two or more at the same time to enhance the effect of the manipulation; for example, in the case of a stock trading in NYSE and NASDAQ at the same time.

ii) two or more, different, but co-related securities are traded in different markets in the same jurisdiction and one is a target of manipulation with the intention of altering the other(s); or where two or more securities are manipulated at the

same time to enhance the effect of the manipulation; for example, when a stock is manipulated with the intention of altering its derivative or vice versa; or where the stock is manipulated to affect the index that it is part of.

iii) one security is traded in one market and in a jurisdiction but is the target of manipulative trades originating from investors based in markets within the same jurisdiction; for example, when manipulators in France and Italy buy stocks of an British company listed in the FTSE, i.e. within the jurisdiction of the EEA, with the objective of pumping its price.

It is also possible to identify some specific cases, such as, *Cross-product manipulations* where co-related products traded in the same exchange or venue, and thus the same jurisdiction, are manipulated; for example, where different series of a stock (for instance, common and preferred series) are manipulated, or where fixed income instruments such as bonds and money markets are manipulated to alter the price of the company's stock, when all previous instruments are traded in the same market.

2.2.5 Regulatory Frameworks Related to Cross-border Trading

Cross-border market manipulations' differs across different regions. In the US, CFTC is involved in knowing how physical and futures market interact and monitor possible occurrence of price manipulation in one market that could affect the prices in another market. The CFTC is in collaboration with foreign regulatory to fight cross-border fraud and other practises considered as illegal that could have adverse effects on customers and serve as threats on the integrity of the market. This arrangement between the CFTC and the enforcement authorities is usually made through Memoranda of Understanding (MOUs) and other relevant arrangements. An example of memorandum which shows that CFTC supports information exchange is the Memorandum of Understanding concerning cooperation and the exchange of information related to the supervision of cross-border clearing organizations that was arranged by Alberta Securities Commission (ASC) Canada, which took effect on June 10 2010. The purpose of information

sharing is to help detect potential manipulative trading practices in the controls of trading between exchanges (Abioye, 2011).

On July 1 2005, the EU Market Abuse Directive (MAD) came into action. Part of the regulation aims at fighting cross-border market abuse by setting an approach, with common interest among EU member states. Under the MAD legislation, for instance, regulatory authorities shared jurisdiction with regard to abusive actions taken in one member that affects regulated markets or other members. However, although basic legislative framework are in place is still not clear how homogeneous the implementation and enforcement of MAD provisions are (Conceicao, 2006). Moreover, there are several issues in which advances are still required in order to respond effectively to market abuse and to achieve a common approach across all member states. According to (Mayhew and Anderson, 2007) *'if the intention behind MAD of a common law for market abuse for Europe is to be achieved, then challenge lies with its interpretation, application and enforcement'*. Table 2-2 presents an extract of the implementation of MAD, including what cross-border investigatory arrangements EU regulators have in place as compiled by (Mayhew and Anderson, 2007).

Table 2-2 Extract on the implementation of MAD across EU countries

	Belgium	France	Germany
Regulator	Commission Bancaire Financiè re et des Assurances/Commissie voor het Bank-, Financie en Assurantiewezen (CBFA)	Autorité des marché s financiers (AMF)	Bundesanstalt fü r Finanzdi-enstleistungsaufsicht (BaFin)
Principal legislative/regulatory source	Act of 2002 on the supervision of the financial sector and financial services, as amended in 2005 and 2006 (the 2002 Act)	General Regulation of the AMF—Book VI—Market abuse: insider dealing and market manipulation Articles L621-13 *et seq*. of the Monetary and Financial Code	Securities Trading Act (Wertpapierhand-elsgesetz–WpHG) 1994, as amended in 2004
Has MAD been fully implemented?	Yes. Main provisions in force as of September 19, 2005	Yes. Main provisions in force as of October 12, 2004	Yes. Main provisions in force as of October 30, 2004

What are the Regulator's investigatory powers?	Powers to obtain information from persons involved in suspicious transactions, including: • undertaking inspections of premises; • hearing witnesses; • the CBFA may also ask auditors to file reports on specific matters	Powers to obtain information from persons involved in suspicious transactions, including to: • compel attendance (but not answers); • compel production of documents; • search with warrant/court order. Obstruction of investigation is an offence	BaFin may request information from anyone, if there are indications of insider dealing or market abuse or the request is necessary for monitoring compliance. BaFin has the power to enter business premises and, subject to more restrictive conditions, residential premises
What were the Regulator's disciplinary powers prior to MAD?	Cessation orders. Admonition, suspension and exclusion from financial markets. Administrative fines on legal entities up to a maximum €1,250,000,	Cessation orders. Admonition, suspension and exclusion from financial markets. Administrative fines of up to €1,500,000 million or 10 times unlawful profits. Publicity	Cease and desist orders for any action in violation of the rules. Administrative fines on legal entities up to a maximum €1,500,000, including for non-compliance with the market abuse provisions
Does the Regulator have any additional disciplinary powers under MAD?	Penalties of up to €2,500,000. Power to appoint a Special Commissioner who must assent to the issuer taking certain actions. Publicity	Wider scope: • no need to prove impact of the offence on prices; • also covers attempted insider dealing	Temporary suspension of trading in financial instruments. Administrative fines, up to a maximum of €1,000,000 on legal and natural persons for a wider range of offences (involving wilful or, in some cases, negligent, breaches of the rules)
Role of the criminal law	The MAD prohibitions are both criminal and administrative offences. The CBFA may file a complaint with prosecutors, but has a discretion not to do so	The AMF must inform the Prosecutor of the Court of Appeal of Paris of any suspected offences. The AMF may file a complaint with prosecutors. The MAD prohibitions are both criminal and administrative offences. Criminal and administrative actions can be taken concurrently	Many of the MAD prohibitions are both criminal and administrative offences. Public prosecutors (not BaFin) prosecute insider dealing or market abuse. BaFin must report suspicion of a criminal offence to the prosecutors, who determine whether to prosecute

The Financial Services Authority (FSA) announced that it would be monitoring more closely, trading practices that have the potential to be abusive to the market. FSA designed the Market Abuse Directive (MAD) and according to them, the purpose of MAD is to 'improve cross-border cooperation, boost the confidence in the integrity of the integrated European markets'. The implementation of the MAD

is also to provide a proper framework for the flow of information through the markets (Mayhew and Anderson, 2007).

In the EU review of the Markets in Financial Instruments Directive (MiFID), equity markets and trading centres are obliged to publish a trade report in real time of their current orders, their prices and amounts, for the interest of sales and purchases. This is referred to as pre-trade transparency which usually applies to Multilateral Trading Facility (MTF) and regulated markets. Also every time a transaction in a share has been finalized, markets and trading centres are obliged to publish a trade report. This is referred to as post-trade transparency and it also applies to MTFs, regulated markets, investment firms and trading that is carried out in or out trading venues. Unlike pre-trade transparency, this trade report gives information about the history of executed transactions and not trading opportunities. The use of transaction reporting was stated in the EU consultation paper and was referred to as a major priority to deal with market manipulation and market abuse. It was also stated that all transactions, including equity and non equity markets, should be reported. Nevertheless, only few cross-border fraud cases have been prosecuted up to this date due to difficulty in collecting evidences given the lack of consolidated pre and post trade information.

2.2.6 Definition of market surveillance and monitoring

For the purposes of this thesis, market surveillance follows the definition presented in the work of (Polansky et al., 2004). Market surveillance is defined *'to encompass the processes and technologies that support the detection and investigation of potential trading rule violations, whether defined in statute or marketplace rules'*.

Market monitoring, therefore, can be understood as the sub-set of processes and technologies that support the detection and investigation of potential trading rule violations with special emphasis on marketplace rules, and the codes of conduct of exchanges and trading venues. According to the same authors, the components of market monitoring and surveillance can be broken down into 'detection', 'investigation' and 'enforcement' tasks, which are supported by different processes and technologies. The term enforcement *'means initiating a formal due process*

proceeding to charge a violation against a legal or natural person and, if found guilty, to impose some form of sanction' (Polansky et al., 2004).

2.2.7 Market monitoring and surveillance: allocation of responsibilities

(Polansky et al., 2004) identify three key criteria to be considered in allocating responsibilities as follows:

***Jurisdiction*.** As a general principle, surveillance responsibilities should be allocated in a manner that reflects the entities' jurisdictional reach. This considers two key aspects: i) jurisdiction to conduct an investigation of a particular person or firm, especially regarding jurisdiction to request and receive information and question individuals regarding a potential violation, and; ii) jurisdiction to carry out an enforcement action. Normally, market regulators have the authority to investigate any person or entity within their jurisdiction; by contrast, exchanges can only investigate their members. In addition, regulators have broad regulatory powers to obtain personal or sensitive information, including customer names, insider trading information and job role information that could facilitate the effective detection and investigation of certain types of manipulation. Disclosure of customer identities and positions or other sensitive information to exchanges should be limited to avoid personal data protection issues. With respect to enforcement, regulators also have a broader reach than exchanges, and they can sanction private investors, company insiders and their close circles, companies, and trading venue employees.

Deterrent effect. This refers to the potential impact that allocation can have on market manipulators. The rationale behind it considers that if an entity is perceived as having a superior understanding of the market complexity and the ways in which the market can be manipulated, then the knowledge that that entity is performing detection activities may have a greater deterrent effect than if another organisation performs the task. This is the rationale behind several configurations in which the exchange, and not the regulator, performs first-line detection activities, since the exchanges have the ability and the proximity to understand the way in which markets are manipulated. According to (Polansky et al., 2004), when the

issue is enforcement, however, a government regulator with broad sanctioning authority may be a stronger deterrent to aberrant behaviour than organisations with comparatively less stringent remedies available.

Organisational capacity and efficiency. The capabilities and efficiencies of an organisation are also a key factor. For instance, detection, surveillance, investigation and enforcement efforts can be greatly diminished if the organisation does not have the right number of suitably trained people or the right know-how, or it has poor access to cutting-edge technology. In practice, different jurisdictions have allocated market monitoring, surveillance, detection, investigation and enforcement responsibilities differently, dividing them according to their relevant legal frameworks and particular market characteristics. In general, most exchanges bear the responsibility of first-line detection for abuses that have features of trade-based manipulations. Given the broader and more complex access to sensitive information needed to detect information-based manipulations, the front-line responsibility for its detection usually remains with the regulator. Exchanges and venues are also responsible for the monitoring aspects, ensuring that the trading systems are functioning properly and are not used in a manipulative manner. This includes monitoring data integrity in terms of accuracy, completeness and sequence, as well as the entry of fictitious orders, volatility of prices, and 'fat-fingers' errors (human errors).

The following are some examples of the allocation of responsibilities in countries around the world:

United Kingdom. According to (Polansky et al., 2004) in the UK exchanges are given the responsibility to *'deter, detect, and monitor the incidence of market abuse'*. The actual investigation and enforcement lies with the Financial Services Authority (FSA or other agencies where appropriate). This means that exchanges make referrals to the FSA at a very early stage, having performed only basic analysis to confirm that a surveillance alert was not generated by error. The FSA have developed several memoranda of understanding with various exchanges, and in some cases the FSA and the exchange may investigate a case jointly. Is up to the exchanges to investigate breaches of their own rules of conduct that do not

constitute market manipulation (for instance, late reporting of trading activities) and to impose disciplinary actions when needed.

Singapore. Based on a speech made by the Deputy Managing Director of Monetary Authority of Singapore (MAS) at the Investment Fund Awards in 2001, the Monetary Authority of Singapore (MAS) enjoys *'broad regulatory powers over the Singapore Exchange (SGX). 'SGX has direct front-line responsibility for the securities and futures markets and for the broker-dealers who trade on the exchange'.* In fact, '*both MAS and SGX perform surveillance but SGX performs the primary surveillance role with MAS performing selective surveillance'. 'The exchange will take action against a broker who has breached exchange rules, but only after MAS informs the exchange that it may do so'. 'The MAS has the power to pursue civil prosecution of listed companies that fail to make timely disclosure of material information, and of any participants suspected of market misconduct'. 'In Singapore, the exchange has demutualised and the existing allocation of regulatory responsibilities is intended, among other things, to address the conflicts of interest that can arise in for-profit exchanges'* (Polansky et al., 2004).

United States. SEC enjoys broad regulatory jurisdiction, but delegates substantial responsibilities to exchanges which are then subject to SEC oversight. *'This arrangement reflects, in part, the historical role of exchanges as the front-line regulatory organisations'. 'The SEC's powers specify statutory authority to grant, deny or withdraw SRO status; approve all SRO rules and rule changes; impose new rules on SROs; discipline SROs for not complying with US securities laws or for failing to enforce SRO rules; hear appeals of SRO disciplinary actions and membership denials; inspect SROs and oversee the SROs' examination of their members with the authority to conduct examinations directly; and enforce SRO rules directly if the SRO is unable or unwilling to take appropriate actions*' (Polansky et al., 2004).

'The SEC has very limited internal equity market surveillance capabilities. Instead, it relies on each SRO to perform front-line surveillance and investigation. In those instances where the SRO has jurisdiction over the individual or firm alleged to have committed a violation, the SRO will typically take enforcement action. Nonetheless,

the SEC may institute parallel proceedings if it deems such action warranted. In cases where the SRO does not have jurisdiction over the firm or person, the SRO will refer cases to the SEC (or Justice Department) for further investigation and enforcement action. This is most likely with respect to insider trading cases' (Polansky et al., 2004).

In addition, as noted above, *'the SEC has broad oversight powers and will inspect SROs to determine the adequacy of their market surveillance capabilities. If the SEC identifies serious or systemic shortcomings in an SRO's detection, investigation or enforcement programme, the Commission will typically impose a sanction under which the SRO agrees to undertake a series of measures to improve its surveillance, sometimes including a commitment to spend a prescribed minimum amount of money on upgrading surveillance capabilities. Otherwise, the output of the SEC's inspections may be limited to recommendations as to improvements in the processes or technology dedicated to regulatory activities'* (Polansky et al., 2004).

2.2.8 Allocation of market monitoring and surveillance activities

Based on the principles outlined in the previous section, it is possible to identify the need for effective monitoring and surveillance that is ultimately the responsibility of a jurisdiction authority. However, given the increased interconnectedness of markets and venues, it is necessary to consider a broader definition of jurisdiction, in other words, the identification of cross-jurisdiction monitoring and surveillance responsibilities. This is clearly the case in economic regions such as the European Union, in which markets are interconnected in order to operate as a single market, but in which monitoring and surveillance responsibilities are shared among exchanges, country regulators, economic unit directorates and authorities.

Table 2-3 expands the allocation proposed by (Polansky et al., 2004) to describe the taxonomy and set of definitions used in this thesis, including a new dimension for allocation, i.e., the cross-jurisdiction regulating authority. The cross-jurisdiction regulator may enjoy broad surveillance, enforcement and investigative powers, but whenever possible, the front-line responsibility for detection should lay with the

exchange or with the local jurisdiction (country) authority. As a general rule, information-based, cross-product, or cross-market types of manipulation should be monitored by the local regulator, who has the necessary access to sensitive information. For other types of abuse, especially trade-based manipulations and violations of exchanges' rules and codes of conducts, the front-line responsibilities should lay with the exchanges.

Table 2-3 Allocation of Market Monitoring and Surveillance Detection Responsibilities

Market abuse	Cross-Jurisd. Reg.	Regulator	Exchange
Abnormal prices of related products at delivery locations	X	F	
Advancing [decreasing]the bid [ask]			F
Boiler room		X	F
Capping [pegging]			F
Churning			F
Commodity flows to delivery points (1)		X	F
Commodity flows to delivery points (2)		X	F
Contract prices at different expirations	X	X	F
Corner	X	F	
Dissemination	X	F	
Flash quoting		X	F
Front running client orders			F
Front running research	X	F	
Hype and dump		X	F
Insider trading	X	F	
Market timing			F
Marking the close			F
Marking the open			F
Matched orders	X	X	F
Mini manipulation			F
Option expiration date stock price or volume changes			F
Option introduction date stock price or volume changes			F
Parking or warehousing	X	X	F
Pre-arranged trade	X	X	F
Scalping		F	
Short sales			F
Slur and dump			F
Spoofing/painting the tape			F
Spot and futures prices at different delivery points		F	
Squeeze	X	F	
Stub quoting			F
Stuff quoting			F
Trade through			F

Trading away			F
Wash sale	X	X	F
Year end /as of trades			F

F. Indicates front-line monitoring and surveillance responsibilities. Shaded box indicates cross-jurisdiction responsibility was added by the author.

2.2.9 Monitoring and surveillance: internal organisation and process

In terms of organisation, an exchange may be self-regulated or non-self-regulated, in which case regulation is usually carried out by a government or regulatory body. (Pritchard, 2003) argues in favour of self regulation, because for-profit exchanges with poor market integrity will see their turnover diminished, and therefore have greater incentives to reduce market manipulations. (Pirrong, 1995) argues against self-regulation on the basis that exchanges face a conflict of interest in acting as both market operator and regulator; as market operator, the exchange generates its income from turnover, while as regulator it must potentially prohibit trade that would have otherwise increased turnover. As explained in the previous section, a self-regulated exchange, or SRO, could perform most front-line detection, monitoring and surveillance efforts, but should still be subject to joint (or independent) supervision from both local jurisdiction and cross-jurisdiction authorities. This is so as to balance the trade-offs between the pros and cons of self-regulation and non-self-regulation.

In order to present an example of the way in which surveillance processes are organised inside an SRO, the following provides a brief explanation of the Australian regulatory framework and describes the Australian Stock Exchange (ASX) supervisory arrangements, as presented in the work of (Comerton-Forde and Rydge, 2006). This work is particularly valuable since most of these processes and organisational procedures are usually guarded with special confidentiality and secrecy by the authorities. There is a concern that providing this type of information to the public may encourage or help manipulators by revealing the potential weaknesses of the systems to them. In fact, one of the most difficult tasks undertaken by the author was to obtain information regarding these issues.

According to (Comerton-Forde and Rydge, 2006), in Australia 'the Corporations Act defines the organisation and responsibilities of surveillance. The ASX and the Australian Securities and Investment Commission (ASIC) are the two bodies primarily responsible for the regulation of equity markets. The Corporations Act, Section 792A, requires the ASX to ensure a fair, orderly and transparent market. The Act states that the ASX must establish and maintain appropriate surveillance processes, including conflict management, monitor the conduct of its participants and enforce compliance with its market rules, codes of conduct and listing rules. The Act also requires that the ASX produces an annual account and report. In 2006 the ASX consisted of several operating units, including Surveillance, Investigation and Enforcement (I&E), Compliance Services, Risk Management, and ASX Companies (ASXC). It also contained a Disciplinary Tribunal (DT) formed of market participants. The ASX imposes several operating rules on its market participants and listed entities. Operating rules are denominated Market Rules and rules regulating listed companies are denominated Listing Rules. Figure 2-2 presents a flow diagram of the surveillance process.

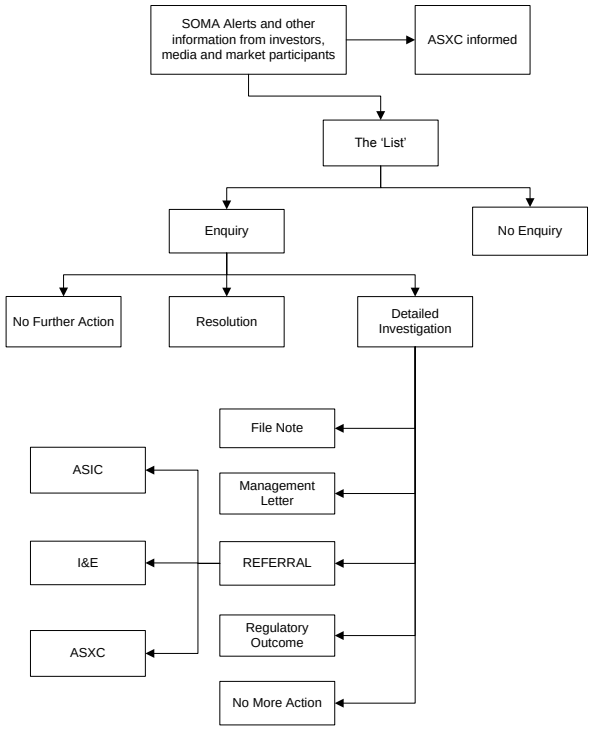


Figure 2-2 Surveillance Process ASX-SOMA (Comerton-Forde and Rydge, 2006)

The Surveillance unit operates 'Surveillance of Market Activity (SOMA)', an information system that provides data mining and pattern recognition to aid in the detection of unusual trading activity. SOMA monitors the appearance of outliers in price, volume and other manipulative patterns. It does this in real time and generates alerts that start an investigative process. SOMA uses not only trading data, but also information obtained from investors, market participants and the media as inputs. The alerts are further examined by a group of analysts; if the observed pattern is not explained by market activity, ASXC may be informed. In addition, some flagrant behaviours are added to the 'List'. Leads on the 'List' are further studied and allocated an enquiry number if more detailed analysis is needed. Matters on the list can have any of three possible outcomes: first, No Further Action (NFA) if it is clear that no market rule was broken; second, if Surveillance is able to resolve the matter with the party concerned, no investigation will commence; and third, a detailed investigation could be initiated (Comerton-Forde and Rydge, 2006).

Investigation results in one of five potential outcomes. As a first outcome, the investigator may send a File Note (FN) that stores the unusual trading activity. Surveillance will maintain monitoring of FNs and may take further action at a later stage if a flagrant breach of rules becomes evident. Secondly, it may send a Management Letter (ML), which is a warning letter written by the Surveillance team. An ML issues a preliminary warning to the involved parties, but does not continue with the enforcement procedure. Again, if the activity continues, Surveillance will take a stronger action. Thirdly, a Regulatory Outcome may be reached, for instance, when the trading activity that led to the alert ceases, trades are cancelled or a full disclosure is made. This is also known as a 'soft outcome'. As a fourth outcome, the investigator can also decide to take No More Action, if after further scrutiny the analysis shows that the matter does not deserve additional examination. Finally, if Surveillance is unable to resolve the matter, a Referral is written and sent to ASIC, I&E and/or ASXC depending on the regulation allegedly breached (Comerton-Forde and Rydge, 2006).

Referrals to ASIC are investigated according to the powers enjoyed by the regulator, which are broader and more complete both in terms of investigative power and enforcement of disciplinary actions. In particular, all suspected insider trading activities are handled by the ASIC, which is solely responsible for the enforcement of insider trading prohibitions. If the regulator finds wrong-doing it can take several actions against ASX staff, market participants, retail customers, investment brokers and internet brokers among others. Referrals to I&E are further investigated, and if the matter warrants enforcement action this is referred to the DT. The DT can impose several penalties; these include suspension, censure, completion of an education programme, fines of up to 250,000 Australian dollars and expulsion from the market (Comerton-Forde and Rydge, 2006). Table 2-4 shows the descriptive statistics on the ASX surveillance process, summarising the number and possible outcomes for cases in the 'List'.

Table 2-4 Descriptive Statistics of the ASX Surveillance Process. Adapted from (Comerton-Forde and Rydge, 2006)

Number	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
The List	288	347	286	317	194	414	444	443	510	495
	100.0%	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*
Enquiry	126	131	145	126	112	123	112	172	225	208
	43.8%	*37.8%*	*50.7%*	*39.7%*	*57.7%*	*29.7%*	*25.2%*	*38.8%*	*44.1%*	*42.0%*
Investigations - Total	41	50	65	64	58	69	94	154	200	168
	14.2%	*14.4%*	*22.7%*	*20.2%*	*29.9%*	*16.7%*	*21.2%*	*34.8%*	*39.2%*	*33.9%*
Soft outcome achieved	-	-	-	-	-	-	-	49	61	56
	n/a	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*11.1%*	*12.0%*	*11.3%*
Referrals - Total	49	53	40	50	59	45	48	51	45	53
	17.0%	*15.3%*	*14.0%*	*15.8%*	*30.4%*	*10.9%*	*10.8%*	*11.5%*	*8.8%*	*10.7%*
Referrals - Manipulation	20	25	19	26	40	35	26	29	35	23
	6.9%	*7.2%*	*6.6%*	*8.2%*	*20.6%*	*8.5%*	*5.9%*	*6.5%*	*6.9%*	*4.6%*

Shaded box indicates proportion was added by author.

2.2.10 Monitoring and surveillance: scope, performance and cross-market capabilities

In terms of the scope of the detection and analysis performed by regulators in comparison with SROs, (Cumming and Johan, 2008) reports on evidence from exchanges and securities commissions from twenty-five jurisdictions in North,

Central and South America, Western and Eastern Europe, Africa and Asia. In general, SROs are capable of performing surveillance of a wide range of market manipulations but, contrary to expectation, the scope of cross-market and cross-product surveillance is very low. (Cumming and Johan, 2008) also report that market activity is facilitated by surveillance efforts, and that cross-market surveillance is correlated positively with turnover velocity and market capitalization of SROs. They find that cross-market surveillance is much more effective when different jurisdictions have information-sharing arrangements, and when such arrangements are broader in scope. Typically, regulators are more keen to subscribe to such arrangements than SROs, hence they argue in favour of a prominent function for regulators in assisting cross-market surveillance.

In particular, (Cumming and Johan, 2008) indicate that the most common types of manipulation subject to single-market monitoring include wash trades, matched orders, spoofing/painting the tape, pumping and dumping, and marking the close. Capping and pegging, insider trading and dissemination are the most common type of manipulation monitored with a cross-market scope. On average, an SRO carries out surveillance on fourteen or fifteen types of manipulation on a single-market basis, and only two or three cross-market types of manipulation. Interestingly, SROs rank their surveillance efforts very poorly, with an average response of 2.15 out of 5 for effectiveness for the different types of market manipulations. The proportion of actual manipulations detected in relation to the number of trades (excluding false positives) is in the order of 1.29% for SROs and 0.61% for regulators (2005 figures). This compares closely with the figures reported in (Putnins, 2009), which states that approximately 1.1% of closing prices are manipulated. However, (Putnins, 2009) also emphasises that for every prosecuted closing price manipulation there are approximately 300 instances of manipulation that remain undetected or not prosecuted.

Using a dataset representing 34 security markets worldwide, (Aitken et al., 2010) report on whether a reduction in trade-based manipulation (higher market integrity) is associated with tighter spreads (higher market efficiency). They find that reducing the incidence of manipulations by half reduces the effective spreads 31 to

59 basis points in the middle liquidity deciles worldwide. These authors also find that particular market technologies, in combination with real-time surveillance and market integrity regulations, reduce ramping alerts. Circuit breakers, closing auctions in combination with real-time surveillance, and more security regulations lower spreads. Moreover, regulations prohibiting short selling markedly increase spreads across all liquidity deciles. In addition, real-time surveillance is more likely to be adopted the lower the volatility, the smaller the foreign direct investment, the faster the execution speeds, and the greater the risk of market manipulations. Circuit breakers, regulations specifically prohibiting ramping and the codification of violations in civil law-based securities all raise the probability of the adoption of real-time surveillance (Aitken et al., 2010).

2.3 Market monitoring and surveillance: Technologies and Systems

2.3.1 Fraud and Data Mining

The amount of data collection on financial markets has been increasing exponentially during the last ten years, and it has reached unprecedented levels of data generation, scarcely comparable with any other fields. The huge amount of trades that one single stock can have in a day makes intraday databases very hard to manipulate. For quotes or orders of trade data, the situation is even worse, given that for each trade there are at least two or more interactions quoting for bid and asks prices.

Finding an appropriate way of analysing and managing this huge amount of data is an important part of trading and surveillance technologies. In this sense, the Knowledge Discovery on Databases (KDD) process appears to be the natural solution, being a *'nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data'* (Fayyad and Piatetsky-Shapiro, 1996).

The KDD or data mining operation can be seen as a three-stage process: the *data preparation stage* deals with improving data quality and summarising the data to facilitate the analysis and discovery process; the *model derivation* stage focuses on choosing learning samples, testing samples and learning algorithms; while the *usage and maintenance phase* is concerned with the monitoring of databases updates and continued *validation* of patterns learned in the past (Kopanakis and Theodoulidis, 2003).

The data mining process is able to use a wide range of analytical methods or algorithms to extract data for the purpose of knowledge discovery, module generation and result prediction. Among these the main methods are classification analysis, association analysis and cluster analysis, which are used widely in different applications (Xia, 2007).

Classification analysis is based on learning functions that map (classify) a data item onto one of several predefined classes. Typical applications of classification

algorithms include credit approval, direct marketing, fraud detection, and technical analysis (Witten and Frank, 2005).

Association is the analysis that helps to find certain relationships among data. It is frequently used in the areas of gene analysis, disease detection and web mining to identify the association rules between the objectives. An association rule is the main part of association analysis, which discovers elements that co-occur frequently and the rules, such as implication and correlation, related to those occurring elements (Xia, 2007).

Clustering analysis compares and groups data by certain criteria. To some extent, it can be considered as a branch of classification analysis, since the clusters are normally formed based on the resemblance between entities. However, clustering analysis is unsupervised. That is to say, there is no classification model generated from the training set before grouping the target data (Xia, 2007).

There has been a considerable amount of work on the application and research of data mining for fraud detection. (Xia, 2007) carried out a literature review in which he identifies research undertaken in recent years, which has focused mainly on money laundering ((Jensen, 1997), (Zhongfei et al., 2003), (Salerno et al., 2005)), credit card fraud ((Yufeng et al., 2004) and (Leung et al., 2004)), financial crime ((Neville et al., 2003, Neville et al., 2005)) and cellular phone fraud ((Fawcett and Provost, 1997, Taniguchi and Haft, 1998)).

(Phua et al., 2005) presented a comprehensive survey of data mining techniques used by fraud detection systems including neural networks, Bayesian networks, statistical modelling and decision trees, and critically analysed their appropriateness as fraud detection methodologies.

(Edge, 2007) investigated the concept of *proactive* fraud management within financial environments to develop a real-time fraud detection model capable of identifying fraudulent account behaviour prior to transaction completion. He also proposed a financial fraud modelling language framework for conceptual-level definition, which offers significant advances in the design and deployment of proactive fraud controls within multi-channel service infrastructure.

In one of (Edge, 2007) main arguments he explains how in *reactive* fraud management, analytical methods such as data mining are implemented on stored transactional data. Fraudulent activities, then, are identified against pre-defined fraud patterns, enabling extraction of previously unknown data sequences for future detection (Edge et al., 2007). *Proactive* fraud management, by contrast, evaluates incoming transactional data prior to being placed within disc-based storage, extracting suspicious activities using predefined fraud libraries or as anomalous account undertakings based on maintained transactional history. Fraudulent transactions may, therefore, be declined and preventive actions triggered before transaction completion and the movement of any associated monetary value (Edge et al., 2007).

Following the work carried out by (Phua et al., 2005), (Edge et al., 2007) also categorised current data mining approaches as *supervised, semi-supervised* and *unsupervised.* The first two categories rely on training data from which the employed solution may construct evaluative models against which newly arriving data instances may be assessed. Thus, since supervised and semi-unsupervised approaches utilise off-line data and are consequently retroactive, unsupervised approaches are a more desirable solution for proactive fraud management.

2.3.2 Existing systems and frameworks

The typical functionalities and processes of a market manipulation detection system can be found in the National Association of Securities Dealers (NASD) Regulation's Advanced Detection System (ADS), discussed in (Kirkland and Senator, 1999) and revised in the works of (Senator, 2000), (Goldberg et al., 2003), and (Donoho, 2004). More recent works describe the deployment of analogous systems in emerging markets, as well as the study of novel data mining algorithms within the behavioural analysis context (Mongkolnavin and Tirapat, 2009, Ogut et al., 2009). Although these monitoring systems were deployed in different countries, in different years, and with different regulatory concerns, there are several very important commonalties, features, and differences that can serve

as a foundation for a common understanding of their fundamental parts and for the development of a market monitoring and surveillance systems framework.

The work of (Mongkolnavin and Tirapat, 2009) introduced a monitoring system used on the Thai Bond Market, which was commissioned by the Thai Bond Market Association (ThaiBMA). In this system, the trading analysis component is pictured as the central part of the system, and two approaches to the trading analysis are further expanded. The first approach, namely Economic Analysis, is based on financial econometrics and compares current trading prices with the historical time series properties of the bond prices using GARCH models (Bollerslev, 1986). The objective of this system is to detect trading divergences from the mean, using past trading, thus raising alarms when unusual trading outliers are detected. However, in the presence of scarcely traded bonds this approach is not effective enough, as it relies on the historic behaviour of prices only. In the second, Behavioural Analysis approach, an association rules data mining application is integrated with the economic analysis. This application focuses on the analysis of trader behaviour, and warning signals or alarms are issued in conjunction with the alarms raised by the first application based on unusual trader conduct (Figure 2-3).

This process flow uses input data in the form of prices, volume and trader information, as well as a collection of market integrity rules that are implicitly present in the system. The output of the system (warning signals or alarms) is then passed to an information management unit, which continues the monitoring workflow, deciding whether the flagged trades warrant further investigation. If appropriate, an investigation process is started, followed by the corresponding report and enforcement (if applicable) of market rules (Figure 2-4).

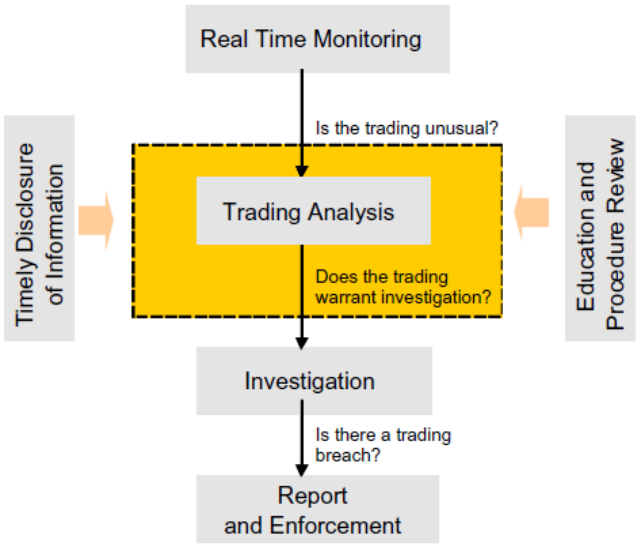


Figure 2-3 Surveillance Workflow of the ThaiBMA (Mongkolnavin and Tirapat, 2009)

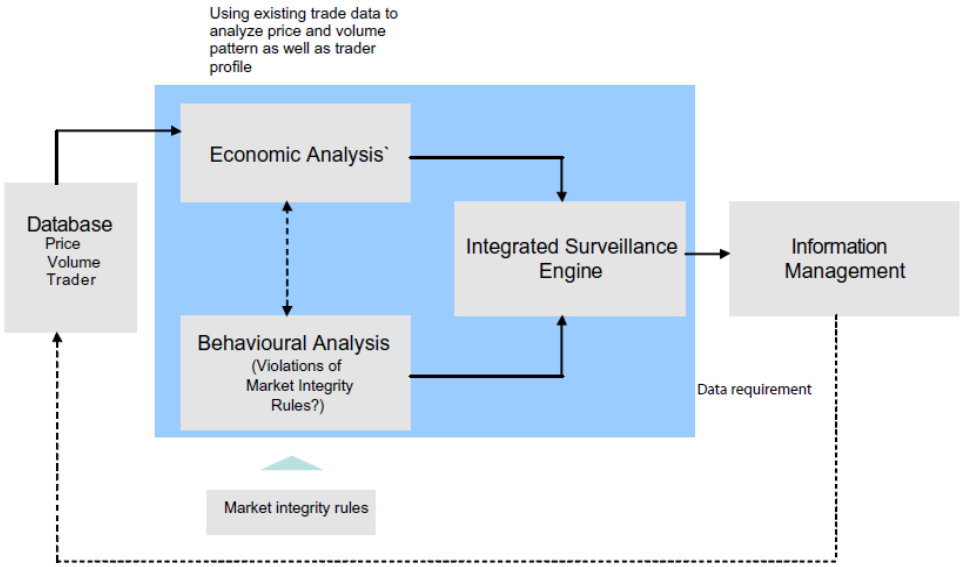


Figure 2-4 The Combination of Economic and Behavioural Analysis Approaches to Trading Analysis (Mongkolnavin and Tirapat, 2009)

Other examples include the detection of cyclical price manipulation (Westphal and Blaxton, 1998) and the monitoring of trading activity on the NASDAQ stock market using the NASD Regulation's Advanced Detection System (ADS) (Kirkland and Senator, 1999). Further extensions to the ADS investigated the detection of insider trading manipulations which combined the derivative and the securities market

(Donoho, 2004) and additional extensions investigated its deployment as a complete data mining system, including components for data preparation such as loading, cleansing, transforming, summarising, and integrating data; data mining, text mining and pattern-matching; and finally, components for visualisation of the analysis and its results (Goldberg et al., 2003), (Senator, 2000).

(Goldberg et al., 2003) work presents one of the most complete conceptual architectures for the Securities Observation, News Analysis, and Regulation (SONAR) system (Figure 2-5). In its conceptual design, it is possible to identify two types of Effects or Economic Analysis, data mining and text mining analysis, which are also organised into historic and real-time monitoring based upon the use of different market data and textual sources such as financial news, SEC corporate filings, spam e-mails, websites, chat rooms, and many of NASD's internal documents such as SEC referrals, complaint data, and disciplinary history data.

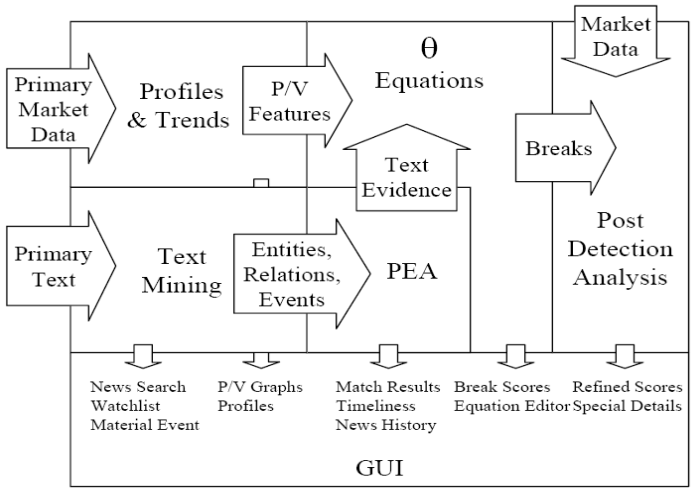


Figure 2-5 Conceptual Architecture of the SONAR System (Goldberg, Kirkland et al. 2003)

According to (Aitken et al., 2010), the latest development in some exchanges, e.g. the London Stock Exchange and DirectEdge, is an effort to launch Enhanced Liquidity Provider Programmes (ELPs) which offer an integrated view of both displayed and dark pool order books. The author also reports on the gradual adoption of real-time market surveillance, mentioning the SMARTS surveillance platform from Smarts Group International Ltd. a NASDAQ OMX company

http://www.smartsgroup.com, as the most complete and advanced technology available. SMART systems have been deployed in more than 50 national securities, exchanges and regulators around the world. Another recent technology is Direct Market Access (DMA), defined as electronic facilities which allow brokers to offer clients direct access to an exchange trading system through the broker's infrastructure without manual intervention by the broker.

Other companies, such as Business Systems http://www.businesssystemsuk.co.uk offer commercial applications for market surveillance. For instance, the *Market Detect* solution is presented as *'A cost effective, real time data analytics application which can be fully customised to meet the requirements and compliance needs of the most demanding financial institutions. Designed as a powerful tool to help eliminate market abuse and fraud, Market Detect intelligently collects and analyses organisational data including agent, trader and client activity data to uncover and present hidden patterns.*

Market Detect functions by proactively taking a direct feed of the trading data and telecommunications data. The data is then passed through Market Detect's state-of-the-art detection engine which monitors trading activity. The user determines the behaviours they wish to be notified of and the system will trigger an 'Alert' when the pattern is identified'. However, the company does not provide either scientific evidence or a formal description of their application's performance and capabilities.

2.3.3 Characteristics and Requirements

According to (Cumming and Johan, 2008), effective surveillance and monitoring systems, which utilizes a broad range of knowledge discovery and data mining techniques, should:

- minimise false positive and maximise true positive manipulative alerts. To do so, the system should be able to recognise normal trading activity and differentiate it from abnormal alert parameters;
- enable a surveillance department or equivalent unit to reconstruct all trading activity to replay the full order/quote schedule;

- be able to recognise all trading activity from each market participant;
- involve surveillance staff who are knowledgeable in the issues that need to be investigated;
- keep market participants informed about the surveillance activities, in order for the system to be more effective;
- take into account the fact that a crucial point in the effectiveness of cross-market surveillance is the extent to which information is shared across jurisdictions
- finally, acknowledge that its effectiveness will always ultimately depend on the regulatory framework.

(Polansky et al., 2004) also recommend that solid market systems should have:

- a mature infrastructure connecting the major entities in the market, i.e., exchange, central depository of securities, brokers, issuers and regulators;
- a solid exchange platform;
- a well-developed depository system and extranet capabilities, at both the exchanges and the depository, to enable communication among parties; and
- a strong set of transactional data and well-developed archival data collection systems.

In order to implement such a system the following steps are needed. First, it is necessary to determine which off-line alerts or manipulative scenarios are to be detected by the exchange. The second requirement is to collect the data necessary for surveillance in a timely manner; this includes order and trade data, as well as audit trails. It may also be necessary to acquire additional data to support detection, e.g. material news and events, or beneficial ownership data. Because this data will come from different sources, incentives and accountability must be created for vendors to provide this in a timely manner and fashion. Third, a data pre-process should be implemented in order to create any derived data necessary

to support the detection. Ideally, all derivation should occur on demand, to maintain maximum flexibility, but realistically some variables would be better computed only once. The fourth requirement is that a query-based component should be available for the creation and visualisation of data relevant to a particular model or manipulation. The objective of this component is to enable the analyst to quickly identify abuse related to a specific breach of regulations and rules. Reports should be parameterized by key variables so that the analyst can control the precision of the report. These reports will support two uses, firstly to scan for abuses, and secondly to build a knowledge base that could help the automatisation of the detection based on previous validated cases of manipulation. Fifth, a component should be built to generate alerts, as an extension to the reports generated in the previous component. The purpose of the alerts is to reduce the amount of material that the analyst has to scan manually. Sixth, online systems should be created to automate the collection and storage of non-transactional data, such as licensing data, issuer financial listings, disclosures and complaints. This system should be accessible and usable by all filers, issuers, brokers, or others. Such a system should be able to formally document data and analyse data elements, frequency of collection, filer data, and relationships among this data. A seventh requirement is to capture, store and catalogue material news by identifying types of market news, capturing this information electronically and cataloguing it to indicate the presence of specific material events that are fed to the detection systems. Eighth, knowledge engineering team(s) that operate in co-ordination with analysts and IT developers should be established. Table 2-5 presented in the next section summarises these and further requirements identified in the literature.

2.4 Gaps in the literature and requirements for a conceptual MMSS

In this sub-section the author's intention is to discuss and comment on the limitations and gaps of the literature in order to come up and identify a set of requirements needed for a modern market monitoring and surveillance system. The author’s interpretation is that in order to monitor the financial markets their stakeholders, such as financial authorities and portfolio managers, have developed

a number of market monitoring and surveillance systems with different characteristics including the human-machine interaction aspects and response-time requirements. Market monitoring approaches typically include special task teams, which manually monitor the transactions and profile of traders, and also a wide range of more or less sophisticated, semi- or fully-automated information systems with various analytical capabilities. Moreover, the heterogeneity of such systems covers a wide range of usage scenarios, which in turn involve numerous processes and require a similar diversity of information systems and system architectural components necessary for the market monitoring tasks. These approaches typically focus on certain markets and consider certain market manipulation scenarios, so it can be argued that they adopt a piecemeal approach rather than adopting a more integrated and holistic perspective.

The main components of market monitoring and surveillance systems that are mentioned in the literature are: inputs in the form of different types and sources of data; monitoring engines with different degrees of integrated analysis applications; and some form of output management, visualisation and evaluation of results. Nonetheless, the existing set-up can be considered limited both in terms of the techniques that are used and in terms of the description of the data processing and data structures that are needed for the system to work. Moreover, existing configurations do not explicitly consider market-to-market interactions that could result in potential inter-market or cross-border manipulation scenarios.

Further to these limitations there are a number of other current and potential processes that are required in order to perform behavioural analysis which are seldom mentioned in the literature. It is also possible to consider the potential effectiveness of combined real-time text mining analysis of incoming news and financial events, or even more sophisticated scenarios using multimedia data with appropriate analysis functionalities, such as speech recognition of telephone conversations, email and spam analysis or other digital content, as part of the behavioural analysis of the traders.

The input for such a conceptual system will also require a wider variety of types and frequencies of data, coming from multiple sources and in different formats, and

additional corresponding processes for data preparation and information management. This could include intraday data, such as trades and quotes of the securities being observed and its derivative, closing prices and daily volume of trade, company profile with financial states, and employee information, including lists of key personnel with access to privileged information, trader profiles, and trader past trading behaviour. Inputs could be found in the form of abundant textual sources, such as financial news, internet forums, blogs, and financial events in the form of filings to the relevant authorities, or even coming from the instant message facilities that traders' platforms incorporate.

These conceptual systems will also need to manage different types of market integrity rules, or more generally speaking, customer rules, given the fact that a system could not only be used by different types of regulating authorities, but also by different types of market agents, such as private investment funds looking for arbitrage opportunities, internet trading platforms interested in monitoring the behaviour of their clients, market operators or market agents in general.

Other key components of the inputs to this conceptual system will be previously used patterns and trading analysis models, and previous breaks and cases of confirmed manipulations. In this sense, the information management flows need to be dynamic and recursive; they should incorporate output management components that are capable not only of raising alarms and warning signs, but also of creating and validating new patterns (adaptive learning), and storing and updating old patterns and cases with new instances.

Finally, it is necessary to consider the interaction of different markets and venues and to distinguish the need for conceptual systems that integrate the analysis of historical data (reactive systems) with the analysis of real-time streams (proactive systems). Several challenges arise when considering the analysis of real-time data streams, especially concerning its integration with historical analysis components, and concerning the design of an adequate system architectural apparatus to support storage, analysis and communication capabilities between them.

Table 2-5 presents a collection of requirements for a conceptual MMSS, as well as its classification as either a functional or a non-functional requirement, and a

broader class. Functional requirements define what a system should do, whereas non-functional requirements define how a system should be. Broader classes are useful to distinguish between different types of analysis and functionalities that the systems may have. When necessary, the original wording of the requirement has been changed or broken down into smaller requirements in order to transform non-functional requirements into functional requirements. This was done to improve the precision and representation of concrete characteristics and tasks that MMSS should do in order to improve the integrity and efficiency of markets.

Chapter 3 begins to address these issues by describing the methodology used to design an information system artefact capable of solving the problems indentified above, which is presented in extension in Chapter 5. Chapter 4 introduces a group of taxonomies and definitions that help organise the concepts that are necessary to understand and address these requirements. In Chapter 6 different case studies are introduced. For each case a list of addressed requirements is provided, explicitly stating the capabilities of the prototype detection engines that were built based on the set of requirements.

In summary, this chapter has presented a review of market manipulations and market surveillance and monitoring systems. It began by presenting a brief overview of how the problem of market manipulation has been historically defined and investigated by the financial and economic community, and has highlighted that manipulations do indeed exist, and that they can be very profitable for manipulators but detrimental to other market participants. The chapter continued by providing a taxonomy of, and standard definitions for, different types and categories of market manipulations, and then moved to the definition and description of the market monitoring and surveillance systems, components and processes, and the allocation of responsibility for monitoring and surveillance. The chapter also reviewed the technologies that facilitate the tasks performed by the MMSS and concluded enumerating requirements, limitations and gaps in the existing MMSS.

Table 2-5 Summary of Requirements of MMSS

Number	Requirement	Class	Type
1	Minimise false positive and maximise true positive manipulative alerts	Analysis	Functional
2	Recognise normal trading activity and differentiate it from abnormal alert parameters	Analysis	Functional
3	Include a complete and clear list of off-line alerts or manipulative scenarios to be detected	Analysis	Functional
4	Perform economic and behavioural analysis*	Analysis	Functional
5	Manage and support monitoring for different types of market integrity rules, or more generally speaking, customer rules*	Analysis	Functional
6	Support real or near real-time text mining analysis of incoming news and financial events*	Analysis	Functional
7	Support real or near real-time multimedia analysis, such as speech recognition of telephone conversations, email and spam analysis or other digital content, as part of the behavioural analysis of the traders*	Analysis	Functional
8	Support integrated analysis applications, for example behavioural and economic or surveillance and monitoring*	Analysis/ Integration	Functional
9	Support a wide variety of types and frequencies of data, coming from multiple sources and in different formats*	Data capture and pre-processing	Functional
10	Collect the data necessary for surveillance, including order and trade data, as well as audit trails, in a timely manner. It may also be necessary to acquire additional data to support detection, e.g. material news and events, or beneficial ownership data	Data capture and pre-processing	Functional
11	Capture, store and catalogue material news, identifying types of market news, capturing this information electronically and cataloguing this information to indicate the presence of specific material events that are fed to the detection systems	Data capture and pre-processing	Functional
12	Data pre-process should be implemented, in order to create any derived data necessary to support the detection	Data capture and pre-processing	Functional

13	All derivation of data should occur on demand, to maintain maximum flexibility.	Data capture and pre-processing	Functional
14	Automatic collection and storage of non-transactional data, such as licensing data, issuer financial listings, disclosures, and complaints	Data capture and pre-processing	Functional
15	Non-transactional data storage system should be able to formally document data, analysing data elements, frequency of collection, filer data, and relationships among this data	Data capture and pre-processing	Functional
16	Reconstruct all trading activity to replay the full order/quote schedule	Information Management	Functional
17	Be able to recognise all trading activity from each market participant	Information Management	Functional
18	Have sufficient information-sharing capabilities across jurisdictions to support cross-market surveillance*	Information Management	Functional
19	Support market-to-market, cross-jurisdiction, or cross-border surveillance and monitoring*	Information Management	Functional
20	A query-based component should be available for the creation and visualisation of data relevant to a particular model or manipulation. The objective of this component is to enable the analyst to quickly identify abuse related to specific breaches of regulations and rules	Information Management	Functional
21	Manage and support previously used patterns and trading analysis models, and previous breaks and cases of confirmed manipulations i.e. previous models manage and support	Information Management	Functional
22	Integrate the analysis of historical data (reactive systems) with the analysis of real-time streams (proactive systems)*	Information Management	Functional
23	Give functionality for output management, visualisation and evaluation of results	Information Management/Output Management	Functional

24	Reports should be parameterized by key variables so that the analyst can control the precision of the report. These reports will support two uses, firstly to scan for abuses, and secondly to build a knowledge base that could help the automatisation of the detection based on previous validated cases of manipulation	Information Management/Output management	Functional
25	A component should be built to generate alerts, as an extension to the reports generated in the previous component. The purpose of the alerts is to reduce the amount of material that the analyst has to scan manually	Information Management/Output management	Functional
26	Grant access to non-transactional data storage system for all filers, issuers, brokers and other customers	General conditions of the system and environment	Functional
27	In order for the system to be more effective, market participants should be informed about the surveillance activities	General conditions of the system and environment	Non Functional
28	Effectiveness will always ultimately depend on the regulatory framework	General conditions of the system and environment	Non Functional
29	Infrastructure connecting the major entities in the market, including: a solid exchange platform; a well-developed depository system; extranet capabilities; a strong set of transactional data; and archival data collection systems	General conditions of the system and environment	Non Functional
30	Incentives and accountability must be created for the data vendors to provide their services in a timely manner and fashion	General conditions of the system and environment	Non Functional
31	Surveillance staff should be knowledgeable in the issues that need to be investigated	Skills and capabilities of the customer	Non Functional
32	Knowledge engineering team(s), that operate in co-ordination with analysts and IT developers	Skills and capabilities of the customer	Non Functional

Shaded box indicates category was added by the author. * Indicates requirement was added by the author.

3 Research design and methodology

This chapter discusses the methodology that was used to design an innovative solution that addresses the gaps identified in the understanding and functionality of market monitoring and surveillance systems, given the context and changes that are taking place in financial markets. The chapter starts by providing a general definition of the research approach that was chosen, followed by a justification of why this methodology is appropriate for the task at hand. It continues with an in-depth description of how the methodology was applied.

3.1 Description of the methodology

This work uses the Design Science in Information Systems (IS) Research methodology. According to (Hevner et al., 2004) design science creates and evaluates information and technology (IT) *artefacts* that address organisational problems. The design science is part of, and feeds from, behavioural science, i.e. they are intrinsically interrelated. Design science addresses the problems through a *cycle of activities* that generate an *innovative solution*. Behavioural science addresses research through the *development* and *justification* of theories that explain or predict phenomena related to the problem and problem space.

For (Hevner et al., 2004), behavioural science is predominantly concerned with understanding and predicting dimensions, whereas design science is more concerned with utility and problem-solving dimensions i.e., how the artefact works. However, according to (Vaishnavi and Kuechler, 2004), this apparent dichotomy may well be associated with certain communities of IS research and not be present in others.

This thesis assumes that there is symbiotic relationship between design science and behavioural research rather than a dichotomy. In this symbiotic relationship, **relevance** is achieved by developing artefacts and rhetoric/theories that address important and relevant problems. **Rigour** is achieved by correctly applying the knowledge base throughout the research process.

The phases of this methodology can be classified in two major categories: *develop theory / build artefact* and *justify predictions / evaluate utility.* However, these are two extremes of an iterative and continuous reasoning process. (Takeda et al., 1990) describe the reasoning activities in design research as illustrated in Figure 3-1.

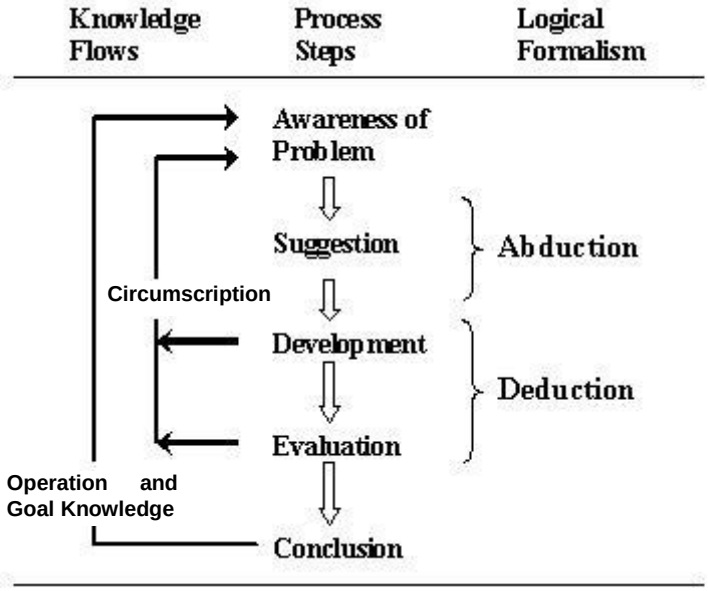


Figure 3-1 Reasoning in the Design Cycle. Taken from (Takeda et al., 1990) as in (Vaishnavi and Kuechler, 2004)

Reasoning starts with an *awareness of a problem*; unfolding situations in which there is a need to improve performance, address a deficiency or solve a problem. In the next step, a *suggestion* for a solution is put forward from the knowledge base, abducting and abstracting theories to accommodate them in the problem space. In turn, a first trial or attempt is carried out, implementing a solution with the *development* of an artefact. Later, trial artefacts are subject to *evaluation* according to their utility in solving the problem. The suggestion-development-evaluation stages jointly form a sub-cycle which is normally iterated until the artefact reaches a 'good enough' performance level. *Conclusion* indicates an arbitrary cutting point in which this iterative process is terminated.

Knowledge is produced continuously during the process of design, although it is said to be produced in two types of feedback loops. On the one hand, the

circumscription loop generates *understanding* that can only be gained from constructing the artefact. It is possible to exemplify this with an analogy of the development of the first telescopes. At the beginning, nobody fully understood why this artefact could make faraway things look closer, but during its construction it was possible to test and adapt parts, lenses, mirrors and materials that were thought competent until they were put to together with other parts of the artefact. Thus, it was possible to detect contradictions and things that did not work according to theory, complementing or rejecting existing knowledge. On the other hand, the *operation and goal knowledge* loop generates insights associated with understanding that is gained through the use and operation of the artefact. Initially, the fundamental laws that make telescopes work were not clear, but they indeed represented an incomplete but material solution to the need for distant exploration. It was only after years of using and perfecting the artefact that humans finally understood the laws of optics and the reasons why the artefact worked.

3.1.1 Steps and outputs of design research methodology

(Vaishnavi and Kuechler, 2004) summarise the general research methodology that is intrinsically related to the process of reasoning in the design cycle, as illustrated in Figure 3-2. Each process step is associated with a research output.

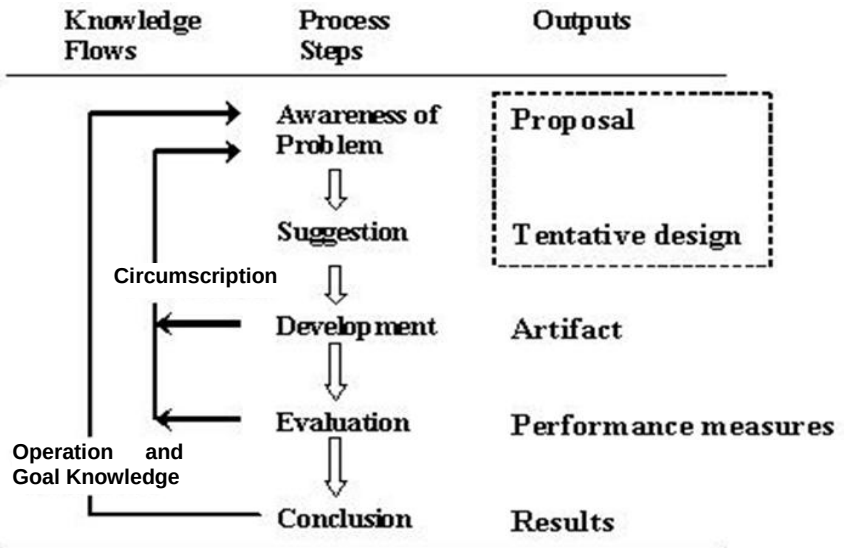


Figure 3-2 General Methodology of Design Research. Taken from (Vaishnavi and Kuechler,

As illustrated in the figure, *awareness of a problem* is expressed as a research proposal in the form of an informal or formal description of the problem and context, while the *suggestion* is expressed as a tentative design. The research proposal and tentative design stages are drawn together inside a dotted box because poor research proposals or bad tentative designs will be discarded before at early stages. Moreover, if after *awareness of a problem* a tentative design is not clear to the researcher, the research proposal will never be put into action.

In the *development* phase, the tentative design is implemented in the form of an artefact. There are four main parts or potential outputs of the artefact: *constructs, models, methods* and *instantiations* (March and Smith, 1995). Given the symbiotic relationship between design science and behavioural science, (Vaishnavi and Kuechler, 2004), identify *better theories* as a fifth potential output of the artefact. In detail:

- *Constructs* are understood as the language and definitions that provide the space in which problems and solutions are characterised and shared (Shon, 1993).
- *Models* make use of the constructs to represent reality (Simon, 1996), i.e. the design problem and its possible solutions. Models are also used to study the effects of design changes on reality (Hevner et al., 2004).

- Processes and sequences of activities are defined as *methods.* Methods, and the artefacts that are created from them, can be formal, informal or a combination of both, and as such can take various representations. This includes mathematical formulas, computational algorithms, software solutions, informal descriptions, or 'best practice' approaches, including frameworks and industry standards.
- *Instantiations* show that constructs, models and methods can be implemented in a working system, demonstrating feasibility and enabling the evaluation of performance and utility of the artefacts (Hevner et al., 2004).
- *Better theories* are the result of the understanding gained though the reasoning cycle. This is appropriate firstly because the construction phase is an object of theorising in different communities: for instance, according to (Vaishnavi and Kuechler, 2004), the construction phase can be an experimental proof of method or an experimental exploration or both. Secondly, it is appropriate because the artefact can depict relationships between its constituents, and if these relations are made evident, then the understanding is increased, therein potentially proving or rejecting previous conceptions and theories about them. In this respect, (Gregor and Jones, 2007) propose a broader view of theory in which constructs, models and methods can also be regarded as components of theory, or theory itself, identifying six core components and four additional components of an information systems design theory. According to the nature of these components (Gregor, 2006), proposes a framework for classifying the theories into one of five classes: (I) theory for analysing, (II) theory for explaining, (III) theory for predicting, (IV) theory for explaining and predicting (EP theory) and (V) theory for design and actions.

Artefacts are built in order to address problems, and should be subject to *evaluation* of the utility they provide in solving the issues. Usually, computational and mathematical approaches are used to evaluate the quality and effectiveness of artefacts, but evaluation could also be carried out with the help of data collection and empirical analysis techniques following behavioural science methodologies

(Hevner et al., 2004). Different communities will prefer and interpret a certain set of criteria over others, and this selection can be indeed explicitly expressed or be part of the paradigms that are passively used. The selection of performance measures, then, is the result of this implicit-explicit relationship. As explained before, the *conclusion* represents an arbitrary point of culmination at which the research process ends. The output of this phase is the result of satisfying, and not necessarily completely solving, the problem. In this sense, the artefact can still be in prototype form, and deviations or anomalous behaviour that defy explanation may well still be present, and moreover, expected and useful.

Figure 3-3 presents a conceptual framework for understanding, executing and evaluating IS research, combining behavioural-science and design-science paradigms as introduced by (Hevner et al., 2004). The figure depicts the concepts and aspects that are to be considered when using the design science methodology.

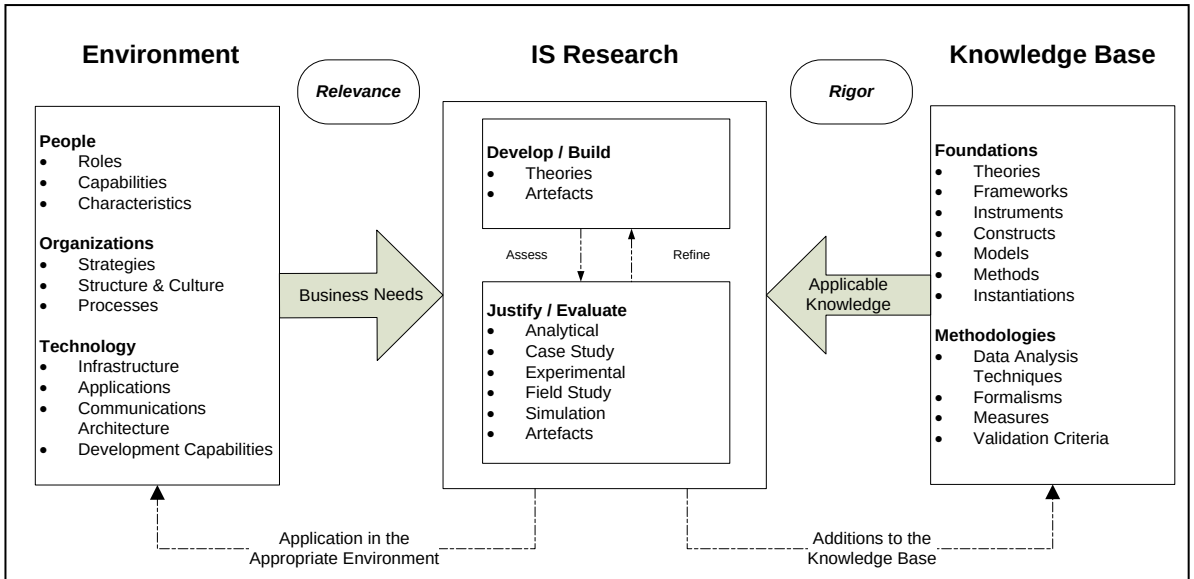


Figure 3-3 Information Systems Design Research Framework (Hevner et al., 2004).

The process of building (developing) and evaluating (justifying) an artefact (theory) is done with the contribution of knowledge gathered in a knowledge base. This contains foundational theories, frameworks, instruments, constructs, models, methods and instantiations used in previous research and problem solutions. As a

result, artefacts are born in the intersection of an *environment* that gives them context (Simon, 1996) and a *knowledge base* which contains the previous and accumulated intellectual and non-intellectual resources through which research is accomplished (Hevner et al., 2004).

The problem is composed of people, businesses, organisations and existing and planned technologies (Silver et al., 1995). The problem considers goals, tasks, barriers and opportunities that define the gaps as perceived by their incumbents. The problem is also positioned relative to existing technologies, applications, communication networks, architectures and related capabilities (Hevner et al., 2004).

According to this framework, one particular concern is intrinsically associated with the design-science side of the research; this is the difficulty in differentiating *routine design* from *design research*. The key for the differentiation is that design research gives a clear identification of a contribution to the archival knowledge, whereas routine research is the application of existing knowledge to organisational problems.

3.2 Justification of the selection of the methodology

There are three main arguments why this methodology is appropriate for the research task at hand. Firstly, this methodology is deemed appropriate because of the type of problem this thesis is dealing with. Table 3-1 summarises the different aspects of the problem that make it appropriate for design research.

Table 3-1 Characteristics of the Addressed Problem

Design Science	Addressed Problem
Considers people, technology and organisation	Considers traders, investors, general public, trading platforms, monitoring and surveillance systems, regulators, stock markets, dark pools, ECNs, etc.
Suitable to understand, explain and improve *information management* problems	Dark liquidity, weakening of public trust, and the existence of manipulations and other systems failures, could all be solved in one way or another by making available the right information, in the right place, at the right time, to the right person or organisation.
Suitable for *wicked problems*	Problem is ill-defined, and constraints and characteristics are ever-changing due to fast changes in technology and regulation. The solution to the problem still depends on people's cognitive and social abilities to perform trading decisions, regulatory changes, to perform surveillance and detection of market manipulations. There are no 'right' or 'wrong' solutions, only 'good enough' solutions.

To expand on the first argument, the new configuration of financial markets has made it increasingly difficult to preserve market integrity, which makes evident the need for a better understanding and enhanced functionality of the systems that support the markets. This need is also related to the way the components of the system interact within the boundaries of a jurisdiction, but progressively more importantly, to the way they interact across multiple jurisdictions. These systems require the orchestration of multiple agents, systems, and technologies, which form an entangled web of systems, and as such the problem is complex and difficult to comprehend. In this sense, it is clear that the problem is composed of people, businesses, organisations and existing and planned technologies.

Furthermore, design research is expressly suited to problems that involve analysis of the use and performance of designed artefacts to comprehend, describe and improve *information management* (Vaishnavi and Kuechler, 2004). Hence, it is especially appropriate for researching flows and interactions of information, regardless of the sophistication of the technology and processes that support them. One way to appreciate the subtleties of the appropriateness of this methodology is to recognise the deterioration of market integrity as an information management problem. Dark liquidity, the weakening of public trust and the existence of manipulations and other systems failures could all be solved in one way or another by making available the 'right information', 'in the right place', 'at the right time', 'to the right person or organisation'. In order to achieve this *magical reality* (Theodoulidis, 2011), developments of new market monitoring and surveillance systems could use new technologies and methods to improve functionalities, but perhaps more importantly, an organised body of knowledge or framework would contribute immensely to the understanding of the problem by making evident relationships and solutions that were not part of the knowledge base before.

Another aspect of the problem that makes it appropriate for researching design science is that the constraints and requirements that define the way financial markets operate are evolving rapidly due to fast changes in business models, technology and regulation, and thus the nature of the problem itself is changing at

a similar pace. In a way, the problem is not fixed, but is rather constantly evolving. In addition, the solution of the problem still depends on people's cognitive and social abilities to perform a great number of tasks, ranging from trading decisions and regulatory changes, to the surveillance and detection of market manipulations.

In the literature, problems with these characteristics are considered to be *wicked problems* (Rittel and Webber, 1984, Buckingham Shum, 1997). In fact, (Hevner et al., 2004) argue that design science is especially concerned with problems *'...characterised by unstable requirements and constraints based upon ill-defined environmental contexts; complex interactions among subcomponents of the problem and its solution; inherent flexibility to change design processes as well as design artefacts (i.e., malleable processes and artefacts); a critical dependence upon human cognitive abilities (e.g., creativity) to produce effective solutions; and a critical dependence upon human social abilities (e.g., teamwork) to produce effective solutions.'*

Secondly, the methodology is appropriate because the proposed IS research framework and methodology, as illustrated in Figure 3-3, incorporates a complete setting, which not only allows the problem to be appreciated and studied in all its complexity, but which also offers an explicit and organised method for building and evaluating solutions for the problem. Moreover, the methodology acknowledges the complexity of the environment in which the solution is developed and the previous knowledge that is available when performing the task. Complexities in the environment include, for example, the rapid developments in technology and regulations, embedding organisations and new forms of human-technology interaction.

Previous knowledge includes, but is not limited to, foundational theories, such as financial and economic theories, systems science theories, service systems worldview, in addition to regulatory frameworks, laws and rules. It also includes previously developed artefacts as well as existing methods and techniques, such as the Cross Industry Standard for Data Mining Process (CRISP-DM) (Chapman et al., 2000) or current and future market monitoring and surveillance systems, as part of its components and procedures.

Thirdly, the methodology is appropriate because the philosophical perspective of this thesis resembles that of design research and cannot be derived from other philosophical perspectives. Design science philosophy is based on the belief in an objective, *evolving* but *single* underlying physical reality that limits the multiplicity of outcomes. An artefact by construction is intended to change the state of the world, and in contrast, *positivist* ontology, for instance, assumes that there is only *one* reality and truth. Moreover, this positivist reality is objective and can be discovered, but not altered. At the other extreme of the philosophical spectrum, the *interpretative* ontology posits that are many subjective realities and that the world is socially constructed. Furthermore, the interpretative researcher engages with and alters reality by the mere act of studying it.

The ontological perspective of design science is also quite unique in the sense that its reality encompasses people, organisations, and technology, and thus it can be considered a socio-technical perspective. In terms of the creation of knowledge, design science, as depicted by the Abduction phase in Figure 3-1, assumes the existence of physical laws and knowledge previously gained from them that is to be used in the tentative design, and it assumes that knowledge is extracted not only through observation of the world, but also through the making and use of the artefact.

In other words, the design researcher believes that the understanding that is gained through the design research process reflects physical and socio-technical laws, through which different configurations and outcomes of the artefact can exist. This thesis uses the idea that the problem can have multiple but single 'good enough' solutions, and that *a single* final solution/truth can never be achieved, due to the complex characteristics of financial markets, and thus the thesis relies on the unique features of the epistemological and ontological views of design science.

Table 3-2 presents a parallel between these three different philosophical perspectives

Considering this philosophical standpoint it is clear to see how the design science is helpful in contributing to the aims of this thesis, in the sense this methodology helps building a body of knowledge that could facilitate the deployment and

acceptance of market monitoring and surveillance systems that are not only effective, but also that are understood in both their human and technological dimensions.

Table 3-2 Philosophical Viewpoints of Three Research Perspectives. Taken from (Vaishnavi and Kuechler, 2004)

	Research Perspective		
Basic Belief	**Positivist**	**Interpretative**	**Design**
Ontology	A single reality. Knowledge, probabilistic	Multiple realities, socially constructed	Multiple, contextually situated alternative world-states. Socio-technologically enabled
Epistemology	Objective; dispassionate. Detached observer of truth	Subjective, i.e., values and knowledge emerge from the researcher-participant interaction	Knowing through making. Objectively constrained construction within a context. Iterative circumscription reveals meaning
Methodology	Observation, quantitative, statistical	Participation; qualitative, hermeneutical, dialectical	Developmental. Measure artefactual impacts on the composite system
Axiology: what is of value	Truth: universal and beautiful, prediction	Understanding: situated and description	Control, creation; progress (i.e. improvement); understanding

3.3 Methodology application

This thesis is organised following the 'IS Systems Research Framework' and 'Design Science methodology' and as such, chapters and chapter sections can be mapped directly to them. Figure 3-4 presents a graphical representation of the research methodology aspects applied in this thesis. In particular, Chapter 1 describes and discusses extensively the *environment* and the problem setting, presenting how people, organisations and technology are affected, but also how they are involved in the configuration of an information management problem with characteristics of a wicked problem. Chapter 2 also touches on the *environment*, summarising and identifying the gaps and limitations of current systems that frame the need for an innovative solution. Moreover, Chapter 2, partially cover the *knowledge base* on which this thesis is built, reporting on previous literature,

technologies, foundations and methodologies. This includes previous financial and economic theories and empirical findings, as well as definitions of market manipulations, legal and regulatory frameworks, and comparisons of these across different jurisdictions, in addition to existing MMSS characteristics, procedures, performance, capabilities and the allocation of responsibilities.

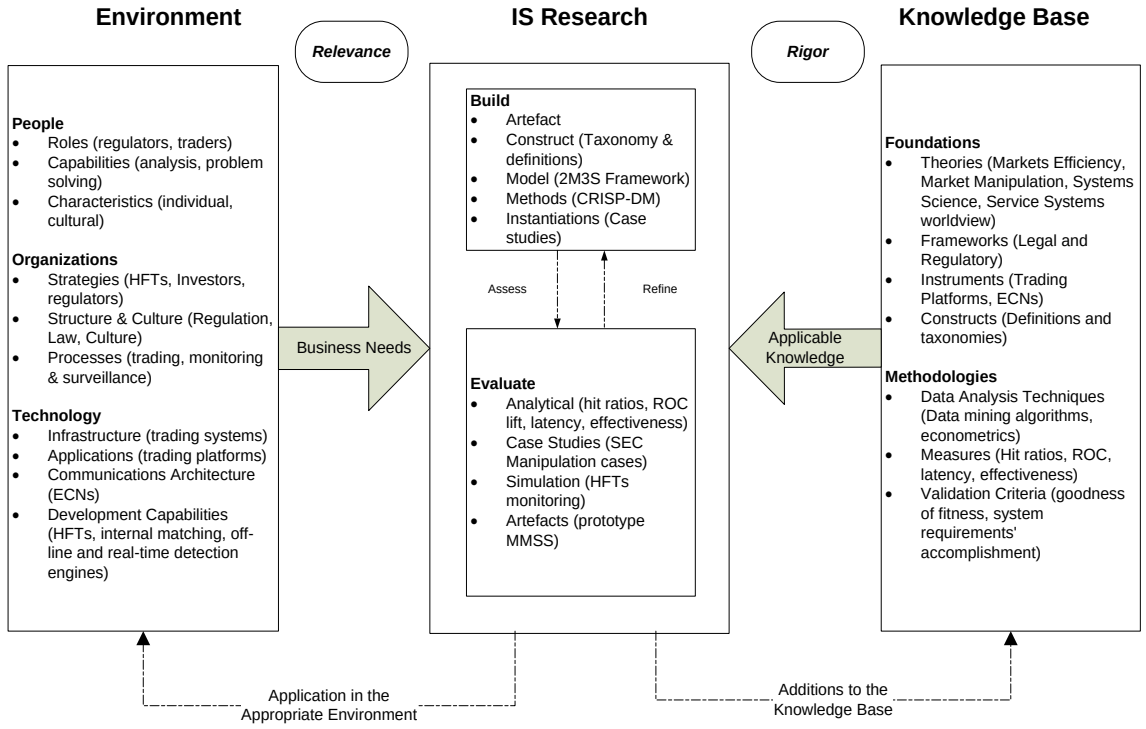


Figure 3-4 Adopted Information Systems Research Framework. Adapted from (Hevner et al., 2004). Examples in parenthesis ( ) were added by the author.

In particular, Table 2-5 formally states the desirable characteristics of MMSS and systems requirements. Chapter 3 provides a complete review of another important part of the *knowledge base*, describing and justifying in detail this research methodology and its philosophical perspective.

Chapter 5 discusses the *building* of the artefact, namely the Market Monitoring and Surveillance Service Systems ('2M-3S') framework. In particular, it presents the *constructs* of the framework, and introduces the representations of the problem domain or *models* in the form of a *'customer-driven value co-creation service value network'* (Kwan and Yuan, 2011). *Constructs* are also discussed and introduced using taxonomies in Chapter 4. *Methods* for building such models draw heavily on

other foundational theories, mainly service dominant logic (Vargo and Lusch, 2004), systems science (Spohrer et al., 2007), and systems thinking (Boardman and Sauser, 2008), which are described in Chapter 5.

Chapter 6 presents individual *instantiations* of the artefact by means of *case studies.* Each case study presents individual tools and salient characteristics of components and systems for the detection of information and trade-based manipulations and provides details of the prosecution cases or simulation analyses that give them context. Based on the previous cases, *evaluation* is carried out on a case–by-case basis, judging the utility of the proposed framework for increasing understanding of a part or the whole of the problem and the use of the models introduced in Chapter 5. *Conclusions* and results are given on two levels. At the lower level, each case study presents its own results and conclusions, and knowledge gained is abstracted away to be integrated later into a final 'conclusions and lessons learnt' chapter following the *circumscription* and *operation and goal knowledge* loops. Moreover, in Chapter 6 each case study follows the design research methodology in order to create independent artefacts that solve the particular problems that are introduced in that specific case.

In all, the chapters and sections represent a *relevant* and comprehensive revision of the business needs from which the problem identified this thesis stems. Together, they also cover a *rigorous* review of the applicable knowledge as represented by respective arrows in Figure 3-4.

In terms of theory type and theory contribution, this thesis can be classified as Type I, theory for analyzing. According to (Gregor, 2006), *"analytic theories analyze "what is" as opposed to explaining causality or attempting predictive generalizations"*. This usually includes classification schemas, frameworks or taxonomies, as it is the case for this thesis. This kind of theory makes contribution to knowledge by describing and analyzing when little is known about some phenomena. The criteria for the evaluation of this type of theory, then, is related to implicit claims that the classification system is useful in aiding analysis, that the category labels and groupings are appropriate, that the logic for placement phenomena into categories is clear, and that no important categories or elements

are omitted from the classification system. By means of case study instantiations, this thesis evaluates the appropriateness of the framework in relation to the degree to which it satisfies these criteria.

Finally, Figure 3-5 presents a schematic view of the implementation of the methodology mapped to chapters of this thesis. Boxes shows the respective sections in which the different research activities are mainly described. The next chapter further organises the concepts, definitions, and properties of the MMSS by presenting a collection of taxonomies relevant to this work.

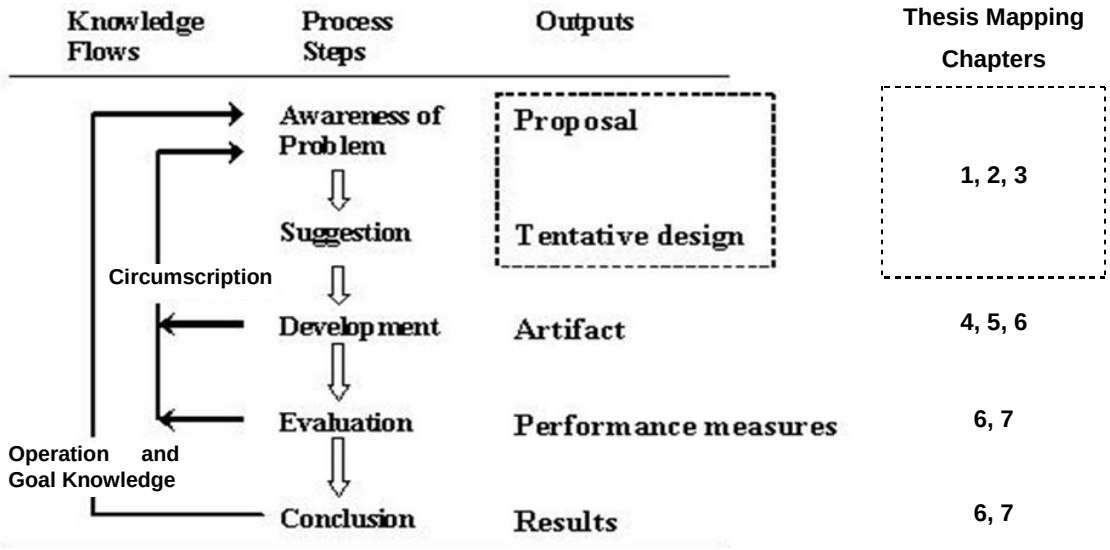


Figure 3-5 Research Mapping. Adapted from (Vaishnavi and Kuechler, 2004). Thesis Mapping Chapters column was added by the author.

4 Taxonomies of market manipulation and market monitoring and surveillance service system concepts

This thesis makes use of a large set of concepts and definitions originating from multidisciplinary research communities together with terms that are commonly used in financial markets. In order to manage the definitions and basic relationships of these concepts, this chapter present a collection of working taxonomies, which are not necessarily complete or accurate, but which are of great help to build an organised body of knowledge. Hopefully, they will also help as a working reference for future developments of a full ontology (Gruber, 1993), an endeavour that is beyond the scope of this project.

These collection of taxonomies are needed to help building a common understanding of market monitoring and surveillance systems, not only because of the large number of concepts that are used when discussing such systems, but also, because of the lack of appropriate terminology derived, for example, from the introduction of constant innovations, such as, high frequency trading. This need to improve the *awareness of the problem* is also reflected in the fact that different communities refers to certain concepts with dissimilar levels of details, and that it is also possible to find different, and sometimes even contradictory, interpretations of concepts and/or relations among them.

In order to build the taxonomy a 'top-down' approach was followed. For this a complete literature review was undertaken which considered the revision of previously developed taxonomies and working dictionaries. Furthermore, these were expanded and enhanced with the aim of increasing the coverage of concepts, as well as to help organising them into three major categories. These categories considered a taxonomy of basic elements and entities present in different manipulation scenarios; a taxonomy of the interacting agents in financial market systems; and a taxonomy of the sub-systems, components and information management concepts of a model detection engine.

In order to present and exemplify the taxonomies, different manipulation scenarios and the financial terminology used to describe them are discussed: through these

scenarios the different taxonomies are introduced gradually so that previously established concepts, properties or relationships can be reused or recalled when necessary. Finally, in order to evaluate the taxonomies these were instantiated by means of case studies. Correspondingly, each of the cases presented in Chapter 6 are analysed also in terms of the coverage of the concepts and terminologies used. The chapter starts with a presentation of the taxonomy of the basic elements and entities present in different manipulation scenarios. It continues with the introduction of a taxonomy of the interacting agents in financial market systems, and finishes with a taxonomy of the sub-systems, components and information management concepts of a model detection engine.

A final clarification is necessary before presenting the taxonomies. Although Figure 2-1 Taxonomy of Market Manipulations and Abuses. Adapted from (Putnins, 2009) and (Cumming and Johan, 2008); and Table 2-3 Allocation of market monitoring and surveillance responsibilities, are not physically present in this chapter, these sub-sections should also be considered part of this proposed collection of definitions and taxonomies.

4.1 A taxonomy of elements in manipulation scenarios

A typical market manipulation problem scenario can be analysed by considering the interaction of six fundamental elements:

- the *agents* or manipulators,
- the *actions*,
- the *venues* or markets involved,
- the *time* of the manipulation,
- the *target asset*, and
- the *effects* of the manipulation.

Figure 4-1 presents a graphical representation of the elements involved in a manipulation.

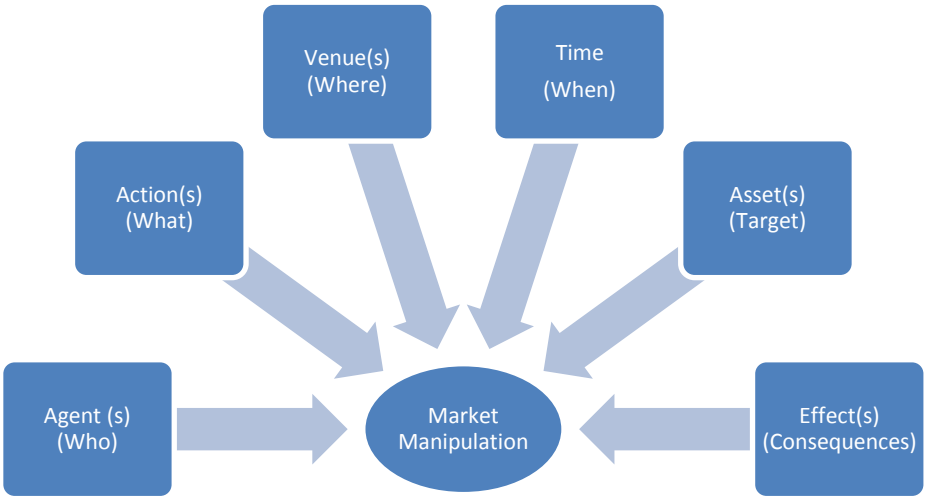


Figure 4-1 Elements Involved in a Manipulation

Conceptually, an agent corresponds to the element that performs a manipulation by taking an action. Agents can be outsiders or insiders of organisations: they can act on their own or in collusion with other agents, exerting their social networks, and may act for their own benefit or on behalf of a third party (Figure 4-2).

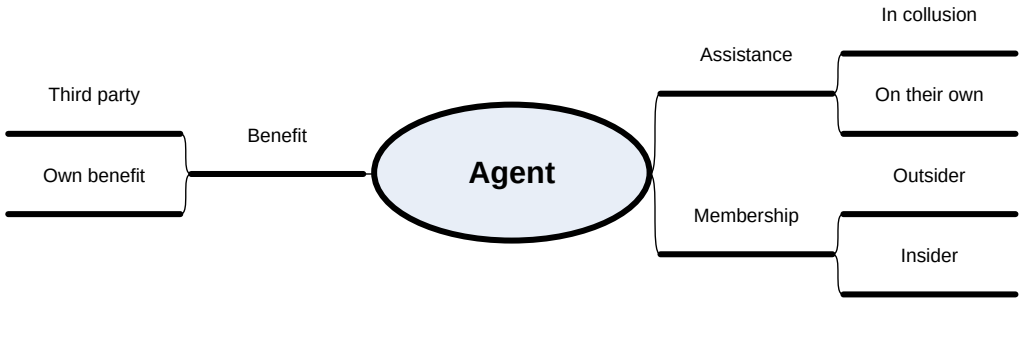


Figure 4-2 Taxonomy of Agents

Agents' actions can be categorised into three groups: the release of false or misleading information to the public; the placement of trading orders that could result or not, in a trade; and any other activities that modify the future cash flows of the companies (Allen and Gale, 1992) (Figure 4-3).

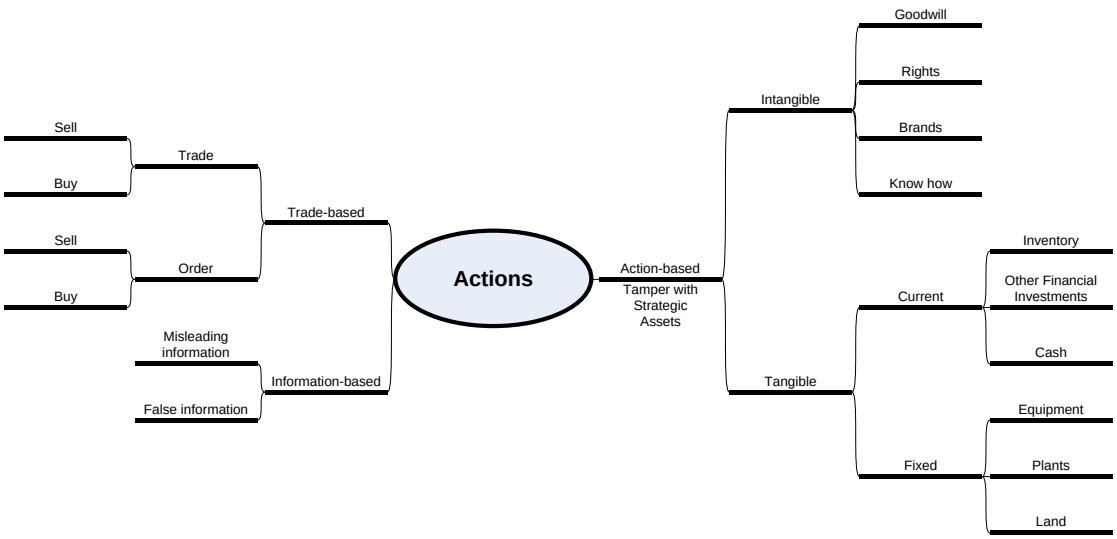


Figure 4-3 Taxonomy of Actions

Venues in which investors and manipulators trade can be classified into self regulated organisations (SROs) and non-self regulated organisations (non-SROs). Venues may allow or disallow cross-product and cross-market trading, and may be 'dark' or 'light' according to the type of liquidity they provide. Venues can support electronic trading, algorithmic trading, high frequency trading or floor trading. They can perform real-time and off-line monitoring and surveillance, and can support cross-product and cross-market surveillance (Figure 4-4).

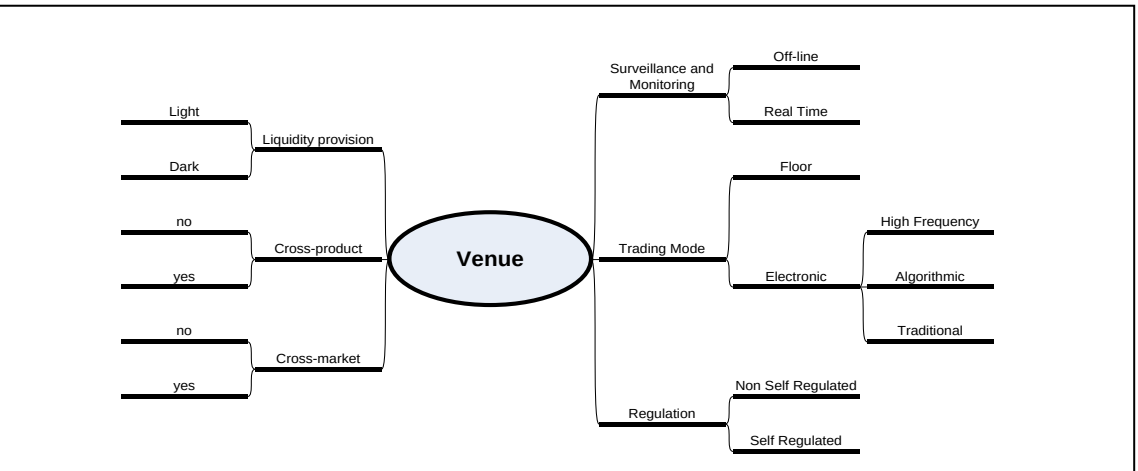


Figure 4-4 Taxonomy of Venues

Agents could target any type of asset or security that is traded in the markets, and thus the target asset set includes a full range of financial instruments, usually stocks but also penny stocks, money markets, bonds, derivatives, structured notes, and other securities. Manipulations can also involve some non-financial assets, mainly commodities such as precious metals, mining products, fruits and vegetables, farm animals and sub products, etc., which are traded with standard delivery, settlement and quality formats (Bodie et al., 2008).

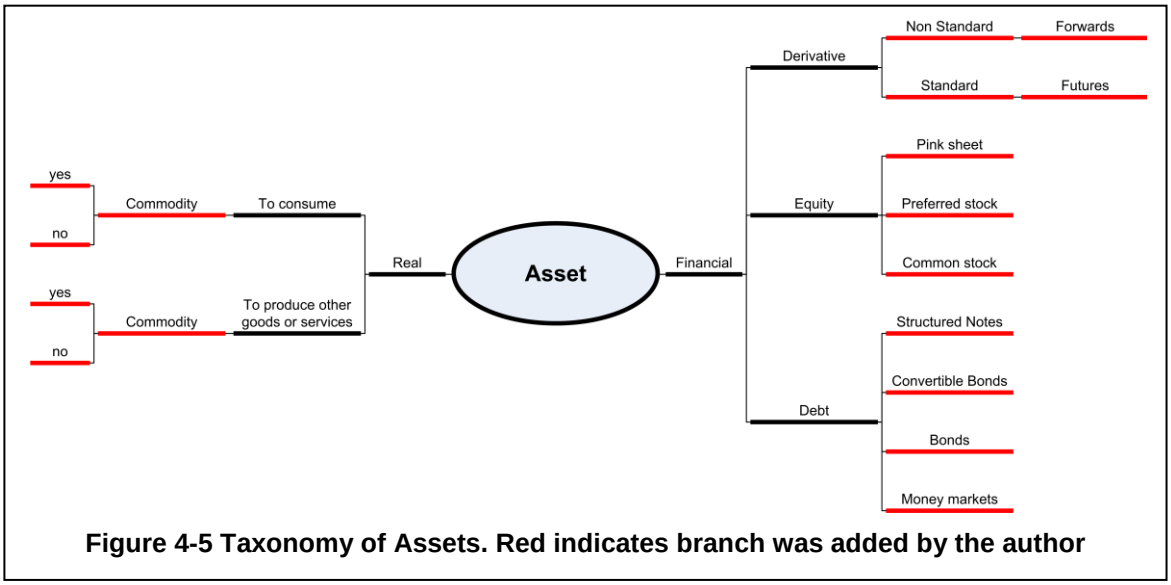


Figure 4-5 Taxonomy of Assets. Red indicates branch was added by the author

These issues also have a temporal dimension which could be reflected in unusual trading activities during critical periods, such as quarter ends, year ends, or at the end of the trading day. In general, they could also occur during normal trading hours, or outside normal trading hours. Trading can happen at high frequency — at the millisecond, microsecond, picosecond levels and so on — or at normal speed (i.e. if it takes more than a second to complete) (Figure 4-6).

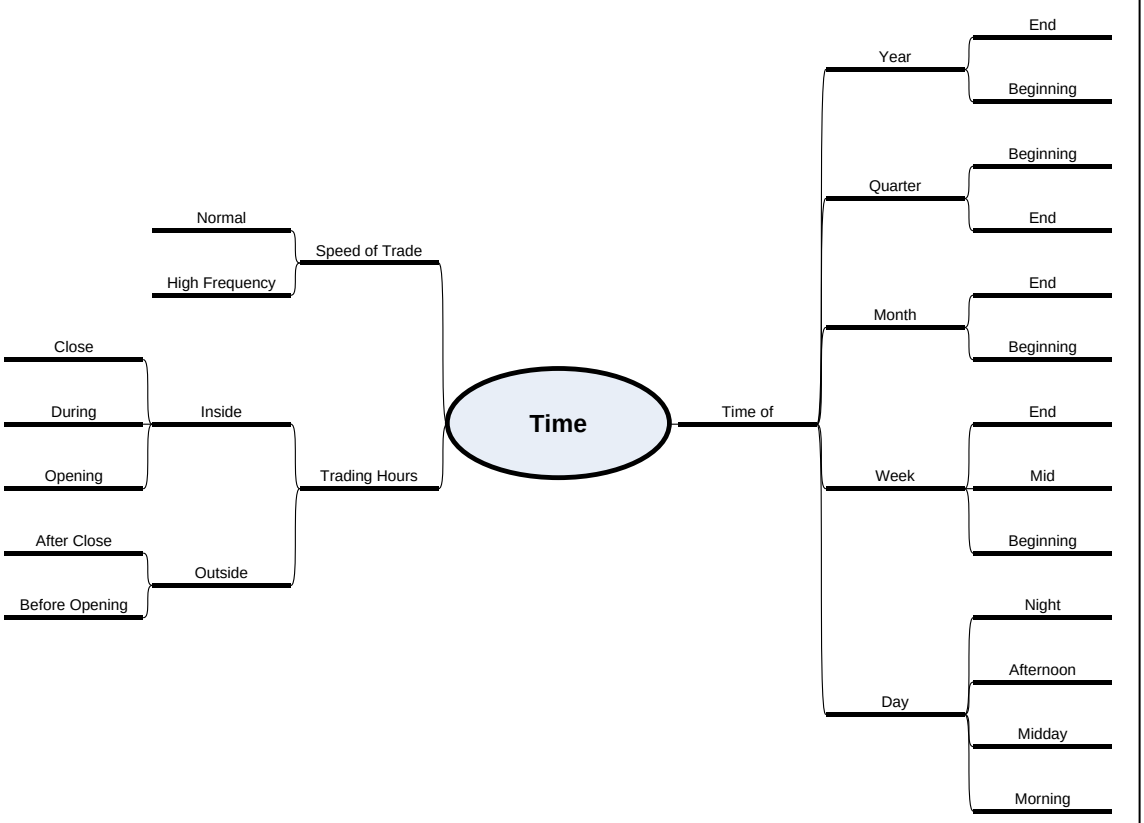


Figure 4-6 Taxonomy of Time Concepts

Actions taken by manipulators will induce an effect or effects on the behaviour of target assets that will directly or indirectly benefit the agents by creating a profit for them. A direct benefit is one that is generated in the venue in which the target is manipulated; an indirect benefit is one generated elsewhere (Figure 4-7).

The effects of the manipulation could be reflected mainly in the market in which the target asset is traded, but also could induce reactions in related assets, other markets in which these are traded, or even outside the markets.

The effects of manipulation create patterns that can be spotted in the behaviour of the agents, or in trading orders and executions that otherwise would have occurred differently. Indeed, agents' behaviour and consequent actions leave traces in both structured and unstructured data. Their behaviour is registered in, for example, the number of order trades they make, the number of orders they cancel, their wire transfer records, phone call records, or even chat logs. Moreover, the effects of

their actions leave even clearer prints in trading data, from the location from which the trader orders originated to jumps in the volume of trading or episodes of unusual price volatility.

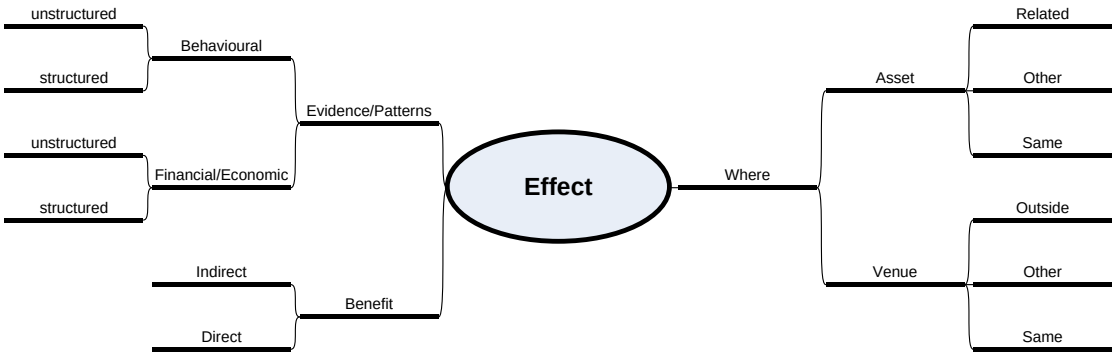


Figure 4-7 Taxonomy of Effects

Some key questions that the customer or user of monitoring and surveillance services should be able to answer include: Who is(are) the agent(s) involved in the manipulation? Which asset is being targeted? In which venue is the manipulation taking place? Which action has been performed or is planned? Is it a trade-based or an information-based action? Which pattern or content can be identified in trading data, financial indicators, or other media such as news, press releases or corporate fillings that can be associated with this manipulation? When was this manipulative action performed? When was it possible to spot the effect of this action of manipulation? Where is the manipulator getting his profit from?

4.1.1 Examples

In order to exemplify this first set of taxonomies, we can start by considering the most common type of manipulations, namely 'marking the close' schemes. These are defined as '*the transactions near the end of day's trading, which affect the closing price of the security in order to give an artificial price to the security in the next day*' (Financial Services Authority, 2007). In these schemes, it is possible to identify an *agent* and a *time*: a portfolio manager trying to inflate quarter-ends portfolio prices (Carhart et al., 2002); an *action*: last-minute purchases of stocks

already held; a *target asset*: the stocks he or she already holds; and its *effects*: stock prices, volume and volatility increasing beyond its normal range of fluctuations.

In terms of information-based manipulation scenarios, consider manipulations such as insider trading, in which fraudsters look to profit from pieces of privileged information that have not yet been disclosed to the public. In insider trading manipulations, agents commonly buy options contracts written on the expected value of an underlying security instead of buying the stock itself. Options are usually low in price compared with stocks, and thus manipulators with limited budgets could buy the rights to buy or sell a larger number of stocks than they would be able to buy or sell if they target the stock, hence leveraging the piece of privileged information they possess (Donoho, 2004).

It is also possible to identify more complex scenarios of manipulation, which include targeting one security by using an intermediate market and/or combinations of target securities. For instance, at present a large proportion of call-option contracts are priced using the implied volatility of prices that can be estimated from the bid and ask spread of underlying asset prices for which the call-option contract is written. Thus, it is theoretically possible to manipulate the call-option contract prices, which are traded in a derivative market, by placing sequences of ill-intended orders on a stock market. In this manipulation, a trader – the agent – places a sequence of trade orders in a stock market without necessarily fulfilling a trade – the action –, which will alter the bid and ask spreads – secondary effect. That will also alter the estimation of the implied volatility of the stock prices – tertiary effect – which will finally result in an increase or decrease in the price – main effect – of the call-option contract – the main target security. Furthermore, this strategy could also be applied in the opposite direction, i.e. manipulation of the stock market via the derivative market.

As can be seen, these combined mechanisms of manipulation can be extended to several target securities, markets and even countries. Moreover, the actions can be the result of the coordinated efforts of different agents (groups of individuals) from inside or outside the company or companies. Actions can also consider

combinations of both information-based and trade-based manipulations. For example, consider the hypothetical case of three coordinated agents, one with insider information about an overseas company of which its American Depository Receipts (ADR) trades in a US market; one with a licence to trade in the same US market; and one with access to a large emailing list outside the USA. The agent inside the company could pass information to the agent with access to the emailing list, who could in turn disseminate a misleading rumour about what will happen with the company quarterly earnings. If the fraudulent emailing campaign is successful, local stock prices will fall after victims of the scam email sell their local stocks, and as a result, ADR prices in the US market will fall in order to reflect the loss in value of their underlying stocks. The coordinated agents could profit from this scheme if the one with a licence to trade in the US market buys the ADR when the price falls, and sells it after the real quarterly earnings are disclosed to the public.

4.2 A taxonomy of stakeholders of financial markets

From the perspective of the market monitoring and surveillance systems, four key stakeholders of financial markets can be identified: these are *investors*, *intermediaries, regulators* and other *support organisations.* Figure 4-8 presents a graphical representation of them.

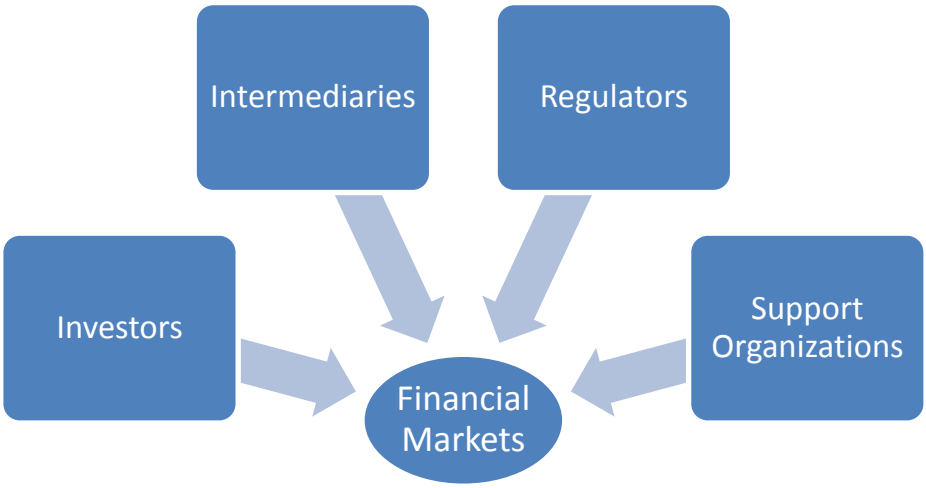


Figure 4-8 Financial Markets' Stakeholders

In particular, investors can be net borrowers or net savers, according to the position they typically adopt in financial markets. Business firms are usually classified as net borrowers, and individual investors, such as households and individuals, are considered to be net savers. Governments, in turn, can usually be both net savers and net borrowers at the same time. Governments can issue debt instruments, such as treasury bills, notes and bonds, to borrow funds when confronting budget deficits. They can also retire outstanding debt when facing fiscal surplus (Bodie et al., 2008). Investors can also be classified as proprietary or intermediary. Investors are proprietary when they trade using their own funds or capital, bearing the systematic and non-systematic risk on themselves. Proprietary investors include, but are not limited to, households and individual investors, business firms, institutional investors or investment firms such as banks, hedge funds, high frequency trading firms; governments; insurance companies, and credit unions, as illustrated in Figure 4-9.

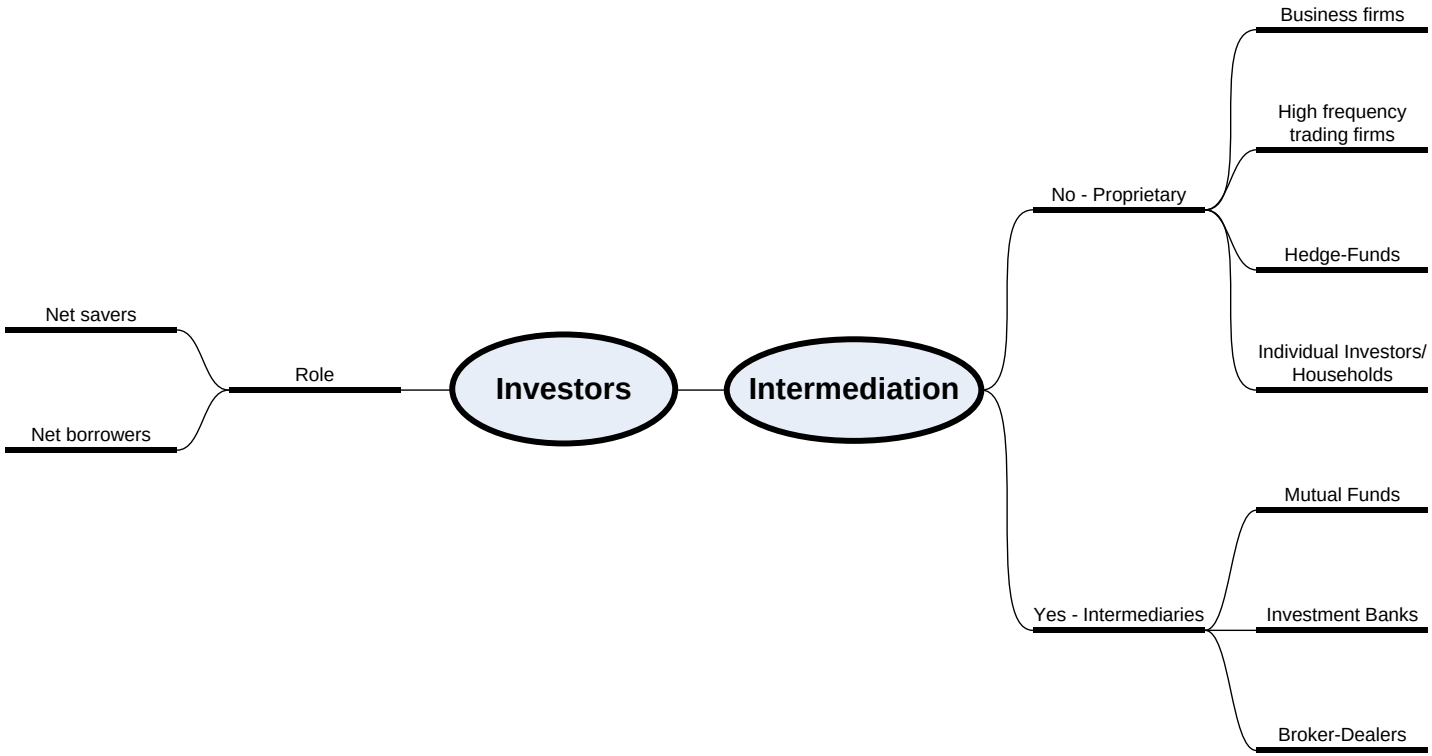


Figure 4-9 Taxonomy of Investors - Intermediation

Notably, most of the organisations listed above can be both proprietary and intermediary investors, depending on whether they are trading the equity of their shareholders and owners or not. However, there are some organisations, such as broker-dealers, investment banks and mutual funds, that are mostly dedicated to financial intermediation *per se*.

Regulators and regulating authorities perform market surveillance and market monitoring activities, which include detection, investigation and enforcement tasks among others. The regulators' reach and activities will depend on the jurisdictional reach that their respective regulatory framework has established. Regulators can be considered local if their responsibility is limited to one jurisdiction, or cross-jurisdiction if their responsibility entails more than one local jurisdiction. Regulators can perform the monitoring and surveillance activities on their own, or can partially (or totally) transfer the responsibility to an SRO, non-SRO or other third party(ies) if regulations allow it or do not forbid it explicitly.

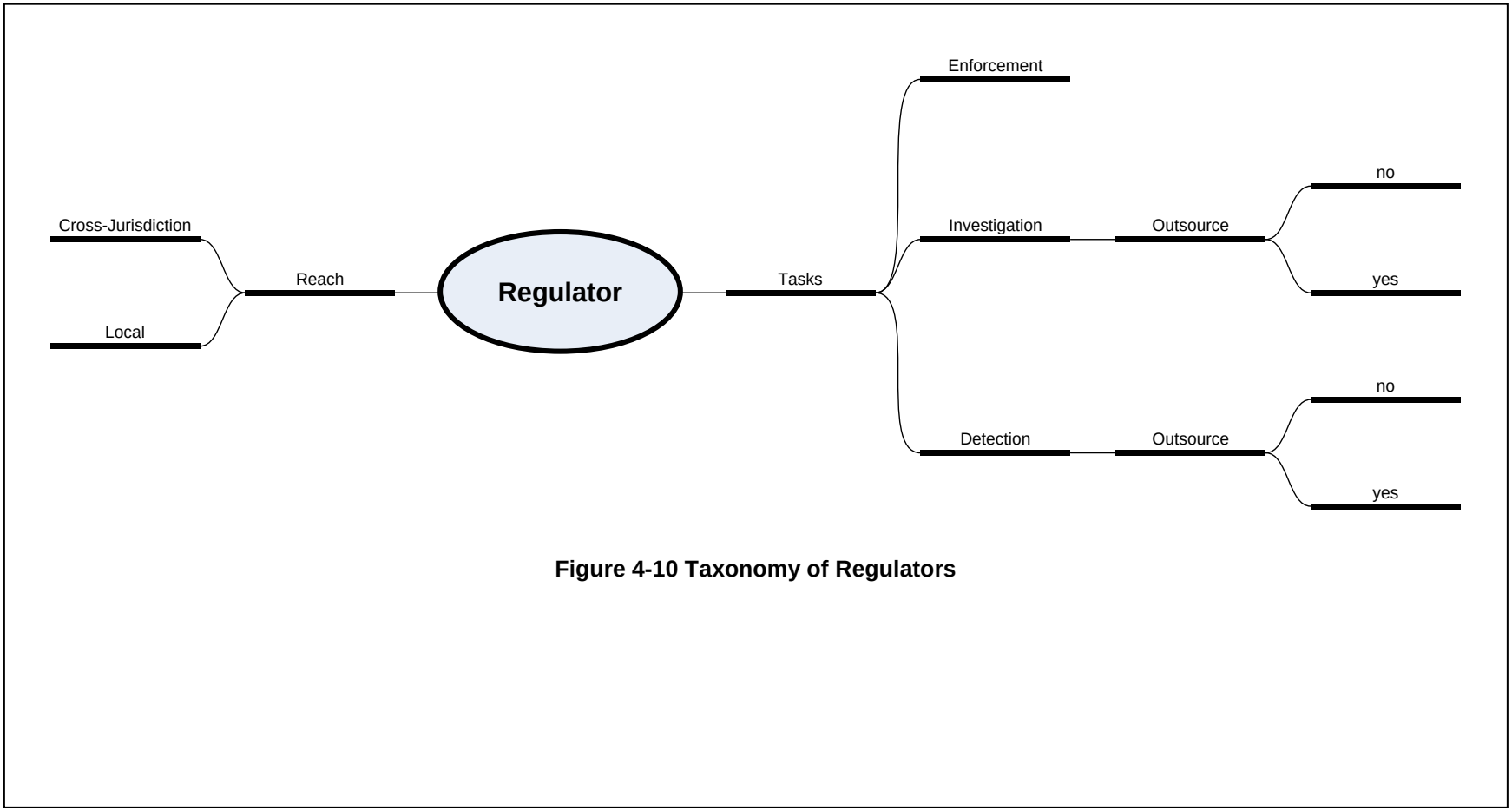


Figure 4-10 Taxonomy of Regulators

Support organisations perform, facilitate, or provide services or goods that other stakeholders need and are willing to acquire from them. Support organisations can provide, for instance, central tape and management services, central depository services, general information services such as news, contingency and political analysis services; financial services such as rating agencies and insurance services; data provision and hosting services; or even commercial, on-demand or historical monitoring and surveillance services. Support organisations can also provide hardware, such as networking, trading and communicating facilities; software, such as business analytics solutions; and other services such as know-how.

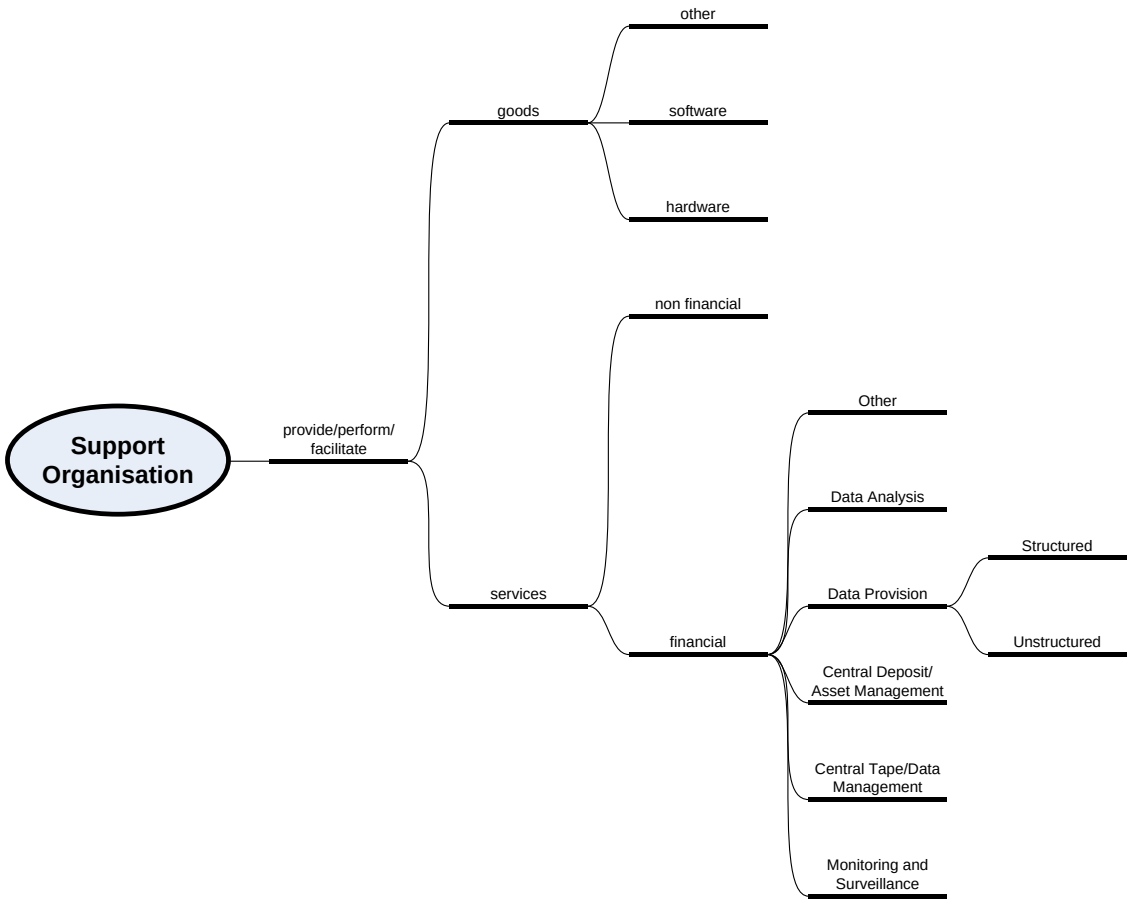


Figure 4-11 Taxonomy of Support Organisations

4.3 A taxonomy of market monitoring and surveillance detection' engines and components

Detection engine components and sub-systems can be grouped in three main categories, namely 1) *collect & integrate,* 2) *analyse* and 3) *report* sub-systems. Correspondingly in the taxonomy presented in Figure 4-12, each component is given a colour (purple, light blue, and orange respectively).

In the engine it is possible to identify three main sub-systems: a) information management, b) off-line detection engine and c) real-time detection engine sub-systems. Customers or users of the system can be also manifold, including the general public, investors, commercial market monitors, regulators, SROs, non-SROs, etc.

Connection links between venues can support (or not) cross-jurisdiction, cross-product or cross-market monitoring and surveillance. In the information management sub-system different processes take place, such as data preparation, data pre-processing, and textual source (e.g. complaints) handling. Accordingly, these input sources can be either structured or non-structured. With regard to the form of information that is transmitted and communicated between systems, these are: logic data flows, if they follow a sequence arbitrarily established by the customer or user; data flows, if they are fixed or always present links; or combined data flows if they possess characteristics of both.

Analyses are further grouped into two categories, namely *behavioural analysis* and *economic analysis*. The key difference between categories is that the first type of analysis focuses on the actions and characteristics of the agents, and the second focuses on the effects of these actions. Figure 4-13 expands the examples of analysis components and their classification.

A knowledge base can contain knowledge related to previously known cases, or to previously known patterns of manipulation. The output management component performs the task of updating the knowledge base and other reports and case management tasks. In the case of a regulating authority this could be deciding whether the flagged trades warrant further legal analysis, starting (or not) a formal

investigative process, and/or following processes of the reporting and reinforcement (if applicable) of market rules. In the case of an arbitrage seeker or market trader, this could be used to decide whether to invest in a stock that presents symptoms of manipulation or any other alternative investment actions.

The off-line monitoring and surveillance engine supports the real-time engine. In the former, components of analysis and or data pre-processing and integration take place, including data polling, circuit breakers and exchange rules monitoring, using for example 'complex events processing' components (Sadoghi et al., 2010).

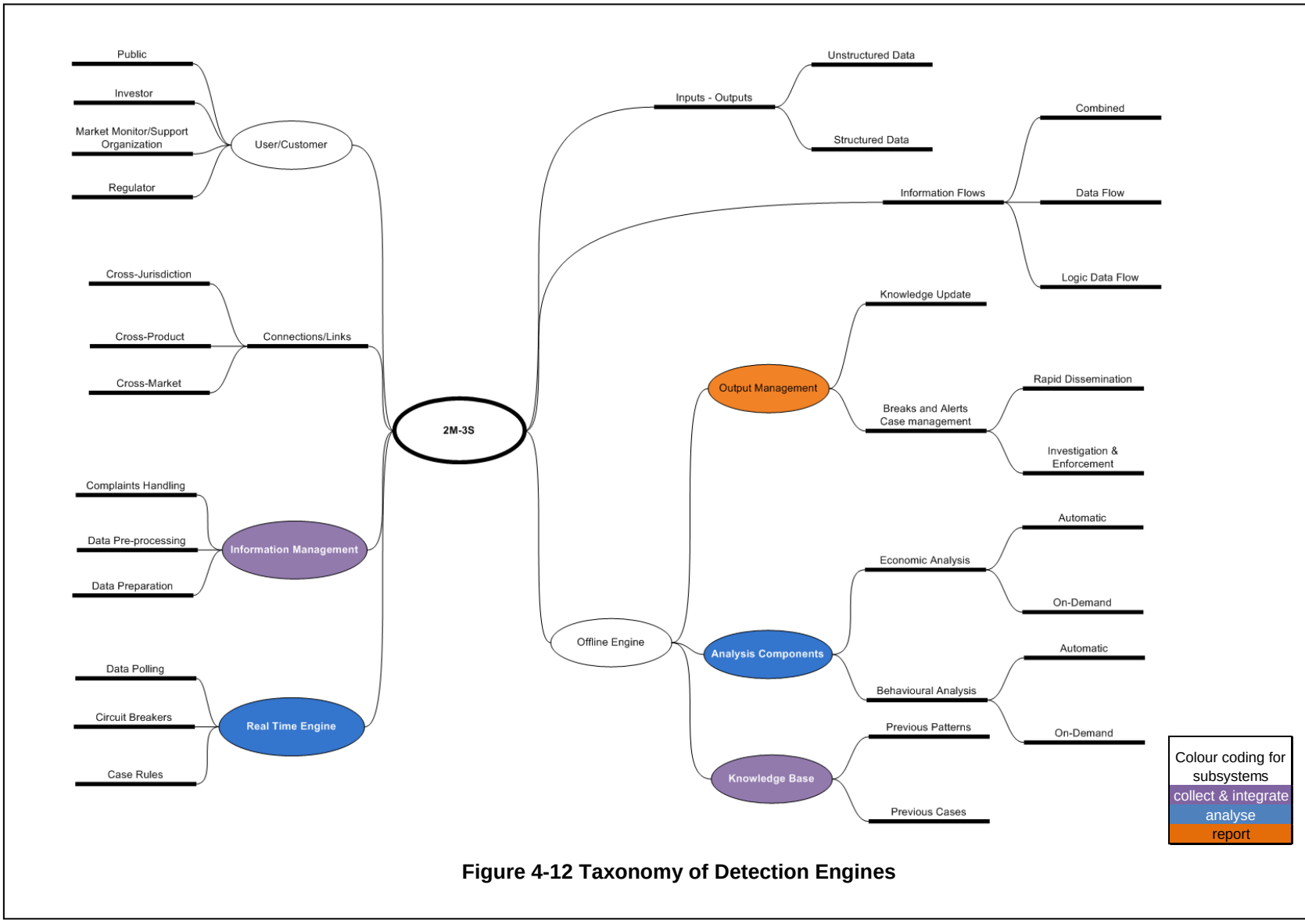


Figure 4-12 Taxonomy of Detection Engines

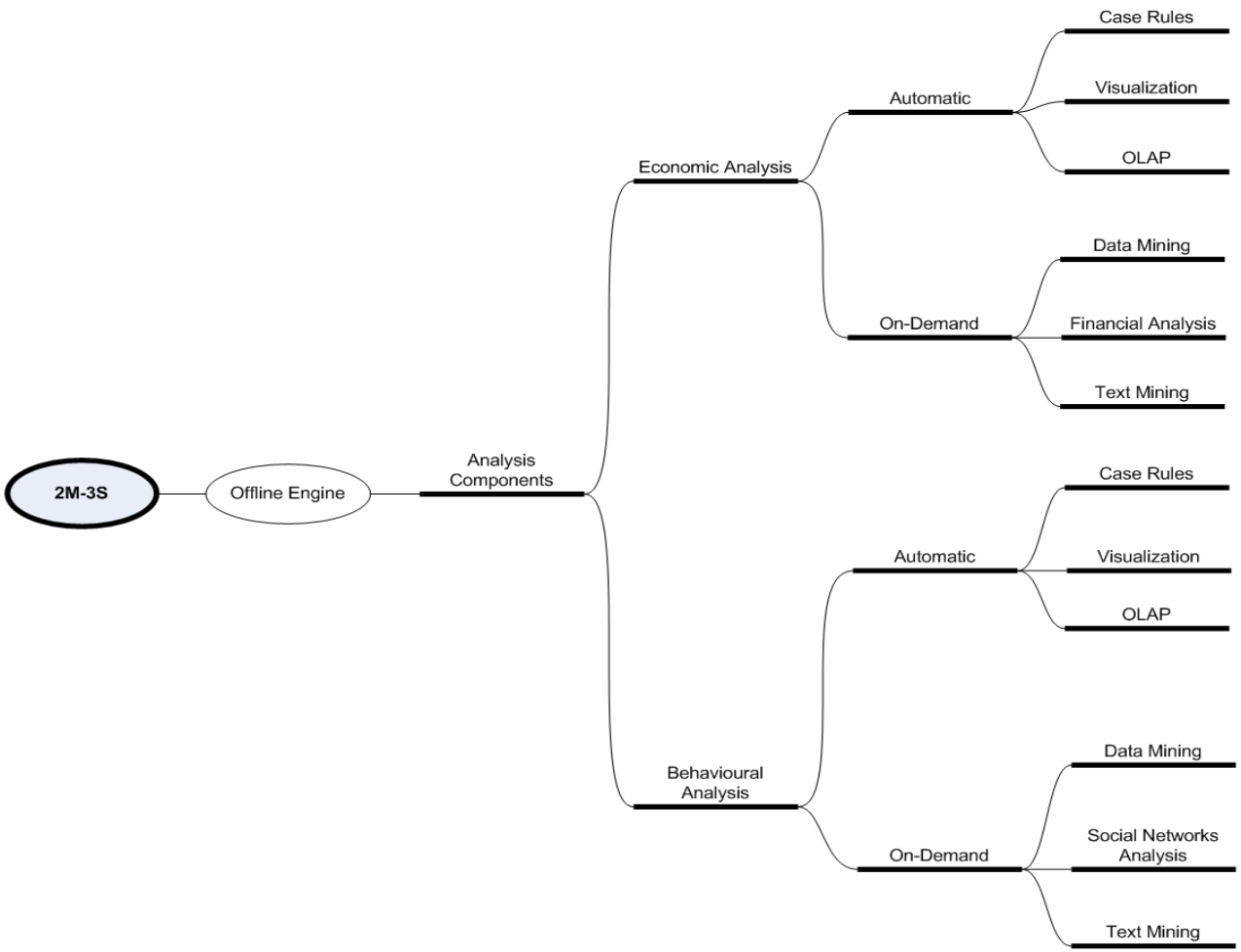


Figure 4-13 Taxonomy of Sample Analysis Components

5 The proposed Markets Monitoring and Surveillance Service Systems Framework

This chapter discusses the *building* of the artefact. In particular, it presents the *constructs* of a framework, and introduces the representations of the problem using *models* in the form of *'customer-driven value co-creation service value networks'* (Kwan and Yuan, 2011). Formally, this framework is termed the 'market monitoring and surveillance service systems' framework or '2M-3S' framework.

The chapter begins with a brief review of service-dominant (SD) logic (Vargo and Lusch, 2004), service science (SS) (Spohrer et al., 2007) and systems thinking (ST) principles (Boardman and Sauser, 2008), which provide the foundations on which the framework is designed. It continues with a discussion of the alternative set of *models* that encompasses the '2M-3S' framework, drawing together principles from SD logic, SS and ST. Finally, a model engine for market monitoring and surveillance is presented, describing the view of the sub-systems, components, tasks and flows of information necessary for a modern detection engine that meet the systems requirements.

5.1 Contextual theories, principles and their application to the framework

5.1.1 Service Dominant Logic

In this thesis we argue that every exchange in the economy can be understood as a service, and in fact, that all economies are service economies. As a corollary, it follows that all market monitoring and surveillance activities are services activities which are delivered within a contemporary service economy. In this sub section the word *exchange* is a generic term borrowed from economic theory, and does not refer to the physical or electronic venues in which trading in financial markets takes place.

In the product-centric worldview businesses are conceived of as manufacturing tangible goods. The goods are the carriers of a certain value, for which customers are willing to pay a price. The cost of producing the good was lower than or equal

to the price paid by the customers, and consequently, companies were willing to manufacture the good using their resources to later participate in an *exchange* of goods for money.

However, in the modern economy, in which more than two thirds of world's wealth is generated by the service sector, the traditional worldview is challenged, as the concepts of 'good', 'value', and 'exchange' become increasingly difficult to delineate. In the service-centric worldview companies are not able to prescriptively 'manufacture' a service with the freedom they enjoyed while the focus was on producing goods. In contrast, customers are seen as active participants in the provision of the peculiar type of goods we call services. Customers bring their own expectations, past experiences, and resources to the service delivery moment, thus shaping the characteristics of the good which has been created and delivered. Similarly, companies can no longer claim that they manufacture a service that provides a known value to the customer. Customers enhance or modify the service using their resources, and in doing so transform the value that is provided. The exchange concept is also distorted, as the process of *'I give you money, you give me a good with a known value'* becomes more complex. Companies are now only able to offer a value proposition in exchange for the price paid.

In particular, (Vargo and Lusch, 2004) and (Vargo and Lusch, 2008) were the first authors to propose an evolution and shift of logic, arguing that **every** exchange process needs to be understood as a service, and thus that every economy is a service economy in its own right. In this view, goods become only the means for service provision, and *'service is what is always exchanged'* (Barile and Polese, 2010).

According to SD logic (Barile and Polese, 2010), service is the application (through deeds, processes, and performance) of specialised operant resources (such as technology, skills, know-how, and wisdom) over operand resources (such as natural resources, brands, and patents) for the benefit of another entity or the entity itself. Operand resources are those that are acted upon to produce a certain output or effect while operant resources are those that are used on operand (as well as other operant) resources. Moreover, service is the fundamental basis of exchange

in the interactions among economic, social and systems entities (Vargo and Lusch, 2004, Vargo and Lusch, 2008, Vargo et al., 2008).

According to the SD logic there are ten fundamental principles which give substance to these ideas. This thesis posits that these ten fundamental principles are closely related to this work. These fundamental principles and their application to the framework are presented in Table 5-1.

Table 5-1 Principles of SD Logic and Application to the 2M-3S Framework

Principles of SD logic	Description	Application to the Framework
Service is the fundamental basis of exchange (FP1)	All services are the application of specialised operant resources over operand resources for the benefit of an entity; Moreover, all exchanges can be understood as a process with the ultimate aim of creating value for some entity. It follows that all exchanges then incorporate services, and that services are the basis of those exchanges.	Market monitoring and surveillance activities are the application of operant resources, such as fraud detection teams, detection engines or data mining technologies, over operand resources, such as trading platforms, data, communication networks, or others. Hence, they configure services process, which are the fundamental basis of this thesis.
Indirect exchange masks the fundamental basis of exchange (FP2)	According to SD logic, the traditional exchange of goods for money only represented a particular type of exchange. In the traditional exchange, money received by the company will eventually be used to buy a good or service for the direct benefit of the shareholders. The last good or service in this chain will indeed provide value, thus goods, money, and services interchanged in between, are only intermediate exchanges that mask the final exchange of services.	A monitoring and surveillance activity could be mandated by a regulator but executed by a third party, such as an SRO. The fact that the process itself does not directly give value to final stakeholders, such as the general public or society as a whole, only masks the final value that it is delivered to them when financial markets operate within normal ranges of efficiency and integrity.
Goods are distribution mechanisms for service provision (FP3)	Goods are a kind of resource that can be used for service delivery. Thus, they can be understood as a physical representation of service provision and efficient knowledge transfer.	In financial markets, the delivery of goods such as commodities can be understood as a kind of resource that is a mean for service delivery.
Operant resources are the fundamental source of competitive advantage (FP4)	As operand resources are closely related to things that are available in nature, competitive advantage can only be gained by making use of them in ways that cannot be easily replicated or transferred. Operant resources give the entities this capability, as they represent distinctive and strategic resources, such as experience,	Operant resources are key resources to achieve and sustain competitive advantage. For instance, from the regulator's perspective, the only way of staying ahead of manipulators is to use strategic resources such as the latest detection technologies in HFT, or well-trained surveillance teams and systems developers to

	relational trust, wisdom, strategic information, social connexions, and others, that make possible the use of operand resources in unique and difficult to replicate ways.	fight new and sophisticated market manipulations.
All economies are service economies (FP5)	As a corollary of FP1, FP2 and FP3, it is possible to argue that all economies can be understood as service economies. This is patent in the contemporary economy, in which services are provided as the result of networked relationships, continuing interactions and value co-creation networks (Grönroos, 2008, Kwan and Yuan, 2011)	Market monitoring and surveillance activities are only part of broader systems, namely venues and financial markets, which are in turn parts of society. As all exchanges at all levels are services, the whole economy can also be understood as a service economy.
The customer is always a co-creator of value (FP6)	In the traditional worldview, value was created in the manufacturing process and was embedded in the good itself. In SD logic, firms can only make value propositions and provide resources to the customer which will co-create the service with them. Furthermore, value is realised in the simultaneous use and production of the service. In addition, from the company's perspective, the customer can be considered as a special type of resource, one that actively participates in the provision and co-creation of value i.e. the service delivery process.	This thesis posits that there can be more than one definition of customer or user of market monitoring and surveillance services. However, regardless of the specific definition of the customer or the user, he/it will always participate in the service delivery process by shaping the service and the value that is provided from it. For instance, regulators could use text mining analysis on-demand services, from a synchronised network of providers, in order to detect and capture certain market manipulation activities, enhancing their monitoring and detection capabilities.
The enterprise cannot deliver value, but only offer value propositions (FP7)	From FP6 it is clear that companies are not in the position of delivering value, rather they only offer propositions of potential value for which they could charge a price. The notion of value is also subjective, as this is perceived differently by customer, companies, or in general, any entity that participates in or observes the exchange.	This thesis posits that there can be multiple types and definitions of market monitoring and surveillance service providers. However, regardless of the type and definition of the enterprise, it will be limited to offering a value proposition. The provider cannot secure any level of value, as this is always ultimately co-created with the customer.

A service-centred view is inherently customer oriented and relational (FP8)	From FP5 it is possible to argue that all companies are members of a service system. All exchanges are services, and thus companies integrate and communicate with suppliers in reticular ways. The success of the value propositions, in consequence, depends not only on the direct provider of the service, but also on the firm's ability to align its whole network of resources to meet the customer's needs and expectations.	In order to be faithful to their value propositions, service providers of market monitoring and surveillance activities will leverage their own service value networks and resources, but ultimately the value delivered will depend on the customer as stated in FP6. In this sense, the market monitoring and surveillance services are always customer and service-centred, and also relational.
All economic and social actors are resource integrators (FP9)	From FP5 and FP8 it is possible to argue that not only companies, but all social actors and entities, are part of service systems. They co-create value and integrate operand and operant resources at different levels, making use of different value generation processes and networks.	All actors, participants, and entities of the market monitoring and surveillance services are, by definition, resource integrators. This is true for small units of resources, i.e. individual investors, as well as for bigger units, such as regulators, venues, or institutional investors. They all have and leverage operand resources via their operant resources. Entities also interact and communicate continuously with one another, and as such, services are created in the integration of resources of multiple entities and stakeholders.
Value is always uniquely and phenomenologically determined by the beneficiary (FP10)	Although planning for value creation involves both internal and external resources and collaborative relationships and networks, the value perceived is ultimately determined as a unique phenomenon by the beneficiary; perfect prediction and measure of perceived value is simply not possible (Spohrer et al., 2008).	The value of the market monitoring and surveillance services will always be determined by the beneficiary. However, the designation of the beneficiary is manifold, as this comprises all entities and stakeholders. Beneficiaries can represent individual investors, as well as investment firms, regulators, or the general public.

5.1.2 Service Science

According to service science, service is value co-creation, i.e. a useful exchange that results from communications, planning, or other purposeful and knowledge-intensive interactions between distinct entities (Maglio and Spohrer, 2010b). Moreover, service is, takes place in, and is part of, a *system.* This system is *'a configuration of people, technologies, organisations and shared information, able to create and deliver value to providers, users, and other interested entities through service'* (Maglio and Spohrer, 2008).

This thesis posits that market monitoring and surveillance activities are services, and thus these activities co-create value as the result of the interactions and interdependency of different stakeholders, technologies and businesses helping to achieve a common goal. Moreover, market monitoring and surveillance activities are *service systems* that are part of broader systems whose entities interact to co-create value to providers (e.g. SROs), customers (e.g. regulators) and other stakeholders (e.g. individual investors).

In particular, SS proposes that service system entities dynamically configure four types of resources: *i) people; ii) technology; iii) organisations; and iv) shared information*, and that these resources can be further classified into *physical, non-physical*, and *with-* or *no-access rights* categories*.* Figure 5-1 presents a simple classification of resources in the context of this framework.

	Rights	**No-rights**
Physical	1. People Individual investors, manipulators, traders, surveillance teams, broker-dealers, head of SEC or FSA, etc.	5. Technology ECNs, trading platforms, servers, data warehouses, etc.
Non-physical	3. Organizations Regulators, broker-dealer companies, investment companies, stock markets, surveillance units, etc.	4. Shared Information Trading information, news, corporate information, etc.

Figure 5-1 Named Resources as building Blocks of Service Systems

Everything that can be named and is useful can be considered a resource. This includes actors, organisations and stakeholders of all types, as well as tools,

technologies, data, information and natural resources in general. An entity is defined as any resource configuration that can initiate service actions, i.e. 'service system entities' (Barile and Polese, 2010). The basic unit of service can be, for example, a single person, or even, a single algorithmic function encapsulated as a service unit. Intermediate units take into account all possible arrangements of connected services within and/or across organisations, and large units may cover a whole society (Qi et al., 2007).

Access rights are norms and legal rules that resolve access to and the use of resources. As with resources, access rights are categorised into a number of types: a) *owned outright* (e.g. property); b) *lease/contracted* (e.g. data provider services, brokerage services, insurance); c) *shared access* (e.g. trading information, trading platforms) and d) *privileged access* (e.g. personal wisdom, skills, regulator agreements, co-location services).

According to (Spohrer et al., 2008), it is the configuration of access rights associated with entities and resources that shapes the *value propositions* offered by service providers. These propositions are continuously evolving to accommodate customer needs and interventions in value co-creation.

Entities calculate how much each value proposition is worth from multiple perspectives. A value proposition, then, can be viewed as a request from one service system to another; it is evaluated from the perspective of multiple stakeholders and is socio-culturally determined. Four primary stakeholder perspectives can be identified: that of the *customer or user*, *provider*, *authority* and *competitors (substitutes)*.

Moreover, it is possible to identify four types of service-level measures: *quality, productivity, compliance* and *sustainable innovation* (Maglio and Spohrer, 2010a). Figure 5-2 summarises the configuration of the different perspectives, values, and measures in the market monitoring and surveillance services context.

Stakeholder perspective	Measure Impacted	Pricing Decision
I. Customer (User) Individual Investor Investment Firm Local Regulator Cross-jurisdiction regulator	**Quality - Revenue** Latency, integrity	**Value-Based** Value, efficiency, integrity
II. Provider Regulator SROs Commercial Provider	**Productivity - Profit** Cost, latency, integrity	**Cost Plus** Cost, efficiency, integrity
III. Authority Local Regulator Cross-jurisdiction regulator	**Compliance** Market efficiency, market integrity	**Regulated** Cost, efficiency, integrity
IV. Competitor (substitute) Other regulators Other SROs, Non-SROs Other Commercial Provider	**Sustainable Innovation - Market Share** Price, latency, integrity	**Strategic** Value, efficiency, integrity

Figure 5-2 Value Propositions and Multi-Stakeholder Measure of Value

5.1.3 Systems Thinking

It is possible to identify a common denominator between SD logic and SS. Specifically, a central concept is the notion of *service system*, as this considers the interrelation or networked behaviour of several entities in a socio-technical context and the relationships that give them purpose.

In Systems Thinking a system is considered to be *'... a collection of parts together with their relationships that forms a whole that serves a purpose that is meaningful to the system alone, that is, not to its parts or their relationships' (Boardman and Sauser, 2008)*.

In SD logic (Maglio and Spohrer, 2008), a service system is *'a configuration of people, technologies, organisations and shared information...'* -a collection of parts- *'... which are able to create and deliver value to providers, users, and other interested entities through service'*, together with the relationships that give them purpose.

As the notion of system is the common denominator of SD logic and SS, the systems thinking approach can be used as a valid perspective to study and analyse their properties. According to (Boardman and Sauser, 2008), ST can be helpful in two ways: firstly, because it helps us to think *about* the properties of the

systems, i.e., to use our capacities and reasoning skills for cognizing, analysing and synthesising a system's characteristics. Secondly, it helps us to think *from* systems, i.e., to use their intrinsic features to identify patterns and problems, and more importantly, to find solutions to these problems that are formed with and for systemic entities and with systemic concepts.

In the words of (Boardman and Sauser, 2008):

> *'... thinking about systems means making systems the focus of our thinking, and our thinking tool kit provides the lenses that constrain and shape our thinking, which might otherwise be chaotic. Thinking from systems means to use systems, more correctly systemic descriptions of a problematique and any accompanying treatments of this, as the lenses, with the issues—and issuers—as the focus of our thinking'.*

In the following sub-sections, the problematic of financial markets, their properties, and potential solutions will be discussed using both the *about* and *from* systems perspectives.

5.1.4 Thinking *about* market monitoring and surveillance service systems

When thinking *about* systems in general, and *about* market monitoring and surveillance service systems in particular, it is possible to identify many lists of concepts of systems and systems properties. (Boardman and Sauser, 2008), for example, talk about system properties in *trios*, identifying no less than eight trios. (Wilby et al., 2011), follows the 'four fingerprints of complex systems' proposed by (Casti, 1992) to help classify them. This classification considers irreducibility; instability; adaptability and emergence as the relevant categories. Table 5-2 presents this last classification scheme and its application to the '2M-3S' framework

.

Table 5-2 Four Fingerprints of Complex Systems in the 2M-3S Framework

Fingerprint description	Application in the 2M-3S framework
Irreducibility The study of 'wholes', 'holisms' and 'holistic view', are central to the concept of *irreducibility.* It means that the focus of the study of systems science, and system thinking, is the analysis of the properties of a system as one indivisible unit. In other words, ST is not concerned with the study of the parts of the system, i.e. *reductionism,* but rather, to the interesting characteristics of the whole, and the relationships among its parts that give rise to the whole and that cannot be deducted from the study of its parts by separation. However, ST recognises that the selection of boundaries to what determines the whole, i.e., what is in or out of the system, is arbitrary.	**A market monitoring and surveillance service system is irreductable** In terms of service systems, and of this research, the principle of *irreducibility* is adopted when considering the scope of the study, specifically in the definition of what constitutes a market monitoring and surveillance service system unit, and what its boundaries, relationships, components, inputs, outputs, and environment are. This concept is closely related to the allocation of responsibilities in a jurisdiction, and thus to the definitions of what we understand by cross-product, cross-market or cross-boundary trading and manipulation. Also, the principle of irreducibility is applied by subscribing to the idea that it is impossible to fully understand market monitoring and surveillance service systems if one only focuses on their parts, and misses the relationships and interactions of the parts and whole between themselves and with the environment.
Instability Systems are instable, i.e., they constantly change from one state to the next, showing different behaviours. This instability is necessary, as it is a reflex of the capacity of the system to react and adapt to its environment. These principles are sometimes also studied as equilibrium, disturbance, homeostasis, self-regulation, variety and critical situations properties(Wilby et al., 2011).	**A market monitoring and surveillance service system is unstable** Market monitoring and surveillance service systems are, by definition, unstable as they change behaviour, e.g. relationships between its parts, continuously. For instance, the customers of the MMSS are capable of re-organising external and internal processes, i.e. they have the ability to direct system behaviour up to a certain degree. To an external observer this continuous reorganisation is perceived as a change in the system, and thus, an indicator of its relative instability. The principle of instability is also reflected in the changing environment, for example, in the constant innovation that creates new products, new venues and new technologies of trading. This instability is also evident in the regulatory and policy changes that try to maintain the integrity and efficiency of markets.
Adaptability Complex systems incorporate intelligent entities with the ability to take	**A market monitoring and surveillance service system adapts to its environment**

decisions and actions based on their resources. These decisions and consequent actions are usually taken under conditions of uncertainty, and consequently systems are intrinsically tied to probabilistic behaviour and adaptability to ever-changing conditions. This requires flexibility which allows the systems to react and adapt to external conditions, reflecting learning and increasing knowledge of its intelligent entities.	Actors in financial markets and market monitoring and surveillance service systems are intelligent, but not necessarily rational, and as such, their behaviours cannot be anticipated with absolute confidence. This requires the capacity for re-organising external and internal capabilities in order to adapt either to market changes, or to accommodate new functionalities, gain extra efficiency, lower costs, etc. For instance, actors adapt to the introduction of electronic trading by creating new investment strategies, such as the semi-market making behaviour of HFT firms. To an external observer, this continuous reorganisation is perceived as a change in the system and evidence of its adaptability.
Emergence Given the existence of the previous characteristics, it is possible to perceive additional and extraordinary system behaviours that were not predictable (Wilby et al., 2011). According to (Casti, 1992), this occurs because the study of parts in isolation will always fail to acknowledge the synergies that are established in the interaction of the components and processes of the system. These extraordinary characteristics give *emergence* to the system, i.e., define them as unique objects in space-time, with their own particular and peculiar characteristics.	**A market monitoring and surveillance service system has properties that are intrinsic to its whole** In market monitoring and surveillance service systems it is also possible to perceive emergent behaviour. For instance, new ways and strategies of manipulation can indeed occur, with properties and consequences difficult to foresee. Moreover, as will be explained in detail in the following sub-sections, MMSS encompasses several components and sub-components and their relationships, which together make possible the peculiar existence of other emergent properties. For example, the system could have the characteristic of supporting, or not, the detection of insider trading, and this property could not be observed if one only analysed the components in isolation (i.e. it is the combination and coordinated work of different components that make the detection of insider trading possible)

5.1.5 Thinking *from* market monitoring and surveillance service systems

As explained before, the thinking *from* systems approach is especially useful because parts and relationships configure known patterns of systems dynamics that are recurrent in real-life situations. In particular, the ST community refers to stereotypical systemic configurations of entities and relationships as *archetypes*. Archetypes are composed of one or more types of relationships among variables, but in general it is possible to identify two main types of relationships (Senge, 1994). These are:

- Reinforcing relationships, which consider feedback interactions or loops
- Balancing relationships, which consider trade-offs between variables, i.e., inverse relationships between them, for example when one increases the other decreases; when one is activated, the other is deactivated, etc.

In the business and strategy context, all archetypes represent problems that are regularly tackled by organisations, but with solutions that are not definitive. Hence, problems tend to linger, sometimes even causing the collapse of the organisations. However, on the positive side, the *principle of leverage* states that each archetype has a fundamental and definitive *solution* that can be applied (Senge, 1994). These solutions unleash major changes in structure, processes, and attitudes inside organisations. Remarkably, these types of solutions tend to be straightforward and definitive because their significant fixes are 'leveraged', i.e. transmitted, and replicated with increasing force to the whole system.

Typical archetypes include:

- 'Limits to grow'. In this archetype one reinforcing relationship interacts with one balancing relationship. For example, in companies growing rapidly, an increase in sales comes with an increase in net income, which in turn implies more money available to pay incentives to the sales forces, which leads to more sales, starting a cycle that is repeated over and over. This reinforcing relationship, however, puts pressure on other parts of the company, such as the support division or the production line. In order to

cope with the increasing demand, support levels are diminished and quality standards are dropped (balancing relationship).

- 'Shifting the burden'. In this archetype two balancing relationships interact with each other. This is the case in systems that have an underlying problem but only tackle its symptoms. There are three main reasons why this could happen: a) it is easier or cheaper to fix the symptoms than the underlying cause; b) the underlying problem is not evident; or c) because there is a delay between the actions taken to solve the problem and the time it takes to appreciate the results. Regardless of the reason or reasons, system entities tend to 'shift the burden' by doing one of those three things: hiding the problem, putting it aside, or fixing it superficially. Each of these alternatives only contributes to perpetuating the underlying problem, and not to tackling its fundamental cause; an example would be a firm that continuously hires external consultants to help it define the company strategy instead of building strategic thinking and planning capabilities in-house[7].

The application of archetypes represents one of the most interesting arguments of this thesis. In particular, the researcher believes that it is possible, and indeed adequate, to use system archetypes to identify and categorise the current problematic of the financial markets. Moreover, using archetypes gives the opportunity to evaluate the stereotypical solutions associated with them. As mentioned before, these solutions use the *from* perspective of ST approach, and they are consequently presented and described in system terms, using systems terminologies and concepts which facilitate their understanding and dissemination.

The question is, then, to be able of identify and relate the right archetypes to the right problems. From the researcher's point of view, it is possible to find characteristics of the two archetypes described above in the current problematic of financial markets.

---

[7] Other famous archetypes include 'Growth and under-investment'; 'Fixes that Fail; 'Escalation; 'Success to the Successful; and 'The Tragedy of the Commons', from (Senge, 1994)

Firstly, there are several features of a 'Limits to grow' system dynamic, i.e., it is possible to identify a reinforcing relationship interacting with a balancing relationship. Following Senge's recommendation, in order to represent this archetype it is necessary to identify a reinforcing process, i.e., what is getting better and what is the action and condition leading to this improvement. It is also necessary to identify a balancing process, i.e., the trade-off that the positive reinforcement process is triggering.

A 'Limits to grow' archetype is characterised by particular actions, conditions and activities which are called: *growing actions*; *conditions*; *slowing activities;* and *limiting conditions.* Together they configure the existence of the archetype as presented in Figure 5-3. Because they are arranged as an infinitive loop, this dynamic can be initially analysed by taking any of the elements of the archetype as a starting point. Arbitrarily choosing *conditions* as a starting point, the argument of this archetype goes as illustrated in Figure 5-3:

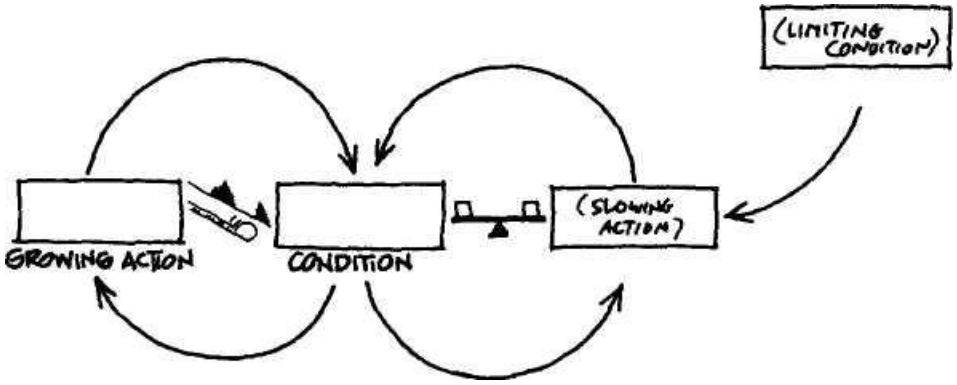


Figure 5-3 Limits to Grow Archetype. Taken from (Senge, 1994)

5.1.6 'Limits to grow' in financial markets

As explained extensively in the introduction chapter, there have been several significant changes in financial markets. In particular, it is possible to identify important new *conditions*, such as better connectivity among markets and among market participants, more competition due to changes in regulation, faster trading due to new technologies, lower transactions costs due to high frequency trading firms' provision of continuous liquidity and other conditions. These have increased

the volume and the number of trades across financial markets, which has in turn fostered the creation of a great number of new products and financial services. Also, greater volume and more products have created the need for more venues, and new exchanges specialising in particular products or offering particular trading standards, such as very low latency capabilities, have been born.

Certainly, this reinforcement process has brought with it an escalating complexity, which has affected market integrity. Increased connectivity has made the creation of dark liquidity venues possible, and the increased sophistication in trading technologies and financial products has brought more manipulations, flash crashes, rogue algorithmic trading, and less transparency with it. In this balancing process, it is possible to identify three limiting conditions: *i) unequal access to technology*; *ii) a lack of openness and excess of secrecy and confidentiality surrounding the market monitoring and surveillance activities*; and *iii) insufficient or obsolete regulation*.

According to (Senge, 1994), in order to find a fundamental solution in this archetype, it is necessary the weakening or removal of the limiting conditions that will have the greater impact with the lesser cost.

In this case, it is possible to think of several alternatives or combinations of solutions, many of which have already been considered by the regulating authorities, such as new circuit breakers; greater restrictions on trade; compulsory registration of all investment actors; and more comprehensive consolidation of trading data.

In the researcher's opinion, the solution with the greatest leverage, however, entails the opportunity to involve new actors in the system. Similarly to the 'open data' initiatives[8], there is the possibility of allowing independent providers to provide market monitoring and surveillance services, which could use and access data feeds coming from venues and exchanges on an 'open source', 'on-demand' and 'on-the-cloud' basis. Currently, although commercial providers are the ones

[8] http://en.wikipedia.org/wiki/Open_data

who in practice develop the monitoring and surveillance platforms, these are almost exclusively mandated by SROs or regulators, which pay considerable sums for these services.

This approach has several disadvantages. For instance, only regulated venues are obliged to perform monitoring and surveillance of trading activities, and little monitoring is done in non-regulated venues. In addition, only regulated venues in developed countries have access to the latest monitoring and surveillance technologies, and small exchanges in third world countries have barely started using modern technologies due to the high investments needed to deploy a system with these characteristics. Moreover, SROs are usually compelled by regulators to monitor only equity exchanges, and little, or no, efforts are dedicated to monitoring other types of instruments, such as bond markets, money markets or derivative markets.

Clearly, a simpler way to increase both the quality and the coverage of market monitoring and surveillance activities, for all types of venues, countries and instruments, across-markets and across-products, it is to facilitate a culture of ‘many-to-many’, open and transparent surveillance and monitoring services. This does not mean that regulators will lose or weaken their faculties, but only that all users or customers could complement the efforts of the authorities by having the opportunity to access a market monitoring and surveillance service when needed or considered more important given certain market conditions.

For example, an individual investor could pay on a query basis to analyse the likelihood of a given stock being manipulated in a particular market, regardless of the type of the venue, i.e. whether this is a regulated, non-regulated, private or 'dark' venue. Moreover, customers could customise the characteristics of the service, for instance, by choosing the period, the quantity and quality of the data, and the type of data and text mining analysers used in the analysis. Furthermore, regulators could also use these services, and could, for example, request an analysis of cross-market manipulation for different venues, or even run a daily report on the quality, type and number of manipulations being detected in a set of markets, in order to compare their efficiency and integrity characteristics. Multiple

providers of market monitoring and surveillance services will also foster competition among them, and it is expected that, following the typical business models of web-based services, these services could have affordable prices, or even be free of charge for individual customers. For example, providers could make a profit on alternative or complimentary services, such as financial advice, brokerage or even pay-per-click advertising. Figure 5-4 presents the instantiation of this archetype in financial markets.

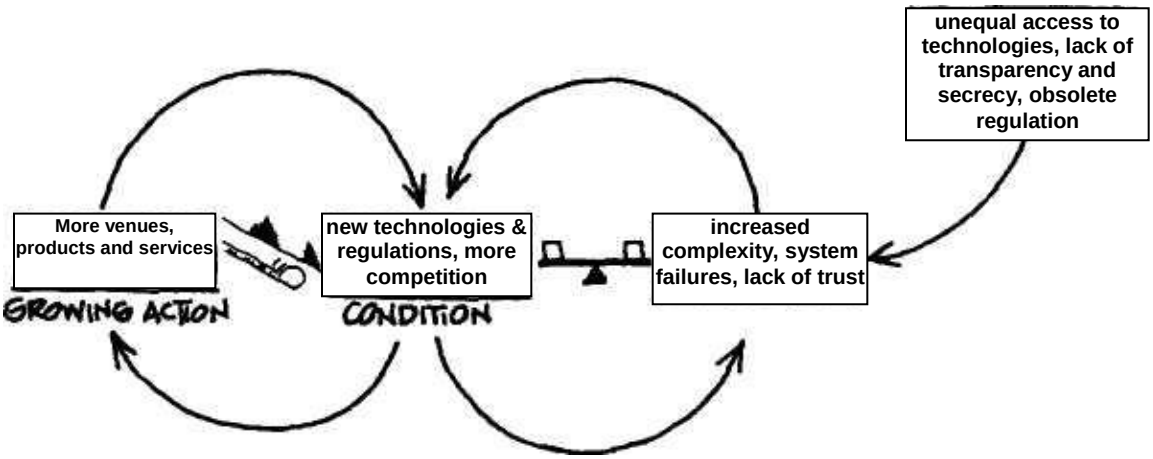


Figure 5-4 Limits to Grow in Financial Markets

5.1.7 'Shifting the burden' in financial markets

From the researcher's point of view, the current situation in the financial markets also posses several features of this archetype (Figure 5-5). For instance, the authorities have focused on enhancing the circuit breakers as a way to stop trading quickly when markets become volatile. This does not tackle the fundamental problem, as collapses can originate everywhere, not only in the regulated markets in which the circuit breaker programme is implemented. Moreover, superficial solutions like the circuit breaker programmes produce several unexpected and critical side effects, ranging from a false sense of security to world famine.

In particular, trading halts and other kinds of circuit breakers arguably contribute to a false sense of security, because their presence can be interpreted by market participants as a definitive solution and the end of market crashes, which in the midterm attracts trading to venues which are still at risk of being manipulated or

collapsing at any moment. In fact, and despite the pilot programmes started by the US authorities, there have been more than thirty mini-flash crashes after the one which occurred on 6th May 2010 (Nanex, 2011).

In addition, other solutions proposed both by the US and the EEA regulators, such as the forced consolidation of pre- and post-trading data, are experiencing a significant delay in implementation and actual results could take years to materialise. Furthermore, the solutions once again prioritise equity markets, and fail to deal with non equity markets and venues which are as or, more important, than the former, for example the derivative markets.

In the derivative markets, for instance, the lack of transparency is of particular importance, especially considering the structural changes to the futures market in recent years. In 1999, banks were permitted by the US Commodities Futures Trading Commission to hold large positions in commodities securities, and in 2004 the US SEC further expanded the bank's options by allowing them to hold positions for as much as forty times the capital they held in collateral. This resulted in an increase in the volume of index fund speculation (a 2,300% increase between 2003 and 2008), which in turn has fostered a worrying distortion in food market prices and agricultural commodities in general (Knaup et al., 2011). According to the FAO's Food Price Index (FAO, 2011), food costs rose by 39% and grain prices rose by 71%, as did prices for cooking oil and fat during that period. According to the World Bank, since June 2011 higher food prices have driven another 44 million people below the poverty line, and more than a billion people are starving worldwide (The World Bank, 2011), caused in no minor part by index fund speculation.

As can be seen, symptomatic solutions not only fail to solve the fundamental problems, but in fact can create side effects and contribute to even bigger problems, such as world famine, thereby creating conditions in which new crises will inevitably impact global markets time and again.

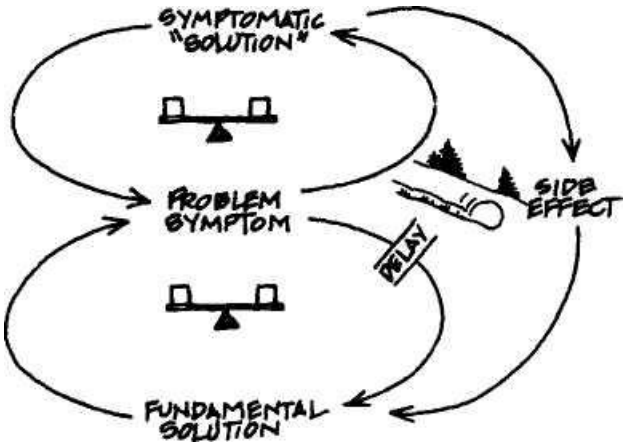


Figure 5-5 'Shifting the Burden' Archetype. Taken from (Senge, 1994)

As presented in the previous archetype, one possible leveraged fundamental solution for this problem is the possibility of allowing independent providers to offer market monitoring and surveillance services. These actors, for example, would not need to wait until a single formal consolidation data tape is created, and could instead create their own mechanisms of consolidation by acquiring the necessary data directly from the venues, regardless of their regulating status. Also, better consolidation requires not only access to the necessary data, but also infrastructure and knowledge of information management, which could be easily found across several IT communities. For instance, consolidation services could be provided 'on-the-cloud', and market monitoring and surveillance providers could fetch the necessary data on an 'on-demand' basis, driven by the customer's needs. Figure 5-6 presents an instantiation of the archetype as applied to modern financial markets.

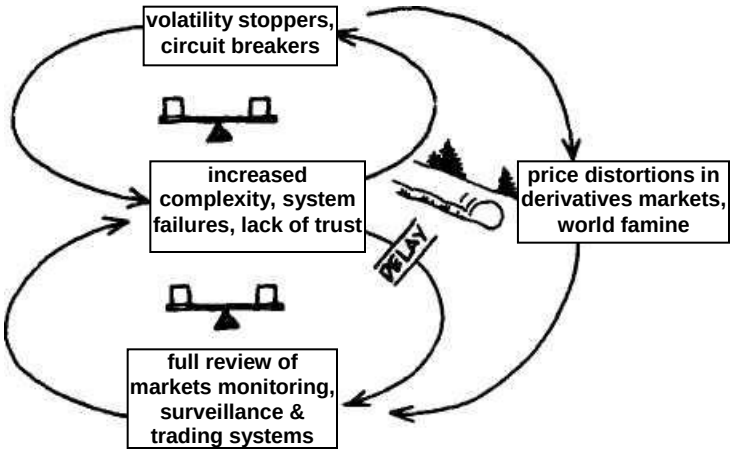


Figure 5-6 'Shifting the Burden' in Financial Markets

5.1.8 Summary

In this sub-section the principles of SD logic, SS and ST have been introduced. Drawing on SD logic, this thesis posits that market monitoring and surveillance activities need to be considered as service systems. Moreover, all exchanges in the modern economy are to be understood as service exchanges, and thus we are immersed in a service economy. From the SS perspective, this section has also described the main characteristics of a market monitoring and surveillance service system, defining the concepts of value co-creation, value proposition, resources, access rights, entities and their different perspectives on how value can be perceived and measured.

Finally, the ST perspective has been used to think *about* service systems, describing their properties using the 'four fingerprints of service characteristics'. The ST perspective has also been used to think *from* services, by using archetypes to describe the current problematic of financial markets, and to propose a leveraged solution. The following section will continue with the development of the framework by giving details about how the *service value networks (Kwan and Min, 2008)*, and *'customer-driven value co-creation service value networks'* (Kwan and Yuan, 2011) concepts are used to build the proposed framework.

5.2 Service Value Networks and Customer Driven Value Co-creation

5.2.1 Basic concepts

The previous section has established why and how market monitoring and surveillance activities can be considered service systems. The aim of this section is to provide further details on exactly how these activities configure and provide value to customers and other entities in the system. In particular, according to (Kwan and Min, 2008), SD logic stresses the generic service provider-customer interaction, and SS similarly emphasises an static value proposition, for which the provider's role is central. In contrast, the 'Service Value Networks' perspective attempts to step into the details of the value process, providing a more symmetric view of the provider-customer service interaction. In this perspective, the customer is explicitly recognised as a co-creator of value, and thus the value proposition is dynamically enhanced through the creation and choice of value networks by both the provider and the customer (Kwan and Min, 2008, Kwan and Yuan, 2011).

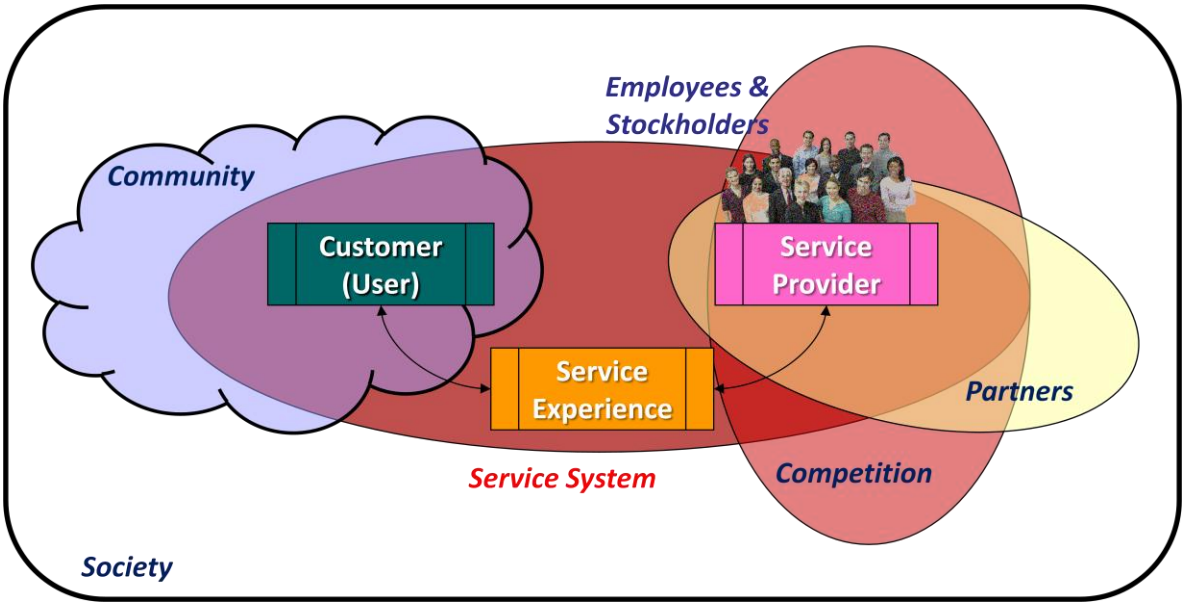


Figure 5-7 A Service System and its Entities, Stakeholders and Worldview. Taken from (Kwan and Min, 2008)

In order to describe how this dynamic value co-creation occurs, let us start by defining the components and participants of the system, and its relationship with the environment. Figure 5-7 shows the worldview of a service system and its stakeholders. In particular, customers or users represent individuals or institutions which embody entities that essentially enjoy the service experience. In contrast, service providers are the fundamental purveyor of the service experience to the customers. Both are members of a society with rules and regulations that define *access rights*. Not shown, but also present, are entrepreneurs and criminals (Kwan and Min, 2008).

In particular, service experiences are intangible and provided (usually) for a finite period of time. Although it is common to pay for a service experience, this does not mean that money is exchanged along with every provision. Regardless of the time and fashion of payment(s), expenses are incurred as resources typically have opportunity costs.

In order to organise service experiences, customers and providers agree on formal or informal *value propositions*, which are presented initially by the service providers. These normally include the expected price, quality and extent of the service experience. Value propositions can also take place between a regulator and a regulated entity, i.e. services can be provided on an obligatory basis, with the quality, time and fashion that the regulating authority decides. This sets up a special case in which the price paid by the authorities can be as low as zero, although it is expected that it should cover at least the opportunity costs of the resources used in the provision of the service (though this is not necessarily the case).

In successful service experiences, the delivery of what was agreed in the value proposition is the basis of *value co-creation*, hopefully resulting in a surplus at the end of the service experience. This surplus can take the shape of a monetary remuneration, a regulatory compliance, or more generally, any improvement in the level of satisfaction or happiness of the parts.

In terms of service production and in order to fulfil the value propositions, service providers configure their resources, i.e. people, technologies and shared

information, through interaction with partners, which might or might not be competitors. As in Figure 5-8, a service provider can configure complex service supply chains, namely Service Value Networks (SVNs). For example, a customer of service system A enjoys a service experience which involves the coordinated action of several interlocked chains of services.

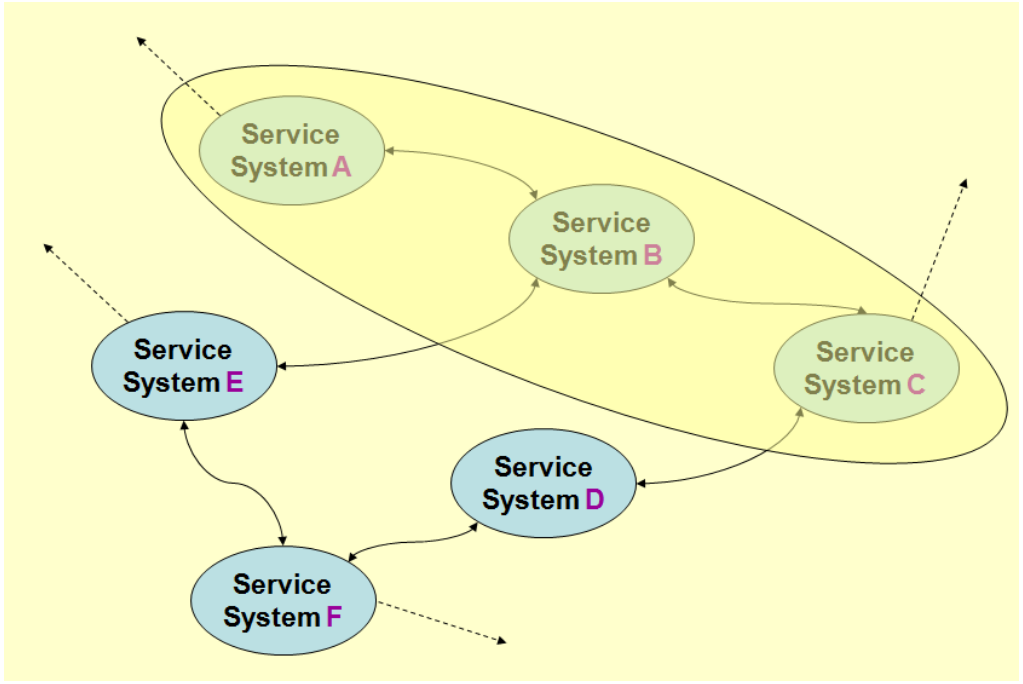


Figure 5-8 Service Supply Chain

Remarkably, customers configure and extend the SVNs shown in Figure 5-8 in order to improve their service experience. Drawing a logical connection between the last two figures, it is possible to assert that the customers' configuration and extension of SVNs are also implicitly present in the community in Figure 5-7 (represented as a cloud in Figure 5-7). Specifically, the community allows customers to exert their own private and public resources every time they dynamically help to create the service configuration they need or want. For instance, a customer of a given service could use a social networking platform like Facebook or Twitter to ask her friends and family their opinion about a service, to receive comments that would help her to decide whether to engage in service delivery with a certain provider, or to ask for the best ways in which she should interact with it.

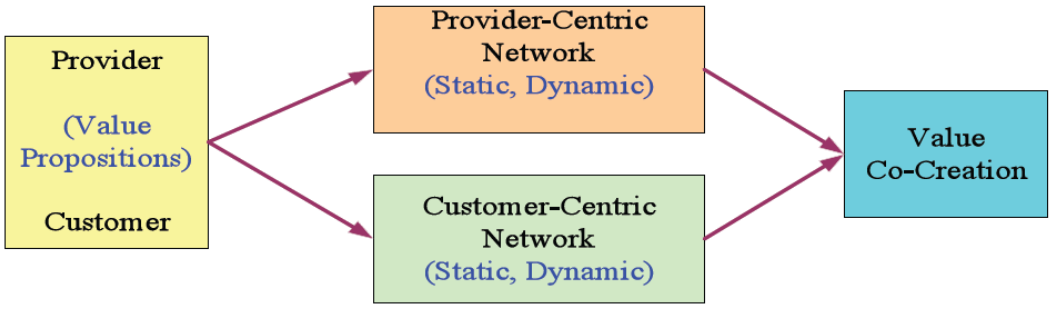


Figure 5-9 SVN Dichotomy and Determinants of Value Co-creation. Taken from (Kwan and Yuan, 2011)

Considering the important role that customers play in value co-creation, then, it is necessary to identify two facets on the SVN concept. If the provider perspective is being observed, then the service value network becomes a Provider-Centric Value Network. In contrast, if the customer perspective is being observed, then it becomes a Customer-Centric Value Network (Figure 5-9). Following SS classification, both the *customer* and the *authority* perspectives represent two particular instantiations of a generic definition of customer or user, and thus of its provider-centric networks. Similarly, *competitors and substitutes* are a particular instantiation of a generic service provider, and thus of its provider-centric networks. In this sense, it is the observer of the system who arbitrarily focuses on one class or the other, depending on the angle being used to analyse or describe the service (irreducibility).

Following this dichotomy, central to the concept of value are the choices that both customers and service providers make to compose their value networks. One main difference in this process is that customers have more options, or extra degrees of freedom to perform this task, as they not only decide on their resources to compose the networks, but also they influence the way providers configure theirs.

In order to explain this concept, it is important to understand that customers have the ability to recreate a given service experience by dynamically combining different service modules as they need them. For instance, customers could obtain the same final service, say the monitoring and surveillance of a given stock, by either entrusting the authority to do it by representation or by hiring a network of

service providers with similar pooled capabilities. For example, the networked service provider could provide service modules for data provision, data pre-processing, data mining analysis, and automatic reporting services.

Four principles help us summarise the determinants of value co-creation as the result of the interaction of value propositions and the two different perspectives on the service value networks (Kwan and Yuan, 2011):

- Firstly, a network is always the primary economic mechanism for value conversion and for describing the value creation dynamics.
- Secondly, the value propositions offered by a provider-centric network may not sufficiently fulfil the needs of the customers in terms of their own value determinations.
- Thirdly, customers might need to expend extra effort to fill this gap, for example with additional self-service or by creating new customer choices.
- Fourthly, this self-service or creation of new customer choices could range from a simple labour effort to driving another network (e.g. a customer-centric social network) to engage in dynamic value co-creation to fill the gap.

To summarise, the combined network could deliver enhanced customer value, provided that any additional service costs do not offset the value gained (Equation 1). Interestingly, the provider's value could also increase concomitantly because of the expected growth in business and service volume (e.g. due to increased customer retention).

Equation 1. Value Formation in Value Co-creation (Kwan and Yuan, 2011)

$$Value = \frac{UserChoices + Results + (self)ServiceExperience}{OverallCost}$$

5.2.2 Variations of Value Networks and Value Propositions

Following (Kwan and Yuan, 2011) there are four ways in which customers can be involved in value co-creation. These configure four value co-creation scenarios with correspondent metrics and customer choices (see Figure 5-10).

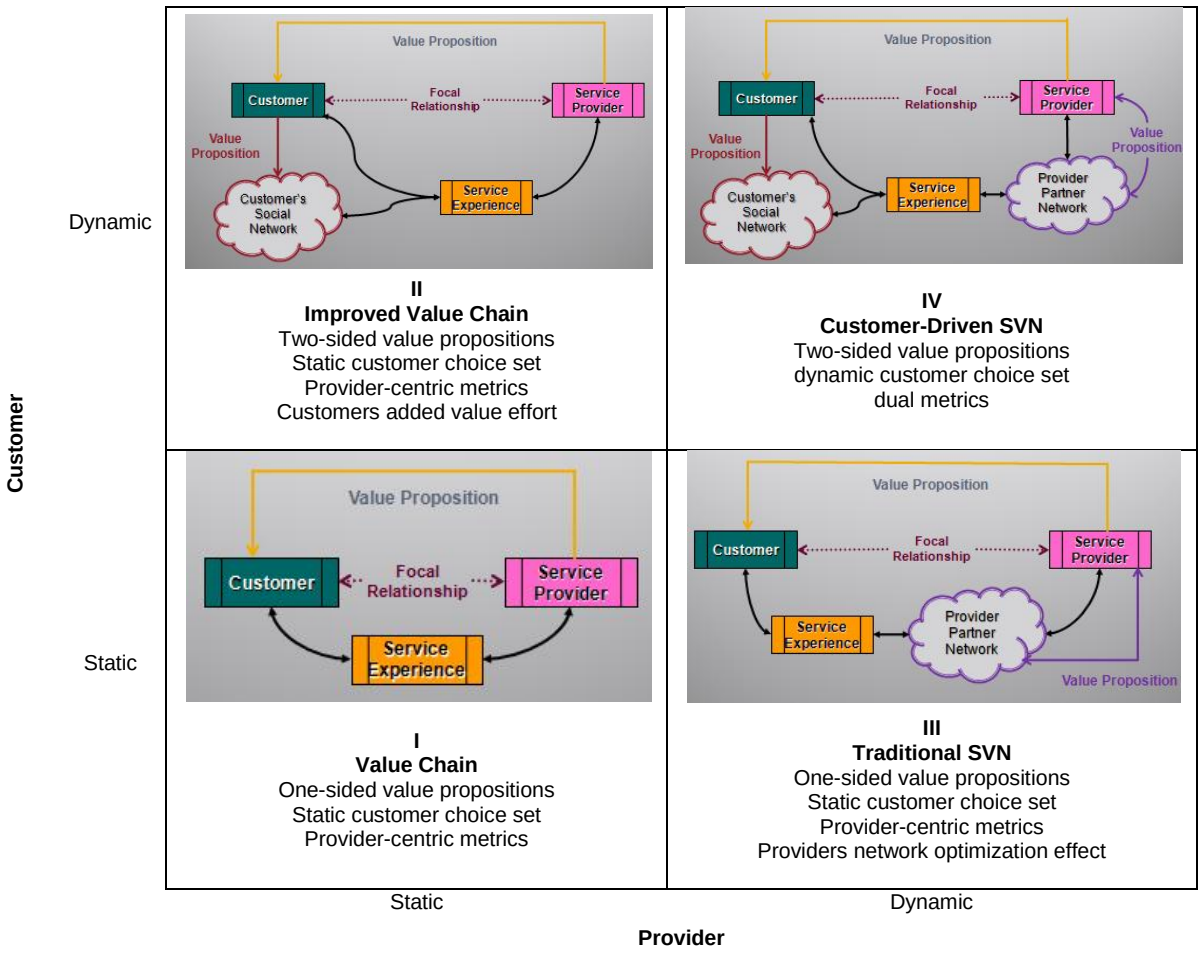


Figure 5-10 Value Co-creation Variation Scenarios and Variants of SVNs. Adapted from (Kwan and Yuan, 2011)

In detail, these can be:

- *Provider Static*, in which the service network is fixed by the provider and the end points of the networks are known.
- *Customer/User Static,* in which the customer does not have any way of altering the end points of the network; they are limited to select for the predetermined end points.
- *Provider Dynamic,* in which the service network is created dynamically by the provider, for example by using a cost optimization algorithm, e.g. reducing the total carbon footprint of the delivery, or minimising the fee of delivery.
- *Customer/User Dynamic,* in which the end points of the network are not known ahead of time. The customer creates these end points to

maximise their value from the services, i.e., the value proposition from the provider only gives some pre-determined value accepted by the customer, who is then empowered to enhance the value based on their dynamic creation of the service network end points.

Accordingly, each quadrant or scenario in Figure 5-10 configures a specific service value network, namely, *value chain* (I quadrant); *improved value* (II quadrant); *traditional SVN* (III quadrant); and *customer-driven SVN* (IV quadrant).

5.3 '2M-3S' Framework' Worldview

The application of the concepts and theoretical foundations discussed in sections 5.1 and 5.2 provide us the basis to develop the underpinnings of the proposed '2M-3S' framework. In particular, Figure 5-11 presents the first step in this construction, i.e. the developing of the systems' worldviews that are proposed in this thesis. It is possible to recognise two main types of customers of monitoring and surveillance service activities, i.e. regulators and investors. It is also possible to identify two types of service providers. The first group corresponds to the 'traditional' service provider, i.e., providers which are typically entitled to develop, deploy and perform market monitoring and surveillance activities. These include regulators and SROs, and in particular cases, non-SROs when these are performing particular surveillance activities, such as surveillance of an insider trading case, or as part of bigger regulatory investigations that involve multiple venues. The second group represents the type of service provider proposed by this thesis, i.e. independent and commercial service providers, either with profit or non profit profiles.

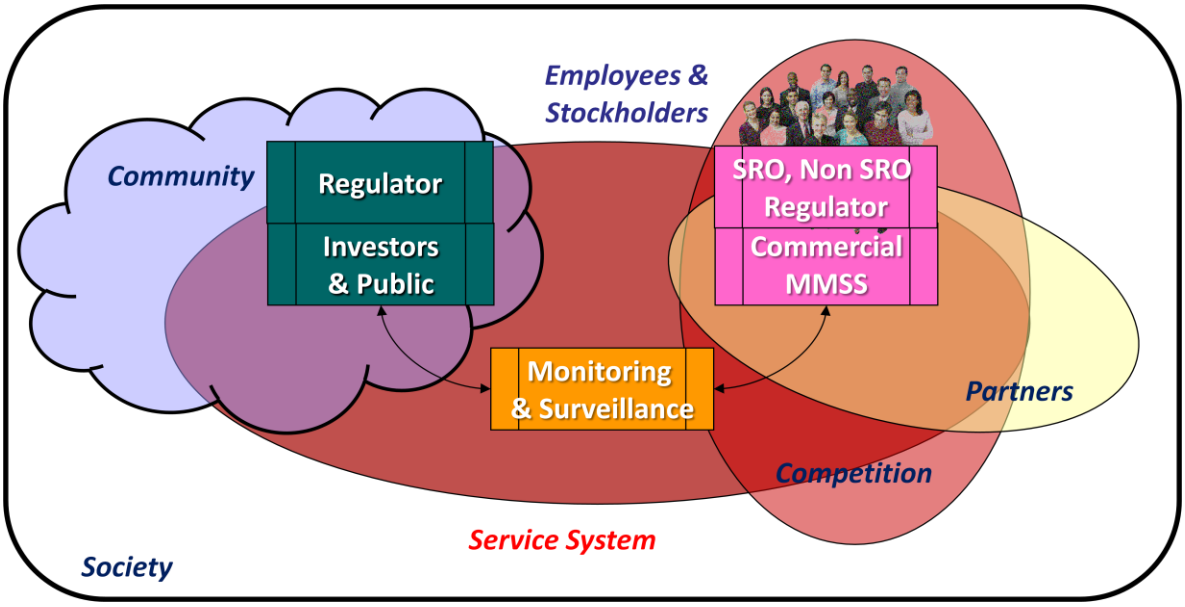


Figure 5-11 '2M-3S' Framework' Worldview

As in the generic case, customers and providers are part of a larger system in the '2M-3S' framework. In fact, customers are part of a community which encompasses other regulators and governmental institutions, inside and outside regional boundaries, other traders, venues, investors, financial institutions, firms and the general public, as well as their social networks. In terms of service providers, a similar set of entities can be identified, including SROs, non-SROs, local regulators, regional regulators, commercial and independent providers of market monitoring and surveillance systems and others.

Providers, of course, have partners to help them purvey the services, including a wide range of modular services, such as data providers, networking providers, trading platform providers, and any others that could contribute to the value co-creation process. Competition is also present in this worldview, especially when considering the new configuration of commercial market monitoring and surveillance services and recognising the customer's ability to choose and exert different service value networks, as well as to use and change services providers with absolute freedom. Not explicitly seen, but indeed very present, are market manipulators, fraudsters and criminals in general.

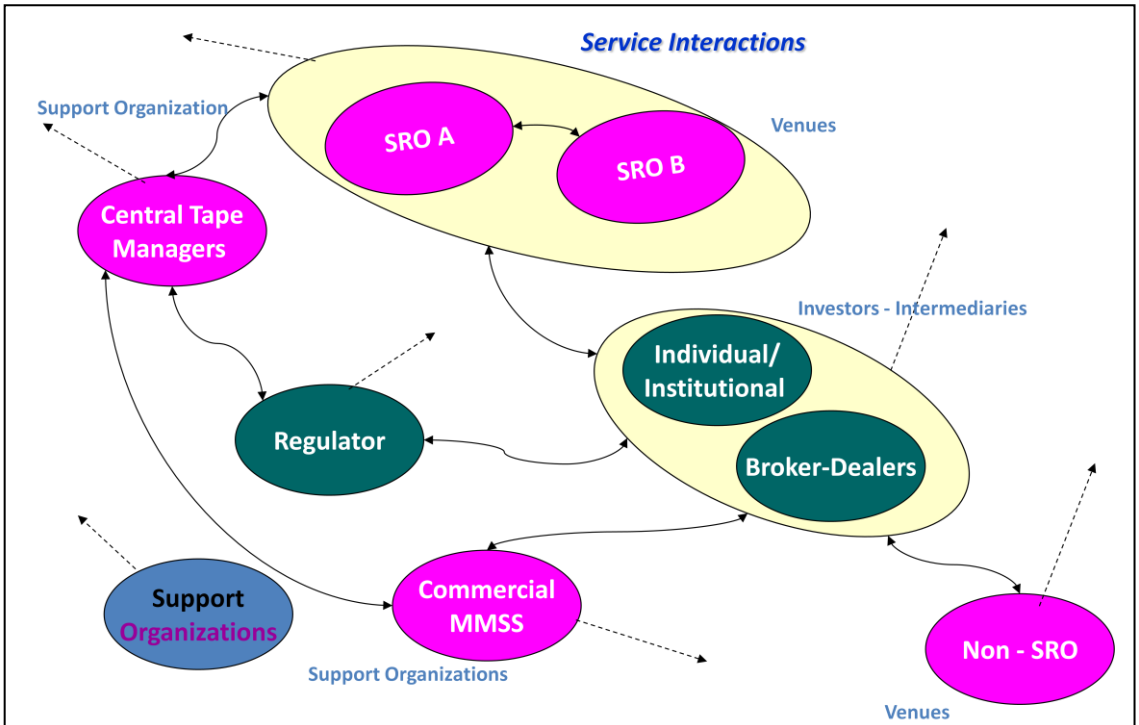


Figure 5-12 Financial Services Systems Worldview

Figure 5-12 presents an alternative worldview, in which typical SVNs are configured inside a single jurisdiction. Following the colour coding of Figure 5-11, green circles represent customers of the market monitoring and surveillance services; purple circles represent typical providers of these services. Continuous arrows represent repetitive interactions among systems, and dotted arrows represent optional and generic connections that can be established if required to increase the adaptability of the entities in the system. Figure 5-12, however, could better demonstrate other kinds of services that are provided in this context, specifically, general financial and trading services. For instance, the figure illustrates how the '*investors and intermediaries'* system, which includes *individuals* and *broker-dealers*, interacts with different types of *venue* systems, such as *SROs* and *non-SROs*. These, in turn, also interact with the *regulator* and with other '*support organisations'* systems. The latter represent different types of supplementary services, such as the *central tape manager* system or the *independent or commercial market monitoring and surveillance* system, presented in purple in the instantiated type, and in blue, for the general case.

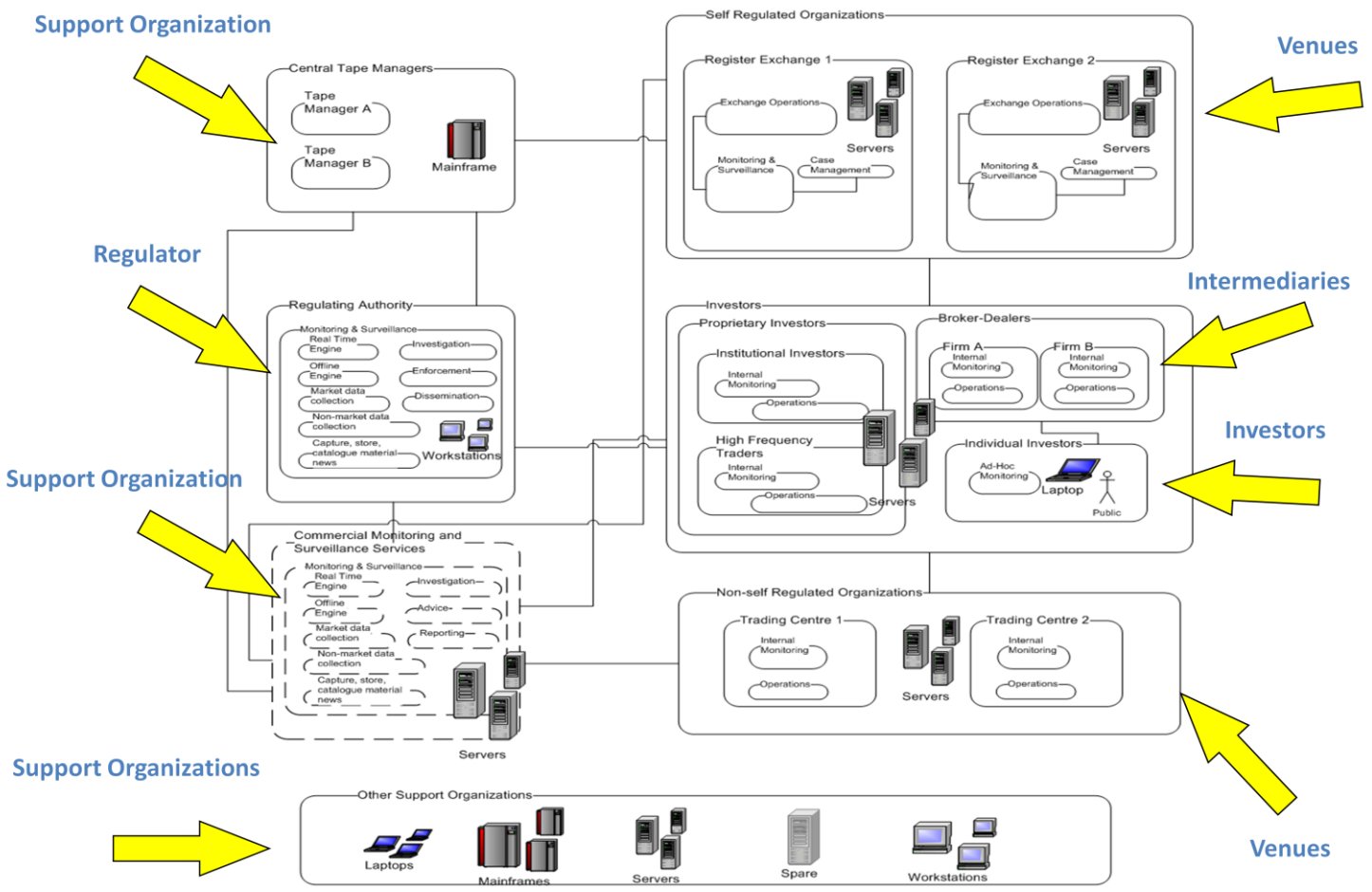


Figure 5-13 Financial Markets Organisation – Single Jurisdiction

In more detail, Figure 5-13 presents the configuration of financial markets following the systemic arrangement for one single jurisdiction, but depicting additional information about its sub-systems and components, this is, Figure 5-13 can be understood as an expanded and alternative version of Figure 5-12. For example, inside the Investors system it is possible to distinguish three types of sub-systems and their components, namely Institutional Investors, Broker-Dealers, High Frequency Trading Firms and Individual Investors.

In Figure 5-14, in addition, multiple jurisdictions are presented. This view explicitly recognises the interactions among markets in order to detect and capture different kinds of cross-market, cross-border and cross-jurisdiction manipulation. At the risk of being prescriptive, this figure shows one tentative organisation and possible interconnection between markets. It shows how two different jurisdictions communicate with one another using their *local central tape managers* as hubs that centralise and transmit all pre- and post-trading data to a *cross-jurisdiction central tape manager*. This information is passed on and used as near to real-time as possible by *cross-jurisdiction regulating authorities* and by *cross-jurisdiction* or *meta-level commercial market monitoring and surveillance services* providers. Both figures also include a set of generic and optional support organisations, providing all kinds of supplementary or complimentary services.

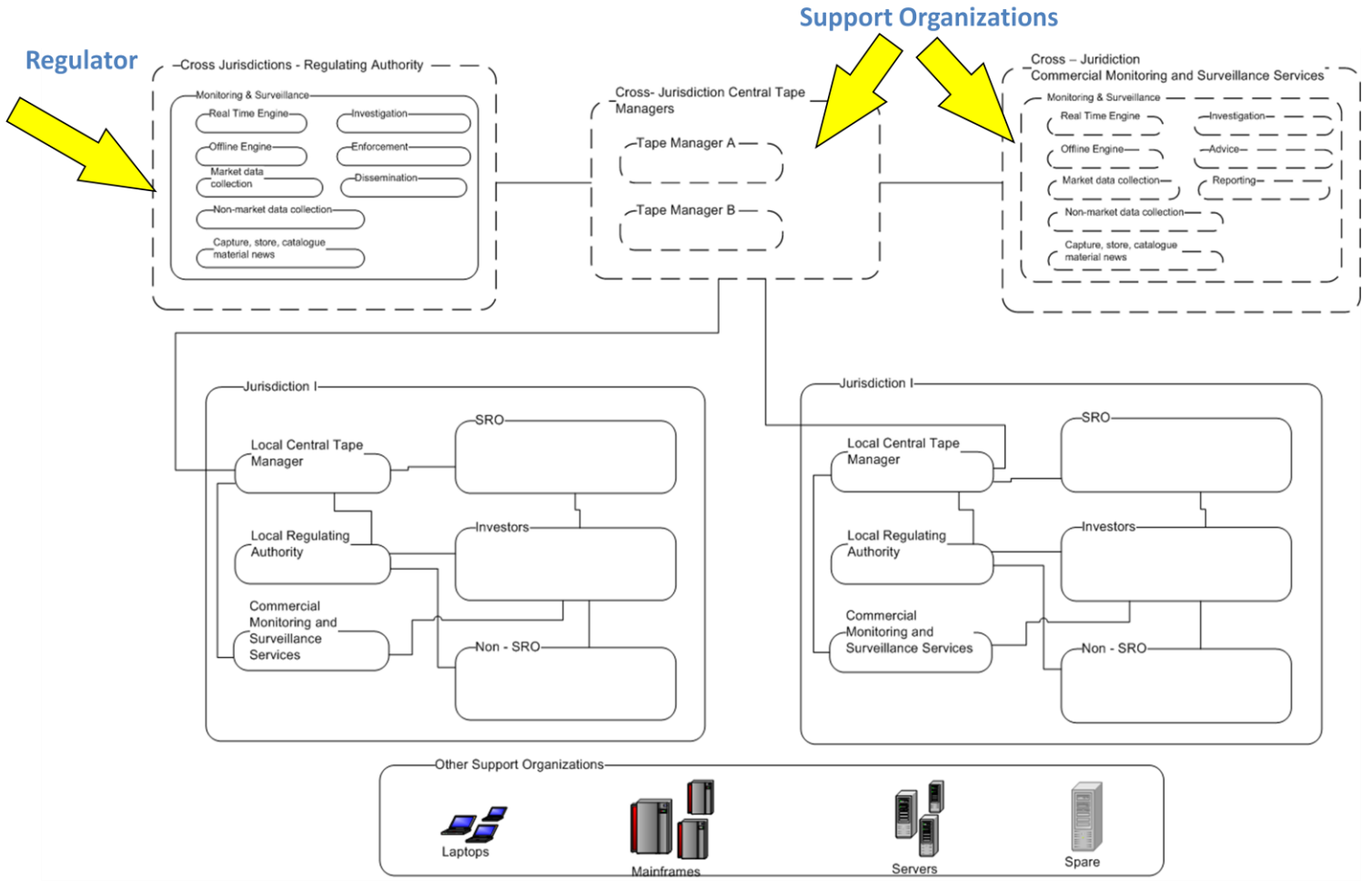


Figure 5-14 Financial Markets Organisation – Multiple Jurisdiction

5.4 '2M-3S' Service Value Networks and Value Propositions

The next step in the development of the '2M-3S' framework is the construction of its value propositions and SVNs. Figure 5-15 presents the generic view on the SVNs and value propositions. The coloured arrows illustrate how alternative value propositions are presented to different types of customers. To be specific, the blue arrow symbolises the presentation of a value proposition to customers by providers of market monitoring and surveillance services. Green and purple arrows additionally represent how customers and providers exert or present their own value propositions respectively to their networks.

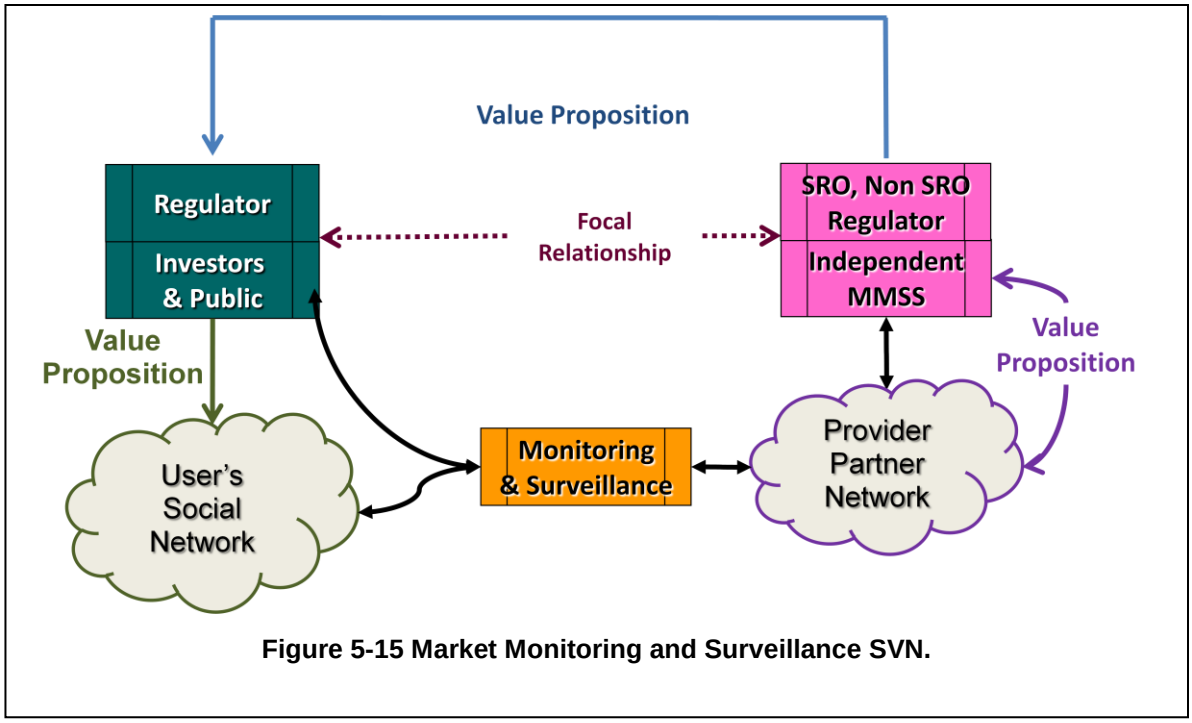


Figure 5-15 Market Monitoring and Surveillance SVN.

In this figure, all service interactions are represented as customer (user) driven value networks, i.e. IV quadrant type in Figure 5-10. This configuration explicitly recognises the absolute freedom of all entities to configure and exert their own, and their providers', networks in a dynamic fashion. Two examples are presented to demonstrate the utility of the framework in the context of the thesis.

In the first example, Figure 5-16 and Figure 5-17 show a 'closed' single jurisdiction and simple SVN configuration, taking into account that a regulator is the entity which demands a market monitoring or surveillance service from an also typical

provider such as an SRO. In particular, the value proposition considers the free provision of real and off-line monitoring and surveillance with a prearranged level of effectiveness and coverage for the manipulations listed on Table 2-1, as well as for particular trading rules established by the trading venue.

Using the provider-centric perspective, the SRO employs a combination of internal and external service providers, including a full providers' and partners' network, which for instance, demands data provision services from a news provider (Factiva), and automatic content analysis from a generic provider of text mining services. Internally, the SRO also uses its own real-time and off-line detection engines, together with the textual sources, to perform monitoring and surveillance activities, the results of which are reported to the regulator.

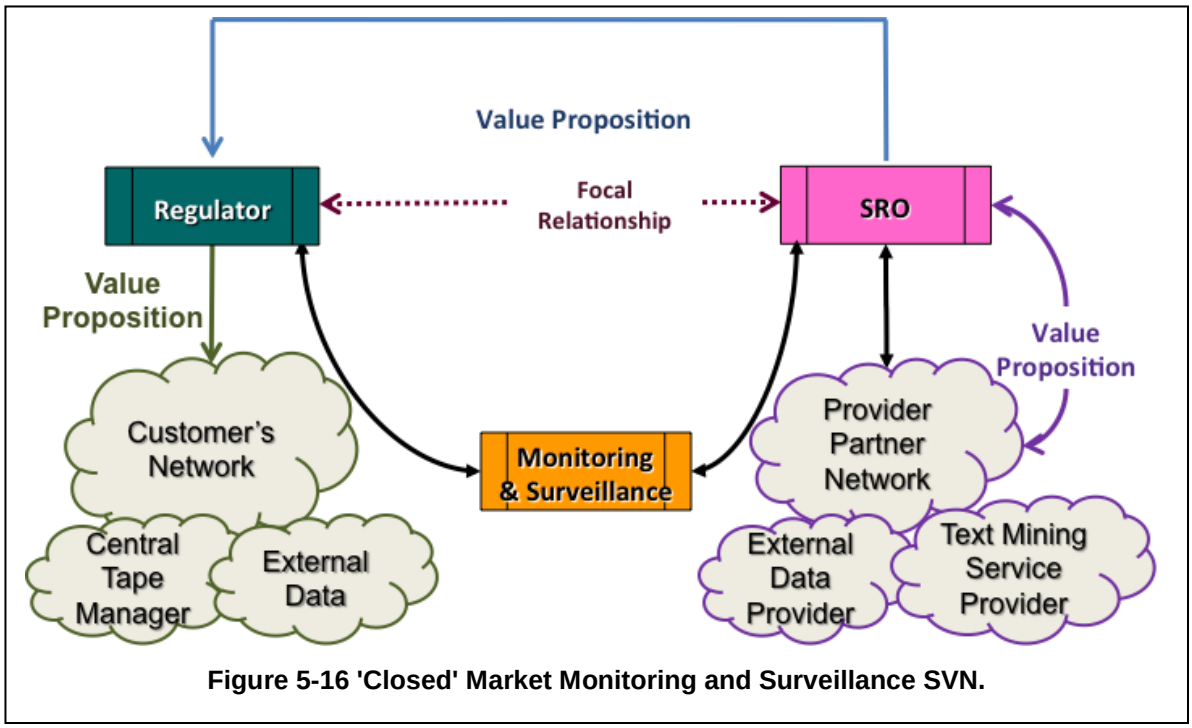


Figure 5-16 'Closed' Market Monitoring and Surveillance SVN.

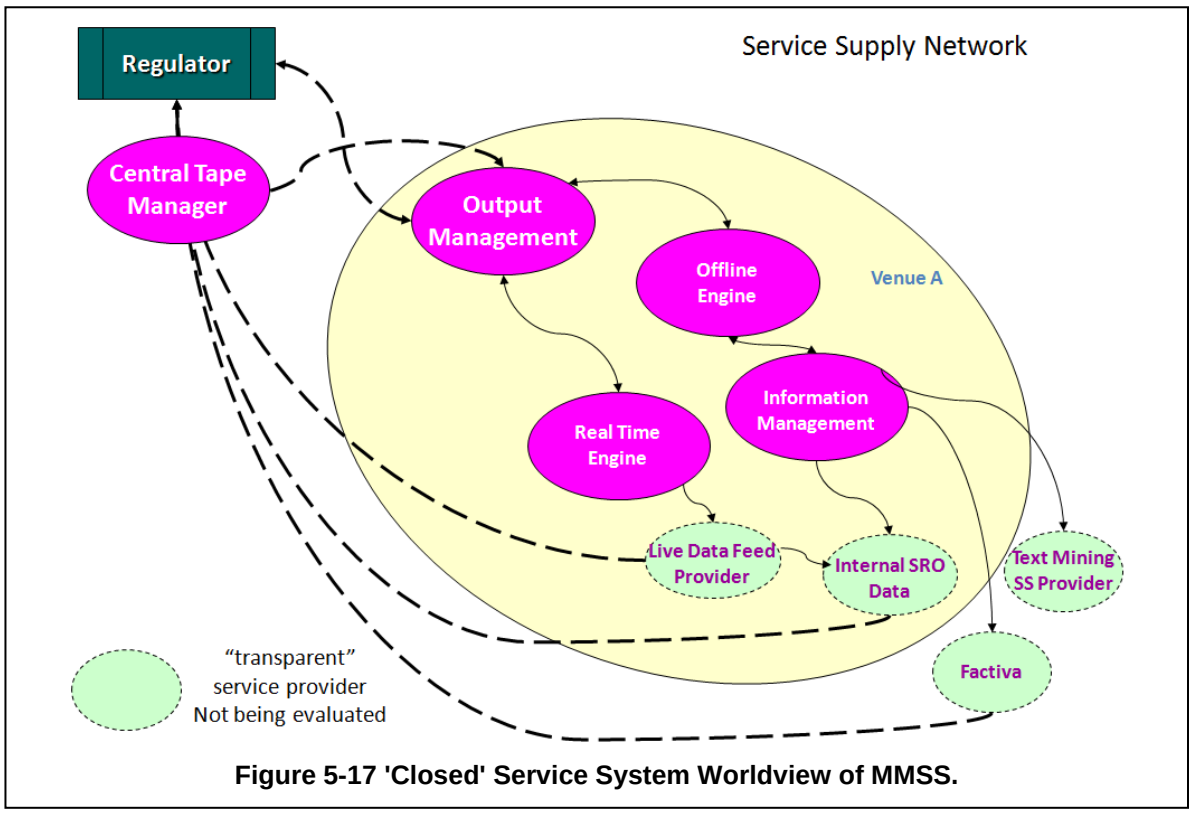


Figure 5-17 'Closed' Service System Worldview of MMSS.

From the customer-centric point of view, the regulator's network exerts its internal capabilities of investigation, enforcement and dissemination to direct an insider trading investigation using data provided by the central tape manager (the entity in charge of integrating pre- and post-trade data coming from the different venues in the jurisdiction).

In the second example, Figure 5-18 and Figure 5-19 show how an individual investor demands a surveillance service from an independent commercial provider after finding suspicious posts on the 'Business & Finance Message Board' of the Yahoo! Finance website. While sending a trading order to its broker service provider, the investor begins to suspect that one or more manipulators are touting a pink sheet stock which is part of her portfolio. She decides to hold the execution of the transaction, and instead she initiates a service request for monitoring and surveillance services. Again, only one single jurisdiction is assumed. The value proposition, specifically, is the exclusive provision of 'runs and raids' detection analysis, excluding other options, such as insider trading detection, which would configure additional services.

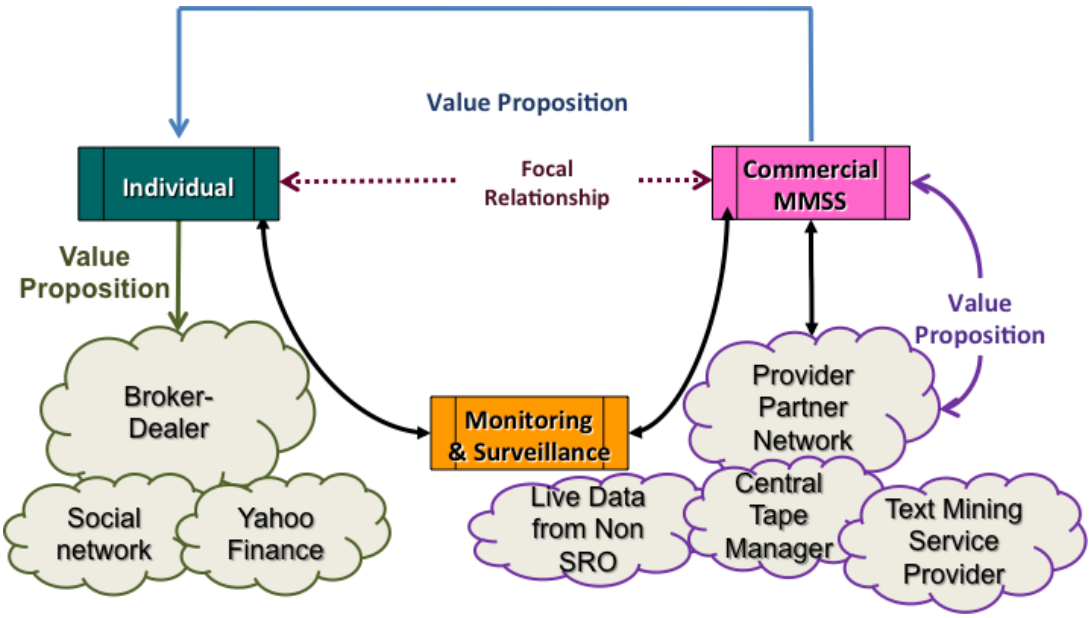


Figure 5-18 'Open Stakeholder' Market Monitoring and Surveillance SVN

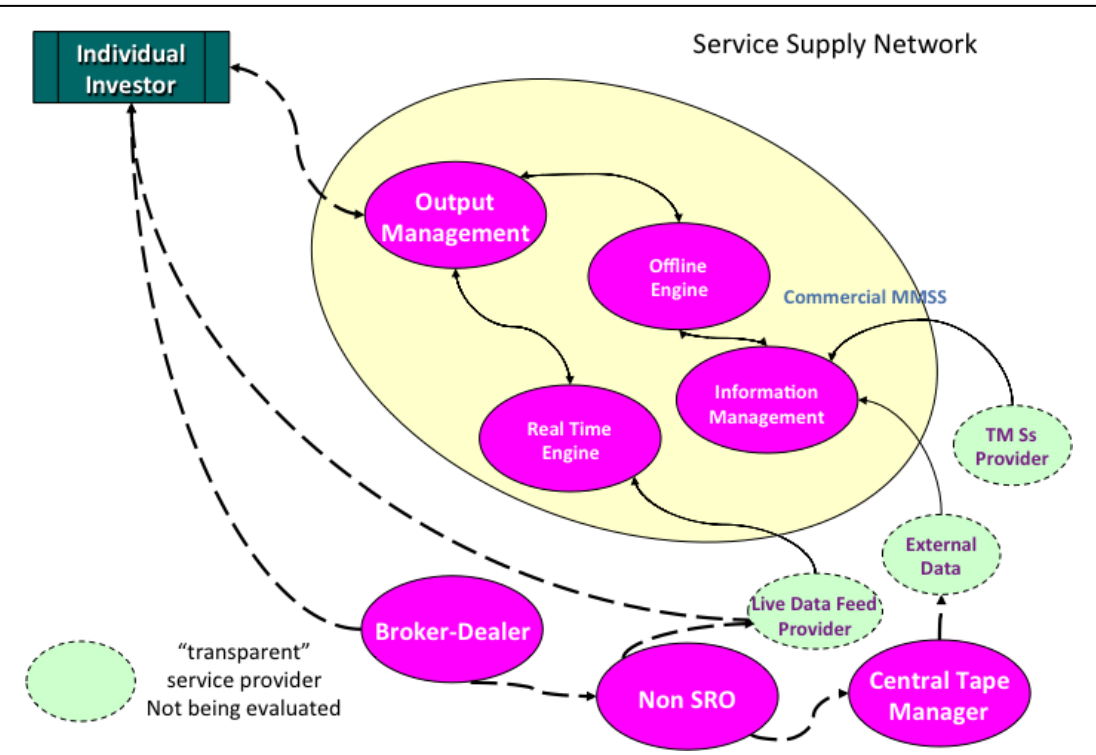


Figure 5-19 'Open Stakeholder' Service System Worldview of MMSS.

Unlike the previous example, Figure 5-19 shows how in this case the independent or 'open' commercial provider contains external live data feed providers, external

textual sources and text mining service providers in its partner network, and thus it depends more intensively on other partners' networks to deliver the service.

5.5 '2M-3S' Framework' Detection Engine

The third step in the construction of the '2M-3S' framework is the development of the sub-systems and components used by service providers to actually deliver service, namely the Detection Engine systems. Figure 5-20 shows a model detection engine, representing my view on the sub-systems, components, tasks and flows of information of a complete financial market monitoring and surveillance system, and taking into account the requirements previously discussed and summarised in Chapter 2.

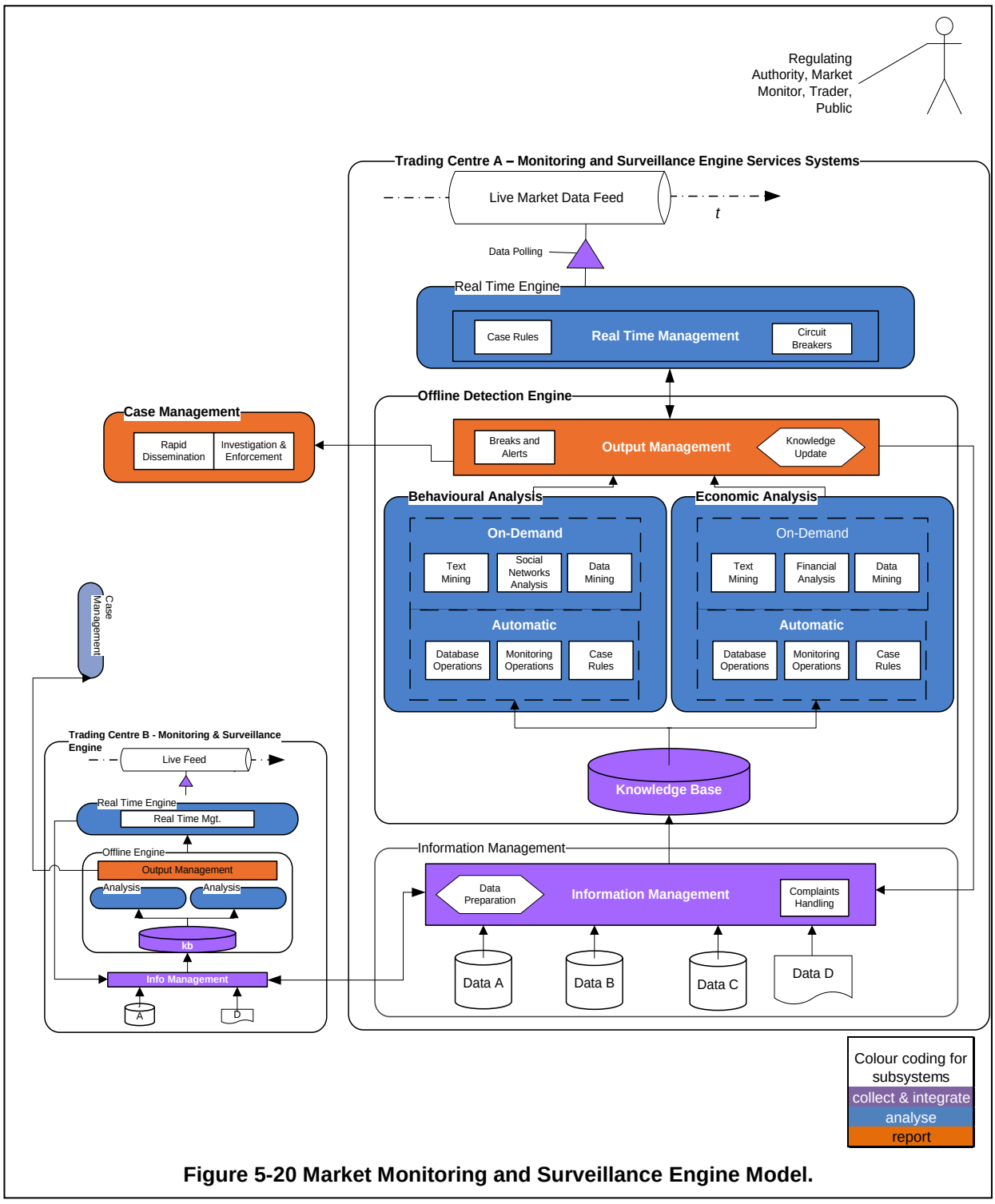


Figure 5-20 Market Monitoring and Surveillance Engine Model.

The wide range of elements and concepts that potentially define a market manipulation present a real challenge for authorities and users in general when designing and specifying the functionalities and expected effectiveness of monitoring and surveillance systems. In general, effective systems should be flexible and able to perform both behavioural analysis and cause/effect analysis on data generated within and outside the boundaries of the market or stock exchange venue. Moreover, they should also have the ability to integrate several types of internal and external data sources.

In addition, effective monitoring systems should cater for other types of analysis that relate to monitoring from a 'market rules and regulations' perspective and which can vary from market to market and from country to country. For example, such analysis could cater for Late-Trade Reporting, Market Integrity and Best Execution aspects (Kirkland and Senator, 1999).

The components and sub-systems can be generally classified into three main categories, namely 1) *collect & integrate,* 2) *analyse* and 3) *report* sub-systems. To aid the understanding, each type has been given a representative colour, i.e. purple, light blue, and orange, respectively.

From bottom to top, it is also possible to identify three main sub-systems: Information Management, Off-line Detection Engine, and Real-time Detection Engine. Within the boundaries of these sub-systems different tasks are executed on an on-demand basis, or can be automated, according to the type of manipulation or monitoring task that is being performed. In addition, from left to right, two venues, B and A, interact with each other in order to consider potential cross-product, cross-market, and cross-jurisdiction manipulation schemes. Markets A and B can be either in the same jurisdiction or in different jurisdictions, although the boundary between jurisdictions is not explicitly presented in this model.

Data is continuously fed into an information management sub-system. Here, different data preparation processes take place in order to combine and integrate the variety of sources and types of structured and unstructured data, putting them in the format required for the analysis components. This preparation includes classifying and categorising the textual data used for complaints handling, or pre-

processing and integrating pre- and post-trading data coming from either internal systems or external providers.

In the analysis sub-systems, restructured data along with updated patterns and models are used to perform a wide range of analyses, components that will be later utilised by the customer or user in a modular or ad-hoc fashion, in sequential or parallel ways. In general, the components for analysis are further grouped into two categories, namely behavioural analysis and economic analysis. In this representation, the behavioural analysis category contains one component application for social network analysis, one for text mining, and one for data mining applications. In parallel, the economic analysis component includes one module for financial modelling, one for data mining modelling and one for text mining analysis. The key difference between categories, then, is that the first type of analysis focuses on the actions and characteristics of the agents, and the second focuses on the effects of these actions. In this sense, the analysis modules can perform independent examinations of the problem, or act in coordination and integration with modules within the same category or across categories following streams or work-flow sequences'.

In the reporting sub-systems the output or outputs of the engines and the information management sub-system, are recursively sent to be stored in a knowledge base, thus making them available to the rest of the sub-systems. Warning signals and alarms are generated for Venue A and Venue B, which for example could be the trading venue for an underlying security or its derivatives; in general the model could be applied to different combinations of market types, and to a greater number of interactions between venues or securities.

Accordingly, in the output management sub-system, several steps are taken in order to process the warning signals and alarms which continually pass information to the information management sub-system. Depending of the type of customer or user, the output management sub-system can execute different actions. In the case of a regulating authority this could be to decide whether the flagged trades warrant further legal analysis, whether to start (or not) a formal investigative process and/or to follow processes of reporting and reinforcement (if applicable) of

market rules. In the case of an arbitrage seeker or market trader, this could be used to decide whether to invest in a stock that presents symptoms of manipulation or any other alternative investment actions.

The off-line monitoring and surveillance engine supports the investigation and enforcement of market rules and controls, for example the ones taking place in the real-time engine. One possibility might be to implement a demand-driven 'active mining of data streams' (Fan et al., 2004), another to utilise 'complex events processing' components (Sadoghi et al., 2010) by embedding them in the real-time monitoring engine. Other necessary tasks and architectural components, as well as complimentary information management processes for live-data integration and pre-processing, also take part in this sub-system. Four examples (one for each case in the study) of how this detection engine model can be used will be presented in the following chapters.

5.6 Chapter summary

In this chapter a framework for Market Monitoring and Surveillance Service Systems, '2M-3S' has been defined. Drawing on concepts from SD logic, SS, and ST, the current problematic and characteristics of the financial markets have been described as service systems which exist and interact in a service-oriented economy. Using the ST perspective, the concepts of archetype and leverage have been used to propose solutions for the current problematic expressed using systems concepts and terminology. Moreover, the market monitoring and surveillance activities have been described as a user or customer-driven service value network, in which value co-creation is achieved when successful service encounters meet the expectations of respective value propositions presented by the service providers.

The chapter also expanded the previously defined concepts and terminology by presenting a model for a typical detection engine. Details of its sub-systems and relationships between them were provided. Together with the taxonomies introduced in Chapter 4, these concepts and theoretical foundations helped defining the characteristics of a modern detection engine that fulfils the

requirements presented in Chapter 2. Figure 5-21 presents an schematic view of the components and underpinnings of the 2M-3S framework scope.

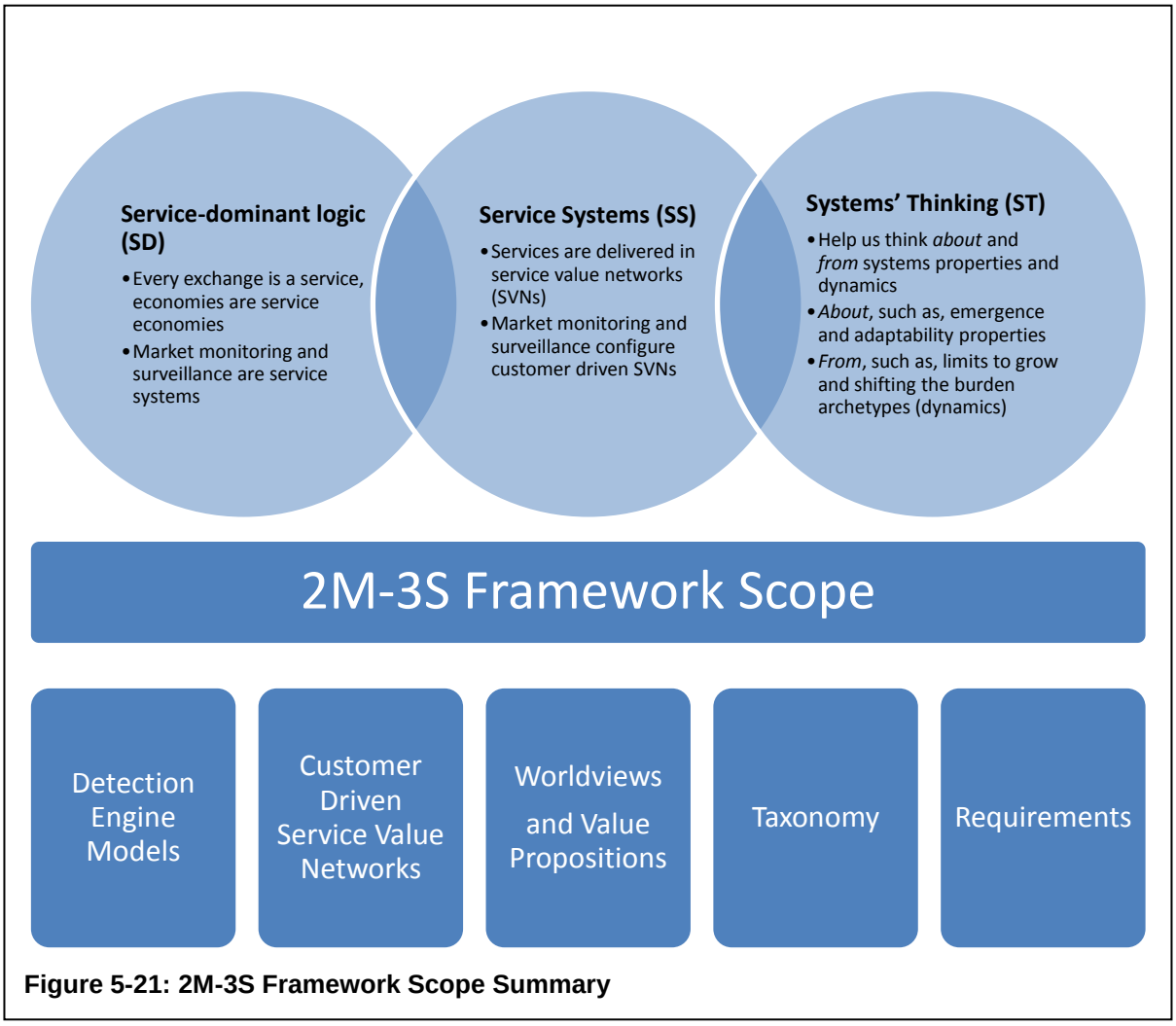


Figure 5-21: 2M-3S Framework Scope Summary

In particular, different SVN configurations were discussed in this chapter, specifically for two types of market monitoring and surveillance service provision' options: a 'closed' and an 'open stake-holder' SVN. These represent two examples of service provision, one in which the regulators are empowered as customers and another where individual investors make use of commercial market monitoring and surveillance services.

6 Case Studies

This chapter presents *instantiations* of the '2M-3S' framework by means of *case studies.* Each case study presents individual tools and salient characteristics of detection engines and provides details of the prosecution cases or simulation analyses that give them context. Chapter 6 concludes with a discussion and lessons learnt section that brings together the knowledge and insights generated in the cases.

Using the general design research methodology, each case is presented following the awareness of problem, suggestion, development, evaluation and conclusions cycle. Table 6-1 describes how each of these steps is covered in the background; data modelling and analysis; analysis and results; and evaluation and interpretation of results sub-sections. Specifically, steps i) and ii) address the awareness of problem; step iii) the suggestion; and steps iv) and v) the conclusion stages of the design research cycle.

Table 6-1 Design Research in Case Studies

Design Research Cycle	Sub-section	Step description
Awareness	Background	i) an introduction is given regarding the context and the type(s) of manipulation(s) or monitoring rule(s) under study
		ii) the proceeding(s) or prosecution file(s) that served as a legal base of evidence are introduced
Suggestion/Development	Data modelling and Analysis	iii) the methodology for the construction of the artefact/model and data used are described in detail
Conclusions	Analysis and Results	iv) the analysis and results are presented
	Evaluation and Interpretation of results	v) results are discussed with the objective of drawing conclusions for each case, and to connect them with its contributions to the overall work

In each of these steps, concepts, models and methodologies of the '2M-3S' framework are used throughout to aid the description of the problems and solutions, enrich the analysis and discussions, and to increase the knowledge generation altogether.

6.1 Case Study One. 'Trade-based' manipulations

6.1.1 Background

This case addresses challenges relating to applying data mining techniques to detect stock prices manipulations, and it extends previous results by incorporating the analysis of intraday trade prices in addition to closing prices for the investigation of *trade-based manipulations* (Allen and Gale, 1992).

Case Study 1 has been published in the *Expert Systems with Applications Journal* under the title "Analysis of Stock Market Manipulations Using Knowledge Discovery Techniques Applied to Intraday Trade Prices". A previous version of Case Study 1 was presented as a working paper at various academic conferences, including the IV European Conference on Intelligent Management Systems in Operation, Salford, UK, 2009; and the Doctoral Conference at Manchester Business School, England, June 2008.

This work helps instantiating the proposed 2M-3S framework and extends previous results on the market manipulation topic by investigating empirical evidence in normal and manipulated hourly data and the particular characteristics of intraday trades within suspicious hours. Additionally, emphasis is placed on the definition of appropriate input variables that make financial interpretation more intuitive and on the explanatory power of the models describing the intraday patterns that can be associated with manipulations. As trader performance is benchmarked against the closing prices, it is not uncommon for traders to try to influence the closing price by making last-minute or ill-intended trades, especially if they have acquired a large net position in a stock during the day (Felixson and Pelli, 1999).

Given the nature of these patterns, different types of data mining techniques, in particular *'black box'* type classifiers, have been identified as suitable techniques to discriminate between manipulated and non-manipulated stocks — see for example (Ogut et al., 2009) — but limited research has been dedicated to developing suitable techniques to discriminate between hourly data, and perhaps more

importantly, to describe and understand the characteristics of manipulated and non-manipulated trades at the intraday level.

This work makes a contribution to the existing body of work in the field by adopting an *'open box'* approach that uses financial variables, ratios and textual sources to discriminate and describe trades at the hourly and at the intraday level. In particular, variables such as 'abnormal returns', volatility and liquidity and the occurrence of financial news and events are used to describe what is '*inside the box',* presenting decision rules in the form of decision trees that can be easily interpreted by a financial expert. The work approaches the problem from the knowledge discovery in databases (KDD) framework perspective, and follows the cross-industry standard process for data mining (CRISP-DM) (Chapman et al., 2000).

The case study used as part of this thesis is based on a group of stock market manipulation cases pursued by the US Securities and Exchange Commission (SEC) that were selected by the researcher. In particular, all SEC litigation releases that contain the words 'manipulation', and '9(a)' or '10(b)', which refer to two articles of the Securities and Exchange Act of 1934 related to market manipulation, were manually identified and classified into different types of manipulations such as 'marking-the-close', 'pump-and-dumps', 'match trades', and 'wash sales'. The information collected for each of the cases includes the trading venue, the company or fund name and symbol, the date(s) relating to the manipulation events, and finally, a brief description of the details of the case and the actions taken by the SEC, e.g. sentences, prosecutions and complaints. In total eight stocks and funds trading on the New York Stock Exchange (NYSE) were chosen on the basis of the available intraday information which was only for the year 2003. This data was named *manipulated cases dataset*.

6.1.2 Data modelling and Analysis

The COMPUSTAT database (Standard and Poor's Compustat Resource Center, 2009) was used to provide supplementary profiling financial information about the selected cases, such as the SIC code, market capitalization and beta. The

COMPUSTAT database was also used to create a control sample of non-manipulated stocks that have similar characteristics to the manipulated stocks. The similarity was based on stocks relating to the same industry (SIC code), having a similar size based on market capitalization, and a similar level of risk based on the beta. In order to identify the specific control sample, four-year average values of the market capitalization and beta were calculated for all the stocks in the same SIC code as the manipulated stock and were ranked in ascending order of market capitalization. Four stocks were selected for each of the manipulated stocks. Two of the selected stocks were the ones that were immediately above the average value of market capitalization, and the other two of the selected stocks were those immediately below. In addition, only stocks that had a positive beta were selected, i.e. the next available stock above or below the manipulated stock with a positive beta was selected. For three of the manipulated stocks, the control sample included fewer than four control cases because it was not possible to identify stocks with similar market capitalization and a positive beta. Finally, in order to make sure that the selected stocks were not manipulated during 2003, the SEC litigation releases database was double checked for any evidence of prosecutions or litigation cases associated with them. This data was collected in what is called the *similar cases dataset*.

It was also decided to create an additional control sample for 2003 that includes information about dissimilar stocks to allow for comparisons between manipulated and non-manipulated cases not only with similar characteristics. Specifically, KLA-Tencor Corp. (KLAC dataset) and the components of the Dow Jones Industry Average (DJI dataset) were used to define the additional control sample. The KLAC stock was used as an example of a non-manipulated stock with a significant volume of transactions that also trades in a different market (NASDAQ). The components of the Dow Jones were selected because they reflect thirty large cap firms in the NYSE market that are among the most monitored and less likely to be manipulated stocks.

An intraday database was constructed starting from the 2nd January 2003 to the 31st December 2003 with trades for the eight manipulated stocks, the twenty six

similar stocks and the thirty one dissimilar stocks. The source of data for intraday transaction and quotes was the 'TAQ' database (Wharton Research Data Services, 2009) and it included more than a hundred million trades.

The intraday database was then populated with information relating to news and events from the Factiva database (Dow Jones Factiva, 2009) that related to each dataset. This included information released by the companies themselves or news from journalists, and it was represented in the database by the number of news items and events that were published during each trading hour. Also, the intraday database was populated with other events collected from the Yahoo Finance (Yahoo! Finance, 2009) website which do not appear in Factiva, such as analyst opinion changes or other company news. The publication time of this information was related to trading hours and the total number per hour was recorded. The news and events information from Factiva and Yahoo Finance was used to investigate how the appearance of such information relates to stock price changes, and a number of additional variables were created for this analysis.

Table 6-2 shows the statistics for the manipulated cases dataset, similar cases dataset, DJI dataset and KLAC dataset. The statistics for each dataset includes sample mean, standard deviation, skewness and kurtosis coefficients for hourly returns, abnormal hourly returns, number and volume of trades, and also the *Riskmetrics* measure of volatility for the returns. Table 6-3 shows the total number of different types of news and financial events from Factiva and Yahoo Finance for each of the datasets.

Table 6-2 Case Study 1: Summary Statistics

Summary Statistics					
SETS	Returns	Abnormal returns	Volatility Riskmetrics	Volume	#Trades
Manipulation	*4 stocks+ 4 funds = 8 total*				
mean	2.7%	0.3%	58.8%	14,908.7	25.9
median	0.0%	-1.9%	17.9%	3,000.0	6.0
std.dev.	93.1%	92.7%	181.8%	66,666.6	115.3
skewness	-1.9	-1.8	10.8	20.2	22.6
kurtosis	83.9	81.3	151.7	578.9	726.7
n	12,748.0	12,748.0	12,748.0	12,748.0	12,748.0
Control_Benchmark	*11 stocks + 15 funds = 26 total*				
mean	1.7%	0.0%	49.4%	24,924.4	33.1
median	0.0%	-1.8%	16.7%	4,900.0	7.0
std.dev.	87.3%	88.6%	93.3%	79,650.6	91.4
skewness	0.4	0.5	5.7	21.9	21.2
kurtosis	28.1	27.0	56.5	1,019.5	1,001.9
n	41,127.0	41,127.0	41,127.0	41,127.0	41,127.0
DJI	*30 stocks*				
mean	1.4%	5.6%	13.2%	910,782.2	665.7
median	0.7%	-0.4%	9.8%	507,200.0	445.0
std.dev.	34.7%	43.1%	13.5%	1,402,839.3	1,309.3
skewness	0.0	1.5	5.5	5.8	6.2
kurtosis	12.9	9.0	70.6	52.6	51.0
n	119,630.0	119,630.0	119,630.0	119,630.0	119,630.0
KLAC	*1 stock*				
mean	1.5%	7.3%	26.4%	616,791.4	1,412.1
median	0.7%	7.1%	22.7%	475,950.0	1,174.0
std.dev.	56.2%	59.6%	15.7%	560,027.0	1,272.1
skewness	0.1	0.1	2.1	1.8	1.5
kurtosis	3.6	4.0	7.0	5.2	3.5
n	2,233.0	2,233.0	2,233.0	2,233.0	2,233.0

Table 6-3 Case Study 1: Distribution of the Type and Number of News Items and Events per Sample

Type	Manipulated 8 stocks	%	Similar 26 stocks	%	DJI 30 stocks	%	KLAC 1 stock	%	Total
Events total	**45**	*100*	**232**	*100*	**1,621**	*100*	**18**	*100*	**1,916**
Mentions in News	40	*88.89*	37	*15.9*	715	*44.11*	4	*22.22*	796
Analyst opinions	0	*0*	28	*12.1*	646	*39.85*	14	*77.78*	688
Dividend Announcements	5	*11.11*	167	*72*	260	*16.04*	0	*0*	432
Events/stocks ratio	5.63	-	8.92	-	54.03	-	18.00	-	-

This case follows the guidelines of the CRISP DM reference model (Chapman et al., 2000). Work was divided into six stages: business understanding, data understanding, data preparation, modelling, evaluation and deployment. Specifically, the business understanding stage was dealt with when converting the trade-based manipulation problem into the present data mining problem. The deployment stage is not discussed here but it is part of future research work. The

rest of the stages are described in the following sub-sections. Appendix 1 presents the knowledge discovery process flow according to CRISP-DM.

Data Understanding

The data understanding phase started with an initial data collection from the TAQ and textual databases and proceeded with other activities in order to become familiar with the data; for example, identifying data quality problems. Specifically, the data contained a 'correction' field indicating changes to prices and conditions in previous transactions made by the traders. As a precaution, the researcher ignored any trades for which the correction field indicated such cases. Although TAQ contains the best bid and offers data on the vast majority of market venues in the USA, the researcher only used data from the NYSE venue.

Data Preparation

The data preparation phase covered all activities to construct the final dataset, i.e., data that was fed into the modelling tool(s) from the initial raw data. As is usual in KDD projects this stage was the most time consuming. Specifically, attention was focused on intraday data observations, but extremely frequent information can also make data mining very difficult, especially when considering real-life scale implementation and computational restrictions. By settling at an intermediate point, the researcher decided to work with hourly prices and volumes as well, although the prototype system was designed in such a way that the span of the analysis can be easily adjusted to the desired unit of time.

Consequently, the work was organised on two levels of granularity, hours and seconds, with the purpose of constructing two decision trees that describe patterns at each level. The work was also organised in sequential terms, where the hourly decision tree was constructed and applied first, and that information was later used by the second tree.

In particular, each level considered several steps of data pre-processing, mainly to join different types of data, and to construct the different families of indicators and explanatory variables that are usually mentioned in the literature. These indicators were designed to reflect changes in returns, abnormal returns, volatility and liquidity, and were expressed either as ratios or *z-scores* in order to make

comparisons across different stocks possible. Indicators that reflected '*preconditions*' were also constructed and they were defined as particular days, and time zones of the day when manipulations usually occur and thus should be closely monitored. Furthermore, news and events extracted from the textual sources were matched with the TAQ data in order to analyse the effects that this information has on the trading patterns.

For the first level of granularity, representative price and size (volume) observations were created from the original trading and quotes data. The process generated evenly spaced, one-per-hour average prices and total volumes. Then, this representative dataset was compiled with information coming from the COMPUSTAT dataset, and was used to construct an abnormal returns indicator. Appendix 3 gives details of the abnormal returns calculations. For the second level of granularity, similar indicators were created, but in contrast to the previous stage, these were calculated for each trade and quote, changing the time span of the analysis from hours to seconds.

Modelling - Regression and frequency of outlier's analysis

In common with the results in previous explorations of the problem — for example (Aggarwal and Wu, 2006) — particular indicators were initially used to study the differences in returns (ZO1 indicator), abnormal returns (ZAR1), liquidity (ZS1) and volatility (RV3) between the different samples. Using hourly data, these indicators were linearly regressed against a constant and a dummy variable for manipulation. The dummy is equal to 1 for the manipulated sample, and to 0 for the similar and dissimilar samples (Equation 2). Thus, if the parameter α_1 is found to be statistically significant, one can argue that there are significant differences in the value of the indicators between different samples. As discussed before, the sample spans from the 2nd January to 31st December 2003.

Equation 2

$$indicator = \alpha_0 + \alpha_1 \cdot Dummy + \varepsilon$$

where *Indicator* refers to the returns (ZO1 indicator), abnormal returns (ZAR1), liquidity (ZS1) and volatility (RV3) indicators; α_0 and α_1 are the coefficients of the

linear model; *Dummy* is a binary variable that takes the value of 1 for the manipulation sample, and 0 in any other case; and ε are the errors of the model which follows a traditional white noise process.

Conversely, testing also checked whether the indicators could be used to discriminate between samples. A logistic regression was estimated in which the dummy variable was used as dependent variable and indicators were used as independent variables (Equation 3, Equation 4). In this case the individual and joint significance of parameters β was studied. Again, the sample period is 2nd January to 31st December 2003.

Equation 3

$$Dummy = \beta_0 + \beta_1 \cdot ZO1 + \beta_2 \cdot ZAR1 + \beta_3 \cdot ZS1 + \beta_4 \cdot RV3 + \varepsilon$$

Equation 4

$$\boldsymbol{f(z) = \frac{1}{1 + e^{Dummy}}}$$

where ZO1 refers to the returns indicator, ZAR1 to the abnormal returns indicator, ZS1 to the liquidity indicator and RV3 to the volatility indicator; *Dummy* is a binary variable that takes the value of 1 for the manipulation sample, and 0 in any other case; $\beta_0, \ldots, \beta_4$ are the coefficients of the logistic model; $f(z)$ the logistic function; and ε are the errors of the model which follows a traditional white noise process.

A frequency analysis was also carried out in which the presence of outliers was counted for each sample and these were related to the frequency and timing of the appearance of news items and events. It is expected that if all the samples behave in the same way, no relationship between the number of outliers and the type of samples should be found, and similarly, no relationship between the number of outliers and number of news items with the samples should be found.

Modelling - Unsupervised techniques

After the basic differences and outlier behaviour of the samples had been analysed, various data mining modelling techniques were applied. For each level of granularity, the work was once again organised in two steps. This was due to the fact that the information on the manipulation cases obtained from SEC was

somewhat vague in describing specific days and hours when the manipulations occurred, and so it was difficult to define what the training examples to be learned by the algorithms were.

In order to work around this issue, an anomaly detection algorithm was used to cluster the transactions according to their similarities, discriminating between outliers and normal trading using different alternative bundles of indicators as inputs. Based on the results of the initial clustering, an abnormality score or flag was calculated counting the times a trading block or transaction was included in a particular abnormality cluster. The flag was 'true' when the score of the trading block was 3 or more, and 'false' for all the other cases.

Modelling - Supervised techniques

In step 2, this flag variable was used as the target variable by the classification algorithms. Several supervised decision trees were trained to discriminate between the true and false values. Due to the higher presence of normal blocks compared with the proportion of suspicious ones, the training dataset was balanced using an under sampling approach, and the best-performing trees were further tested using a *jack-knife* or *bootstrapping* technique (Efron, 1979) and ranked based on measures of unequal importance (Kubat and Matwin, 1999). The training sample was composed of half of the manipulated sample, the similar sample and the KLAC sample. The remaining block hours of trades were used as a testing sample, which also contained data from the DJI sample.

This double-step training strategy, based on the vague description of the SEC files, was also tested by manually constructing a 'watch-list' which included hours and dates of trade that were the closest matches to the periods of manipulation mentioned in the SEC files. Thus, the performance of the hourly decision tree in these periods was specifically monitored. Appendix 1 shows the watch list and the results of this task.

In a similar way, the analysis for the second granularity was carried out as follows. Using the decision rules created during the first stage of the analysis, each of the manipulated blocks was classified as suspicious or not suspicious. Only the suspicious blocks were selected, and again a two-step modelling approach was

implemented. In step 1, an anomaly detection algorithm was applied to create clusters of normal and abnormal transactions. A decision tree was trained to discriminate between normal and abnormal groups, and by doing so, showing the patterns and characteristics of the two groups. The training sample was composed, this time, of half of the data available.

Evaluation Criteria

By putting emphasis on the training of decision tree algorithms, the researcher was especially interested in simplicity of interpretation of the rules behind the classifications. Using this as the baseline criterion, competing models were ranked on 3 dimensions: i) overall accuracy, and measures of unequal importance, ii) sensitivity and iii) false positives per positives ratio (Barandela et al., 2003, Kubat and Matwin, 1999). Finally, the best models were selected based on the joint compliance of high classification power — measured by the three dimensions listed above — and simplicity of interpretation implied in the number of branches of the decision trees.

6.1.3 Analysis and Results

6.1.3.1 Data Understanding and Data Preparation Results

This section presents the results of the analysis and examples of the visualisation techniques that were performed using the SPSS Clementine data mining workbench. As an example, Figure 6-1 shows the behaviour of the hourly indicators for the fund 'BIF', which was one of the funds manipulated according to the SEC litigation release files. As stated in the SEC files, *'from at least June 2002 through December 2003'*, the fund manager engaged in 'marking-the-close' manipulation of four thinly traded funds, benefiting by collecting more management fees from enhanced performance results. In fact, the SEC proceeding stated that:

> *'[The fund manager] placed two trades on December 31, 2003… [he] placed a market buy order for 80,750 shares of BIF at 3.57 p.m. EST, which constituted 76.18% of the day's trading volume in BIF. The price before the order was $5.90, and BIF closed at $6.30 that day'.*

As shown in Figure 6-1, most of the indicators contain jumps for 24th and 31st December; the latter includes the most extreme jumps. It is worth noticing that the news indicators are zero for both dates which means that there was no news or other financial information available on the markets that could explain those jumps.

At the same time, the indicator of volatility RV3 presents a jump for 24th December, but not for the 31st. It is also worth noticing that the second jump was also related to a high value of the indicator of preconditions, due to the fact that it was a quarter- and year-end and that the transaction occurred near the closing time.

This example emphasises the huge gap between what was effectively described as a manipulation by SEC and what can be spotted in the data. In fact, most of the legal cases are vaguer, and talk about 'several manipulations' occurring in an imprecise span of time. Expressions like 'from at least June 2002 through December 2003' were used, which effectively means that the only certainty is that the stock or fund was manipulated during an 18-month period. This fact strengthens the objective of this case, which is to highlight the importance of an independent and scientific description of the manipulation patterns.

Figure 6-2 shows the differences between samples for the return indicator (average value of ZO1 indicator per hour). Bars are coloured red or blue to distinguish between quarter ends and the rest of the days of the year. One can observe a 'U-shape' for the returns which are higher during opening and closing hours. There is also a clear difference when comparing across samples. Although the manipulated sample shares the same U-shape, the returns are extremely high on the closing hours of quarter ends. Compared with the similar benchmark sample, the manipulated sample also has lower returns in the morning, but overcompensates with high peaks in the afternoon. For the DJI and KLAC samples, one can observe particularly high returns in the morning hours of quarter ends.

Figure 6-3 presents the same type of graph for the liquidity indicator (ZS1). It is possible to see that the manipulated sample presents higher liquidity than the

similar benchmark sample in closing hours of quarter ends. Nonetheless, the liquidity is lower for the rest of the hours especially in quarter ends. A common denominator between the samples is that all of them show higher liquidity at closing hours, but this effect is bigger in the manipulated sample. Figure 6-4 shows the behaviour of the volatility indicator (RV3). Particularly, the volatility indicator shows no clear tendency in either the DJI and or KLAC sample, but it presents a positive tendency for the similar benchmark and manipulated samples. Once more, there is a jump in volatility associated with the manipulated sample in closing hours. There is also a clear difference between the volatility on quarter ends compared with the rest of the days. Quarter-end volatility is lower in all hours, except for the closing hour.

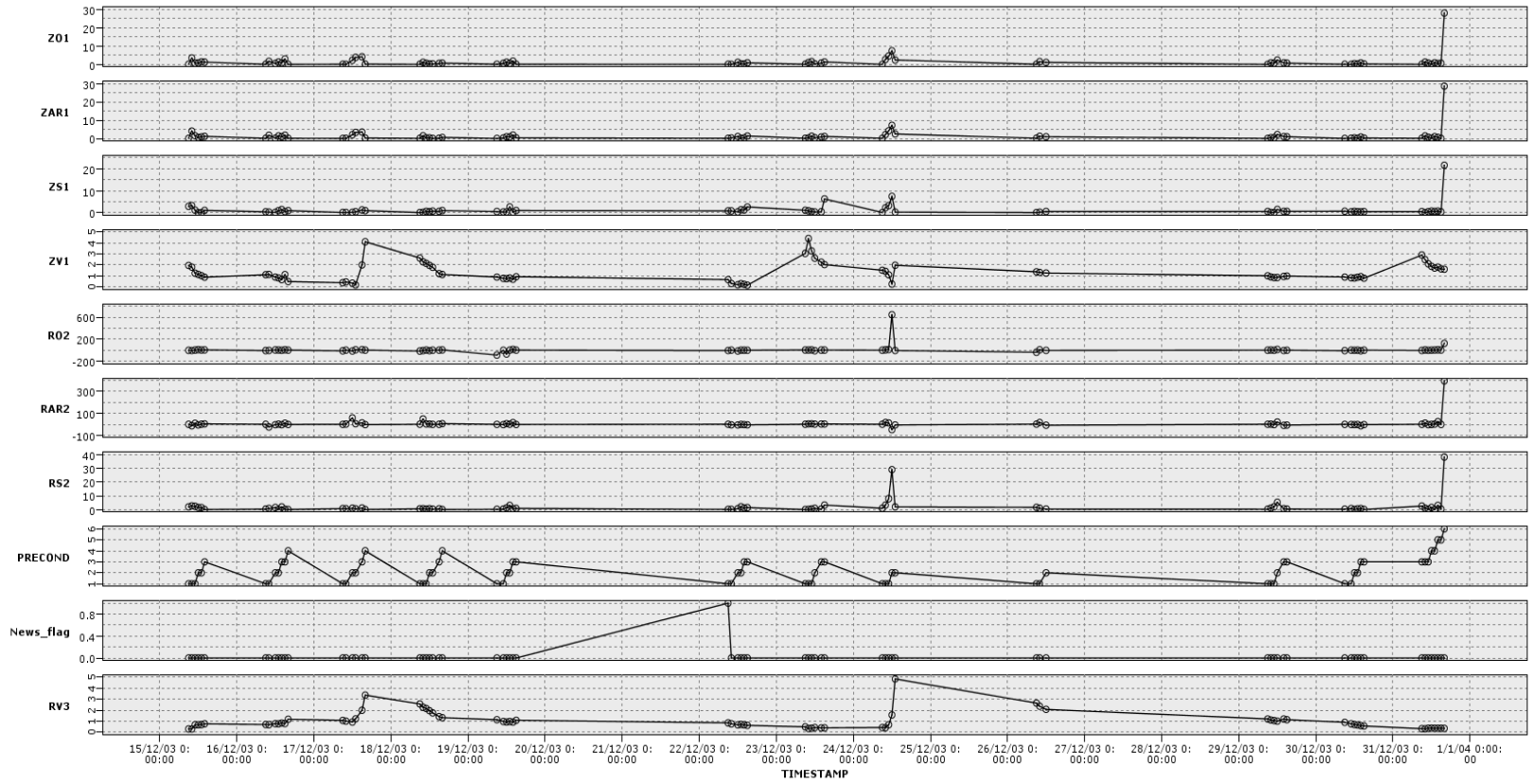


Figure 6-1 Case Study 1: Indicators for BIF, Dec. 2003

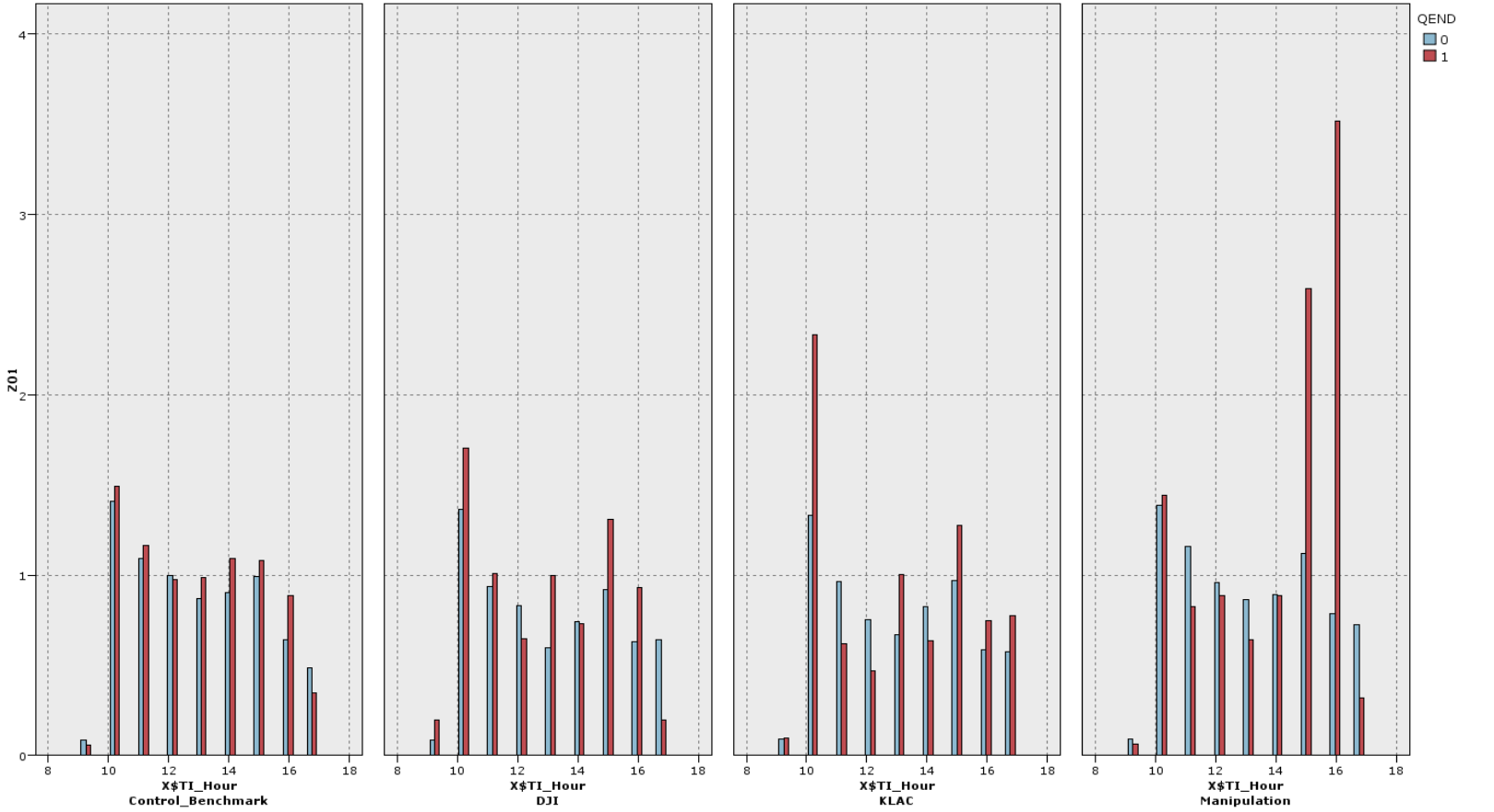


Figure 6-2 Case Study 1: Hourly Average Returns Indicator (ZO1 Indicator)

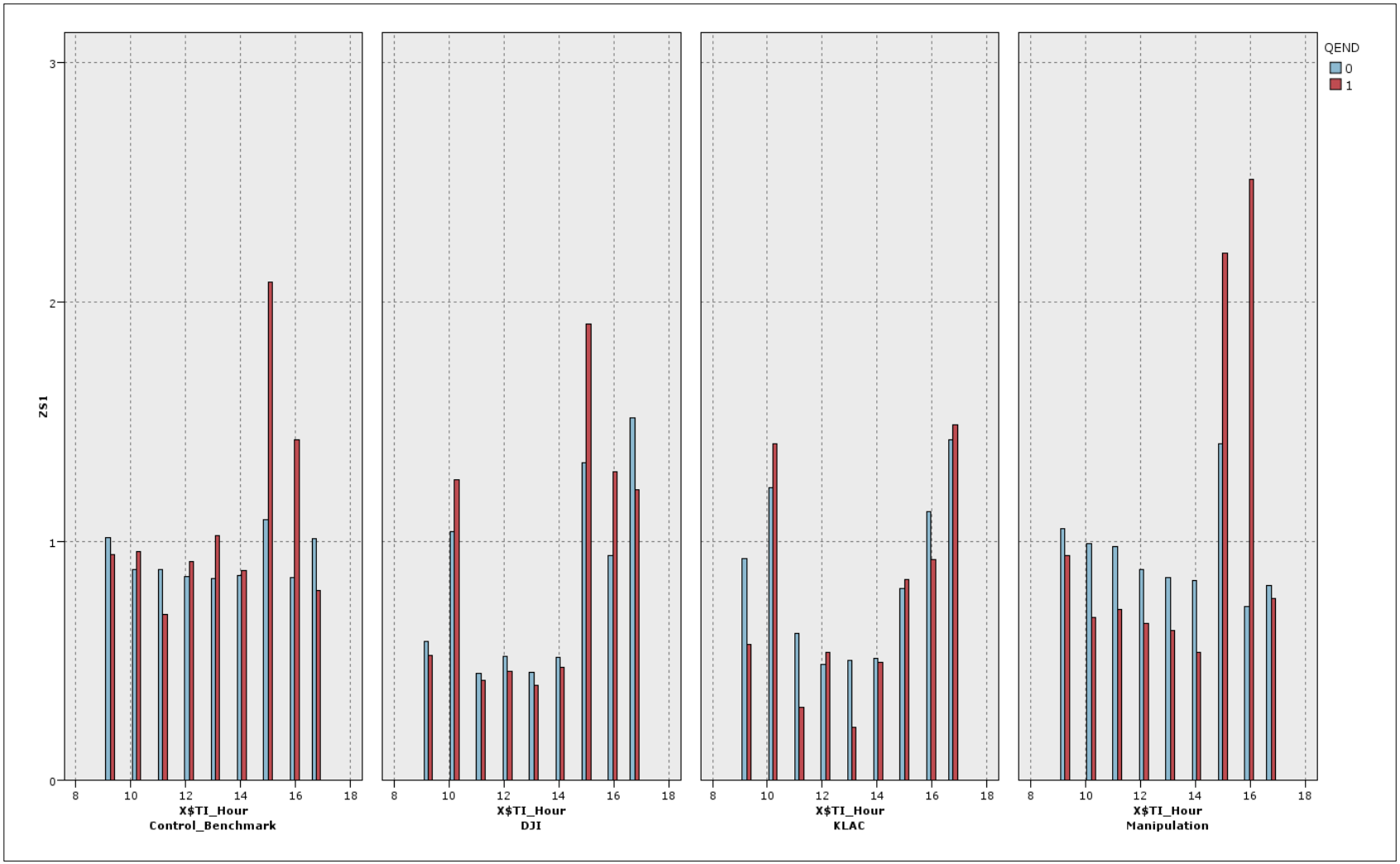


Figure 6-3 Case Study 1: Hourly Average Liquidity Indicator (ZS1 Indicator)

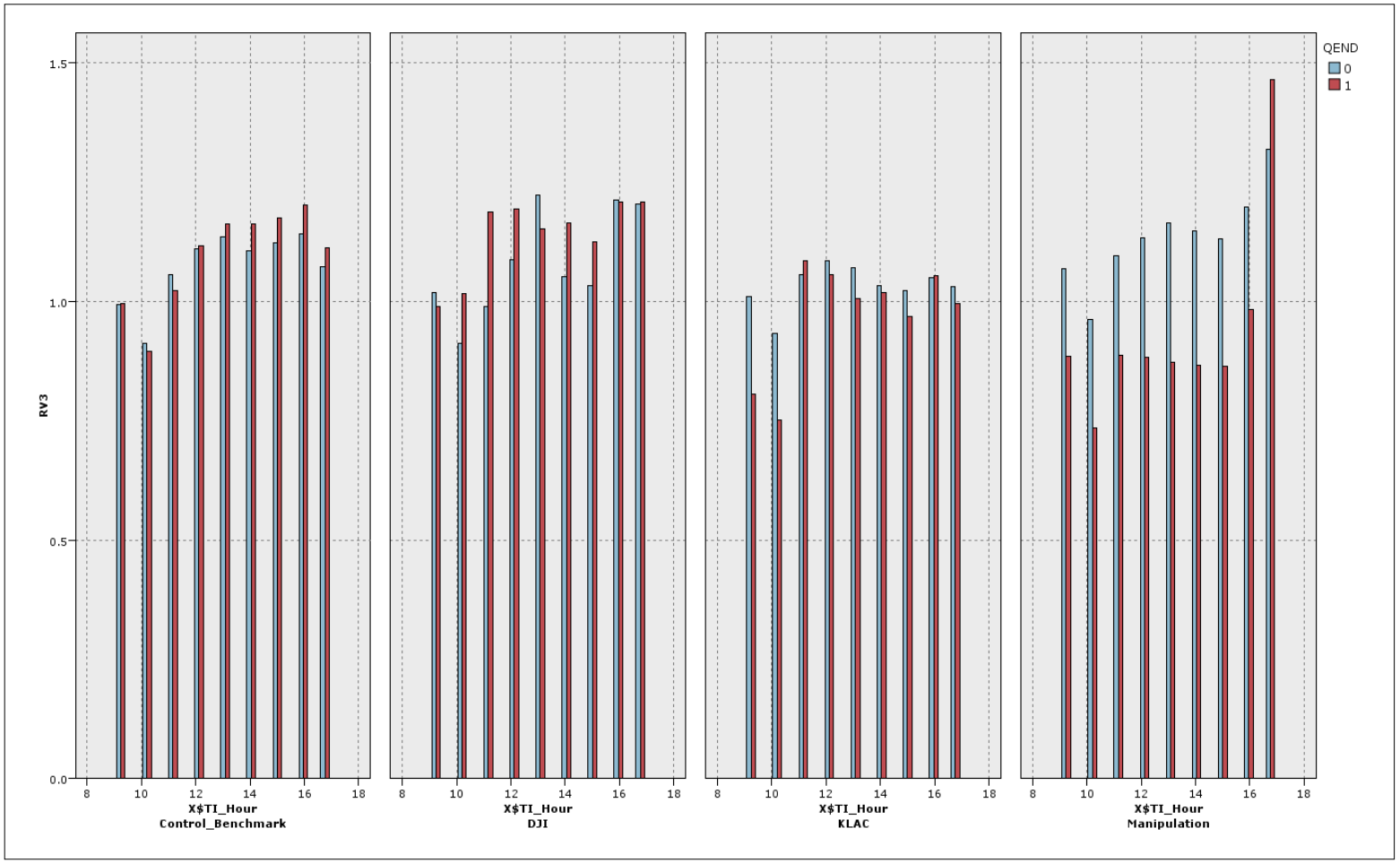


Figure 6-4 Case Study 1: Hourly Average Volatility Indicator (RV3 Indicator)

6.1.3.2 Data Modelling

Analysis of Regression and frequency of outliers analysis

From the visualisation analysis, one can observe that although there is a difference in the indicators between the samples, this difference varies depending on preconditions, such as the hour of the day and the time of the year. In order to study the overall effects irrespective of preconditions and to make our results comparable to the ones presented in previous literature findings, regression models were applied to the data as specified in Equation 2, Equation 3 and Equation 4.

Table 6-4 Panel A shows the regression results for Equation 2. For the manipulated sample the returns, the abnormal returns, the liquidity and the volatility are always higher than for the rest of the samples with coefficients statistically significant at the 1% level. This result confirms previous findings that on average returns, liquidity and volatility increase during the manipulation period (see for example (Aggarwal and Wu, 2006)).

Table 6-4 Panel B shows the results for the logistic regression analysis as specified in Equation 3 and Equation 4. The significance of the indicators as a means of discriminating between samples was tested. It can be seen that all the indicators present a significant relationship with the dependent variable, and all except the abnormal returns indicator (ZAR1) present a positive coefficient with the return indicator (ZO1), while the abnormal returns are the ones that present a bigger effect. The negative coefficient of ZAR1 is the only unusual result which can be explained by the absence of controlling variables, such as the hour of the day and time of the year or the consideration of new information in the form of news items and other financial events.

Finally, Table 6-5 relates the frequency of outliers in each sample to the appearance of news items. Outliers were defined as observations lying in the one percent fractional rank of each indicator. In order to make the results independent of the number of trades of each sample, the count was expressed as a proportion of the total of block hours with one or more trades. Panel A shows these

proportions and as can be seen the manipulated sample presented the highest proportion of outliers in all but the abnormal returns indicator.

It is worth noting the resemblance with the similar cases sample, and the large difference with the DJI and KLAC samples that showed a much smaller proportion. In the case of the abnormal returns indicator, the DJI sample follows the proportion of the manipulated sample more closely. As previously argued, there are other issues to consider with extreme jumps and outlier observations, such as the appearance of new information in the form of financial news. Panel B shows the ratio of the number of outliers per number of block hours with one or more news items. Here the manipulated sample is clearly the one with fewer outliers. This could be related to the absence of news and other information that could explain the outliers.

In contrast, the DJI and KLAC samples present the highest ratios, indicating the extent to which blue-chip stocks are monitored by the market. Once more, the similar control benchmark comes close to the manipulated sample but it keeps higher ratios, signalling that although less monitored, there are still more news items for this type of stock than for the manipulated ones. Panel C and Panel D show the gross count of block hours with more than one trade and the number of block hours with more than one news item respectively. These show the same trends.

Table 6-4 Case Study 1: Return, Abnormal Returns, Liquidity, and Volatility of Manipulated Stocks

Panel A: Linear Model

$$Indicator = \alpha_0 + \alpha_1 \cdot Dummy + \varepsilon$$

	Returns [ZO1]	Abnormal Returns [ZAR1]	Liquidity [ZS1]	Volatility [RV3]
α_0 :	0.707**	0.846**	0.823**	1.091**
st.err. :	(0.003)	(0.003)	(0.004)	(0.002)
α_1 :	0.213**	0.051**	0.160**	0.018**
st.err. :	(0.010)	(0.010)	(0.015)	(0.006)

** 1% significance level

Panel B: Logistic Model

$$Dummy = \beta_0 + \beta_1 \cdot ZO1 + \beta_2 \cdot ZAR1 + \beta_3 \cdot ZS1 + \beta_4 \cdot RV3 + \varepsilon$$

$$f(z) = \frac{1}{1 + e^{Dummy}}$$

	Coef.	St.err.	Exp(B)
β_0 :	-2.683**	(0.019)	
β_1 :	0.264**	(0.013)	1.302
β_2 :	-0.152**	(0.013)	0.859
β_3 :	0.024**	(0.005)	1.024
β_4 :	0.051**	(0.014)	1.052

** 1% significance level

Sample period is from 2nd January 2003 to 31st December 2003.

The sample has 175,608 observations, which is the total of the hourly values of the indicators across the different stocks and funds of each set.

Table 6-5 Case Study 1: Frequency Analysis of the Outliers per Sample

Panel A: Proportion of outliers per indicator				
SETS	returns	abnormal returns	liquidity	volatility
Manipulations	2.6%	1.0%	2.7%	2.9%
Similar	2.2%	0.7%	2.0%	1.8%
DJI	0.4%	1.2%	0.5%	0.5%
KLAC	0.8%	0.2%	0.3%	0.4%
Average	**1.5%**	**0.8%**	**1.4%**	**1.4%**
Panel B: Ratio #news per #outliers				
SETS	returns	abnormal returns	liquidity	volatility
Manipulations	0.06	0.15	0.05	0.05
Control_Benchmark	0.12	0.45	0.12	0.15
DJI	5.92	4.97	5.17	4.75
KLAC	1.06	3.60	3.00	2.00
Average	**1.79**	**2.29**	**2.09**	**1.74**
Panel C: Count of Block Hours with trades>1 and news>1				
SETS	# blocks	# news		
Manipulations	12,748	36		
Control_Benchmark	41,127	157		
DJI	119,630	5,606		
KLAC	2,233	18		
Total	**175,738**	**5,817**		

Unsupervised Modelling

Table 6-6 Panel A shows the results of the abnormality analysis for the hourly granularity level. Cells contain the average value of the cumulative abnormal indicator (CAI), which counts the times a block hour was flagged as an anomaly. The average was calculated for each sample for three intervals of the day: opening (from 8am-10am), midday (from 10am-12pm) and closing (from 12pm-17pm). In terms of the average value of the CAI indicators a U-shape can be observed, and

the opening and closing hours present higher values, i.e. a higher number of outliers or abnormalities can be found in those hours.

Table 6-6 Case Study 1: Unsupervised Modelling Results

Level 1. Time span block hours			
Panel A: Average CAI value			
SETS	Opening	Midday	Closing
Manipulations	0.14	0.14	0.24
Control_Benchmark	0.18	0.17	0.25
DJI	0.16	0.09	0.32
KLAC	0.19	0.13	0.21
Panel B: Number of block hours with more than one news item			
SETS	Opening	Midday	Closing
Manipulations	21	5	10
Control_Benchmark	131	17	9
DJI	2588	1010	2008
KLAC	12	5	1
Level 2. Time span seconds			
Panel C: Number of abnormalities and number of news items			
SETS	Opening	Midday	Closing
#Abnormalities	2988	821	798
#news	3	2	4

Table 6-6 Panel B shows the number of block hours with one or more news item aggregated by interval. It can be observed that the number of news items is more or less uniformly distributed across the day, which is expected given its random character. By studying Panel A and Panel B together, it is possible to relate the average CAI with the number of news items. In fact, for the DJI it is possible to observe that it has one of the highest CAI values and at the same time, it has the more news items. Overall, the anomaly detection algorithm was successful in flagging outlier observations, reflecting the same underlying structures that were discovered using the visualisation and frequency of outlier analysis.

Table 6-6 Panel C reports the results of the abnormality analysis for the second level of granularity. The first row contains the actual number of trades flagged as

anomalous within the particular hours that were identified as suspicious in the hourly granularity level. The manipulated sample was the only sample considered on this level, because the researcher could not argue against or in favour of particular trades that could potentially be flagged as manipulations in the other samples. Panel C second row contains the aggregate number of news items, broken down for each interval of the day. The number of news items is very low compared with the number of abnormalities, which can be interpreted as a preliminary sign of manipulations.

Supervised Modelling

In both by hour (level 1) and by second (level 2) modelling stages, a number of different supervised classification algorithms based on decision trees were used for the analysis. Specifically, the algorithms used were: QUEST, C5.0 and C&RT, which are all available in the SPSS Clementine data mining workbench. Table 6-7 and Table 6-8 show the overall accuracy and error rate of the algorithms for level 1 and 2 respectively.

From Table 6-7 Panel A and Panel B it is possible to observe that in terms of overall accuracy and sensitivity, the different algorithms reach good levels of classification performance with sensitivity (unequal importance) ratios close to the 90% range. This is true across the different partitions of the sample. These results are especially significant considering that trees were trained using a proportionally small, but well-chosen balanced training sample. These results are also robust, as can be seen in Panel C which shows that models maintained their performance when tested with 250 randomly selected smaller samples taken from the original dataset. This made possible to estimate non-parametrically the mean and the standard deviation of the hit ratios. Standard deviations were relatively small so it can be argued that decision trees are good discriminators and that the same could be assumed when they are confronted with new cases.

Table 6-7 Case Study 1: Supervised Modelling Results (Level 1)

Panel A: Test Set Performance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm	Cases	T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	148,713	1,561	171	137,169	9,812	6.71%	93.29%	90.13%	93.32%	86.27%	0.12%	6.29
QUEST	148,713	1,568	164	135,552	11,429	7.80%	92.20%	90.53%	92.22%	87.94%	0.12%	7.29
CR&T	148,713	1,050	682	136,637	10,344	7.41%	92.59%	60.62%	92.96%	90.78%	0.50%	9.85
Panel B: Overall Perfomance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm	Cases	T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	175,738	1,743	180	161,498	12,317	7.11%	92.89%	90.64%	92.91%	87.60%	0.11%	7.07
QUEST	175,738	1,745	178	159,318	14,497	8.35%	91.65%	90.74%	91.66%	89.26%	0.11%	8.31
CR&T	175,738	1,229	694	160,659	13,156	7.88%	92.12%	63.91%	92.43%	91.46%	0.43%	10.70
Panel C: Bootstrapping Performance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm		T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	mean	86.9	9.0	8,070.9	616.4	7.12%	92.88%	90.6%	92.90%	87.64%	0.11%	7.18
	std. dev.	*(9.6)*	*(2.7)*	*(82.4)*	*(23.7)*	*0.26%*	*0.26%*	*2.9%*	*0.26%*	*1.27%*	*0.03%*	*0.87*
QUEST	mean	87.5	8.6	7,968.1	726.1	8.36%	91.64%	91.1%	91.65%	89.25%	0.11%	8.40
	std. dev.	*(9.3)*	*(3.1)*	*(85.7)*	*(26.0)*	*0.28%*	*0.28%*	*3.0%*	*0.28%*	*1.07%*	*0.04%*	*0.95*
CR&T	mean	61.6	35.1	8,036.0	657.0	7.87%	92.13%	63.7%	92.44%	91.43%	0.43%	10.83
	std. dev.	*(7.6)*	*(5.9)*	*(84.3)*	*(25.7)*	*0.29%*	*0.29%*	*4.7%*	*0.29%*	*1.02%*	*0.07%*	*1.39*

Table 6-8 Case Study 1: Supervised Modelling Results (Level 2)

Panel A: Test Set Performance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm	Cases	T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	65,102	2,233	54	60,002	2,813	4.40%	95.60%	97.64%	95.52%	55.75%	0.09%	1.26
CR&T	65,102	2,156	131	57,855	4,960	7.82%	92.18%	94.27%	92.10%	69.70%	0.23%	2.30
QUEST	65,102	2,110	177	55,285	7,530	11.84%	88.16%	92.26%	88.01%	78.11%	0.32%	3.57
Panel B: Overall Perfomance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm	Cases	T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	130,070	4,536	71	119,901	5,562	4.33%	95.67%	98.46%	95.57%	55.08%	0.06%	1.23
CR&T	130,070	4,342	265	115,418	10,045	7.93%	92.07%	94.25%	91.99%	69.82%	0.23%	2.31
QUEST	130,070	4,234	373	110,463	15,000	11.82%	88.18%	91.90%	88.04%	77.99%	0.34%	3.54
Panel C: Bootstrapping Performance												
		Confusion Matrix						Unequal Importance Measures				
Algorithm		T-T	T-F	F-F	F-T	Error rate	Accuracy	Sensitivity	Specificity	False positive rate	False negative rate	False positives per positive
C5	mean	225.5	3.5	5,994.6	278.9	4.34%	95.66%	98.46%	95.55%	55.29%	0.06%	1.24
	std. dev.	*(13.9)*	*(1.9)*	*(78.6)*	*(14.8)*	*0.23%*	*0.23%*	*0.8%*	*0.23%*	*1.97%*	*0.03%*	*0.10*
CR&T	mean	217.0	13.3	5,770.5	502.3	7.93%	92.07%	94.22%	91.99%	69.83%	0.23%	2.32
	std. dev.	*(13.7)*	*(3.5)*	*(77.0)*	*(22.8)*	*0.34%*	*0.34%*	*1.5%*	*0.34%*	*1.66%*	*0.06%*	*0.19*
QUEST	mean	211.3	18.4	5,521.8	751.1	11.83%	88.17%	91.99%	88.03%	78.05%	0.33%	3.57
	std. dev.	*(15.5)*	*(3.7)*	*(70.7)*	*(25.3)*	*0.37%*	*0.37%*	*1.5%*	*0.38%*	*1.37%*	*0.07%*	*0.29*

In addition, it is possible to see in Appendix 1 how the decision tree classified the block hours included in the watch list. Over a total of 55 block hours, 38 blocks were flagged 'T', thus reaching an accuracy ratio of 69%. This is a strong indication that the decision tree can successfully discriminate suspicious trading due to the fact that these particular block hours were selected based on the cases contained in the SEC litigation releases database. In other words, it can be argued that if the decision tree was used, it would have been able to raise an alarm in almost 70% of these cases.

The left panel in Table 6-7 shows the actual tree and decision rules created using the C5.0 algorithm, which was the best performing tree in this stage. Structures in the data can be followed graphically by inspecting the branches of the tree from top to bottom. Also, patterns are described in the set of rules delivered in the right panel of Figure 6-5. In the tree panel, each terminal node contains a frequency table of the number of block hours and the proportion of the training sample classified in each category. The 'T' flag represents the block hours marked as suspicious, and 'F' the ones that are not. In the rules panel, each rule contains the description of the pattern as well as the internal proportion of the class. For example, Rule 3 for 'F' was true for 2.00% of the block hours that were flagged as false in the complete sample (both training and test datasets).

Table 6-8 shows the results for the level 2 analysis. Panel A and Panel B show again that decision trees performed very well. In this case, they correspond to specific trades within the block hours flagged as suspicious for the C5.0 tree of the previous stage in the manipulated sample. From Panel C, it can be seen that the trees performed very well in the 250 subsamples which were randomly tested. Although it is clear that the C5.0 algorithm again created the best performing decision tree, it is necessary to examine how easy it is to interpret the rules implied by the tree. In other words, the C5.0 tree has too many nodes and branches, which make its interpretation more difficult. Due to the fact that this work is more concerned with analysis results that can have a meaningful financial interpretation than with the precise classification performance of the system, the decision tree generated by the CR&T algorithm is preferred. This was able to achieve a similar

level of classification performance in a much simpler and more comprehensible way. In Figure 6-6, the left panel shows the actual CR&T tree and the right panel shows the decision rules derived from it.

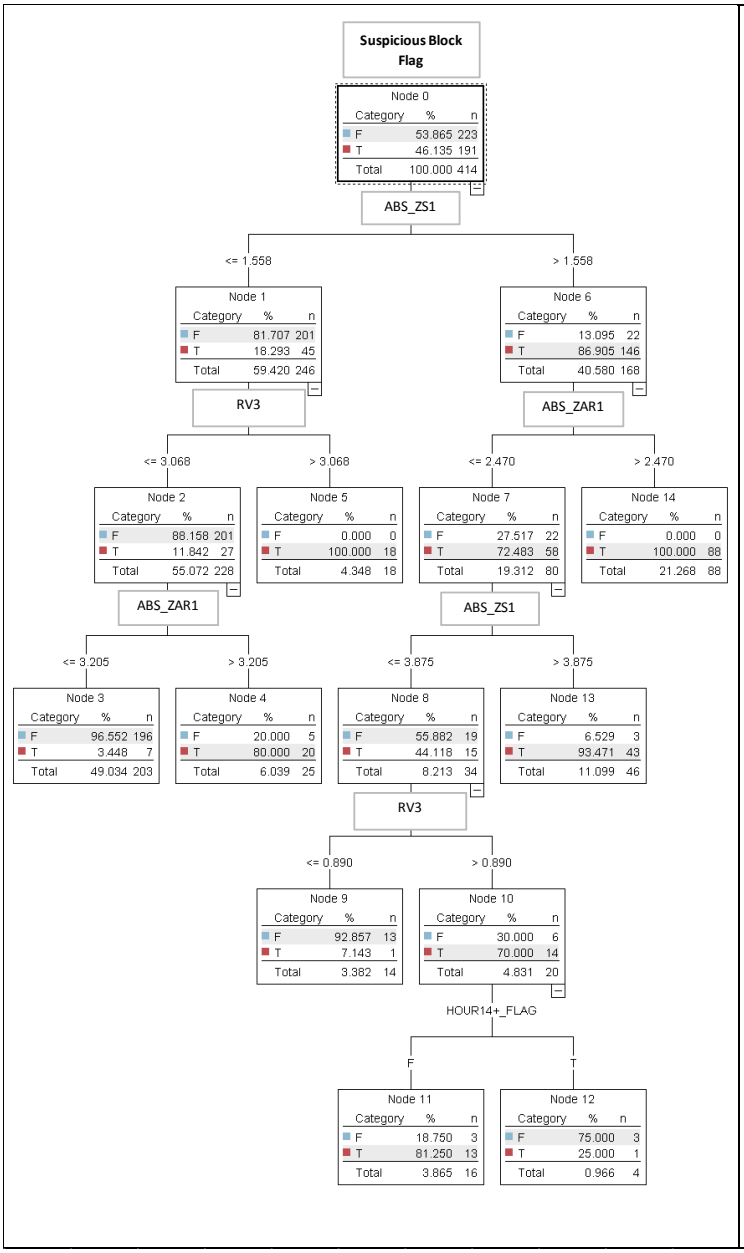


Figure 6-5 Case Study 1: C5.0 Tree – Suspicious Block Hour Flag, Level 1

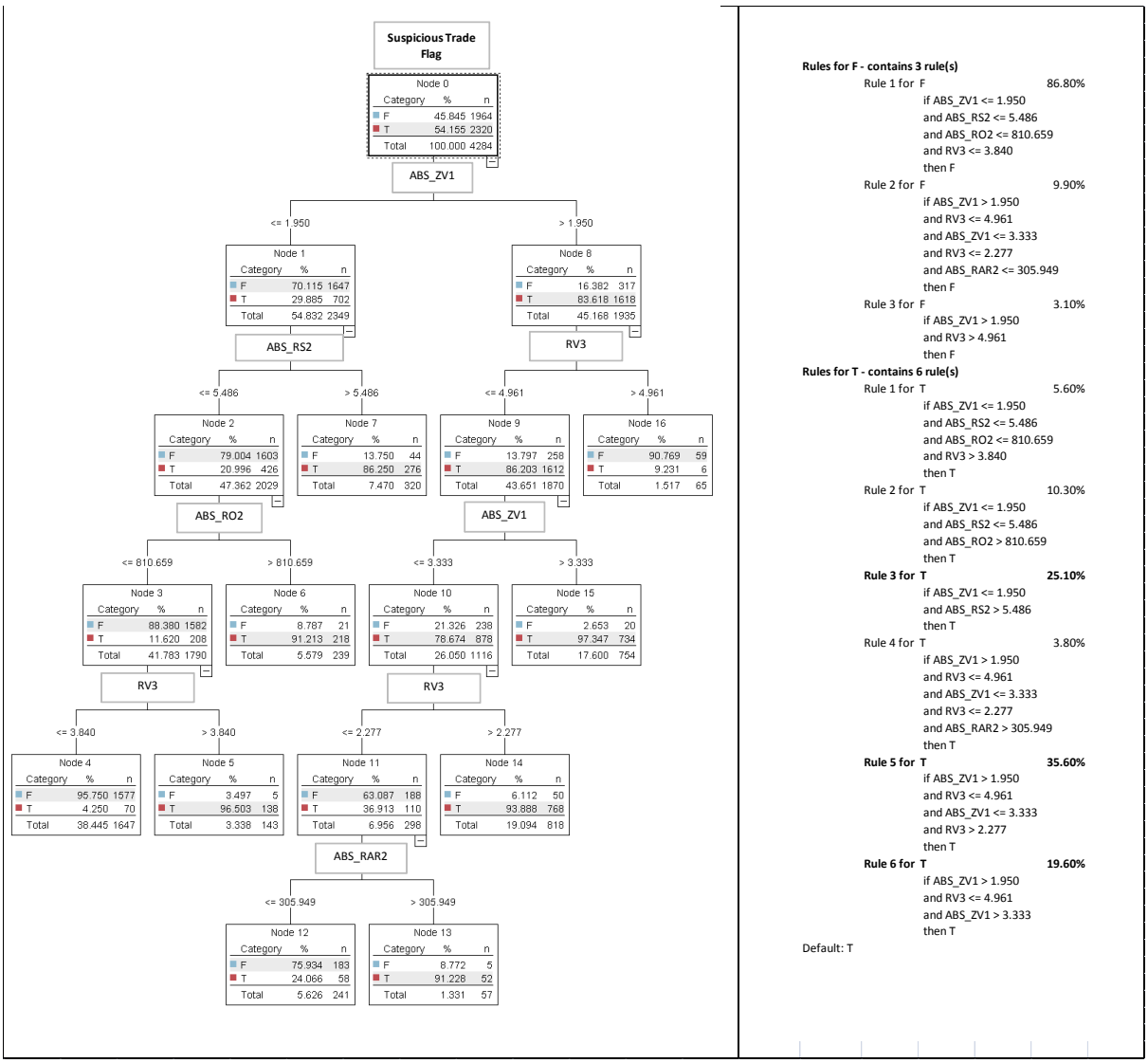


Figure 6-6 Case Study 1: CR&T Tree – Suspicious Trade Flag, Level 2

6.1.4 Evaluation and Interpretation of the results

6.1.4.1 Analysis Discussion

Based on the CRISP-DM methodology, the data understanding and data preparation stages of this study confirm several previously known patterns reported in the financial literature. Similarly to the work of (Aggarwal and Wu, 2006), it found that the average volume was lower for manipulated stocks and that during the manipulation period, liquidity, returns and volatility are higher for the manipulated stocks than for the controlling sample. Additionally, all the cases involve attempts to increase the stock prices, and 'potentially informed parties' fit the profile of the

more common sort of fraudsters. As in the work of (Carhart et al., 2002), this study corroborates that quarter ends and year ends, as well as closing hours, are common preconditions for the manipulations, again confirming the existence of higher liquidity, returns and volatility associated with the manipulated sample. Indeed, a high presence of returns, volatility and volume outliers is directly related to the manipulation sample; nonetheless, not all those outliers are directly linked to manipulations due to the moderating effect of new information as discussed in the work of (Van Bommel, 2003). These results are consistent with theoretical models of successful trade-based manipulations, first described by (Allen and Gale, 1992) and extended by (Aggarwal and Wu, 2006).

In the data modelling stage of the CRISP-DM methodology, the study goes beyond confirming previous patterns, and reports on the discovery of new patterns in the form of decision trees. Specifically, two levels of analysis were defined: on the first level, the work studies the differences in the intraday trades dataset using hourly trading data. These patterns were used to discriminate between normal and suspicious block hours of trading. On the second level, the paper focussed on the characteristics of specific trades during suspicious block hours. To describe them, the unit of time is changed to seconds and both trades and quotes data are analysed.

To understand the knowledge gained from the modelling stage it is crucial to exploit the easiness of interpretation of input variables in the form of ratios and z-scores and decision rules made on the base of them. In this work, commonly reported variables in the financial literature were expressed as ratios and z-scores; thus, the rules represented by the branches of the decision trees can be understood as a combination of the characteristics of the observations of the returns, liquidity and volatility in terms of how far from the mean — measured in number of standard deviations — these observations lay, and in which proportion of real cases these characteristics have been associated with a suspicious trade.

Particularly, analysing the three most active, or common, characteristics of suspicious trading at the hourly granularity it is possible to discover the following patterns: Figure 6-5, Rule 1 states that 22.4% of the suspicious blocks correspond

to observations where the ZS1 indicator was less than 1.6, and the RV3 was less than 3.1, and the ZAR1 was bigger than 3.2. In other words, when liquidity and volatility are within normal ranges (less than three standard deviations away from the mean), jumps in returns are associated with suspicious blocks in 22.4% of the cases. Following the same logic, Figure 6-5, Rule 2 states that 20.8% of the suspicious trades relate to normal levels of liquidity (ZS1<=1.6) but high levels of volatility (RV3>3.1). Figure 6-5, Rule 4 states that 21.5% of the suspicious trades relate to normal levels of returns (ZAR1<=2.5), but unusually high liquidity (ZS1>=3.9)

For the second level of granularity, Figure 6-6, Rule 3 states that for specific trades within hours flagged as suspicious in level 1, extreme values of liquidity (RS2>5.5) together with volatility in normal ranges (ZV1<=2) were associated with 25.1% of the abnormal trades. Moreover, Figure 6-6, Rule 5 states that 35.6% of the abnormal trades were related to the *Riskmetrics* volatility indicator in high ranges (2.3<RV3<=5.0) and the standard deviation volatility indicator in normal ranges (ZV1<=3.3). Lastly, Figure 6-6, Rule 6 states that 19.6% of the abnormal trades were related to high values (RV3<=5.0) for the *Riskmetrics* volatility and to equally high standard deviation volatility indicators (ZV1>3.3).

In terms of policy recommendations, these patterns confirm the importance of monitoring sudden changes in the volume of trading as these are very frequently associated with manipulations. This recommendation is followed closely in order of importance by the monitoring of the presence of high levels of volatility. In contrast, fewer efforts should be placed on monitoring changes in returns or abnormal returns, as these indicators were more susceptible to noise and, thus less important in describing suspicious trading within the context of this case study.

6.1.4.2 '2M-3S' Framework Application

This case addresses the challenge of building market monitoring and surveillance components, especially data mining components, that are working prototypes. These were built using real trading and textual data, and taking as a base real cases of manipulation. In particular, as described throughout the case, the emphasis on the data mining task was to build classification models whose rules

are humanly comprehensible, especially with the aim of allowing the researcher to understand the patterns in data and to compare them with previously reported patterns.

Relating the work to other aspects of the framework, Figure 6-7 presents a view of how an offline detection engine could be organised in order to deploy and use the C5.0 ‘decision trees’ result of the modelling stages. In particular, the respective data sources and data types are used as inputs to the information management sub-system, where different data pre-processing jobs take place. The output of this layer is then used in the offline detection engine by automatic and on-demand processes, which include the automatic scanning and reporting of news items and material information, outlier detection and, of course, pattern-match or rule-match recognition based on the embedded rules of the C5.0 and CR&T models constructed previously. This information is finally handled in the output management layer, which is the layer or sub-system in which suspicious trading is officially declared as a breach and an alarm is raised, initiating a formal investigation.

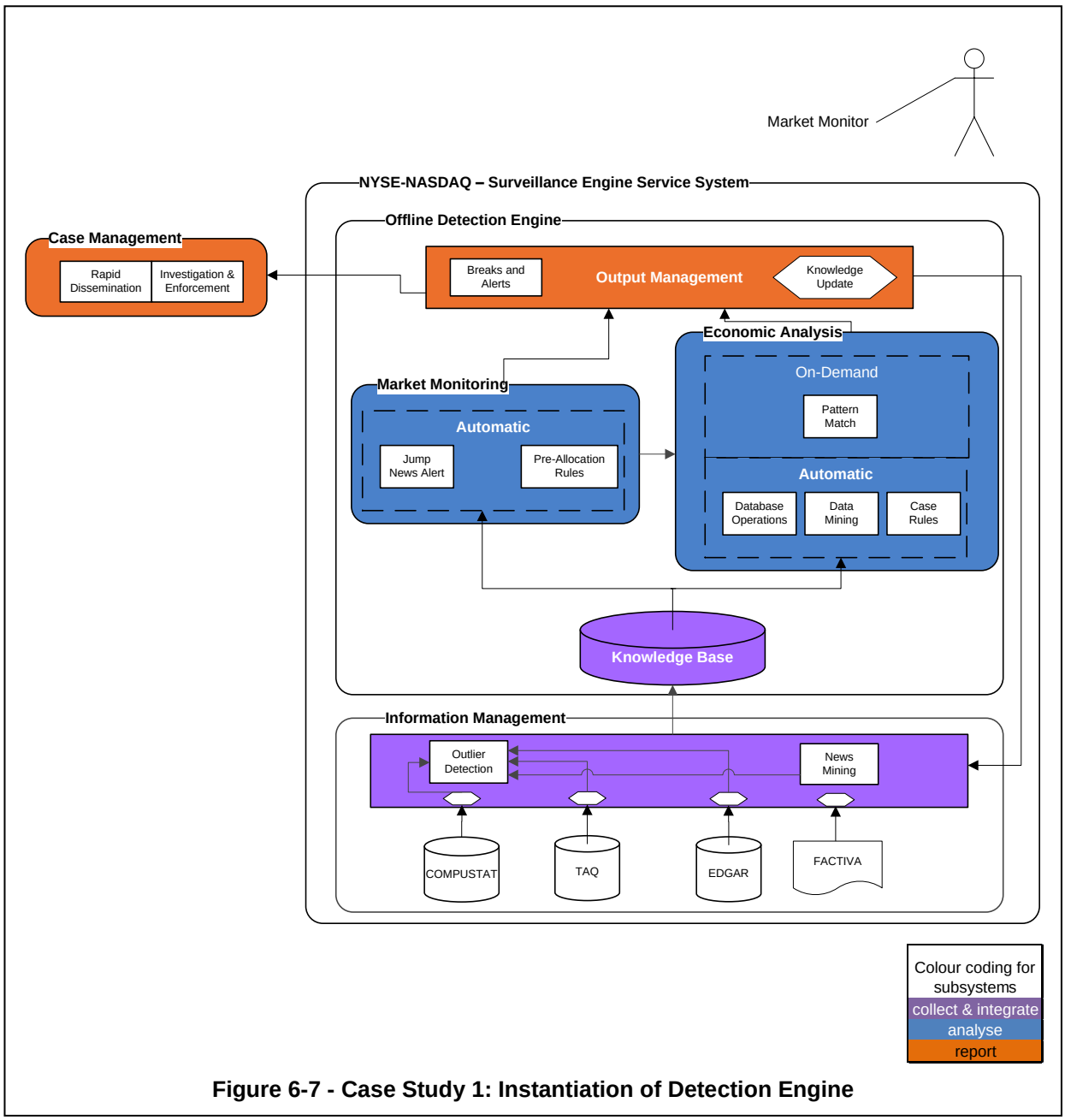


Figure 6-7 - Case Study 1: Instantiation of Detection Engine

In terms of the instantiation of the case as a service system, Figure 6-8 shows how a value proposition is presented from the SRO (the provider) to the Regulator (the customer), in terms of services for monitoring and surveillance that are performed by the exchange using the infrastructure and the decision trees that were described in the case details.

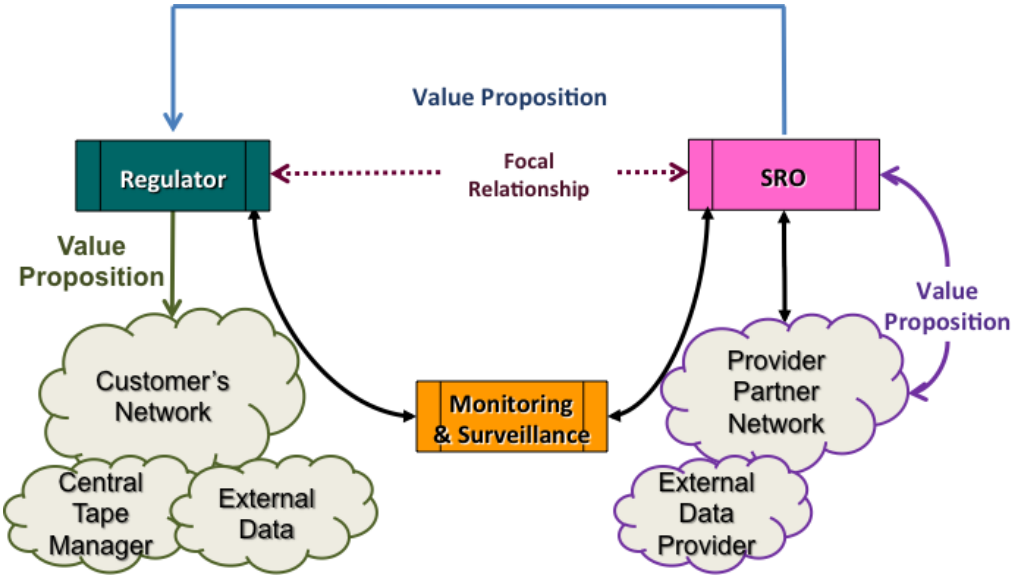


Figure 6-8 Case Study 1: Market Monitoring and Surveillance SVN.

The provider exerts its network of partners, in this case TAQ, FACTIVA and COMPUSTAT, to develop and deploy detection components in the form of decision trees (Figure 6-9). These and its own resources are utilised by the regulator, SEC, in an indirect way, superseding the quality and performance of the surveillance and monitoring that is performed in the exchange. As mentioned in Chapter 2, in those instances where the SRO has jurisdiction over the individual or firm alleged to have committed a violation, the SRO will typically take enforcement action. Nonetheless, the SEC may institute parallel proceedings if it deems such action warranted. In cases where the SRO does not have jurisdiction over the firm or person, the SRO will refer cases to the SEC (or Justice Department) for further investigation and enforcement action. This is most likely with respect to insider trading cases (Polansky et al., 2004).

Moreover, if the SEC identifies serious or systemic shortcomings in an SRO's detection, investigation or enforcement programme, the Commission will typically impose a sanction. Otherwise, the output of the SEC's inspections may be limited to recommendations as to improvements in process or technology dedicated to regulatory activities (Polansky et al., 2004). Figure 6-10 presents an alternative view of the case' world, in which the interactions between the exchanges, the regulators and other typical actors in the system are clearly visible.

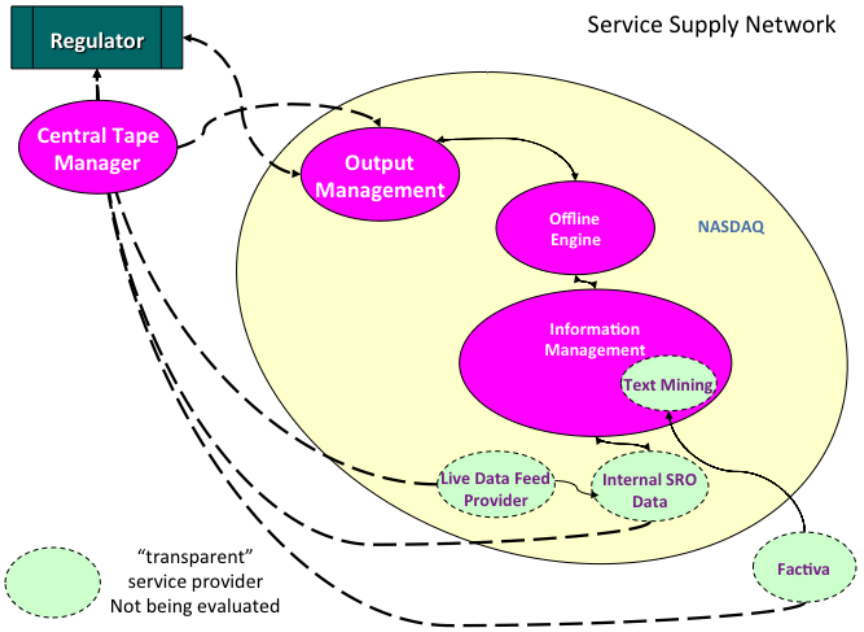


Figure 6-9 Case Study 1: Service System Worldview

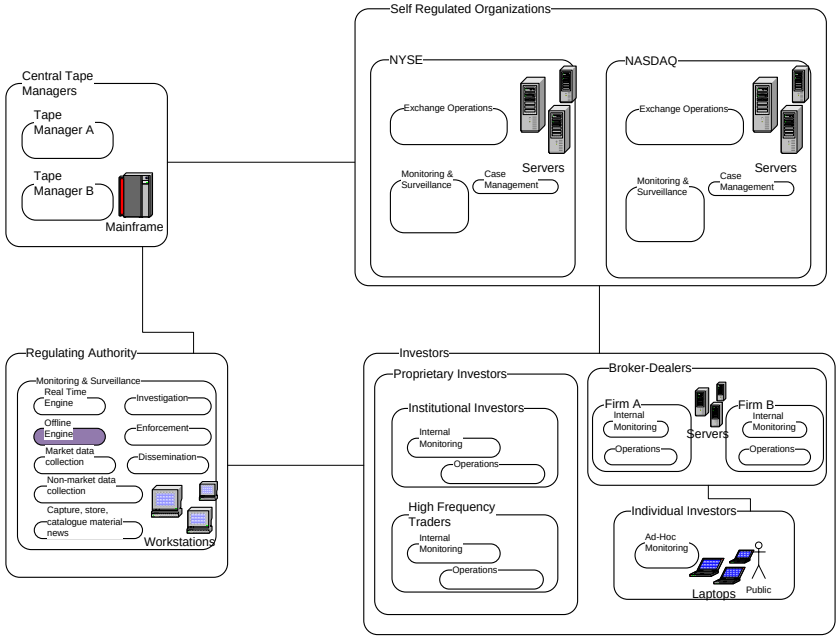


Figure 6-10 Case Study 1: Service System Alternative Worldview

In terms of the '2M-3S' framework taxonomy instantiation, Figure 6-11 helps summarizing the concepts and requirements that were addressed in this case. The terminology used is easily recognisable and mapped to clear definitions of manipulations and other concepts discussed in this thesis.

Manipulation							
Manipulation	**Cross-Jurisdiction**	**Agents**	**Action**	**Venue**	**Time**	**Target Asset**	**Effects**
Marking-the-Close; contract-based; trade-based	No	Outsider; On their own; own benefit	Buy/Sell trades and orders	SRO:NYSE Electronic Traditional and HFT supported	Normal speed, inside trading hours; End of day; Quarter End	Financial; Equity; Common stocks	Financial/Economic structured data; inside venue only

Stakeholders		
Investor	**Intermediary**	**Regulator**
Public; Fund Managers; net saver	Mutual Fund	SEC; FINRA

Detection Engine			
	Collect & Integrate	**Analyse**	**Report**
Information Management	Data preparation; outlier detection; news mining		
Off-line detection engine		Economic Analysis; Automatic & On demand; Decision Trees	Visualization; OLAP; Breaks and Alerts
Real-time detection engine	n/a	n/a	n/a

Requirements		
Requirements	**Class**	**Type**
1 to 3, 5, 8,	Analysis/Integration	Functional
9 to 12	Data capture and pre-processing	Functional
20, 21, 23, 25	Information Management/Output Management	Functional
n/a	General conditions of the system and environment	Non-Functional
n/a	Skills and capabilities of the customer	Non-Functional

Figure 6-11: Case Study 1 Taxonomy Instantiation

6.2 Case Study Two: 'Insider Trading' manipulations

6.2.1 Background

A version of Case Study 2, together with parts of Chapter 2, Chapter 4 and Chapter 5, was presented as a working paper under the title of 'A Systematic Framework for the Analysis and Development of Financial Market Monitoring Systems' at the SRII 2011 Global Conference, March 30-April 2, San Jose, California, USA, and it is expected that proceedings papers will be published by IEEE in a forthcoming special SRII/IEEE journal.

This section discusses a case scenario that considers the perspective of a regulating authority in the form of a market enforcer analyst. This type of user/customer normally has broad access to market data, including several unstructured sources such as news items or financial forms and filings. It is assumed that is confronted with the problem of raising early alarms for potential cases of manipulation, and thus would initially require powerful detection engines to pre-screen thousands of trades generated in a similar number of securities. The customer is interested in filtering those which present abnormal characteristics, such as sudden changes in trading volume, jumps in prices and/or drastic changes in volatility. Additionally, the customer is interested in relating the unusual trading activity to unusual traders' behaviour, such as an increase in the volume of insider trading, or links between coordinated traders performing last-minute purchases in order to 'mark-the-close' of or 'wash sale' a security.

The customer counts on the monitoring engine to create a watch-out list at the end of every trading day. A set of customer rules will then be applied to the list in order to pre-allocate suspicious cases into one of the predefined classes of possible manipulation cases The customer will then perform further investigative analysis utilising the different components available in the workbench of the detection engine. The customer expects the system to be reactive and semi-automatic given the fact that the final status of a case should be subject to interpretation and overruling from the financial authority. The customer also expects the system to

serve as a repository of historic cases and patterns of manipulations, information that could be used to learn from past cases and improve the results of its analysis components in the future.

This case study is based on a stock market manipulation case pursued by the US SEC that was collected by the researcher. In this case, the SEC filed a complaint alleging that *'James T. Anderson (Anderson), Zomax's former Chairman of the Board of Directors and Chief Executive Officer, and his wife, Michelle Bedard-Anderson (Bedard-Anderson), Zomax's former Executive Vice-President of Sales and Marketing, engaged in insider trading by liquidating their 821,250 shares of Zomax stock on the basis of material, non-public information that Zomax's third quarter 2000 revenue and earnings would be lower than current consensus estimates. Specifically, the Complaint alleges that in August 2000, Anderson and Bedard-Anderson sold 365,250 shares of Zomax stock in open market transactions. Shortly thereafter, Anderson and Bedard-Anderson used the Jim and Mikki Anderson Charitable Reminder Annuity Trust to sell their remaining 465,000 shares of Zomax stock. As a result, Anderson and Bedard-Anderson allegedly avoided losses of approximately $9 million. The Commission's complaint also alleges that Anderson tipped his friend, Neil Dolinsky, to sell his shares of Zomax stock on the basis of material, non-public information. As a result, Dolinsky allegedly avoided losses totalling $139,014'*. As alleged in the Commission's Complaint, Zomax violated Sections 10(b) and 13(a) of the Securities Exchange Act of 1934 (Exchange Act) and Rules 10b-5, 12b-20 and 13a-13. http://www.sec.gov/litigation/litreleases/lr19262.htm

The case study uses real trading data as well as a collection of unstructured sources for the Zomax Inc. stock (symbol: ZOMX), traded in the NASDAQ venue. In particular, the source of data for intraday transaction and quotes was the 'TAQ' database (Wharton Research Data Services, 2009) and it included all intraday transactions and bid and ask information for the stock for the year 2000. The intraday database was then populated with information relating to news and financial events collected in the Factiva database (Dow Jones Factiva, 2009). This included information released by the companies as material facts, insider

transaction form news and other types of news totalling 424 news items covering the 1999-2000 period. In addition, the COMPUSTAT database (Standard and Poor's Compustat Resource Center, 2009) was used to provide supplementary profiling financial information about the selected case, such as the SIC Code, market capitalization and the beta of the stock. COMPUSTAT was also used as the source for market returns in the form of the S&P500 daily index. The Thomson Reuters Insider Filings Data Feed (IFDF) database was used as a source for all US insider activity as reported on Forms 3, 4, 5, and 144 in line-by-line detail totalling 942 items covering the 1999-2000 period. The IFDF is available at the Wharton Research Data Services (Wharton Research Data Services, 2009).

Further detailed information regarding the proceedings actions initiated by the SEC regarding this case can be found in the EDGAR online database at http://www.sec.gov/litigation/complaints/comp19262.pdf
http://www.sec.gov/litigation/litreleases/lr19262.htm

6.2.2 Data Modelling and Analysis

Initially, the information management layer received and stored data from the various sources assuming a day-by-day granularity. More specifically, second-by-second intraday data was aggregated into hourly representative observations and data from structured data sources, such as COMPUSTAT and IFDF, were stored in a relational database. Using post-processed data, different indicators such as abnormal returns, average volume, volatility of returns or cumulative number of Form 4 and Form 144 filings, were calculated and then used by the analysis components inside the case monitoring engine. Similarly, unstructured data sources, namely, news items coming from Factiva, were stored in the same database, maintaining whenever possible their original xml tags. Different text mining analysers were built in order to extract key concepts from the textual sources and to find categories or clusters in Form 144 information. Consequently, new standardised and clean data was available on an 'as if daily' basis, updating the knowledge database component inside the engine respectively.

More specifically, as shown in Figure 6-18, the information management layer included a range of data preparation tools components; normal database manipulation operations components, such as query and OLAP reporting; and other data and text mining components. The different tasks and components were developed following the guidelines of the CRISP DM reference model (Chapman et al., 2000) using the IBM-SPSS PASW13 data and text mining workbench. As is usual in any data mining task, work was divided in six stages: business understanding, data understanding, data preparation, modelling, evaluation and deployment.

The economic analysis uses the output from the news mining component that extracts a list of key concepts mentioned in the available news of the day in order to give some insight on the frequency and content of the news and the number, type and frequency of the insider filling forms submitted to SEC as in the work of (Zaki et al., 2010). An additional data mining module performs outlier detection analysis using the financial data. The strategy considers a two-stage mining approach, implementing a sequence of outlier detection algorithms and decision trees (unsupervised and supervised learning techniques) as presented in Case Study 1. This two-stage strategy helps to discover outliers in price, volume and volatility indicators that are not associated with the existence of news items. In addition, the decision trees algorithms allows the customer not only to signal suspicious outliers in financial indicators, but also to better understand the patterns of trading associated with normal and suspected manipulation trades and funds as in Case Study 1.

The behavioural analysis utilises the data from the insider filing forms by performing traditional database operations such as a queries and OLAP cube analysis, and also an association rules data mining component, following a similar mining strategy to the one proposed in the work of (Mongkolnavin and Tirapat, 2009). The objective of these components is to find insight in potential associations between identity tags of insiders, trading information such as number of stocks or derivatives traded or intended to trade, the broker used to execute the transaction, and the time and date of the transactions. If pairs of identity tags are associated

with particular groups of insiders or brokers, in a given time framework, say at or near the closing time of trading days or before a material news release, this could be a preliminary signal of manipulation.

The case scenario assumes as the starting point the moment when the monitoring system triggers an automatic alarm after detecting a sharp drop in the price of the ZOMX stock together with a sudden increase in the trading volume during 22nd September 2000. Predefined customer-rules pre-allocate this suspicious case to the 'insider trading' category after checking that the weekly total of Form 4 filings submitted by insiders was higher than the average for the previous year. Then, the provider's analyst starts an investigation by performing different analysis tasks as discussed in the following stages.

6.2.3 Analysis and Results

6.2.3.1 Economic Analysis

Based on the financial theory, markets respond to the appearance of new information, and thus stocks that were mentioned in the news of the day are likely to show some reaction or effect that could be captured either as a significant move in returns, volume of trading or change in volatility. In this sense, the existence of news items relating to a particular security is relevant to help discriminating between those that react due to the appearance of new information and those that react due to the effects of manipulations. In other words, significant movements in specific financial indicators that are not associated with the appearance of new information can be seen as another symptom of manipulation (Van Bommel, 2003). Nonetheless, insiders with access to non-public material information could trade before the new information is released to the public, gaining an unfair advantage with respect of the rest of the market.

As a first analysis task, then, the analyst produces two graphical views of the trading activity on the ZOMX stock. Figure 6-12 and Figure 6-13 present an example of the type of views that could be produced by the analyst. In particular Figure 6-12 presents the analyst with the price performance of the ZOMX stock overlaid with the values of the alarm indicator represented by the size and the

colour of the points in the graph. A quick view provides the information that this is the only critical alarm (red big circles) that has been raised for this stock during the previous months.

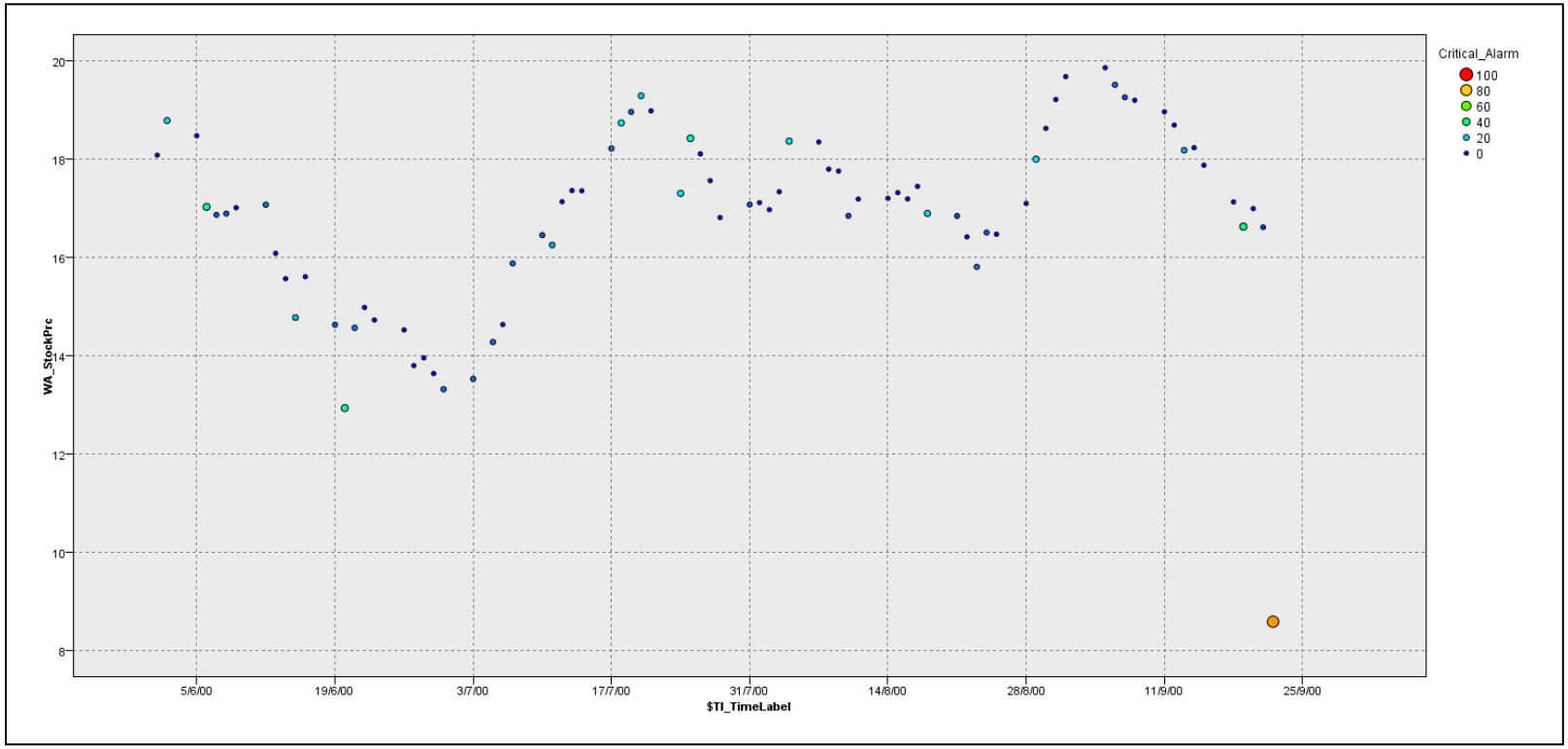


Figure 6-12 Case Study 2: Data Mining Outlier Detection Engine Visualisation

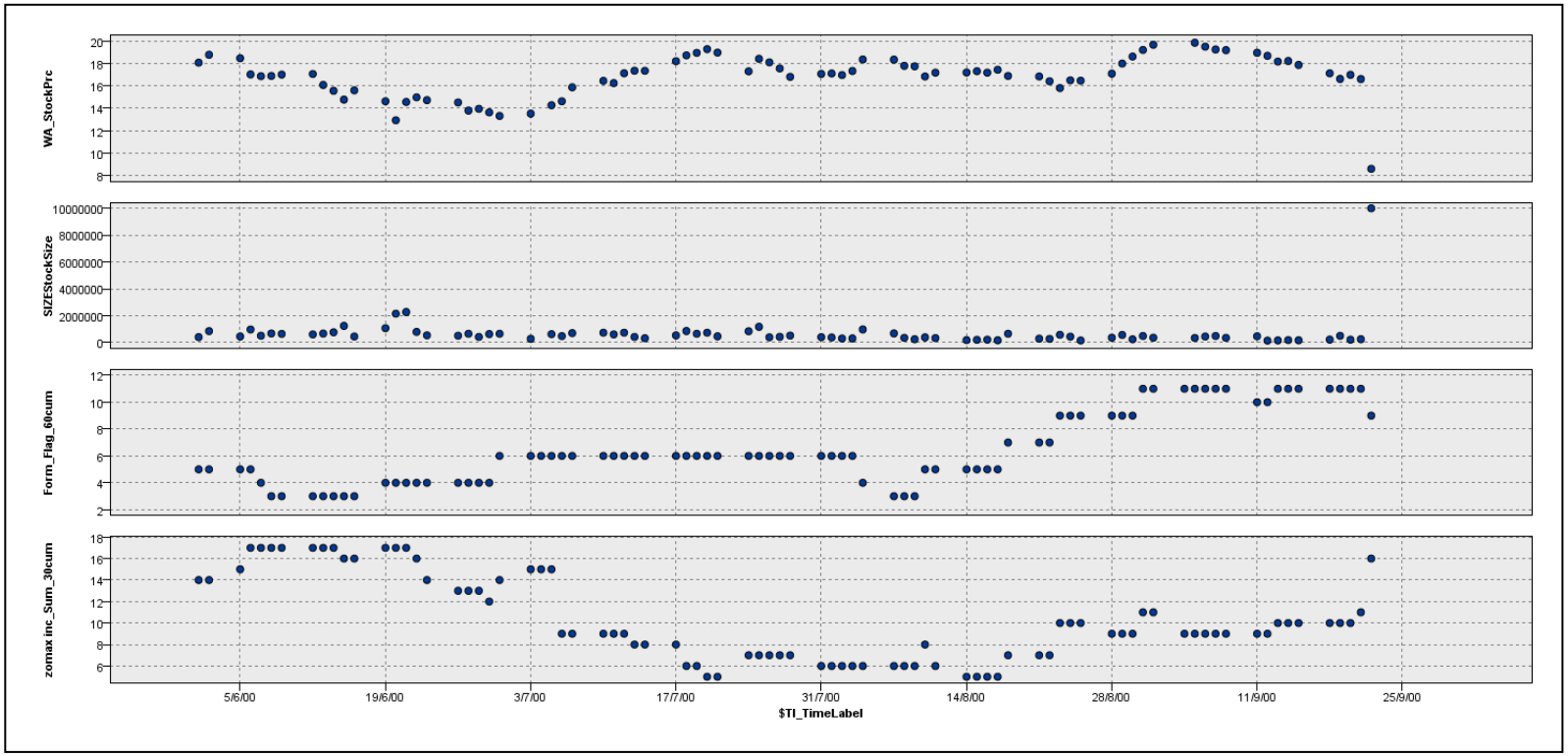
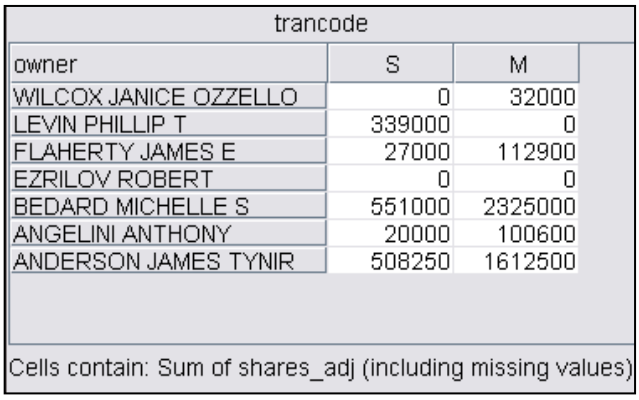


Figure 6-13 Case Study 2: Time Series Visualization of Key Indicators

Figure 6-13 presents a summary view of four time series for the recent months (5th June to 22nd September 2000). From top to bottom, the first two panels show a line plot of the behaviour of the price and trading volume associated with the ZOMX stock. The third panel shows the cumulative 60 days total of insider filing forms submitted to SEC. The fourth panel shows the cumulative total number of news items, of the last 30 days, released up to the corresponding date. It is possible to observe that the drop in price came together with huge increase in trading volume, and that effectively the number of insider filings had been increasing during the last month. The number of news items did also present a peak during the previous day. Based on the expected patterns suggested by financial theory, the analyst studies a sample of the news items released on 21st September (see Appendix 4) and concludes that the reason behind the increase in the volume of trading and the drop in prices was that the company announced that its third quarter 2000 financial results would be lower than current consensus estimates. Additionally, the analyst confirms the customer-rule criteria by also formally considering the case as a suspected insider trading case.

6.2.3.2 Behavioural Analysis

The next step of the analyst's investigative process is to find out which insiders have been actively trading on the ZOMX stock or derivatives before the release of the material news. The analyst performs a database query operation to create an OLAP report. Figure 6-14 presents the output of this query. Based on this output the analyst discovers that Michelle Bedard (BM) Vice President, and James Anderson (AM) Chief Executive Officer, have been the two most active insiders. Column 'S' corresponds to the total number of stocks sold by the insiders, and Column 'M' to the total number of stock options executed during the first 9 months of year 2000, information that was reported mainly in Forms 4.



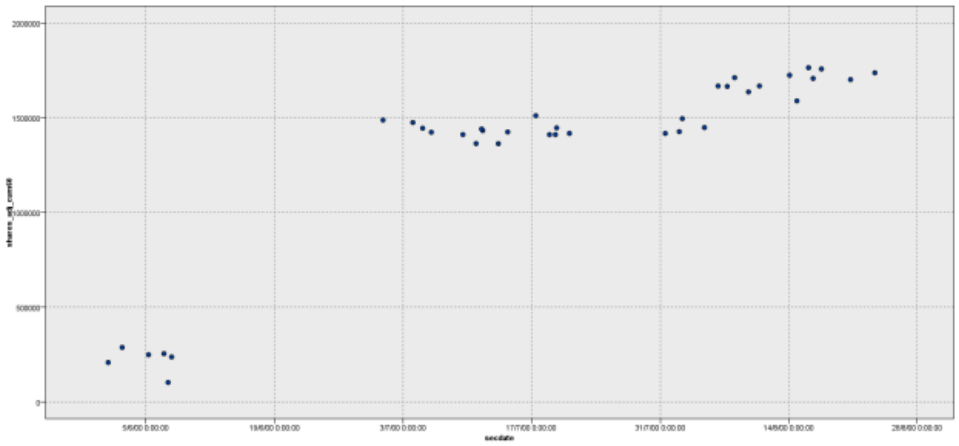
	trancode	
owner	S	M
WILCOX JANICE OZZELLO	0	32000
LEVIN PHILLIP T	339000	0
FLAHERTY JAMES E	27000	112900
EZRILOV ROBERT	0	0
BEDARD MICHELLE S	551000	2325000
ANGELINI ANTHONY	20000	100600
ANDERSON JAMES TYNIR	508250	1612500

Cells contain: Sum of shares_adj (including missing values)

Figure 6-14 Case Study 2: OLAP Analysis IFDF.

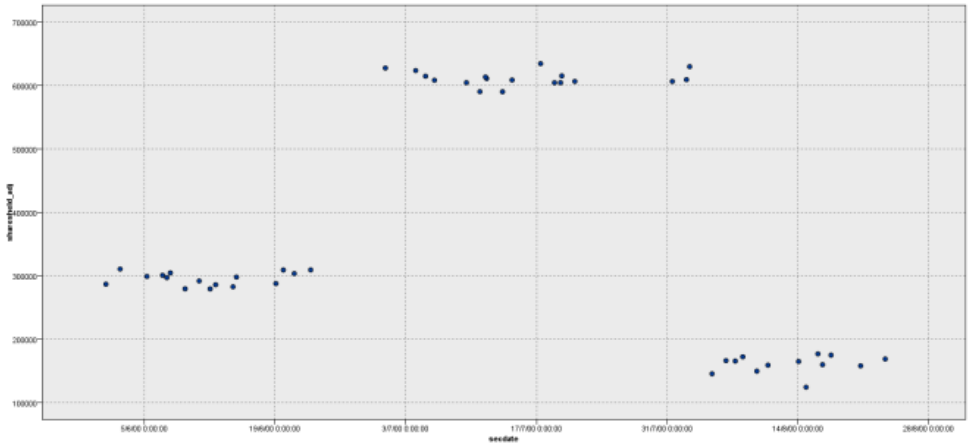
Next the analyst is interested in the trading evolution of the suspicious insiders. The analyst produces two views, one for each insider, summarising the information on total sales and total number of shares held or declared as held by the insiders. Both insiders have been increasing their number of stocks sold as can be seen in panels (a) of Figure 6-15 and Figure 6-16. More importantly, both insiders have declared a drastic reduction in the number of stocks held by them as can be seen in panels (b) of Figure 6-15 and Figure 6-16.

Michelle Bedard



Cumulative Sells Indicator Form 4 - (a)

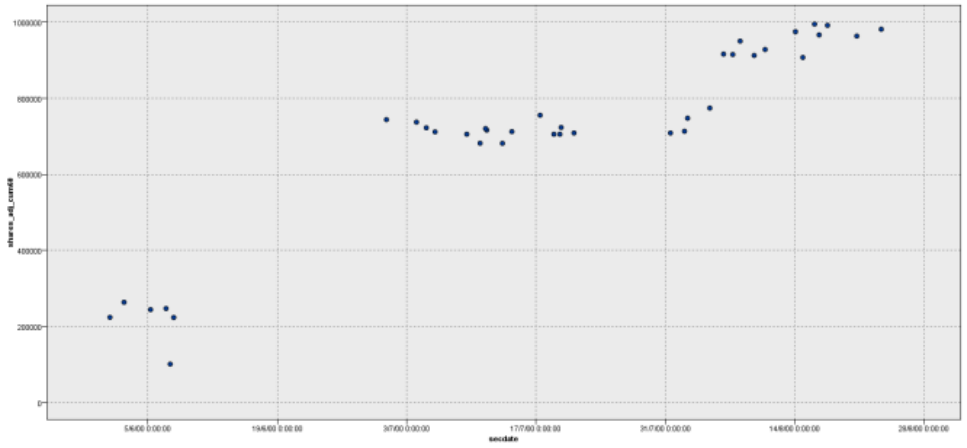
Michelle Bedard



Shares Held Form 144 - (b)

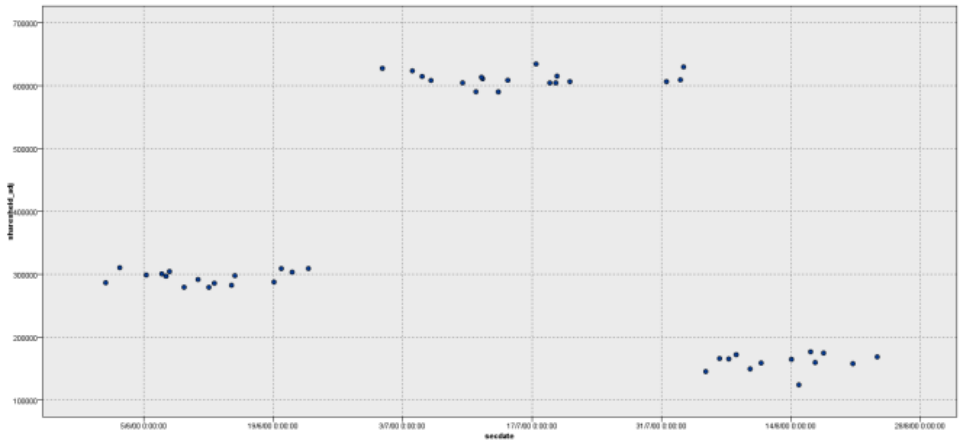
Figure 6-15 Case Study 2: Visualization Analysis of Insiders Activity (BM)

James Anderson



Cumulative Sells Indicator Form 4 - (a)

James Anderson



Shares Held Form 144 - (b)

Figure 6-16 Case Study 2: Visualization Analysis of Insiders Activity (JA)

The analyst is interested to check if the insiders have been acting in coordination with each other or with other agents, for example selling their stocks during the same days or periods. The analyst performs an association and co-occurrence of events analysis using the respective component. Figure 6-17 presents the graphical representation of the association analysis. As can be seen in Figure 6-17, it is possible to observe a strong relationship between the date when the trades were made and the broker used by the insider to place their trading. For

example, the bottom blue connection shows a strong association (from right to left) between BEDARD MICHELLE – ANDERSON JAMES – a group of trades occurred during July and August – and “rj steichen” and “charles schwab” brokers.

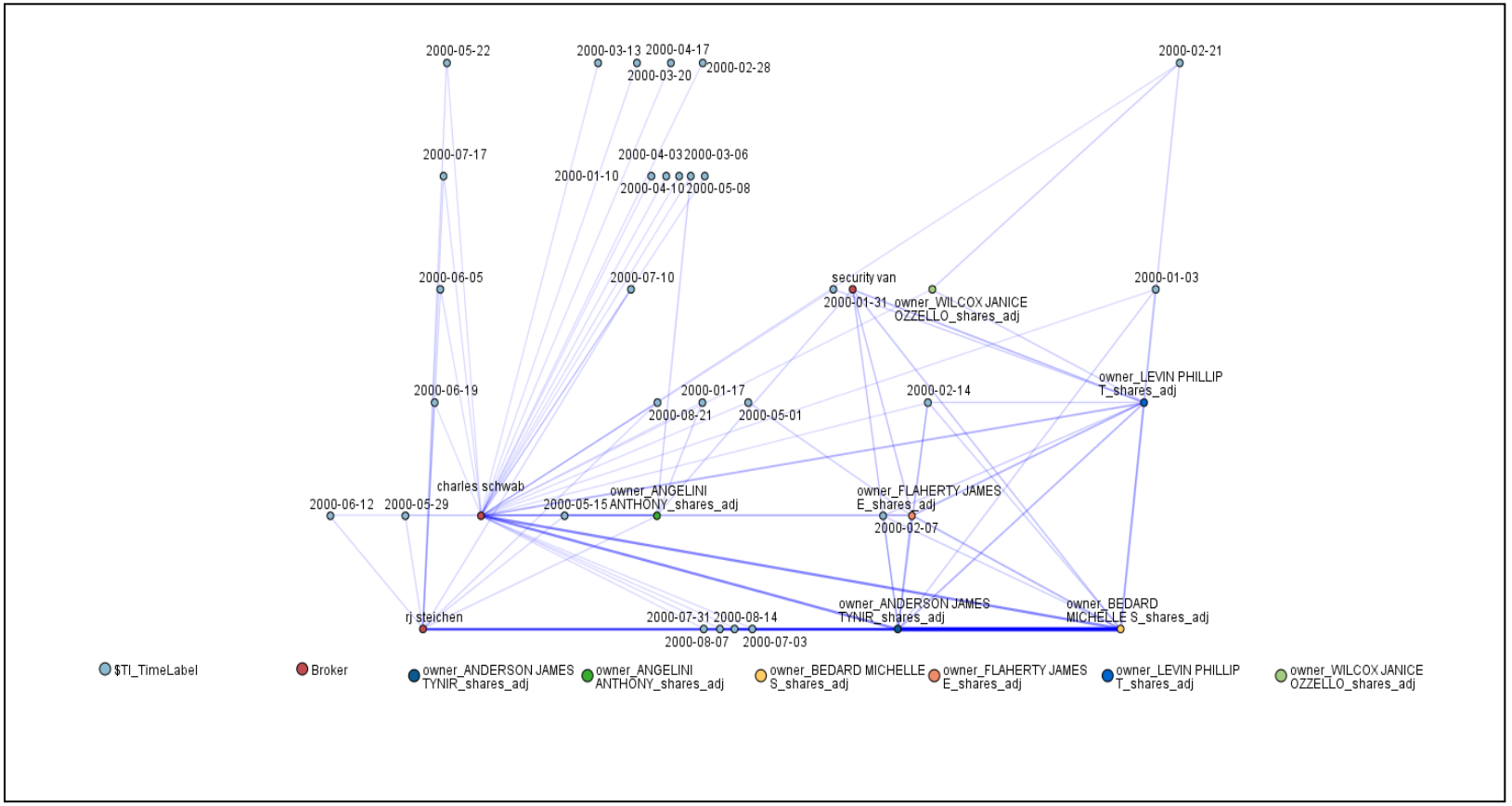


Figure 6-17 Case Study 2: Forms 4 and 144 are Used to Explore Associative Relationships

6.2.4 Evaluation and Interpretation of Results

6.2.4.1 '2M-3S' Framework Application

This case addresses the challenge of building market monitoring and surveillance components, especially data mining components, that are working prototypes for the detection of insider trading manipulations. Again, these were built using real structured and unstructured data, and taking as a base a real insider trading case. In particular, the difficulty of insider trading detection is the ability of the system to discriminate between other types of market manipulations, in particular trade-based manipulations, as these tend to have similar patterns, i.e. jumps in prices or volume. The main difference, then, is that in a typical trade-based manipulation these jumps are not preceded or caused by a material news item impacting the markets. Detection systems, then, should learn to identify jumps which originate from news items, and also, to automatically create 'hot-lists' of insiders who have been trading actively before the release of the material news to the public. Furthermore, if strong associations between insiders can be made, then this could be interpreted as a sign of coordination between insiders to take advantage of the material news they possess. Another important aspect of insider trading is the possibility of manipulators who target both the derivative and the stock of the company for which they possess material information, mainly because of the great leverage and reduced capital requirements when trading in the derivative markets. This aspect, however, is not touched in this case and is left for further research. Relating the case to other aspects of the framework, Figure 6-18 presents a view of how the insider-trading-capable detection engine could be organised.

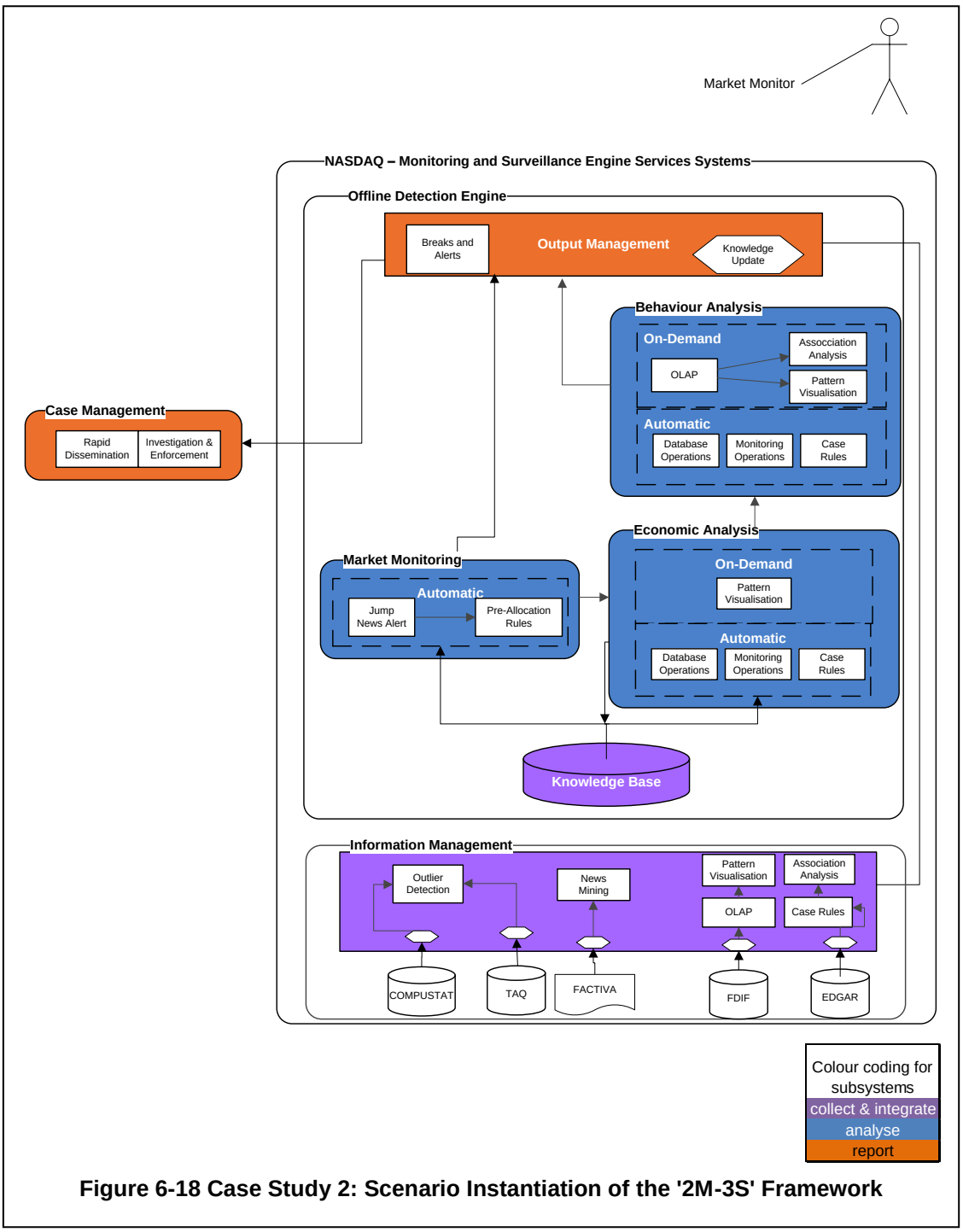


Figure 6-18 Case Study 2: Scenario Instantiation of the '2M-3S' Framework

In terms of the instantiation of the case as a service system, the case considers a similar configuration of value networks and value propositions as in Case Study 1, thus, Figure 6-8 can also be used as a reference for the basic service configuration. Again, a value proposition is presented from the SRO (the provider) to the Regulator (the customer), in terms of services of monitoring and surveillance

that are performed by the exchange, expanding in the case the value proposition with the detection of insider trading. Similarly, Figure 6-19 shows a possible configuration of SVNs for both the customer (regulator) and provider (exchange), which is expanded specifically by considering the inclusion of The Thomson Reuters Insider Filings Data Feed (IFDF) database as part of the providers' partner network.

Figure 6-10 is still a valid alternative view of the world, in which the interactions between the exchanges, the regulators and other typical actors in the system can be observed.

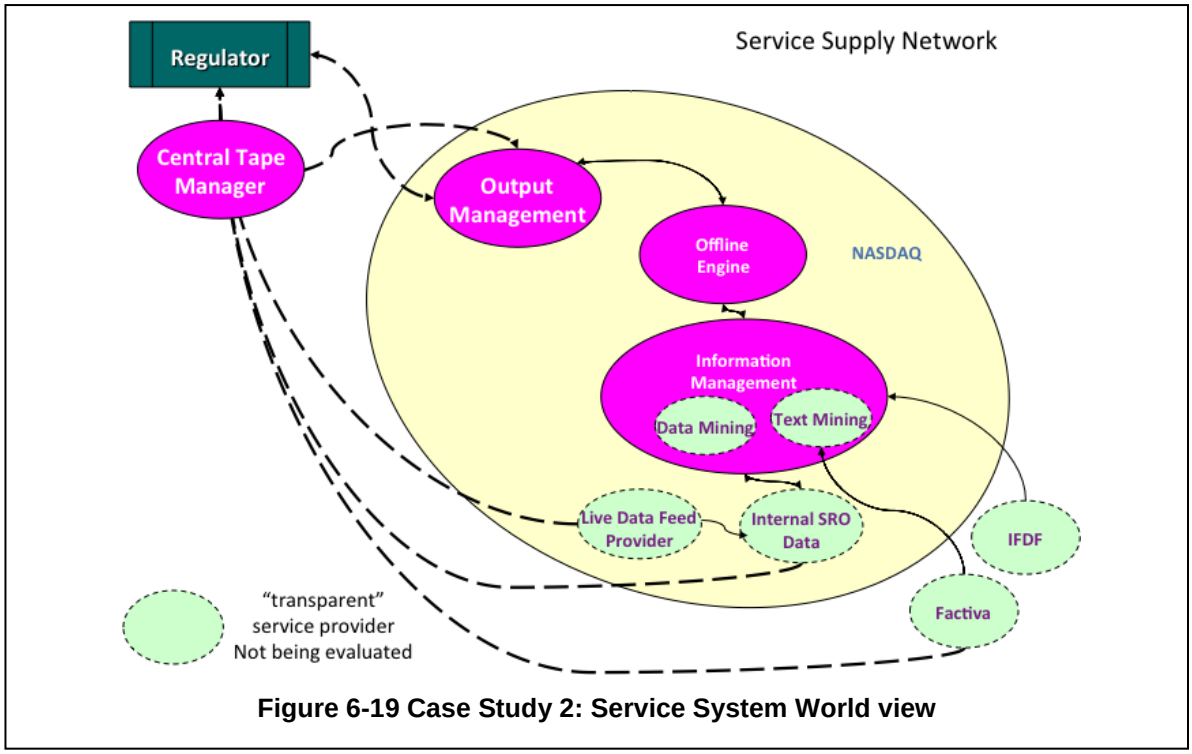


Figure 6-19 Case Study 2: Service System World view

In terms of the use of the 2M-3S framework taxonomy from Figure 6-20 it is worth noticing the inclusion of new requirements, such as, requirements 4 and 17 that relates to the necessary components to support behavioural analysis, in this case association analysis components. Other terminology are again easily recognisable and mapped to the definitions of manipulations and other principles discussed in this thesis.

Manipulation							
Manipulation	**Cross-Jurisdiction**	**Agents**	**Action**	**Venue**	**Time**	**Target Asset**	**Effects**
Insider trading; breach of fiduciary duty; information-based	No	Insider; In collusion; own benefit	Buy/Sell trades and orders	SRO:NASDAQ Electronic and HFT supported	Normal speed, inside trading hours; Quarter End	Financial; Equity; Common stocks. Derivates; stock-options	Financial/Economic; structured and unstructered data; inside venue only

Stakeholders		
Investor	**Intermediary**	**Regulator**
Executive; net saver	Broker-Dealers	SEC; FINRA

Detection Engine			
	Collect & Integrate	**Analyse**	**Report**
Information Management	Data preparation; outlier detection; news mining, previous patterns		
Off-line detection engine		Behavioural and Economic Analysis; Automatic & On demand; Decision Trees; Association Analysis	Visualization; OLAP; Breaks and Alerts
Real-time detection engine	n/a	n/a	n/a

Requirements		
Requirements	**Class**	**Type**
1 to 5, 8,	Analysis/Integration	Functional
9 to 12	Data capture and pre-processing	Functional
17, 20, 23, 25	Information Management/Output Management	Functional
n/a	General conditions of the system and environment	Non-Functional
n/a	Skills and capabilities of the customer	Non-Functional

Figure 6-20 Case Study 2: Taxonomy Instantiation

6.3 Case Study Three. Using social networking tools for the detection of 'Cross-Jurisdiction' 'Wash-Sale' and 'Match-Trade' manipulations

6.3.1 Background

This case introduces features of cross-border/cross-jurisdiction manipulations as well as social network analysis tools used to study them. This case differs from previous case studies, mainly because although the case scenario is real, part of the dataset used to build and demonstrate the detection engine is generated by the researcher. The case assumes that the fundamental task to be executed is the detection of manipulative and fraudulent behaviour by a group of fraudsters using off-shore accounts.

This case follows one of the definitions for cross-jurisdiction manipulations, where this can be understood as a deliberate attempt to interfere with the fairness of market operations that involves any type of 'breach of fiduciary duty', 'runs and raids', 'abuse of market power' and 'contract-based' manipulations, and where a security is traded in one jurisdiction but is the target of manipulative behaviour originating from manipulators based in other jurisdictions.

The challenge, then, is not only the detection of anomalous trading, such as jumps in price and volume of trading, but also the association of groups of investors performing coordinated manipulations based in different jurisdictions with anomalous trading activity. This task is particularly difficult because these groups try to conceal their fraudulent behaviour using different nominees accounts, usually registered across several jurisdictions. Manipulators try to give the impression of unrelated trading, which is in reality orchestrated by one or more coordinators with the intention of controlling a security. The analyst in this case will try to find connections between accounts which in probability should not occur, that are related to jumps in prices and volume. It will also consider a component for social networking analysis and graphical visualisation of these patterns.

This case study is based on a manipulation case pursued by the US SEC that was collected by the researcher. In particular, SEC filed a complaint alleging that from

at least the spring 2004 until 17th September 2008, George Georgiou (the 'manipulator') engaged in a series of fraudulent schemes to manipulate the market for the common stock of four microcap companies. Georgiou, a Canadian citizen, controlled a large percentage of the unrestricted stock of these companies. In separate but related schemes, he orchestrated and directed manipulative trading in these stocks by using numerous nominee accounts held at offshore broker-dealers, and using a variety of traditional manipulative techniques, including executing or directing matched trades, wash sales, or other prearranged trades, marking the close, and paying illegal kickbacks in exchange for purchases.

Georgiou conducted these manipulations with the intention of artificially inflating each company's stock price and to create the false impression of an active and liquid market for these otherwise thinly traded stocks. Georgiou profited from his scheme by selling inflated shares at a substantial profit, and defrauding offshore broker-dealers by obtaining margin loans, using the manipulated stocks as collateral, that further funded his manipulations and allowed him to withdraw cash that he wired to offshore bank accounts. Through his illegal conduct, Georgiou realized at least $20.9 million in ill-gotten gains. As alleged in the Commission's Complaint, Georgiou violated Section 17(a) and Section 10(b) of the Securities Exchange Act of 1934 ("Exchange Act") and Rule 10b-5.

The case study uses real trading data for one of four microcap stocks mentioned in the legal proceedings, i.e. data for Hydrogen Hybrid Technologies, Inc. This stock has been quoted on the OTCBB and Pink Sheets since 26th June 2007 under the trading symbol 'HYHY', and is listed on the Frankfurt Stock Exchange. The reason behind this selection is mainly due to data availability since microcap stocks, which trade in OTC or Pink Sheet markets, tend to be small companies which are thinly traded, and also because OTC and Pink Sheet markets are non-SRO, and thus not obliged to store pre-trade data (quotes). Moreover, post-trade data is usually also scarce and of diverse quality. In fact, the researcher found that consulted data providers for OTC and Pink Sheet markets only kept data starting from 2008, and thus it was not possible to get data for the other stocks mentioned in the case, as these occurred prior to 2008.

For HYHY it was only possible to obtain trades data, not quotes, from 5th May 2008 to 31st December 2008. The data was obtained from NASDAQ on-demand data services (NASDAQ OMX Group, 2011). It included all intraday transactions, and as is common for these types of datasets, did not include a column or label indicating the broker accounts from which these trades were originated. The Thomson Reuters Insider Filings Data Feed (IFDF) database, the COMPUSTAT database (Standard and Poor's Compustat Resource Center, 2009), and the Factiva database (Dow Jones Factiva, 2009) were used to check for insider trading information, profiling information, and corporate news information respectively, but it was found that no relevant records existed that could help the elaboration and presentation of this particular case.

Further detailed information regarding the proceedings actions initiated by the SEC regarding this case can be found in the EDGAR on line database at http://www.sec.gov/litigation/complaints/2009/comp20899.pdf

6.3.2 Data Modelling and Analysis

As in the previous case studies, trading data was pre-processed and stored assuming a day-by-day granularity, and second-by-second data was aggregated into hourly observations. Using post-processed data, different indicators, such as abnormal returns, average volume, and volatility of returns were calculated and then used by the analysis components inside the monitoring engine. In contrast to previous cases, and in order to specifically study the behaviour of broker's accounts, two columns were added to the original dataset. Each trade was manually assigned a simulated 'buyer account' and 'seller account' ID respectively, following as close as possible the description of the coordinated trading activities described in the legal proceedings. As in previous cases, the different tasks and components were developed following the guidelines of the CRISP DM reference model (Chapman et al., 2000) using the IBM-SPSS PASW13 data and text mining workbench. Distinctively, in order to enhance the social networking analysis capabilities NodeXL open source software (Smith et al., 2010) was used in conjunction with IBM-SPSS PASW 13.

The economic analysis component included a data mining module performing outlier detection in the trading data. The modelling followed the two stages approach, implementing a sequence of abnormality detections followed by C5.0 and support vector machine voting algorithms (unsupervised and supervised learning techniques).

Differing from previous cases, the behavioural analysis utilised the combination of real and simulated data to perform social networking analysis on particular days and trading hours that were previously flagged as suspicious by the economic analysis component. In particular, the social networking analysis component was used to look for suspicious formations of groups or coordinated trading activities, which, for example, generated extra volume or moved the prices significantly in specific days.

Although the objective of finding associations between records is similar to the one presented previously, the technique used to find and study these patterns differs greatly. In Case Study 2, for example, the association between traders and broker accounts was discovered in IFDF (insider trading data) with the use of OLAP tools and association analysis algorithms. In Case Study 3, by contrast, the tool used is NodeXL, which comprises a full range of social networking analysis components, not only for graphical representation but also for the calculation of graph metrics, clustering of interesting groups, sub-groups analysis, among other tasks, which gives the researcher a full workbench to study the social networks' patterns found in the data. The objective of this strategy is to gain more understanding of identity and location of the broker and broker account that conducted the suspicious trading.

6.3.3 Analysis and Results

6.3.3.1 Economic Analysis

Firstly the detection engine performed an automatic screening of trading hours in order to look for suspicious jumps in prices or volume. Figure 6-21 presents the price performance of HYHY overlaid with the values of the alarm indicator represented by the size and the colour of the points in the figure. Figure 6-22

presents a summary view of six time series for the period under consideration. From top to bottom, the first two panels show a line plot of the behaviour of the price and trading volume associated with the HYHY stock. Time series three to six show indicators that are used by the data mining indicator to signal a particular trading block hour as suspicious, the result of which is visible in Figure 6-21 as alarms. The third panel presents the behaviour of the z-score returns indicator, while the fourth panel presents the behaviour of the z-score volume outliers indicator and the fifth panel presents the z-score number of trades outliers indicator. Panel six presents an alternative long-term definition of the z-score return indicator.

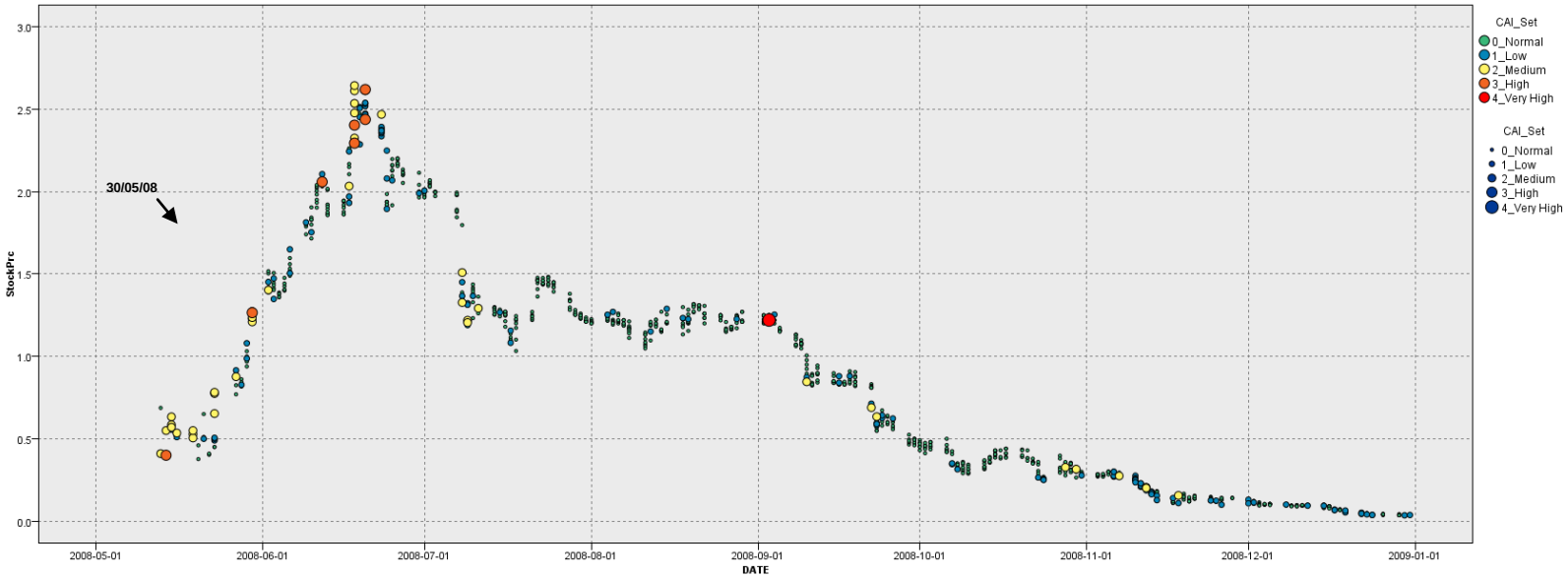


Figure 6-21 Case Study 3: Data Mining Outlier Detection Engine Visualisation

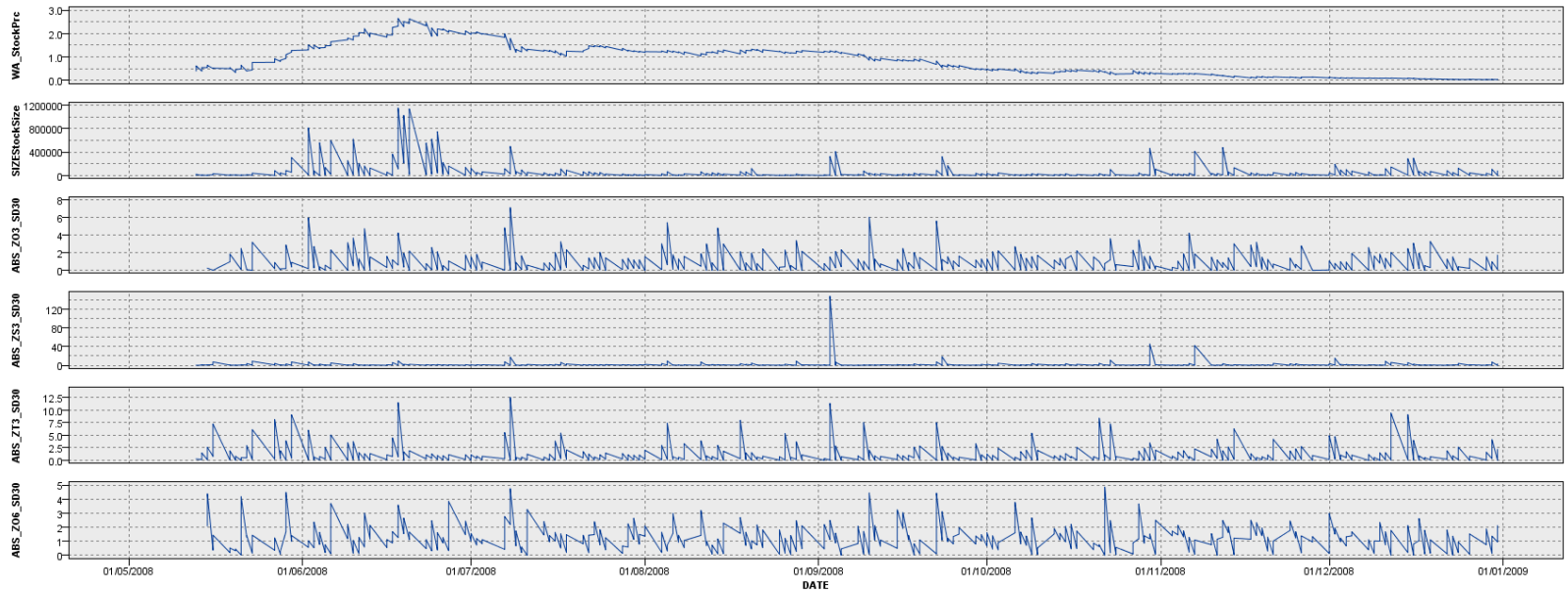


Figure 6-22 Case Study 3: Time Series Visualization of Key Indicators

In terms of case scenario interpretation, it is possible to observe that prices and volumes started to increase abruptly by the end of May, beginning of June 2008. This behaviour is captured nicely by the 'High' and 'Very High' alarms and other indicators which show several peaks and maximum values especially during June 2008.

6.3.3.2 Behavioural Analysis

Most of the description of the details of the legal case is taken literally from the proceedings. In order to be precise in describing what was known and what was assumed by the researcher, the text differentiates between what was factually obtained from the case record and what is the researcher's assumption, stating when necessary this difference. According to the legal proceedings, Georgiou implemented comprehensive and sophisticated schemes to manipulate the stocks of four companies: Avicena, Neutron, Hydrogen Hybrid and Northern Ethanol. Although the manipulations of each of these stocks occurred during different but overlapping time periods, the manner and means employed to conduct the schemes were similar. In particular, Georgiou asserted direct control over several nominee accounts for which he was able to issue trading and wiring instructions directly to broker-dealers. Some of those accounts were held at three Bahamian broker-dealers: Alliance Investment Management ('Alliance'); Accuvest Limited ('Accuvest'); and Caledonia Corporate Management Group Limited ('Caledonia'). Some other accounts were held in the Turks and Caicos Islands. Georgiou, along with an associate, opened the Alliance accounts, the Accuvest account, and the Caledonia account, and Georgiou personally directed the trading decisions in these accounts. Georgiou also had indirect control over certain nominee accounts. For those accounts, he communicated trading instructions to nominees who, in turn, executed Georgiou's trading instructions. At least one of the accounts that Georgiou indirectly controlled was held at Temple Securities. Others were held at various broker-dealers in the name of Fercan Developments ('Fercan'), Starport Landing, Inc. ('Starport') and in the names of individuals identified as 'RB', 'JA', and 'KG'. These were among the most active accounts in placing manipulative trades in

the stocks mentioned above. Another individual, identified only as 'CW', agreed to trade at Georgiou’s direction and to enlist other nominees to do the same in exchange for cash payments and guaranteed trading profits. That individual, who later became a cooperating witness, was also significantly involved in each of the manipulations.

In order to help the reader, the NodeXL software was initially used to depict this social network, including its head and main coordinator, 'Georgiou', as well as the main accounts, nominee accounts, and other individuals mentioned in the legal case.

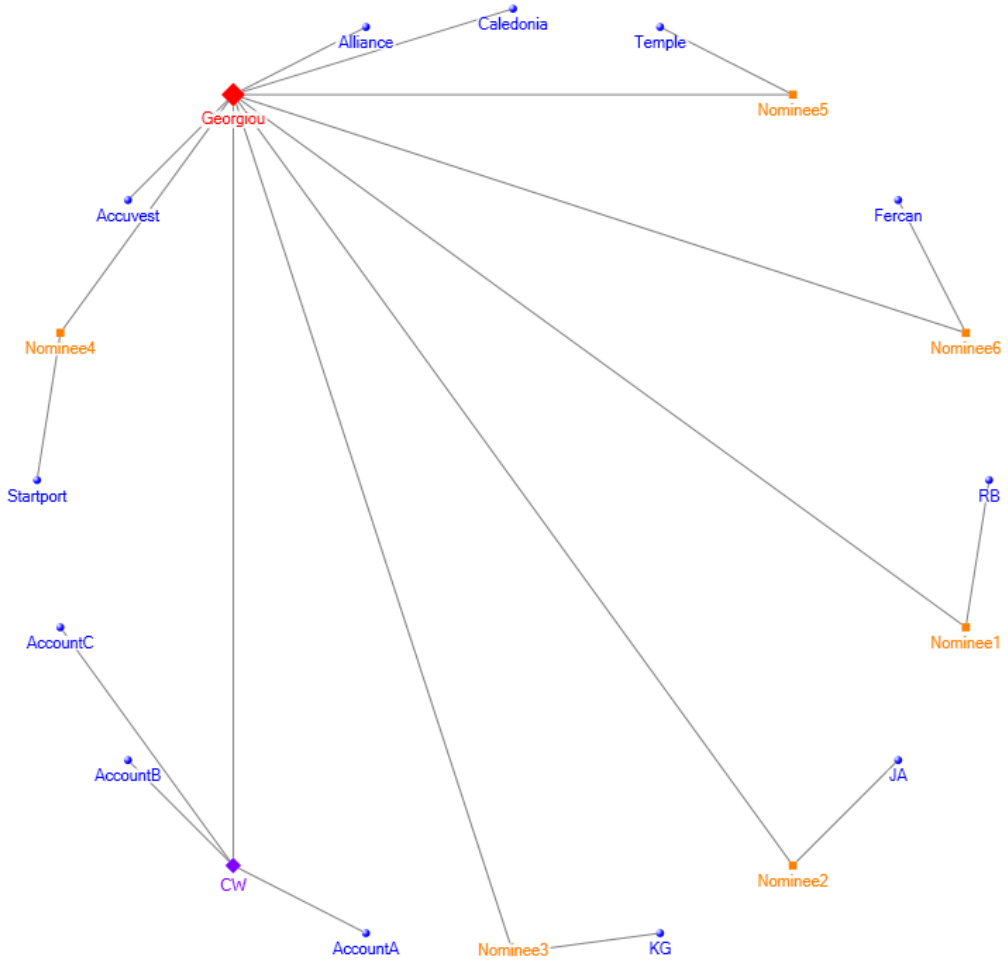


Figure 6-23 Case Study 3: Social Network of Manipulators and Accounts

In Figure 6-23 individuals are identified either by their simulated name or initial, e.g. 'CW' or 'Nominee2'; and accounts as blue spheres. Using the social network analysis terminology, nodes or vertices in the network represent either individuals or accounts, and links or edges represent a connection between nodes. For instance, it is possible to observe that Georgiou had direct control over the 'Accuvest', 'Caledonia' and 'Alliance' accounts, because he was the one executing trades through these accounts held in the Bahamas, although during the period of manipulation he was in fact in Canada.

Similarly, it is possible to appreciate that Georgiou exerted indirect control over other accounts, such as, the 'KG' account, which he controlled by sending instructions to Nominee3. Although it is possible to deduct from the legal proceeding that 'CW' probably knew most, or all, of the nominee account holders, no direct description of these associations is mentioned in the legal proceedings, and thus, it was decided not to include further connections between 'CW' and other individuals. Another assumption is the actual number of unidentified nominees, which was arbitrarily set to six, figure which was clarified in the case. In the same way, the number and name of nominee accounts controlled by 'CW' was three, namely 'AccountA', 'AccountB' and 'AccountC'. In the case of 'RB', 'JA', and 'KG', it was decided to use the initials as account names, rather than individual names, to emphasise the existence of a link between individuals and accounts.

Focusing the attention on the HYHY manipulation, the legal proceeding states that on 17th April 2008, Georgiou first told CW that he intended to manipulate Hydrogen Hybrid. Georgiou also told CW that there were 21 million shares of non-restricted stock and that he had control of all but 400,000 of those shares. He also told CW that he intended to begin the campaign in the near future and, in the short term, wanted to move the stock from $0.40 to $1.80 per share. At the time of this conversation with CW, two accounts controlled directly or indirectly by Georgiou (the 'Accuvest' and 'Starport' accounts) held 5,888,000 shares and 3,385,925 shares of Hydrogen Hybrid respectively, nearly one half of the non-restricted Hydrogen Hybrid stock.

Georgiou also recruited others to participate in the Hydrogen Hybrid manipulation. On May 29, 2008, Georgiou told CW that he had hired a 'group' to participate in the manipulation. He further stated that they began their campaign on 23rd May 2008 and would be going 'full force' the following week. Immediately before the manipulation started, from 19th March to 22nd May 2008, Hydrogen Hybrid traded mostly in the $0.40 to $0.50 range on average volume of approximately 40,000 shares a day.

From the economic analysis outputs, it is, for example, possible to appreciate that the outlier detection engine flagged 30th May as a day with 'High' likelihood of suspicious trading. In fact, the legal proceeding states that on that day Georgiou told CW that he and his group intended to drive the price up to $3 per share. That same day, Hydrogen Hybrid's trading volume increased by more than 1 million shares from the previous day to approximately 1,470,000 shares (an increase of 362 percent) and the stock price increased to $1.31 per share (a 19 percent increase from the previous day's closing price). Moreover, at Georgiou's direction, on the same day the 'Starport' account sold 232,000 shares, the 'Accuvest' account sold 664,500 shares, and an 'Alliance' account bought 775,500 and then sold 157,500 shares. Georgiou made five telephone calls that day to his nominee with trading authority on the 'Starport' account.

As previously mentioned, in order to study the trading behaviour from the social networking analysis perspective, two columns were manually added to the original datasets assigning to each of the trades (354 in total) a buyer and a seller account ID, following the totals stated in the legal proceeding. As the total trading of that day was mostly, but not completely, executed between the manipulators' accounts, fourteen other simulated account holders were considered, each of which was also assigned randomly to either the buyer or the seller side, trading indiscriminately with and without manipulators. It was also assumed that 'CW' exerted control over his three accounts, and thus the total number of nodes or vertices in the network was set to 20 accounts.

It is very important to clarify that in most trading platforms and venues, parties are blind when trading, i.e. they do not, and should not, know who they are trading with

in order to avoid coordination between them. If parties are coordinated or colluded they could drive the prices or volume to whatever level suits them best. If all traders trade blindly, then the pairing of buyer and seller accounts will occur following a uniform distribution, i.e. any kind of pattern in pairing should be merely the result of chance. If the opposite holds true, then it should also be possible to assert that some sort of coordination is taking place among traders, thus there is a high probability that a manipulation or fraudulent scheme is occurring.

Using the original trading data extended with the two columns, the NodeXL software was used to study the occurrence, or not, of buyer-seller pairing patterns. In order to count with a comparative benchmark, initially a mock dataset was used to create a network in which pairing occurred totally at random.

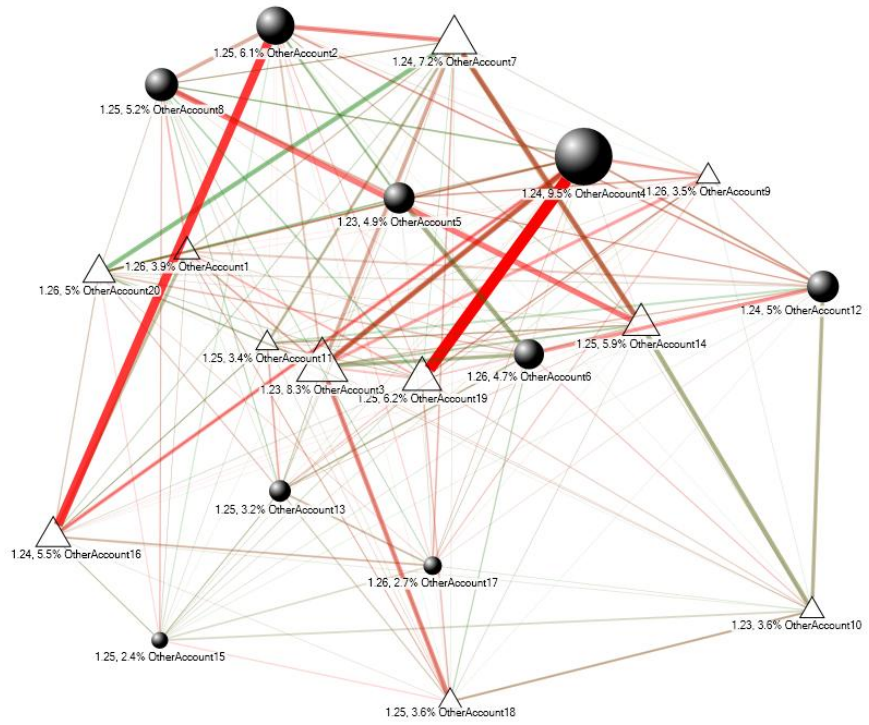


Figure 6-24 Case Study 3: Benchmark Network

Figure 6-24 presents the graphical representation of this randomly generated pairing with the following features: each node represents an account, numbered from 1 to 20. Each node includes a label in which is possible to appreciate the average volume weighted price of the trades that were executed by that account, as well as the percentage of total volume that corresponded to that account. The

shape of the node, either sphere or triangle, represents the net holding position of the account at the end of the day or period; net buyers are represented as spheres and net sellers as triangles. Similarly, the size of the node corresponds to the absolute total of trades executed by that account. For example, the triangle depicted in the bottom right corner shows that 'OtherAccount10' was a net seller, which sold at an average volume weighted price of $1.23, and that this account represented 3.6% of the total volume traded in that day. Moreover, edges or links also present important information regarding the number of trades and the hours in which this trading occurred. Specifically, the greater the number of trades between two nodes, the wider the link. Also, four time zones of trading were utilised. Time Zone 1 corresponded to trades happening before 10:00am; Time Zone 2, to trades happening between 10:00 and 12:00; Time Zone 3, to trades happening between 12:00 and 14:00; and Time Zone 4, to trades happening after 14:00 hours. Time Zone 1 is represented with light green; Time Zone 4, in contrast, is represented in bright red. Time Zone 2 and Time Zone 3 are represented with a combination of green and red, dominating one or the other depending on the respective zone. For instance, 'OtherAccount2' and 'OtherAccount16' traded mainly with each other during Time Zone 4. Taking the network as a whole, the following interpretations can be made: each account represents a more or less even proportion of the total volume of the day, with no more than 10 percent concentrated in one account; the number of buyers and sellers is more or less equal; and trading occurred throughout the whole day and was not, for example, concentrated in the opening or closing hours. Overall, then, it is possible to conclude that this network configuration does not show consistent signs of coordination among traders or accounts, and it can be used as a benchmark for comparison.

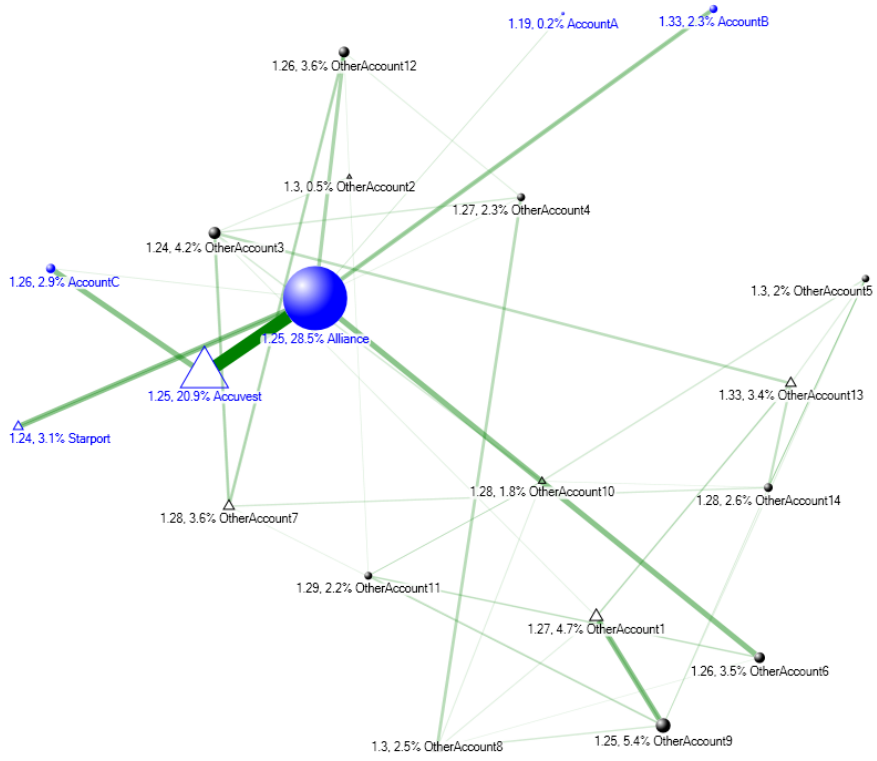


Figure 6-25 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 1

Returning to the HYHY social networks analysis, Figure 6-25 illustrates how the buyer-seller network on 30th May could have looked if the manual assignation of buyer-seller account's IDs were exactly as simulated, bearing in mind that, based on the information presented in the legal case, it was only possible to know the totals of trading for each account, but not the details of each of the transactions. To facilitate interpretation and recognition of the group of manipulators, all the accounts related to them are represented in blue. In contrast, accounts not related to the manipulation are represented in black. It is possible to appreciate immediately the great difference between this and the benchmark network, in the sense that a group of accounts were dominating the morning period, trading at least double the amount that other account holders were trading. In particular, the 'Alliance' account sold strongly to the 'Accuvest' account, and 'Accuvest' also traded a small interest with 'AccountC'. 'Alliance' also traded small interest with 'Starport' and 'AccountB'.

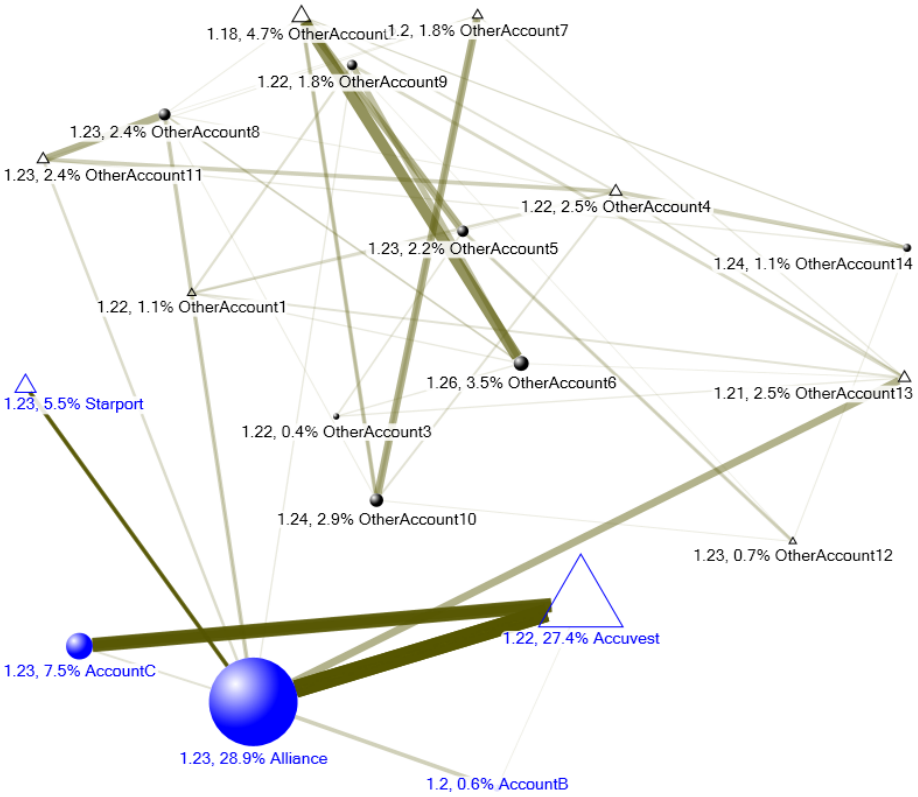


Figure 6-26 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 2

Figure 6-27 and Figure 6-28 it is possible to appreciate the evolution of the configuration of the buyer-seller network for Time Zones 2, 3, and 4. Clearly, the pattern identified in Time Zone 1 it is replicated throughout the day, with a strong association between the 'Alliance' and 'Accuvest' accounts, followed closely by 'AccountA', 'AccountB' and 'AccountC' which dominate the trading mainly between each other, but also sporadically with other accounts. Other accounts trade only small interest with each other, and the pairings among them change continuously without establishing any clear pattern.

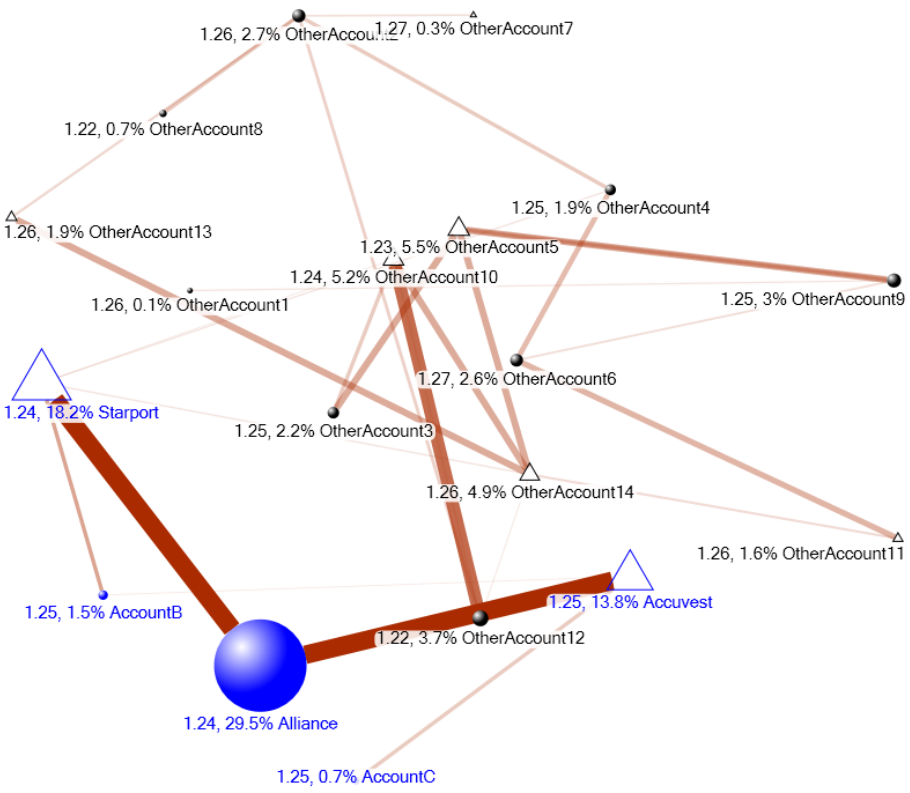


Figure 6-27 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 3

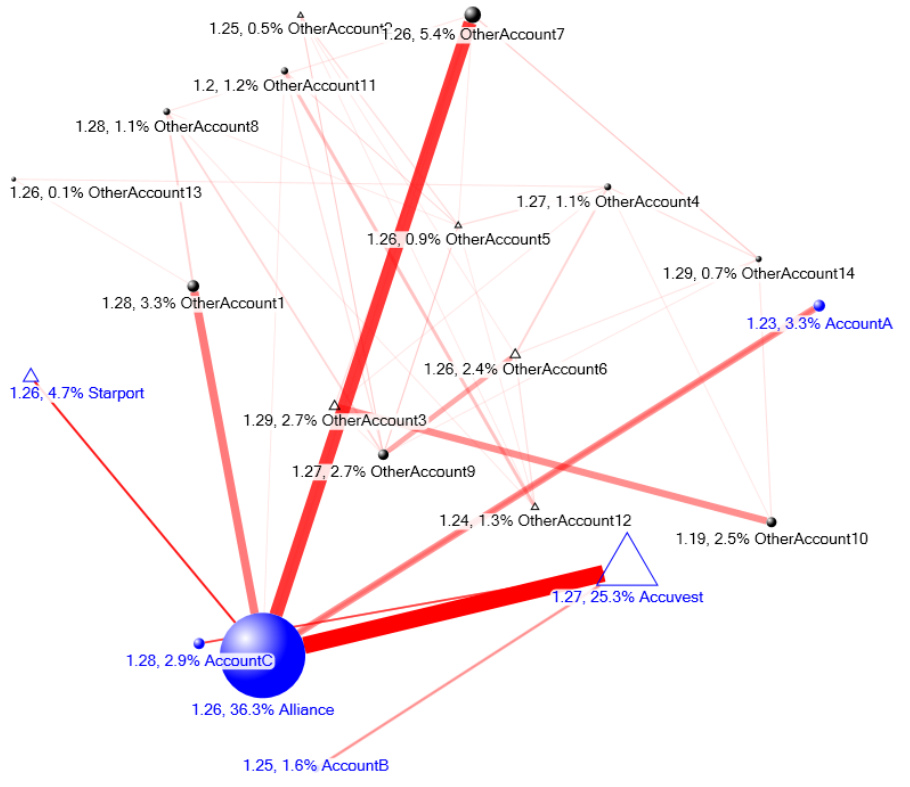


Figure 6-28 Case Study 3: Buyer-Seller Network of HYHY. Time Zone 4

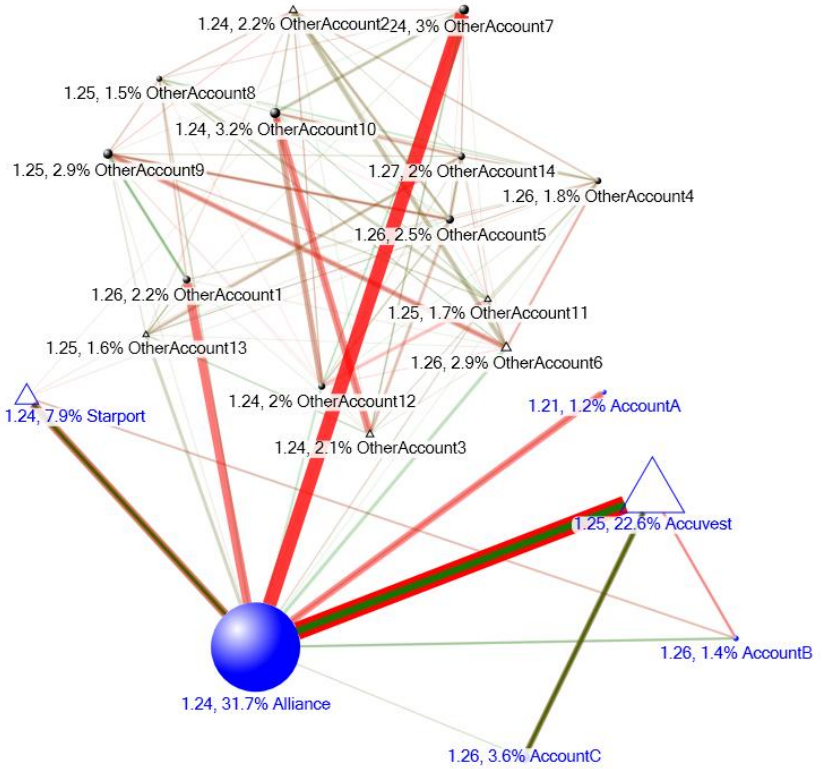


Figure 6-29 Case Study 3: Buyer-Seller Network of HYHY. All Day

Figure 6-29 presents an overall view of the network configuration for the whole day, in which is possible to identify not only the existence of a group of coordinated accounts, but also that most of the trading by the group occurred in the afternoon and near the closing of the day, thus representing a potential 'marking-the-close' manipulation pattern, along with the possibility of 'wash-sale' and 'match-trade' manipulations. Table 6-9 presents the detailed network metrics for each of the nodes, or accounts, as well as the labels and data used to generate the graphs. This table was produced using the NodeXL software. Totals for buy, sell and number of trades are as close as possible to those mentioned in the legal proceedings, and the difference is never more than 26 stocks. It is worth noticing the high betweenness centrality of the 'Alliance' account, due to the fact that 'Alliance' accounted for most of the trading of the group, being at the centre of the sub-graph or 'group' of coordinated manipulators. Also, 'Alliance' acted as a 'hub' from the group to the rest of the network, and thus was located in a strategic position 'between' the group and the rest of the market.

Table 6-9 Case Study 3: HYHY All-Day Nodes Network Metrics

Node	Degree	Betweenness	Closeness	Eigenvector	Clustering Coefficient	Color	Shape	TotalBuy	TotalSell	AvgWgtPrc	TotalTrade	TotalTrade%	NetHolding
AccountA	1	0.000	0.024	0.005	0.000	Blue	Sphere	35700	0	1.21	35700	1.21%	35700
AccountB	3	0.743	0.026	0.008	0.667	Blue	Sphere	40990	0	1.26	40990	1.39%	40990
AccountC	2	0.000	0.024	0.006	1.000	Blue	Sphere	107211	0	1.26	107211	3.64%	107211
Accuvest	3	0.500	0.025	0.006	0.667	Blue	Triangle	0	664526	1.25	664526	22.59%	-664526
Alliance	14	62.375	0.042	0.057	0.341	Blue	Sphere	775525	157500	1.24	933025	31.71%	618025
Starport	5	4.615	0.030	0.024	0.500	Blue	Triangle	0	232010	1.24	232010	7.89%	-232010
OtherAccount1	8	1.810	0.033	0.047	0.786	Black	Sphere	41975	21675	1.26	63650	2.16%	20300
OtherAccount2	12	1.733	0.033	0.071	0.788	Black	Triangle	20300	43600	1.24	63900	2.17%	-23300
OtherAccount3	9	1.396	0.030	0.052	0.722	Black	Triangle	30000	32700	1.24	62700	2.13%	-2700
OtherAccount4	13	8.639	0.040	0.070	0.628	Black	Sphere	32825	20200	1.26	53025	1.80%	12625
OtherAccount5	12	2.439	0.033	0.070	0.742	Black	Sphere	39350	32935	1.26	72285	2.46%	6415
OtherAccount6	11	3.447	0.037	0.064	0.727	Black	Triangle	39500	45575	1.26	85075	2.89%	-6075
OtherAccount7	11	4.238	0.037	0.064	0.745	Black	Sphere	57700	30350	1.24	88050	2.99%	27350
OtherAccount8	12	3.987	0.038	0.070	0.758	Black	Sphere	33275	11850	1.25	45125	1.53%	21425
OtherAccount9	12	4.324	0.038	0.069	0.727	Black	Sphere	58445	26800	1.25	85245	2.90%	31645
OtherAccount10	10	3.193	0.032	0.057	0.733	Black	Sphere	57625	36000	1.24	93625	3.18%	21625
OtherAccount11	11	2.928	0.037	0.067	0.818	Black	Triangle	25200	25675	1.25	50875	1.73%	-475
OtherAccount12	11	2.928	0.037	0.067	0.818	Black	Sphere	34140	23700	1.24	57840	1.97%	10440
OtherAccount13	10	3.410	0.036	0.058	0.733	Black	Triangle	10510	36375	1.25	46885	1.59%	-25865
OtherAccount14	12	4.296	0.034	0.069	0.712	Black	Sphere	30700	29500	1.27	60200	2.05%	1200

Table 6-10 presents the comparison of the metrics for the randomly generated buyer-seller network and the HYHY case buyer-seller network. The main statistics that are normally used in social network analysis can be observed in the table. For instance, both networks are shown to have duplicated links, meaning that it was common for two accounts to trade between each other on more than one occasion during the day; and that the total number of links (354) was set equal to the number of trades that occurred on 30th May 2008. Moreover, the table shows that in the benchmark network it takes fewer steps or links to connect two accounts located at opposite extremes in the graph. This is because it is not possible to find marked groups or clusters that are separated from the main network. If this was the case, it would take extra links to reach the farthest members of these groups, which is what can be observed in the HYHY network. For the same reason, i.e. the existence of groups of accounts separated from the main network, the graph density of the HYHY network is smaller.

In the benchmark network the number or 'degree' of links that each account has is larger, measured either as the average or the median degree. This is because each account traded uniformly with all others, and in contrast in the HYHY network, there were some that only traded with its group.

In terms of betweenness, the HYHY has a larger number again because of the existence of separated groups of accounts, which also implied that some accounts act as 'hubs'. Closeness Centrality is a measure of the average shortest distance from each node to each other node, and thus is also affected by the presence of isolated sub-graphs, as members of these groups have to go through a higher number of links before reaching the main network.

The Eigenvector Centrality measures not only the degree of connectivity of the nodes individually, but also corrects that measurement by the degree of the nodes that a particular node is connected to. In other words, the Eigenvector centrality of a node that is connected to a 'popular' node will be higher than that of a node connected to a less 'popular' one. As the number of links in the benchmark network was established randomly, this measure of popularity is slightly smaller than the

one for the HYHY network, as the former describes the consistent connectivity of sub-group members to hub members that have higher degrees.

The Clustering Coefficient measures how connected a node's neighbours are to one another. In a randomly generated network in which it is equally possible for a node to be connected with all the other nodes, this measure should be higher than one in which sub-groups or 'clusters' are present. The HYHY network contains a group of accounts that are mainly connected to each other, and not with the rest of the accounts, and this breaks the uniform connectivity that is achieved in a randomly generated network.

In summary, the HYHY network metrics indicate the presence of a highly interconnected group of accounts, a pattern that is strong and stable throughout the day. This should be interpreted as initial evidence of coordination between manipulators, and could be used as a complimentary alarm that should be further investigated by the market surveillance team.

Table 6-10 Case Study 3: Graphs Metrics

Graph Metric	Benchmark Network	HYHY All day Network
Nodes	20	20
Unique Links	48	37
Links With Duplicates	306	317
Total Links	354	354
Maximum Geodesic Distance (Diameter)	2	3
Average Geodesic Distance	1.11	1.54
Graph Density	0.84	0.48
Minimum Degree	14	1
Maximum Degree	19	14
Average Degree	15.90	9.10
Median Degree	16	11
Average Betweenness Centrality	1.55	5.85
Median Betweenness Centrality	1.58	3.06
Average Closeness Centrality	0.05	0.03
Median Closeness Centrality	1.58	0.03
Average Eigenvector Centrality	0.05	0.05
Median Eigenvector Centrality	0.05	0.06
Average Clustering Coefficient	0.83	0.68
Median Clustering Coefficient	0.83	0.73

Finally, Figure 6-30 presents an alternative view of the HYHY network in which the node shape has been replaced with an image of the flag of the country from which the trading was originated. As it was not clear from the legal proceedings where accounts were located, the researcher assigned a country of origin as following: 'Alliance' and 'Accuvest' - Canada; 'Starport' - Turk and Caicos; 'AccountA' - United States; 'AccountB' and 'AccountC' - Bahamas; 'OtherAccounts' 1 to 14 - United States. If available, this kind of view could greatly help the user/customer to quickly identify groups of traders or accounts coordinating activities from multiple geographic locations or jurisdictions, thus quickly identifying suspicious cross-border/cross-jurisdiction manipulations.

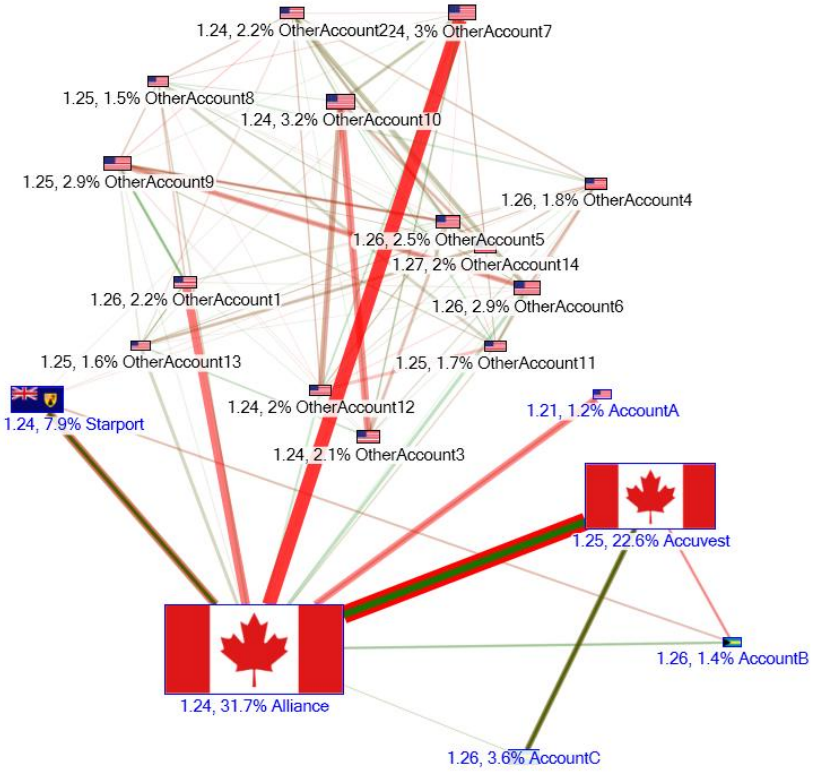


Figure 6-30 Case Study 3: Buyer-Seller Network HYHY Case. All Day Flags

6.3.1 Evaluation and Interpretation of Results

6.3.1.1 '2M-3S' Framework Application

This case addresses the challenge of using social network analysis components for the detection of 'wash sales', 'match trades' and in general, any kind of manipulation in which the coordination of different traders is taking place, even in the case of cross-border or cross-jurisdiction manipulations. In terms of the instantiation of the engine using the ‘2M-3S’ framework model, although the analysis architecture is very similar to that presented in Case Study 2, the difference lies in two aspects: firstly in the availability of fully functional social networking analysis components, such as the NodeXL; and secondly, in the fact that the customer and provider of the detection engine is assumed to be the same agent, i.e. the regulator.

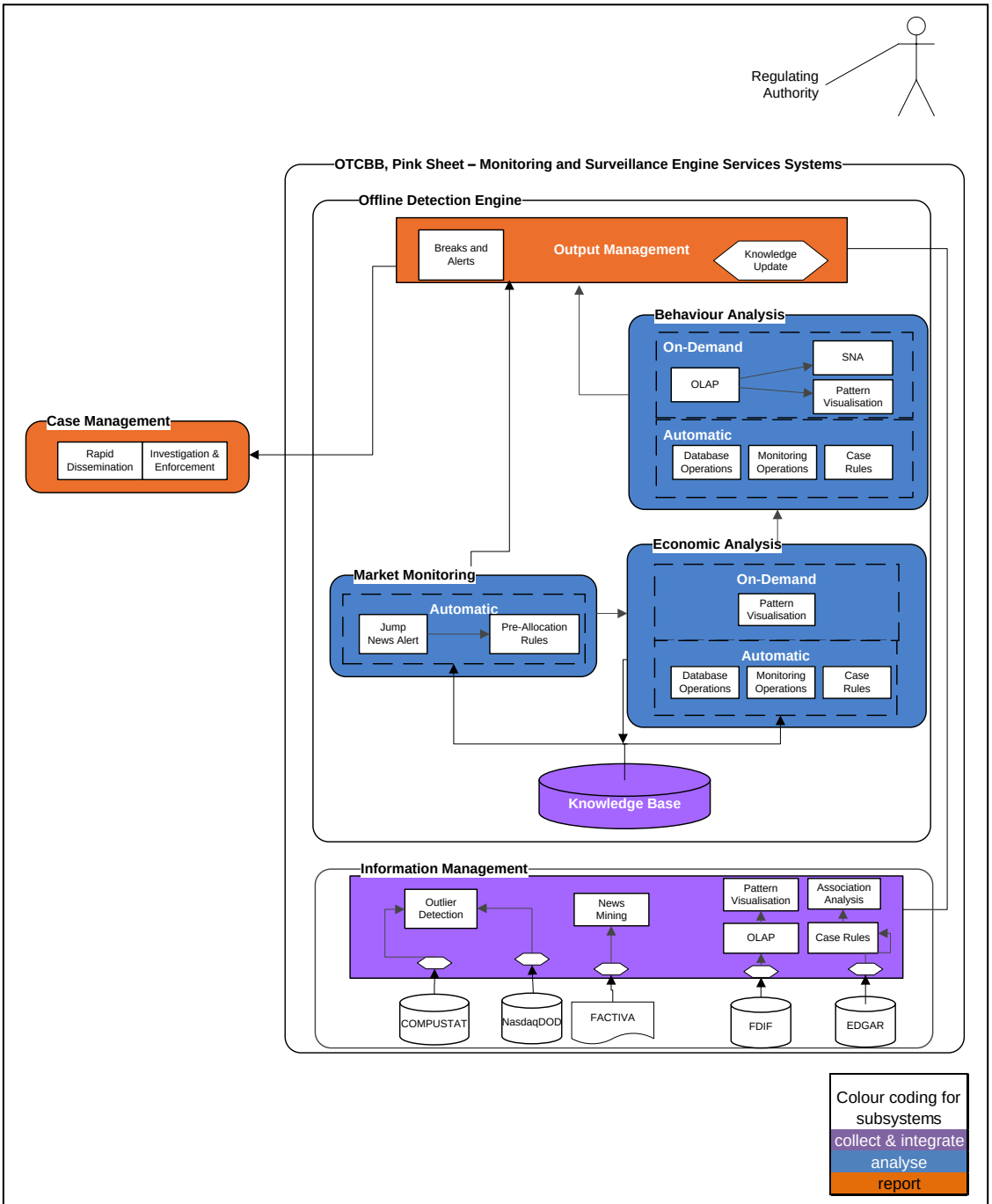


Figure 6-31 Case Study 3: Scenario Instantiation of the '2M-3S' Framework

Figure 6-31 presents a view of how the social networking analysis capable detection engine could be organised. In particular, it is necessary to remember that non-SROs are not obliged to perform surveillance and monitoring activities, and when they do so, this is usually done on a case-by-case basis, and as such, non-SROs do not necessarily invest in or possess automated detection engines.

In terms of the instantiation of the case as a service system, the case considers a configuration of value networks and value propositions in which the provider and the customer of the service is assumed to be the Regulator. Although the agents involved in this service configuration are namely the same entity, in reality it is possible to consider this a special case in which different departments or divisions within the regulating authority perform the detection and investigation of cases separately (Figure 6-32).

Similarly, Figure 6-33 shows a possible configuration of SVNs for both the customer (regulator - investigation division) and provider (regulator - detection division), which is expanded specifically by considering the inclusion of the NASDAQ on-demand data service providers and the social networks analysis (SNA) components.

Figure 6-33 presents the alternative view of the world which illustrates the interactions between two jurisdictions, recognising the cross-jurisdictional aspects of the case. In the figure only the Bahamas and the US markets are depicted, although this could be expanded to include other jurisdictions mentioned in the case, such as the Turks and Caicos Islands. Because there are no agreements on cross-market or cross-jurisdiction surveillance and monitoring in place between the aforementioned jurisdictions, the monitoring and surveillance is done independently in each jurisdiction. In this case, the surveillance and detection engine is based in the US Regulator, which is assumed to have a detection engine which is fed with market data, either using direct links with the venues or through links with the central tape managers support organisations. All the SNA and data mining components, as well as other components for pre-processing, case management and information management in general, are deployed by the local regulating authority, in this case US SEC.

In this case the 2M-3S Framework taxonomy instantiation (Figure 6-35) was used to define and describe a real case scenario of cross-jurisdiction manipulation, in which a group of traders coordinated activities from different countries to target and manipulate a particular set of stocks traded in the OTCBB and Pink Sheets

markets located in the USA. Similarly, new requirements, such as, requirement 19 were addressed that relates to the cross-border detection capabilities.

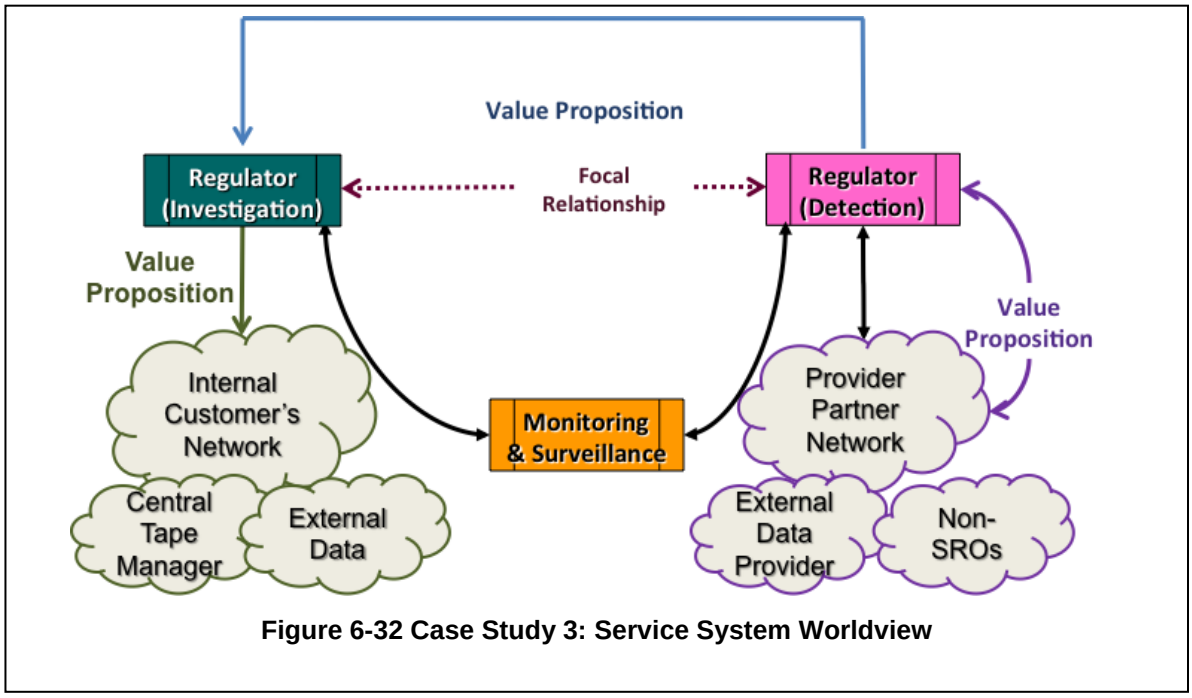


Figure 6-32 Case Study 3: Service System Worldview

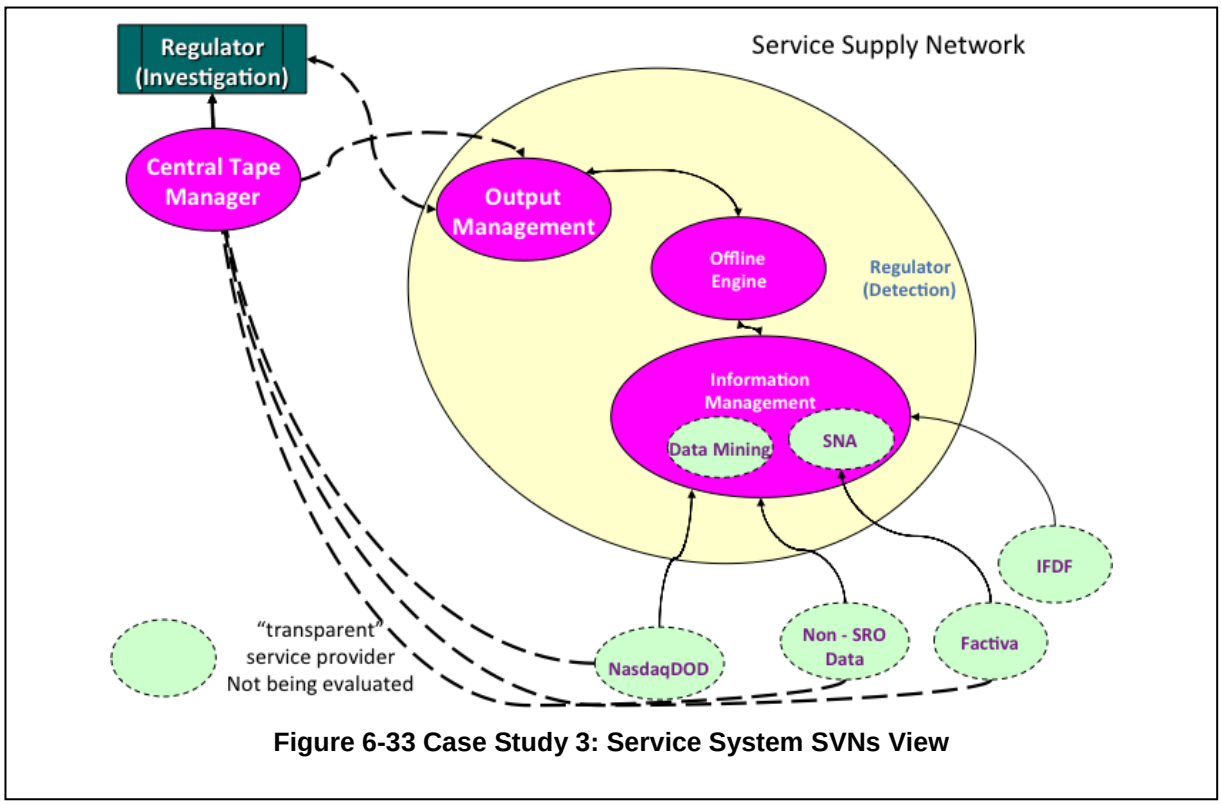


Figure 6-33 Case Study 3: Service System SVNs View

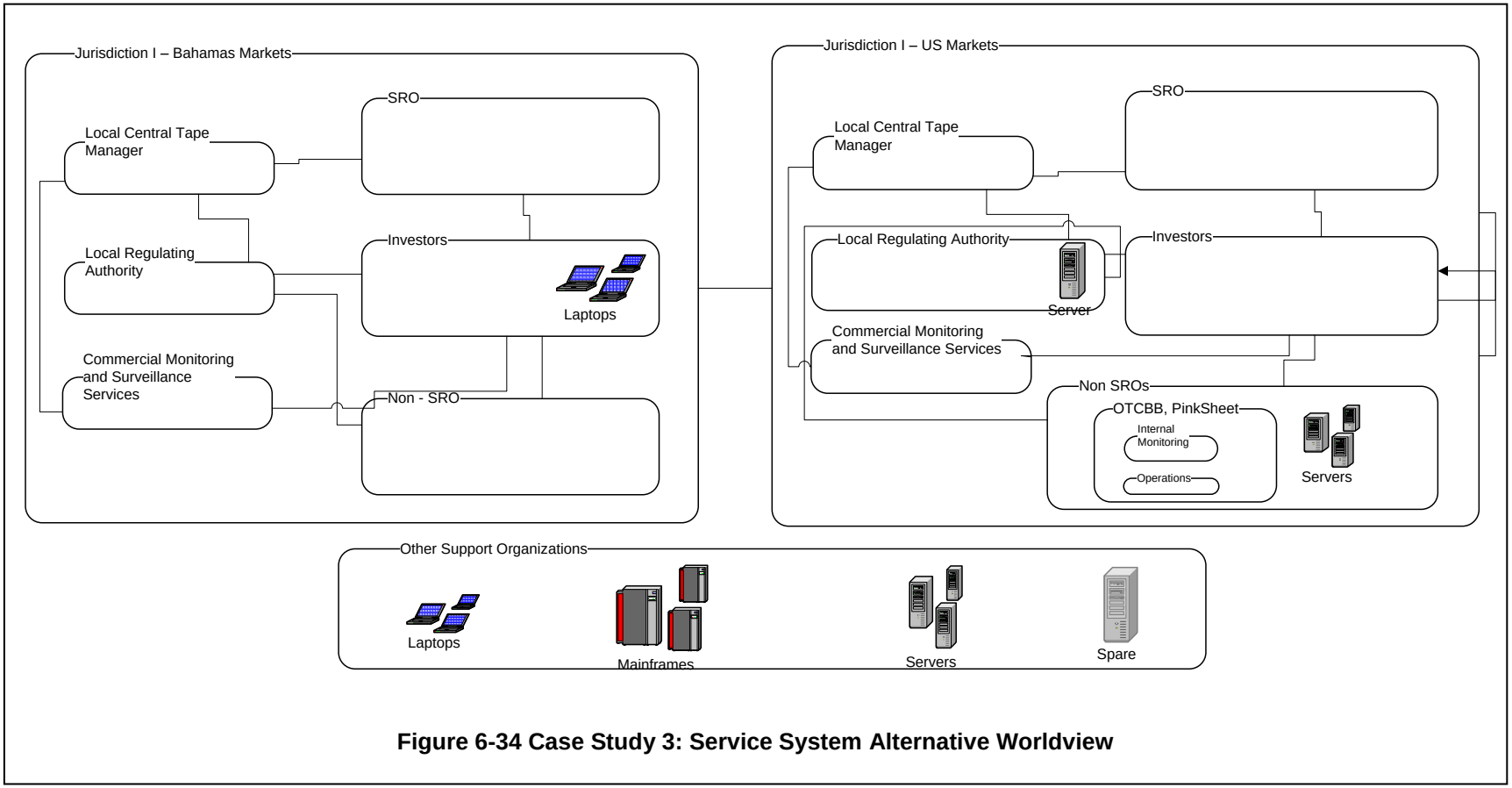


Figure 6-34 Case Study 3: Service System Alternative Worldview

Manipulation							
Manipulation	**Cross-Jurisdiction**	**Agents**	**Action**	**Venue**	**Time**	**Target Asset**	**Effects**
Wash-Sales, match-trades; breach of fiduciary duty; trade-based	Cross-border	Insider and Outsiders; In collusion; own benefit	Buy/Sell trades and orders	Non-SRO: OTCBB and Pink Sheets Electronic only	Normal speed, inside trading hours; all year	Financial; Equity; pink-sheets small caps	Financial/Economic; structured data; inside venue from outside jurisdiction orders

Stakeholders		
Investor	**Intermediary**	**Regulator**
Executives; Manipulators, net savers	Broker-Dealers	SEC

Detection Engine			
	Collect & Integrate	**Analyse**	**Report**
Information Management	Data preparation; outlier detection; news mining, previous patterns		
Off-line detection engine		Behavioural and Economic Analysis; Automatic & On demand; Decision Trees; Social networks Analysis	Visualization; OLAP; Breaks and Alerts
Real-time detection engine	n/a	n/a	n/a

Requirements		
Requirements	**Class**	**Type**
1 to 5, 8,	Analysis/Integration	Functional
9 to 12	Data capture and pre-processing	Functional
17, 19, 20, 23, 25	Information Management/Output Management	Functional
n/a	General conditions of the system and environment	Non-Functional
n/a	Skills and capabilities of the customer	Non-Functional

Figure 6-35 Case Study 3: Taxonomy Instantiation

6.4 Case Study 4: High frequency trading Quote Stuffing

6.4.1 Background

Since electronic trading was introduced, new ways of sending and executing orders to the markets have emerged. In particular, investors have developed systematic methods to process and send orders with the use of computational algorithms, which are sets of automated instructions on how and when to perform trading depending on markets conditions (Aldrige, 2010). Relying on cutting-edge computing technology, this new form of trading has reached incredible levels of performance: currently an order to trade can be generated, sent to an exchange, executed, and sometimes cancelled, within a few milliseconds. As trading messages are generated and processed at very high speed, this phenomenon has been termed High Frequency Trading (HFT).

Because the literature on HFT topics is relatively recent, the scientific community is still debating its characteristics and whether HFT is advantageous or disadvantageous in terms of its contribution to market efficiency and market integrity. According to the (U.S. Securities and Exchange Commission, 2010a), although it is still too early to coin an official definition of HFT, it is at least possible to identify some common characteristics of trading based on this technology, which are: i) a high number of trades with a lower average gain per trade; ii) a large number of orders get cancelled right after submission; iii) the use of high-speed computers to generate, route and execute orders; iv) orders are sent from servers located very close to the exchange that receives them; v) very short time-frames for establishing and liquidating positions; vi) investors using HFT usually aim for flat or near flat holding positions at the end of the trading day.

Similarly, it is possible to find studies presenting preliminary evidence on both the advantages and disadvantages of HFT, starting new lines of research. On the positive side, evidence shows that HFT has added liquidity, reduced spreads and helped align prices across markets: see for example (Hendershott et al., 2011, Jovanovic and Menkveld, 2011). According to (Aldrige, 2010), these advantages

include its cost effectiveness and a significant reduction of errors, as fewer personnel interventions are necessary to complete a trade, thus taking error-prone humans out of the decision loops.

On the negative side, however, allegations of unfairness have been raised by retail and institutional investors because HFT firms have access to arbitrage opportunities which are only possible using that particular technology. Moreover, there seems to be growing evidence of dubious HFT practices, for example brokers or traders sending bursts of quotes (at very high rates) with extremely unusual characteristics (Nanex, 2010).

The potential for these 'quote-stuffing' practices, as these have been termed, to pose problems for markets is a major area of concern, as the objectives of quote-stuffing could be directly related to potential manipulative or even 'anti-social' behaviour on the part of some HFT actors. For instance, the majority of the orders in quote stuffing episodes have no real chance of being executed, either because they are immediately cancelled after submission or are valid for extremely low periods of time (as low as 0 milliseconds), or because they quote prices which are very far away from the best bid and ask offer prices.

Nanex (www.nanex.net), a data provider has collected a large number of examples of the most noteworthy occurrences of quote stuffing. Some of these include very strange trading patterns, so peculiar that they have been collected in a 'crop circle of the day' archive, which lists episodes with equally remarkable names, such as 'the knife', 'the wild thing', and 'the click'.

Some of the potential explanations are that these strange sequences are intended to create confusion among other traders, because the firm that submits them actually ignores them, but for the rest of the market participants it is extremely difficult to work out whether these quotes represent valid interests or are just noise. This could give manipulators an advantage of milliseconds that could be crucial to close positions before other competitors do (Madrigal, 2010). Moreover, quote stuffing can create the impression that there is non-existent demand for a particular security, which can induce other investors to follow a trend that is in reality just an illusion (Katika, 2011).

Quote-stuffing not only has the potential to confuse other market participants, but also puts real pressure on systems since are perceived and processed as valid quotes, thereby adding a significant workload to already heavily used systems. Crucially, increased workloads can create delays and lags in processing times that affect the way valid quotes are processed. These delays can carry very important consequences, one of which relates to the way trading information is disseminated and consolidated.

For example, in the US the 'Order Protection Rule', one of the provisions of Regulation NMS, ensures that investors receive an execution price that is equivalent to what is being quoted on any other exchange where the security is being traded. The order protection rule requires that each exchange establish and enforce policies to ensure consistent price quotation for all NMS stocks, which include those on the major stock exchanges as well as many over-the-counter (OTC) stocks (Source: http://www.investopedia.com). In order to fulfil the requirements of the rule, market participants rely heavily on the National Best Bid and Offer prices tape. This is a consolidated database that collects information about the best prices available in the markets at any given moment. If delays are artificially introduced to market systems, then the accomplishment of the rule by exchanges and broker-dealers becomes almost impossible because participants may think that they are trading with the latest and most up-to-date information when in reality has been delayed or contaminated by quote prices that are just noise.

The purpose of this case is to study the consequences of HFT quote-stuffing practices in market systems. In order to do so, a markets simulator was built and used to model trading systems' behaviour when confronted with HFT quote-stuffing practices. Specifically, two minutes' worth of time-stamped data with millisecond granularity was used to analyse the process of sending, receiving and processing trading orders originating from HFT platforms, studying the effects of the quote-stuffing orders on the display book, best bid and ask prices and trade prices, among other variables.

This case presents some preliminary evidence on how HFT quote stuffing increase the gap of prices between markets, contrary to the evidence that HFT has helped align prices across markets (Hendershott et al., 2011, Jovanovic and Menkveld, 2011). This is very important, as manipulators could profit by artificially creating latencies in trading data feeds that would make arbitrage possible by taking advantage of the HFT induced price differences between markets.

Using these findings, a combination of indicators and visualizations are proposed to show how efficient circuit breaker mechanisms and order control mechanisms could be implemented by an exchange in order to safeguard appropriate behaviour in its trading system. Moreover, suggestions are put forward on how a real-time engine could have been used to detect potentially manipulative HFT behaviour and other misconducts.

Although there have been no officially confirmed cases of manipulation using HFT technologies, there were hundreds of instances during the flash crash in which a single stock had over 1,000 quotes from one exchange in a single second (Nanex, 2010). During the crash the selling pressure was initially absorbed by the HFT semi-market-making behaviour, but eventually the HFT firms ceased to provide liquidity when their positions started to accumulate at one side of the market. As a result, participants found no demand for buying or selling interest and consequently started getting rid of their shares at prices that were close to zero or infinitely high. In particular, this case analyses one such episode that occurred on the 6th May 2010, just after the flash crash started.

Specifically, between 14:36:00 and 14:46:00 hours, thousands of strange quotes for the ProShares Ultra Silver Exchange Trade Fund (ETF), symbol: AQG, were sent to NASDAQ with no fundamental reason or new information that could explain them. For instance, in the two-minute period starting at 14:41:000 and ending at 14:42:59.999, 1,766 quotes were sent to the market with an expiration life of 0 milliseconds (ms). In the majority of these prices were outside the best bid and ask offers, i.e. effectively these were orders that had little or no chance of being executed at available prices. Moreover, at 14:41:18.000 the International Securities Exchange received an ask quote with price and quantity equal to 0.

6.4.2 Methodology and data used

6.4.2.1 Data description

The case study uses real trading data for the ProShares Ultra Silver ETF. The dataset included all quotes data received by the exchanges listed in Table 6-11 for the two-minute period starting at 14:41:000 and ending at 14:42:59.999 on the 6th May 2010. In this period, every quote had a valid quote condition 'R', representing a 'regular' quote.

Table 6-11 Case Study 4: Exchanges for AGQ

Market Centre Symbol	Market Centre Code	Market Centre Name	Quotes
ISEG	I	International Securities Exchange	764
ARCX	P	NYSE Arca	1848
NSDQT	T	The NASDAQ Stock Market LLC	3717
BATS	Z	BATS Exchange Inc.	2091
		Total	**8420**

The data was obtained from NASDAQ on-demand data services (NASDAQ OMX Group, 2011). In total, the dataset included 8,420 quotes, which were fed into the market simulation for analysis purposes. As presented in Table 6-12, the average life of a quote was less than 1 millisecond (0.57 ms), and some ask prices and ask quantities were as low as 0. As a reference for the reader, the last valid trade for AGQ occurred at 14:40:37, only a few seconds before the window in analysis, and it was executed at $52.11 with a trade size of 100.

Table 6-12 Case Study 4: Descriptive Statistics AGQ Dataset

	Count	Mean	Standard Deviation	Maximum	Median	Minimum	Total N
Quotes_BidPrice	8420	56.97	.24	57.23	57.08	56.41	8420
Quotes_BidQuantity	8420	784.31	636.83	5200.00	800.00	100.00	8420
Quotes_AskQuantity	8420	813.91	668.43	3200.00	800.00	.00	8420
Quotes_AskPrice	8420	57.07	.67	57.65	57.18	.00	8420
Quote_Life	8420	.057	.291	7.553	.001	.000	8420

6.4.3 Data Modelling and Analysis

In order to study the effects of quote stuffing, the researcher built a software using Visual Basic for Applications (VBA) that emulates the logic and behaviour of trading systems. In particular, the programme considered the existence of four

interconnected markets which receive and process messages for one stock/security originating from investors and other exchanges. The software is able to execute and manage more than 100 different trading rules including four main types of events: messages quotes to buy or sell; cancellation of orders; corrections of orders; and 'new best' messages sent by other exchanges in order to disseminate changes in the order books.

The simulations considered that each trading event took a limited amount of time for processing, and, for example, if a new message was received while another message was being processed, the new message was added to a waiting queue. Accordingly, each message was time-stamped four times: the time the order was sent to the exchange, i.e. *original time*; the time the order was first received by the exchange (or entered the queue), i.e. *time of arrival*; the time the order was entered for processing (after waiting in the queue or not), i.e. *in time*; and the time the order was finally completely processed, i.e. *out time* (see Figure 6-36).

As a full market simulator, the software calculated and maintained records for order books for each of the exchanges in the simulation, historic information about best bid and ask prices and quantities, and depth of book until the 4th best position.

The software was also able to generate new event messages automatically when certain conditions applied, for example with the intention to fulfil the 'Order Protection Rule'. For example, when an exchange received an order to buy while knowing that the best bid-ask prices were in another exchange, that order was automatically re-routed to the exchange with the best prices. In order to do this, the exchange that received the order in the first place generated a new 're-routed buy or sell quote', preserving the price, quantity and other relevant information of the original quote and changing the destination and arrival time of the order according to its new destination. Also, orders that changed the best bid-ask prices (1st position in the book of orders) automatically generated a 'new best' message that

was dispatched to the other exchanges in order to disseminate the best prices. Although possible, the simulation did not consider any 'flash quoting'[9].

As the objective of the analysis was to study the effects of quote-stuffing, the main parameters in the simulation were the time it took exchanges to process each of the messages and the number and type of messages that exchanges received in a given period. In particular, the processing time consisted of a base processing time plus an additional time for events that generated 'new best' messages.

Because the researcher was also especially interested in studying the effects of cancelling messages, some of the original quotes were cancelled if they met certain criteria. The logic for generating cancelling messages was the following: if the life of a quote was less than the total time of processing in a given exchange, then that quote was considered 'dead' or 'cancelled'. Accordingly, one new 'cancel' message was added to the dataset for each original quote in the dataset whose quote life was less than the processing time assumed in the simulation. Specifically, the time of arrival for cancelled messages was set the same as the time of arrival of the quote it was cancelling, but the sequence of arrival was set as immediately after the original quote. This meant that in the virtual reality, although two messages had the same arrival time, messages were always processed one after the other following the sequence of arrival. Figure 6-36 presents a schematic view of the processing times for one market. In the figure, the top timeline represents the generic case in which messages are queued up while another message is being processed. The bottom timeline shows how two consecutive messages, one 'normal' quote message and its 'cancellation', arrive and are processed sequentially.

[9] For more information about flash quoting refer to footnote 2 on page 27.

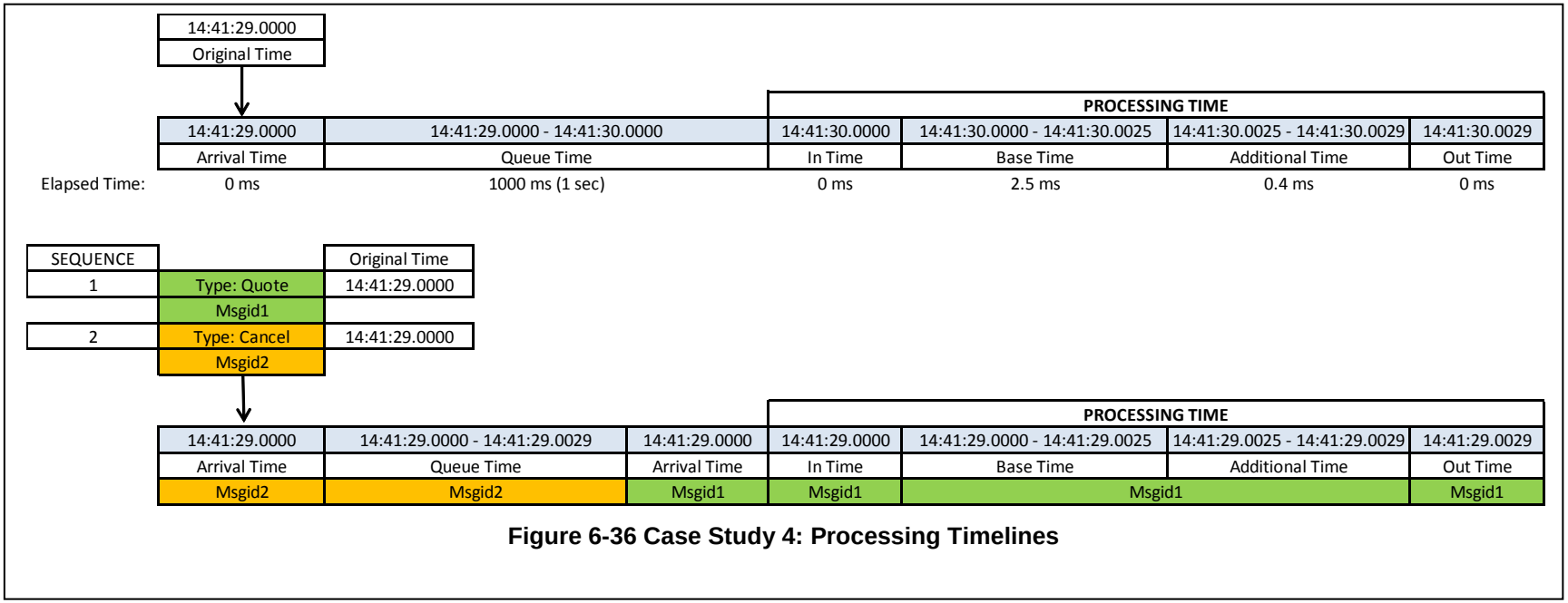


Figure 6-36 Case Study 4: Processing Timelines

In total, four simulations were run, each of which assumed that the order book started with the same base prices (best bid at $57.12 and best ask at $57.13) for all exchanges; that all messages travelled at 0.9 times the speed of light; that all exchanges processed messages at the same speed; that no message was corrected (in the original dataset there were no corrected messages either); and that markets were located equidistant to each other (a distance of 5 km was assumed; consequently, it took 0.019 ms for a message to travel from one market to another).

In Simulation 1 and Simulation 2, it was assumed that all 8,420 messages were valid, and none were cancelled. In Simulation 1, the speed of processing for a single message was set to 0.5 ms with an added time of 0.4 ms if the message generated a 'new best' event. In Simulation 2, the base speed was increased to 2.5 ms with an added time of 0.4 ms for 'new best' events. In Simulation 3 and Simulation 4, processing speeds were set the same as for Simulation 1 and Simulation 2 respectively, but depending of the actual life of the quotes, some messages were cancelled immediately after submission. Due to the speed of processing in Simulations 1 and 3, their exchanges are referred to as 'fast'. Simulation 2 and 4 exchanges are referred to as 'slow'.

Table 6-13 summarizes the basic parameters and values of the simulations. New best messages only consider the base time of processing, as these represent the dissemination of best prices from other markets, and as such, the receiving exchanges only acknowledge the arrival of the new information.

Table 6-13 Case Study 4: Simulation Parameters

Simulation/Processing Time (ms)	1 Fast	2 Slow	3 Fast	4 Slow
Quote	0.5+0.4 = 0.9	2.5+0.4 = 2.9	0.5+0.4 = 0.9	2.5+0.4 = 2.9
Cancel	0.5+0.4 = 0.9	2.5+0.4 = 2.9	0.5+0.4 = 0.9	2.5+0.4 = 2.9
New Best	0.5	2.5	0.5	2.5
Simulation/Number of messages to be processed				
Quotes	8420	8420	8420	8420
Cancels	0	0	3103	5825

In order to study the effects of the quote stuffing in the different simulation scenarios the researcher calculated the best bid and ask prices, as well as the gap or difference in prices that was produced at any moment in time between exchanges. To calculate the total gap in prices at moment *t,* i.e. $TotalGap_t$, the squared price differences were added up for each pair of exchanges, and then the square root of the totals was calculated, as shown in Equation 5. The sum of the squared differences was used to avoid the problem of compensating differences of opposite signs.

Equation 5. Total Difference of Prices or Gap at Moment *t*

$$TotalGap_t = \sqrt{\sum_{i=1}^{4}\sum_{j=1}^{4}(BBid_i^t - BBid_j^t)^2} + \sqrt{\sum_{i=1}^{4}\sum_{j=1}^{4}(BAsk_i^t - BAsk_j^t)^2},$$

for each j>i,

where $BBid_i^t$ and $BAsk_i^t$ were the best bid and ask prices at moment *t* for market *i.*

A moving variable, $GapLife_t$, was also calculated: this represented the duration of a gap from the moment it started, i.e. the last moment in time where $\boldsymbol{TotalGap_t}$ =0, until the current moment *t*.

Table 6-14 presents a simple example of the 'GapLife' variable when a gap started at moment *t=1* and lasted until *t=2*.

Table 6-14 Case Study 4: GapLife Example

t (ms)	Gap ($)	GapLife (ms)
0	0	0
1	10	1
2	15	2
3	0	0
...	...	...

Five other moving variables were created that represented the ratio of a number of different events that generated a change in the 1st position of the order book over

the last 50 quotes received by an exchange. These variables measured the number of cancelling quotes, i.e. $Ratio\_Cancels\_Quotes$; the number of new best messages received, i.e. $Ratio\_Nbest\_Quotes$; the number of quotes that generated a change, i.e. $Ratio\_QwCh\_Quotes$; the number of re-routed messages received, i.e. $Ratio\_RR\_Quotes$; and the number of trades that were executed, i.e. $Ratio\_Trades\_Quotes$. In addition, a sixth variable was created, $Ratio\_BookChg\_Quotes$, that represented the ratio of any generic event that induced a change in the 1st position of the order book over the last 50 quotes received by the exchange.

It is necessary to highlight that these variables counted only the events that produced a change in the best position in the book, because the majority of messages did not produce this type of change. In order to clarify this point, it must be remembered that the order book keeps a record of the best bid and ask prices in descending order of priority, and as such, when the majority of orders submitted in a quote stuffing burst are not intended to be executed, i.e. are very far away from the best bid and ask prices, they will not induce any change in the best places of the book per se. Nonetheless, these may create delays in the book's refreshing processes as the exchanges must firstly process and analyse a quote to determine that it will not produce a change, thereby increasing the length of time that other messages have to spend in the waiting queue and thus slowing down the pace of updates.

In addition, the researcher was interested in measuring the effects of the quote stuffing orders not only in terms of the price differences, but also in measuring the potential effects in the duration of the gap itself. Consequently, there were two working hypotheses as it was expected that quote stuffing had a double effect on the gaps: i) they will increase the gap in terms of price differences; and ii) they will also increase its duration. Both of these effects would result from the delay introduced into the systems by the quote stuffing orders.

The contribution of the different parameters used in the simulations allowed to isolate and measure these gaps under different controlled conditions, namely, the speed of processing of the exchanges, and the number of invalid or cancelled

quotes that exchanges had to process during the quote stuffing episodes. In order to measure these effects, several descriptive statistics and Pearson's correlation were calculated for each of the six gap ratios.

Although the researcher originally planned to run linear regressions to estimate the contribution of each of the ratios to the gaps, this task was abandoned after realizing that the input variables (ratios) were highly correlated to each other — in the best case — and that there was almost perfect co-linearity between some of them — in the worse case. To understand why this was so, consider that at any single moment in time a message can only trigger a limited set of events, and that the majority of events are mutually exclusive. For instance, if at moment $t+m$ the exchange receives a quote message, this message could actually be an original quote message or a re-routed quote message. Moreover, if it is either of the two, it cannot be a cancelling message, a new best message, or a correction message. In other words, at any moment in time the value of one ratio was influenced almost perfectly by the value of the other ratios (if you know one, you can deduct the value of the others). As this breaks the basic assumption of independence of the input variables, any results would have been spurious and so it was decided to continue with simple correlation analysis only. The latter measures one-to-one relations between variables, and thus it is not affected by the co-linearity between variables.

Besides the correlation analysis, several other descriptive statistics were calculated, including the mean, standard deviation, minimum, and maximum values of the gaps both in terms of money and time, as well as the total count of the different events. In addition, a test was run to see whether the mean gap of each simulation was statistically different to the others, in order to determine which set of parameter values created the worse (best) conditions for trading in terms of gap duration and price differences. All statistical, pre-processing, graphs and other data analysis were done using the IBM PASW Modeller v.14 'Clementine' software.

6.4.4 Analysis and Results

The first step of the analysis was to examine the overall effects of the quote stuffing in the gap and gap life. Table 6-15 presents the summary statistics of the

gap and gap life (in seconds) for each of the simulations. It is possible to observe that both gap and gap life means are higher when the processing time at exchanges is higher (simulations with slow exchanges have greater gaps than simulations with fast exchanges), and that simulations with cancelling or 'dead' quotes have both higher mean and higher standard deviation in terms of gap and gap life. In absolute terms, Simulation 4 had the highest mean and standard deviation of gap and gap life. In contrast, Simulation 1 had the smallest indicators. Appendix 5 shows that the differences between the means of the simulations are statistically different at the 1% level, and thus it can be concluded that these differences are not the result of chance. Figure 6-37 and Figure 6-38 show the behaviour of the gap and gap life over time in the different simulations. The time scale is set to milliseconds, starting at second 20 (20,000 ms) and ending at second 120 (120,000 ms).

Table 6-15 Case Study 4: Gap and Gap Life Summary Statistics

	S1: Fast	S2: Slow	S3: Fast	S4: Slow
	All alive	All alive	Dead quotes	Dead quotes
Total Gap Mean	0.009	0.012	0.013	0.021
Total Gap St Dev	0.024	0.024	0.029	0.037
Total Gap Min	0	0	0	0
Total Gap Max	0.191	0.191	0.2	0.22
Gap Life Mean	5.609	2.655	4.322	9.036
Gap Life St Dev	7.353	3.126	5.841	10.295
Gap Life Min	0.000	0.000	0.000	0.000
Gap Life Max	26.540	13.891	22.515	33.924

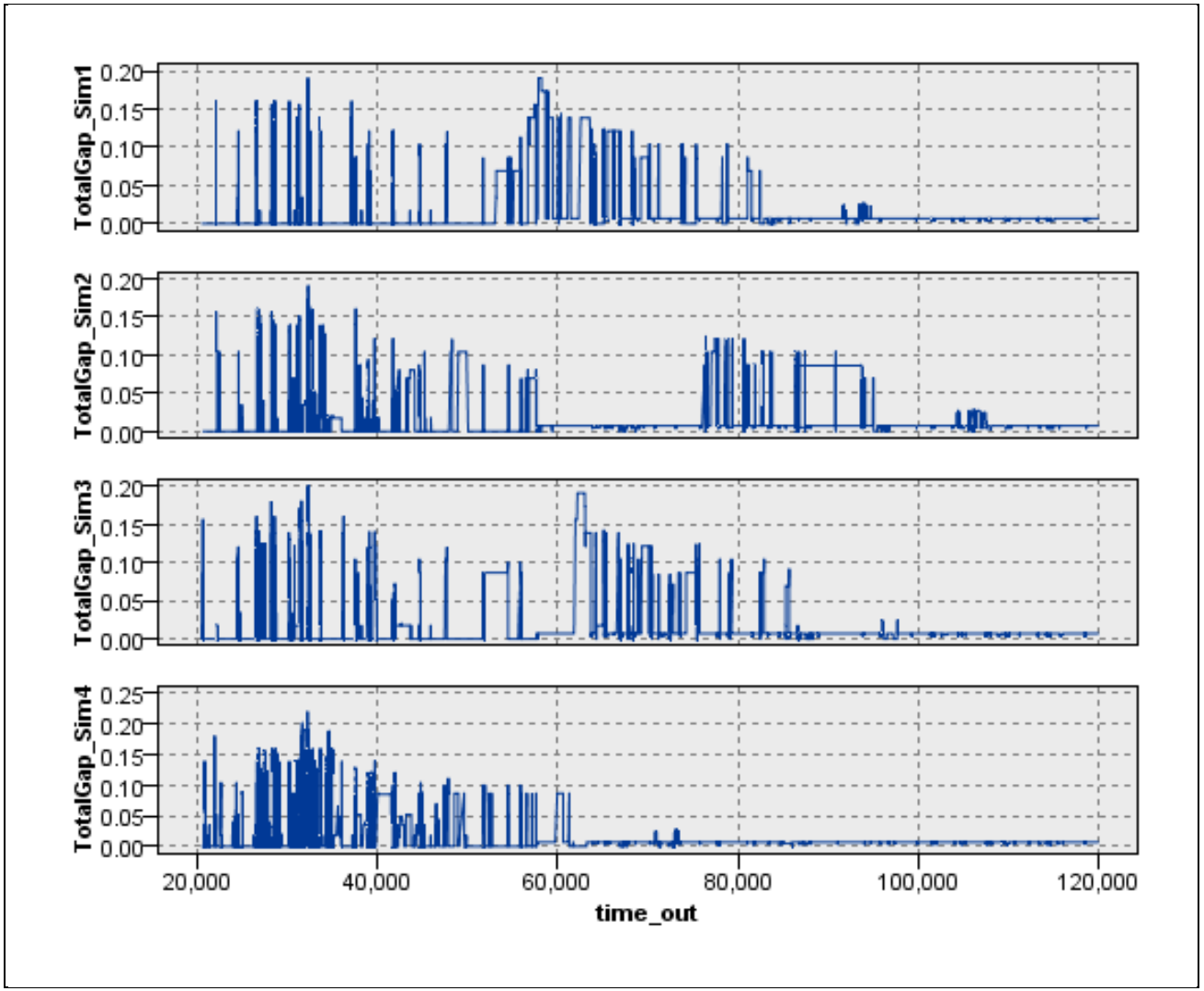


Figure 6-37 Case Study 4: Total Gap

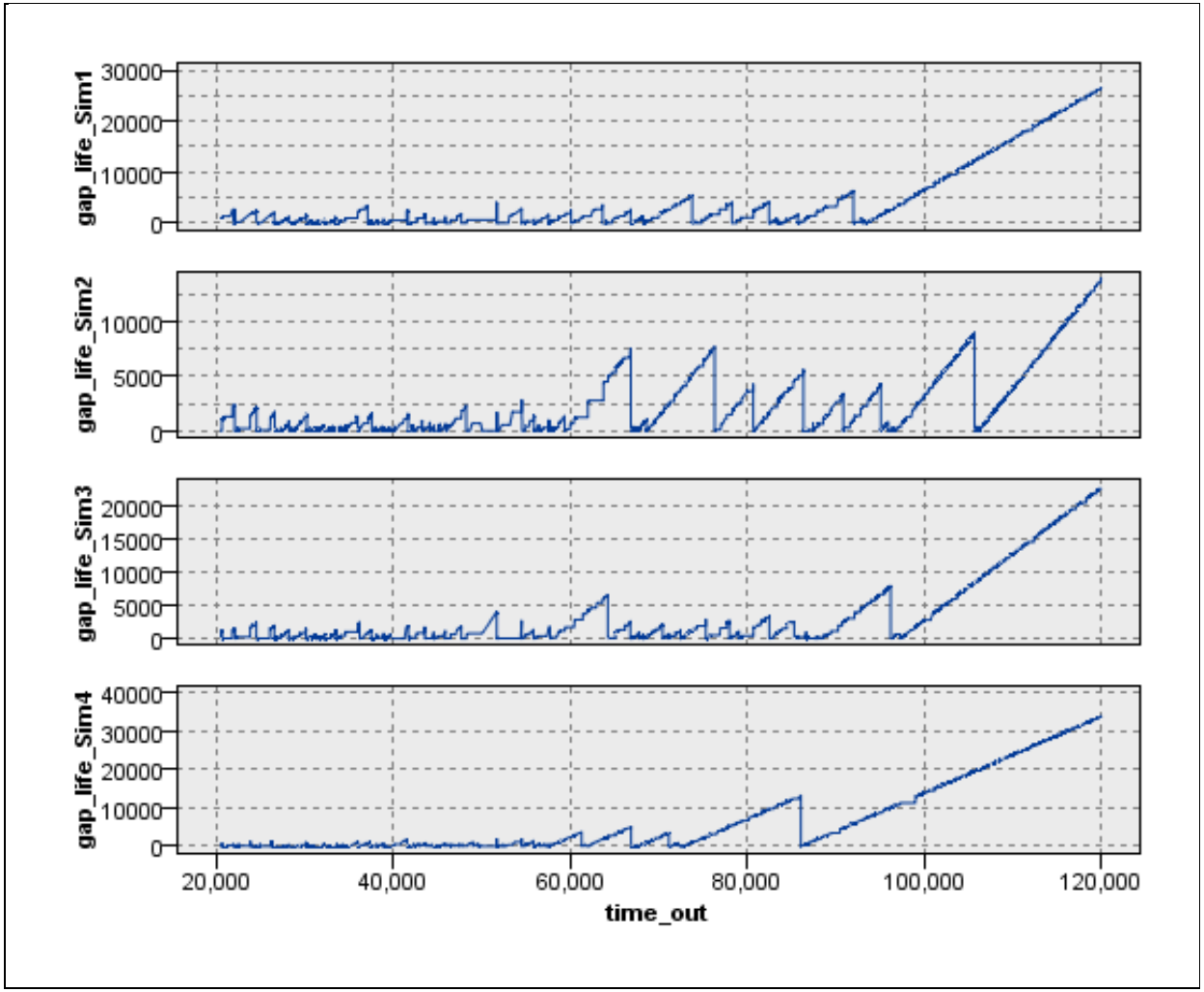


Figure 6-38 Case Study 4: Total Gap Life

Table 6-15 shows the total number of events that were generated and processed in each simulation. The count of events is broken down in terms of the total number of events and the number of those which induced a change in the first positions in the order book. The proportion of the latter is also given in order to facilitate comparison among simulations. The results show how significant the effects of quote stuffing are in putting pressure on exchange systems. In Simulation 1, for instance, where it was assumed that all quotes should be treated as valid, and that the processing speed was as fast as 0.9 ms per message, a great number of spurious trades (4,862 trades) occurred.

Table 6-16 Case Study 4: Gap and Gap Life Count of Events

	S1: Fast		S2: Slow		S3: Fast		S4: Slow	
	All alive		All alive		Dead quotes		Dead quotes	
	Count	% of prev. row	Count	% of prev. row	Count	% of prev. row	Count	% of prev. row
Total No of Cancels	n/a		n/a		3091		5827	
Total No of Cancels w/Chg	n/a		n/a		193	6.24%	600	10.30%
Total No of Nbest	15054		11465		17020		16388	
Total No of Nbest w/Chg	595	3.95%	1116	9.73%	1065	6.26%	1876	11.45%
Total No of Quotes	8420		8420		8420		8420	
Total No of Quotes w/Chg	299	3.55%	621	7.38%	528	6.27%	872	10.36%
Total No of RR	3117		3070		3330		3379	
Total No of RR w/Chg	297	9.53%	448	14.59%	410	12.31%	669	19.80%
Total No of Trades	4862		4732		5142		5301	
Total No of Trades w/Chg	328	6.75%	537	11.35%	460	8.95%	732	13.81%

Table 6-17 presents the results of the correlation analysis. In Panel A it is possible to appreciate that in general, all ratios correlated positively with the Total Gap variable, except for the new best messages ratio in Simulation 1. In terms of ratios, those that contributed the most to the gaps are the Quotes with change ratio, the Cancelled quotes ratio and the New best ratio. Nonetheless, all ratios were found to be statistically significantly correlated with the Total Gap at the 1% level of significance.

Table 6-17 Panel B presents the results of the correlation analysis this time with respect to the Gap Life. It shows that all ratios are inversely correlated with the Gap Life, which can be interpreted as demonstrating that events which did not produce a change in the best place of the order book actually contributed to extending the duration of the gap. As shown previously in Table 6-16, only a small proportion of messages (ranging from 3.55% to 10.36% depending on the simulation) actually induced a change in the order book; consequently the majority of messages left the gap untouched and merely increased its duration. In terms of magnitude, it is possible to appreciate that the gap life is more closely correlated with the ratios when the exchanges are slow and when they are confronted with more cancelling messages. However, the correlation differences between Simulations 1, 2 and 3 are not as significant as for the Total Gap.

Table 6-17 Case Study 4: Correlation of Gaps vs Ratios

Panel A	**S1: Fast**	**S2: Slow**	**S3: Fast**	**S4: Slow**
Correlations w/Total Gap*	**All alive**	**All alive**	**Dead quotes**	**Dead quotes**
Ratio_BookChg_Quotes	0.049	0.138	0.284	0.510
Ratio_Cancel_Quotes	n/a	n/a	0.301	0.458
Ratio_Nbest_Quotes	-0.017	0.094	0.278	0.474
Ratio_QwCh_Quotes	0.121	0.242	0.328	0.537
Ratio_RR_Quotes	0.077	0.032	0.071	0.265
Ratio_Trades_Quotes	0.071	0.048	0.067	0.272
Panel B				
Correlations w/GapLife*				
Ratio_BookChg_Quotes	-0.217	-0.227	-0.256	-0.402
Ratio_Cancel_Quotes	n/a	n/a	-0.198	-0.353
Ratio_Nbest_Quotes	-0.243	-0.205	-0.295	-0.403
Ratio_QwCh_Quotes	-0.235	-0.311	-0.260	-0.426
Ratio_RR_Quotes	-0.075	-0.034	-0.058	-0.163
Ratio_Trades_Quotes	-0.088	-0.051	-0.045	-0.163

*All correlations are statistically significant at the 1% level

As an example, Figure 6-39 shows graphically the behaviour of the ratios and gap in Simulation 4. The graph illustrates how closely they moved.

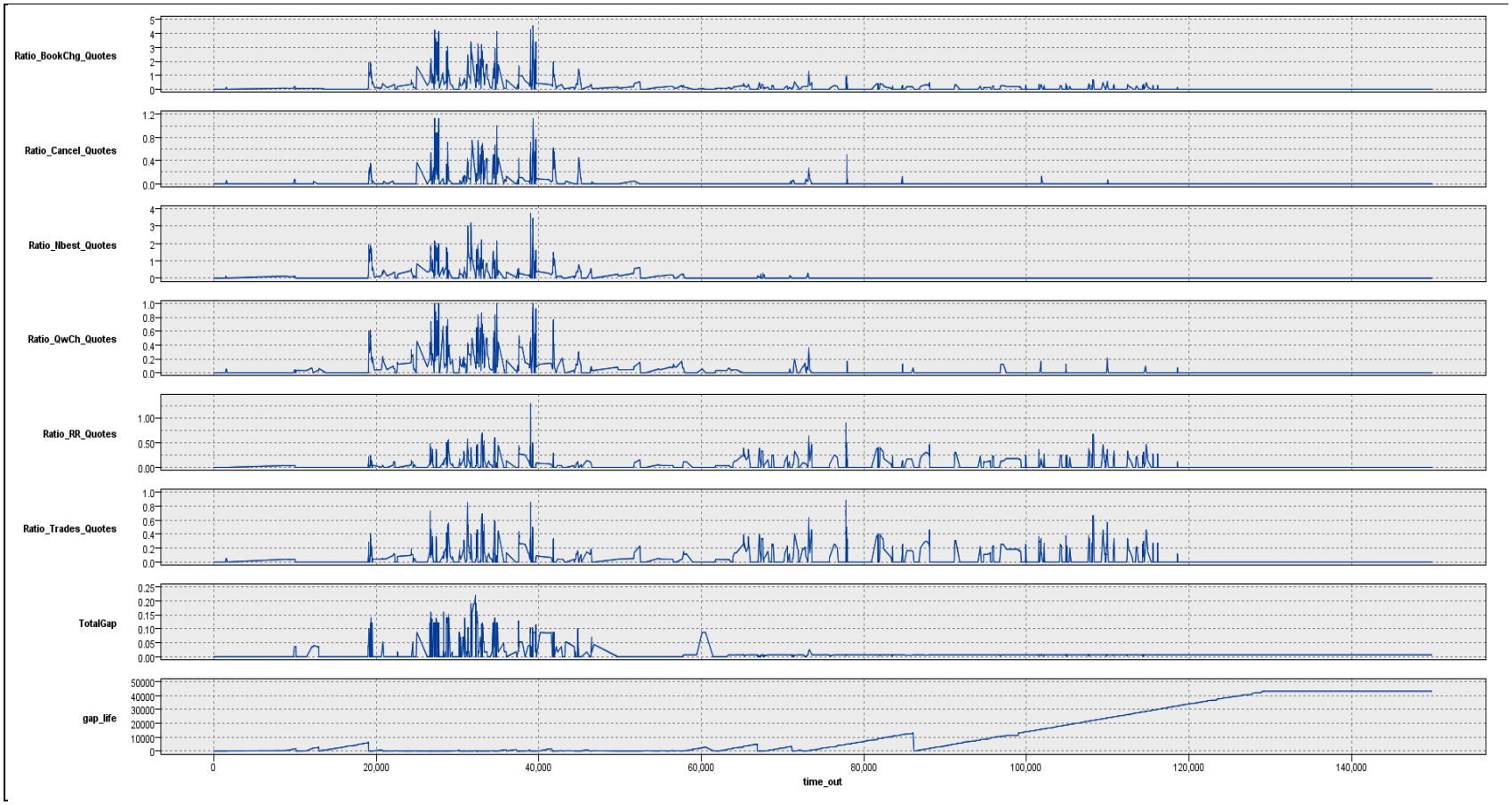


Figure 6-39 Case Study 4: Gap vs Ratios in Simulation 4

6.4.5 Evaluation and Interpretation of Results

In Table 6-15 it is worth noticing that in the worst case the average gap in price can be as much as $0.021 and can last for 9.036 seconds. Given that this simulation covers only two minutes of data, it is easy to see how important this magnitude can be if one thinks of the potential for arbitrage opportunities that could use this difference in prices to the detriment of other investors. Ten seconds is enough for ill-intended HFT manipulators to place and executed thousands of orders, and if manipulators managed to invest $100,000, for example, they could make as much as $25,813 profit in that brief period[10].

The results shown in Table 6-16 can be interpreted as illustrating that even in very fast exchanges strange quotes can induce systems to execute a high number of spurious trades (as many as 4,862 in the two-minute period). These trades should not have occurred if a proper monitoring system had had the chance to reject invalid quotes before they entered the trading systems.

Remembering that a quote triggers a chain of events that is almost impossible to foresee is key to understanding the 'butterfly' effect that is occurring in these simulations. If one invalid quote is executed as valid, then the way in which subsequent events and messages are resolved can be completely different to what would have happened if that invalid quote had been rejected in the first place. For instance, if an invalid quote changes the best bid in one exchange, then that mistake will affect all trading not only in that particular exchange; the error would be disseminated to other exchanges via 'new best' messages, thereby perpetuating and enhancing the original error. This holds true even if the invalid

---

[10] Assuming that $100,000 were invested to buy 1,751 stocks at $57.11 in the market with the cheapest price and later sold at $57.11+$0.021=$57.131 in the market with the highest price. The profit is calculated as 702* $0.021* 1751 = $25,813. 702 is the number of orders that could be proportionally placed in the market in 10 seconds based on the 8,420 orders that were submitted in the two-minute period. The profit estimation does not take brokerage or any other trading costs into account.

quote is later cancelled, as was assumed in this simulation. Table 6-16, then, should be analysed carefully in the sense that these figures represent the interaction of both valid and invalid messages in a simulated reality, which is very different from what actually occurred in real life that day.

However, the contribution of the simulations is very straightforward because it allowed the researcher to properly appreciate which conditions are worse (or better) when markets confront quote stuffing, as well as which types of messages and conditions have the greater impact. If one uses the number of spurious trades as an indicator of the consequences of quote stuffing, then it is clear that both slow and fast exchanges suffer more or less the same, as the number of spurious trades is high in all the simulations. Moreover, cancelling the quotes after they have been already processed (Simulations 3 and 4), actually increases the number of spurious trades, so that the effect is worse for slow exchanges.

In Table 6-17 Panel A, the fact that the majority of the ratios correlate positively can be interpreted as indicating that quote stuffing 'bursts' contain a high number of messages with the potential to des-align the best bid and ask prices in the exchanges, and that these des-alignments are also disseminated to other exchanges. In terms of the magnitude of the correlations, it can be seen that in Simulation 4, the correlations are higher than for the rest of the simulations. It is also evident that Simulation 3 and Simulation 4 both have higher correlations than Simulations 1 and 2, which indicates that not only the speed of processing is important, but also the number of quotes that are later cancelled. In summary, the higher the speed, the smaller the gap that is produced, and the higher the number of cancelled quotes, the larger the gap that is produced. This can be interpreted as signifying that quote stuffing in general produces a number of events that are positively correlated with the Total Gap.

The results of Table 6-17 Panel B can also be interpreted as an indication that regardless of the speed of processing and the number of quotes that are cancelled, quote stuffing practices produce events that extend the life of the gaps, and thus create more arbitrage opportunities for potential manipulators. Overall, the results of the simulations present preliminary evidence in favour of the working hypothesis,

and thus it is possible to assert that under the controlled condition of this case study, quote stuffing episodes: i) increase the gap in terms of price differences, and; ii) they also increase its duration. Consequently, it is possible to deduct that under these conditions, quote stuffing episodes deteriorate market integrity, as prices in the order books do not reflect all publicly available information every time fresh data was delayed by quote stuffing activities.

6.4.5.1 Suggested Solutions and '2M-3S' Framework Application

The previous sections have presented evidence about how quote stuffing practices are detrimental to financial markets as they induce delays in trading systems which create differences in the prices that exchanges believe are correct. It was shown how in a simulated reality, quote stuffing practices induce these effects, and that these are worse when exchanges are 'slow' to process different types of trading messages. Moreover, the case study demonstrated that under controlled conditions a higher proportion of invalid quotes, even when these are later cancelled, also directly affects both the magnitude and duration of price gaps. In the following section, these results are used to discuss how quote stuffing events could be monitored and what type of indicators and actions could be deployed in order to minimize its consequences.

Firstly, in order to properly address these issues let us start by clarifying that even for non-HFT manipulations, in today's configuration of the markets it is still very difficult to fulfil the spirit of the 'Order Protection Rule'. To picture this problem, imagine a world where only four markets exist and that these are all interconnected to each other. Assuming that: all markets process trading orders at the same speed $p=a$; all exchanges are located equidistant to each other; all messages travel at the same speed; and that markets are equal in every other respect except for the speed m at which they are able to create and dispatch messages to each other informing about new price information, the time t it would take for such a message to travel from one market to the other can be exclusively analysed in function of the magnitude m.

In this line, assuming that the *p* and *m* speeds are as shown in Table 6-18, it is possible to estimate a travelling time *t* measured in *a* units. Figure 6-40 presents these travelling times graphically.

Table 6-18 Case Study 4: Travelling Times

Market	Orders processing speed	Messages generation and dispatch speed	Travelling time = *t(m)*
A	p=a	m	a
B	p=a	2m	2a
C	p=a	3m	3a
D	p=a	4m	4a

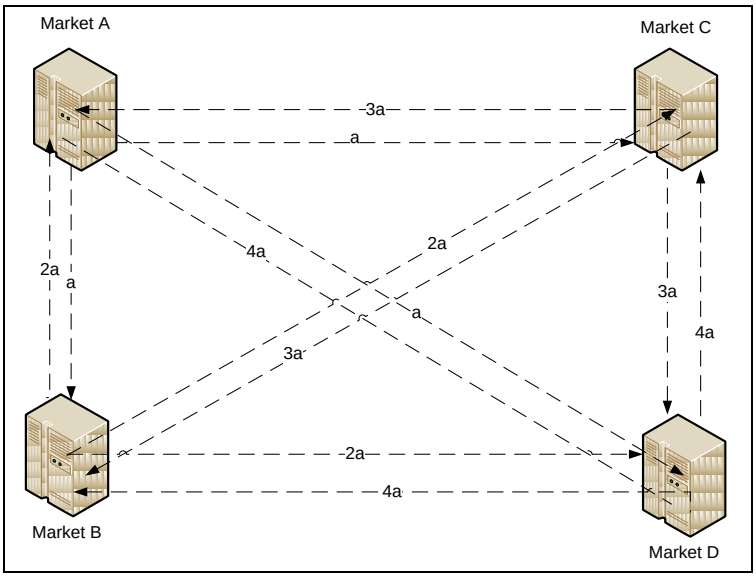


Figure 6-40 Case Study 4: Markets' Travelling Times

For instance, it is possible to appreciate that the travelling time from Market A to other markets is always *a*; for Market B it is *2a*, and so on. Now consider two scenarios in which new orders arrive to markets at a given pace: Scenario 1, where orders arrive at intervals that are considerably bigger than the time it takes the fastest market to process a trading order, e.g. they arrive every *20a's*; and Scenario 2, where orders arrive at considerably smaller intervals, say every *0.1a's.* It is possible to deduct that in Scenario 1, assuming that no order re-routing takes places, it is always possible for markets to know and process incoming messages using the latest up-to-date information of prices. In contrast, in Scenario 2 it is almost impossible for markets to know and process incoming messages using the most up-to-date price information.

In Scenario 1, for instance, if new messages arrive at Market D for example every *20a's,* there will always be enough time for this exchange to process and send the new price information to the other markets before a new order or any other new price information arrives. Mathematically, it will take *p+4m= a+4a =5a* to process and send the new information to other markets, which is always less than the time in will take for the next message to arrive (5a<20a if a new order arrives, or 2a<20a if new price information arrives coming from the fastest market, Market A). In Scenario 2, however, if new messages arrive at Market D for example every *0.1a's,* there will never be the chance for this market to completely process the first message before a new message arrives, and so a message queue will be created. Mathematically, it will take *p+4m=5a* to process and send the new information to other markets, meanwhile 5a/0.1a= 50 new orders would have joined the waiting queue. Consequently up-to-date price information will be always delayed and or distributed unequally among exchanges.

Readers must be wondering why in the case study the minimum TotalGap value is 0, given that sometimes the pace at which original quotes arrived (0 ms intervals) was much less than the time it took exchanges to process them. If this was the case, then the difference in prices (TotalGap) should be always greater than zero. However, the reason for the 0 value is that in the simulations there was always one moment in time where TotalGap=0, which was effectively *t=0*. At the beginning of the simulation all order books started with the same bid and ask prices, and at that point the TotalGap was zero. Moreover, the pace of the arrival of orders exactly matched what occurred in real life; consequently they arrived at different intervals, some of which were long enough to allow the gap to be closed again.

This simple analytical exercise helps us understand that in the past, while trading orders were originated and sent to the markets a 'human' speeds, i.e. at intervals that were greater than the time it took exchanges to process, send and acknowledge orders, it was less likely that price information was delayed and messages would have to queue before they were processed. Thus, it was also more likely for traders and markets to operate in line with the 'Order Rule Protection'.

In today's market, where orders are received increasingly faster than the time it takes exchanges to process them, it is clear that the Rule is becoming less and less respected. Similarly, consolidated trading information exercises, such as the NBBO tape, are also condemned to contain delays and price differences. This is clear when one thinks that in order to estimate a consolidated tape it is necessary first to receive the most up-to-date information before any other exchanges do, aggregate the information, process it, and then disseminate it back to the exchanges before a new trading order arrives, which is almost impossible at current consolidation speeds.

The question is, then, what steps are necessary to safeguard the spirit of rules such as the Order Protection Rule, which were created to preserve market integrity, when advances in market efficiency are in direct conflict with the former?

In the researcher's opinion, first it is necessary to increase awareness of the unintended consequences of increasing the speed of trading, as well as to create mechanisms to guarantee that these consequences are managed and kept below a certain minimum.

One extreme way to solve the problem could be to enforce a fully centralised market, in which all trading orders, without exception, are sent and processed by one entity, which should be in charge of distributing the information to the right markets at the necessary pace, which would also effectively introduce artificial, but controlled, delays so that all market participants have access to the same information at the same time. This however, as plausible it may seem, is actually almost impossible to achieve given the global interconnectivity of markets. It is almost impossible to imagine that it will be ever possible for markets around the world to be centralised and synchronised, especially considering the economic costs and political consequences of such an agreement.

In this sense, a more plausible solution would be to enforce a policy or rule by which exchanges are obliged to compensate investors retroactively for any losses which result from the desynchronization of prices, which would be very similar to the 'clearing' role markets already play. It is expected that in the absence of manipulators, these price differences would have a random, unbiased impact on

both the buy and the sell sides, and thus exchanges could effectively act as trustworthy clearing parties. In order to achieve this, it will be necessary for markets to collect and store all relevant information, for instance, on order book and its depth, making exchanges accountable for any price differences that they help create or disseminate that is not properly compensated at the end of the trading day.

From the monitoring and surveillance point of view, it is clear then, that it is crucial that delays are not introduced intentionally to the systems, and thus it is absolutely necessary that exchanges can filter out invalid and or strange quotes before they are processed by the trading systems. With respect to the quote life, for instance, a minimum life could be enforced, much like the 'time-in-force (TIF)' policy in which all quotes are considered valid for a minimum of, for example, 50 ms. With respect to the price and quantity, collars could be introduced, rejecting for instance any order which is too far from the last valid best bid or ask offer, which would also help to tackle any stub-quoting problems. In addition, cancellation ratios could be monitored, and all cancellations that exceed a certain limit should also be rejected.

Using the data generated in Simulation 4; Figure 6-41 and Figure 6-42 show a view of some indicators that could be implemented to monitor the markets.

Figure 6-41 shows how, based on the best ask price, a price collar might look. The collar was calculated following the new quotation standards for market makers that were introduced in the US after the flash crash. The rule states that all quotes submitted by market makers during market hours should not be more than 8% away from the last best bid or ask offer. This kind of collar could be extended to all orders, not only to market makers. Just before the 20,000 ms mark it is possible to appreciate how the quote with the $0 price and 0 quantity, that was submitted to the International Securities Exchange, should have been rejected (for graphical purposes only, the ask price of that order was set to $42.333). Figure 6-41 also shows an indicator of the quote life, which has been re-scaled here for graphical purposes.

The 40 mark is the equivalent for a 0 ms quote life. At mark 41 an overlay function is plotted with the intention of graphically showing what could have happened if all

quotes with a quote life of less than 1 ms would have been rejected. It easily to see how a significant amount of quotes could have been rejected in this way.

Similarly, in Figure 6-42 based on the last trade price, trading halts could have been triggered following the logic of the 'programme for trading pauses due to extraordinary volatility' which was introduced after the flash crash. The programme states that the price move required to trigger a trading pause shall be 30% or more for securities priced at $1 or higher, and 50% or more for securities priced at less than $1; it affects all trades occurring after 8th August 2011. Using the 10% deviation rule, which was in place at the time of the flash crash, Figure 6-42 shows how trading halts could only have occurred at the beginning of the simulation. Because the simulation only looked at a two-minute period, rather than the five-minute trigger rule stipulated by SEC, the price average calculation used all the prices as they became available in the simulation. This explains why the average prices and the collar in Figure 6-39 have the peculiar shape that can be observed up to the 40000 mark.

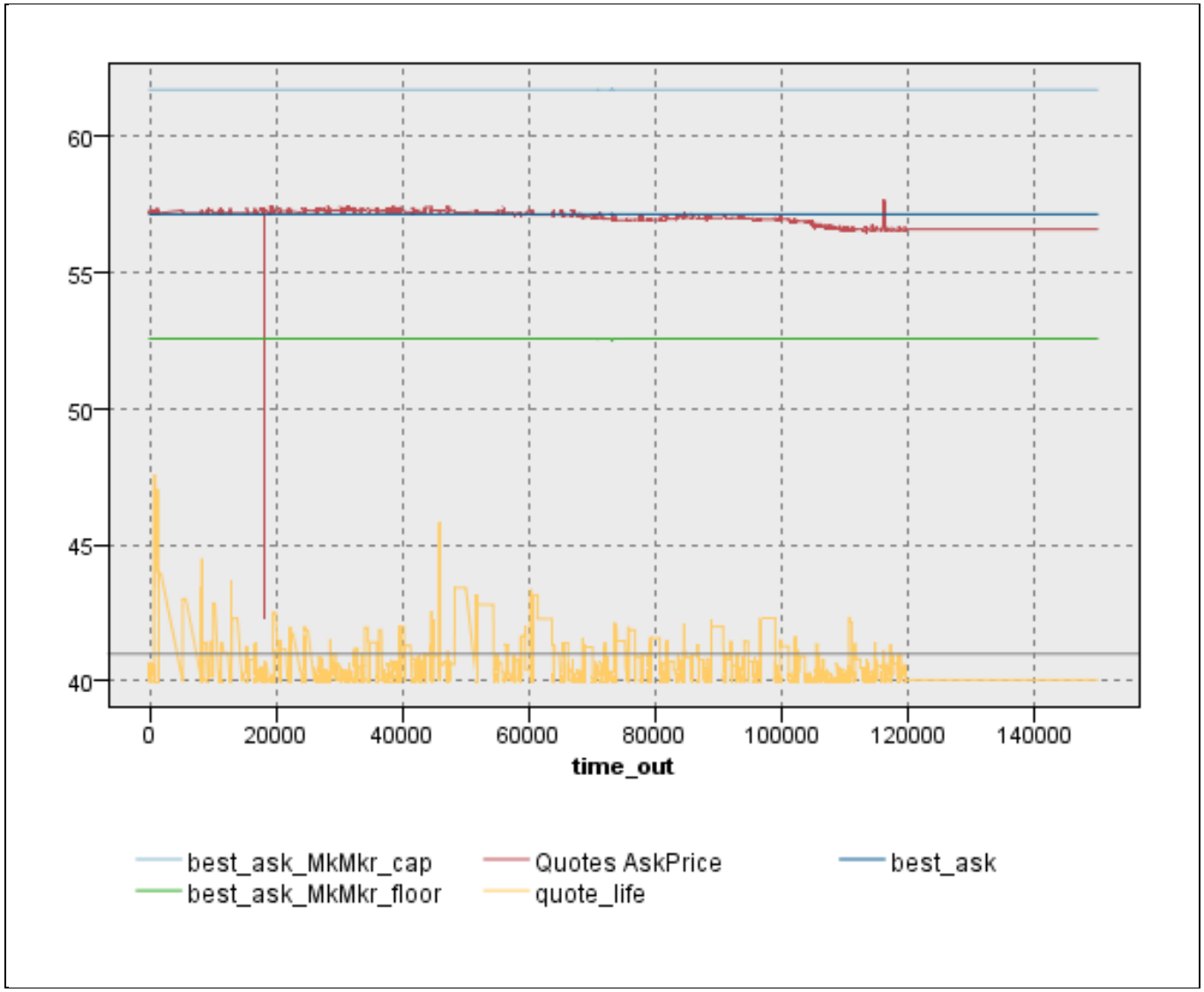


Figure 6-41 Case Study 4: Order Protection Indicators

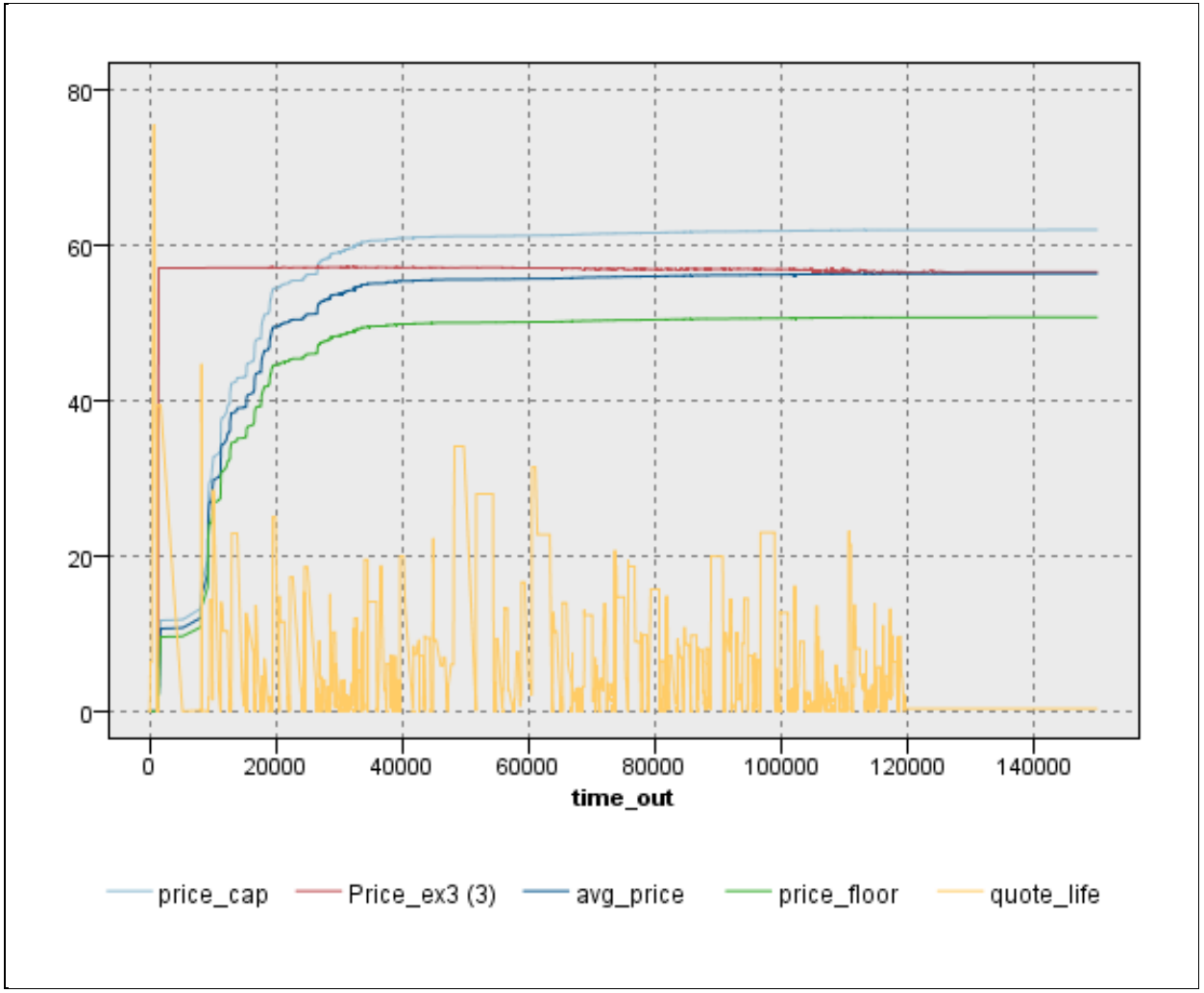


Figure 6-42 Case Study 4: Trade Protection Indicators

Figure 6-43 shows what an ideal market monitoring and surveillance engine organisation should be like. In particular, the 'Case Rules' component in the real-time engine layer could be used to implement the order protection and trade protection rules by creating indicators which use complex event processors, for example, to calculate and deploy filtering rules to reject or accept invalid orders before these are processed in the trading system. This component could also calculate, monitor, and take actions based on indicators such as the order cancellation ratio (see for example Figure 6-41), rejecting all cancelling orders that exceed a certain accumulated limit (for instance, 10 cancels over the last 50 valid quotes). Similarly, the 'Circuit Breaker' component could be used to calculate, monitor, and take actions, also using also complex event processors, that halt trading whenever price collars are exceeded. As in previous cases, the '2M-3S' framework worldviews and SVN views could be used to represent any configuration of customer-provider that satisfies the delivery of monitoring and surveillance services, either in an single jurisdiction or in a multiple jurisdiction setting.

In terms of coverage of concepts and requirements Figure 6-44 presents the taxonomy instantiation, which includes all the necessary HFT related components and terms, which were defined in previous sections of this thesis, such as real time engines, complex event processors, and stuff quoting definition.

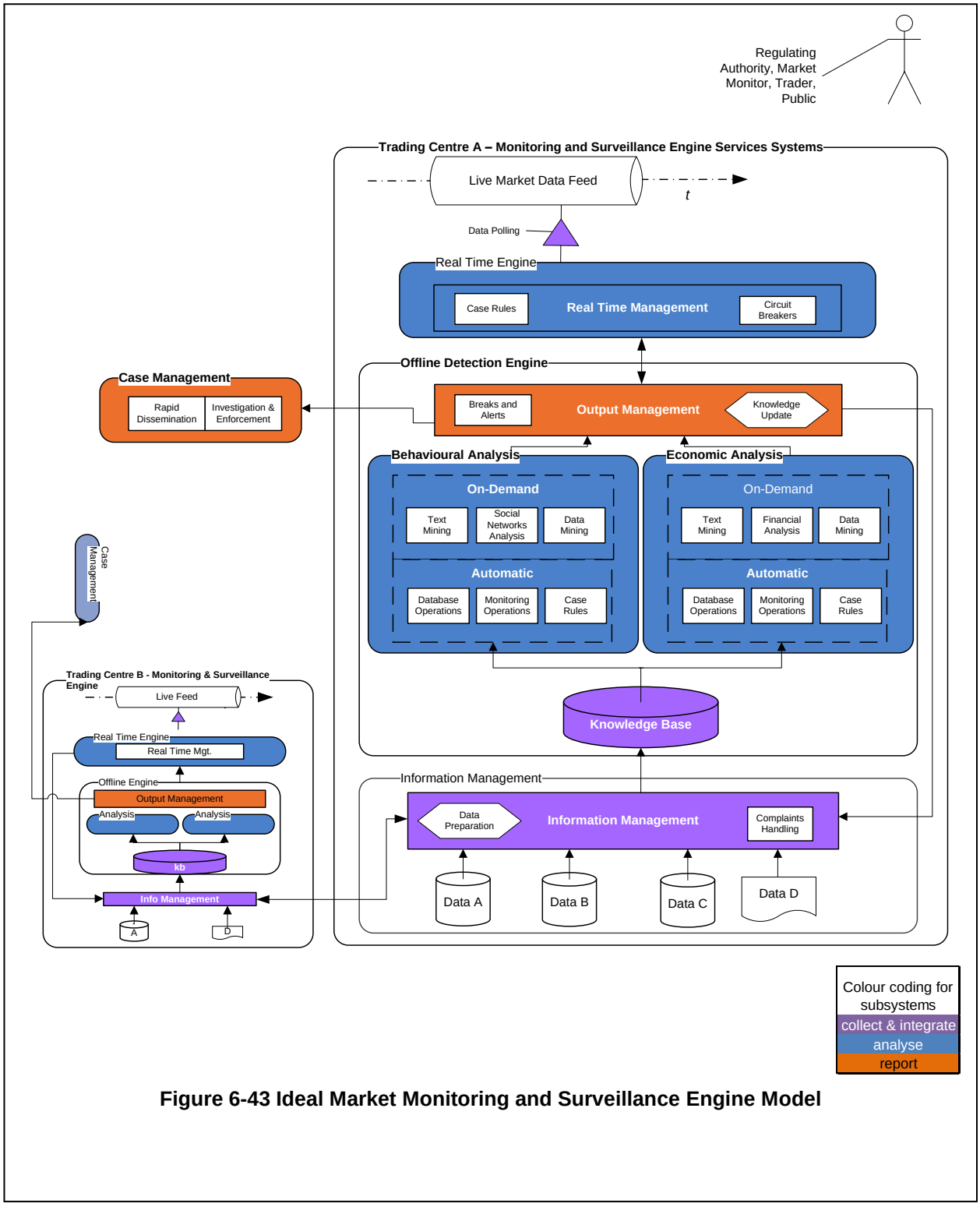


Figure 6-43 Ideal Market Monitoring and Surveillance Engine Model

Manipulation							
Manipulation	**Cross-Jurisdiction**	**Agents**	**Action**	**Venue**	**Time**	**Target Asset**	**Effects**
HFT Stuff-quoting; abuse of market power; trade-based	no	Outsiders; on their own; own benefit	Buy/Sell trades and orders	SRO: ISE, NYSE Arca, NASDAQ, BATS; Electronic HFT supported	High frequency speed, inside trading hours; all year	Financial; Equity; common stocks	Financial/Economic; structured data; inside venue

Stakeholders		
Investor	**Intermediary**	**Regulator**
Executives; Manipulators, net savers	HFT	SEC, FINRA

Detection Engine			
	Collect & Integrate	**Analyse**	**Report**
Information Management	Data preparation; outlier detection; news mining, previous patterns		
Off-line detection engine		Behavioural and Economic Analysis; Automatic & On demand; Decision Trees; Social networks Analysis	Visualization; OLAP; Breaks and Alerts
Real-time detection engine	real-time sampling	complex events processing	visualization; automatic breaks and alerts

Requirements		
Requirements	**Class**	**Type**
1 to 6, 8,	Analysis/Integration	Functional
9 to 12	Data capture and pre-processing	Functional
16, 17, 20 to 23, 25	Information Management/Output Management	Functional
n/a	General conditions of the system and environment	Non-Functional
n/a	Skills and capabilities of the customer	Non-Functional

Figure 6-44 Case Study 4: Taxonomy Instantiation

6.5 Chapter summary

The cases described here were selected on the basis that together they form a set of scenarios that allows the researcher to present the flexibility and competence of the '2M-3S' for the description of cases and prototype detection engines built to deal with different types of market manipulations or rules monitoring. This includes not only different types of market manipulations, but also different configurations of customers, jurisdictions, architectures, data types and analysis components that cover a significant part of the body of knowledge that has been introduced in this thesis.

These case studies can be considered *critical* cases, because they consider a careful and strategic selection of scenarios in which it is possible to observe the characteristics of the '2M-3S' framework that permits logical deduction of the type 'if this (not) works/is (not) valid for this case, then it can be applied to all (no) cases' (Flyvbjerg, 2006). Cases can also be considered as *paradigmatic cases*, in the sense they will hopefully help develop a common understanding for the analysis and future development of market monitoring and surveillance systems, as well as for the detection of different types of manipulations and market abuses (Flyvbjerg, 2006). Table 6-19 presents a summary of the selected cases and their main characteristics:

Table 6-19 Summary of Case Studies

Name	Description	Jurisdiction/ Customer/ Provider/ Definition	Detection Engine main characteristics
Case Study One: Trade-based manipulations	Addresses challenges relating to applying data mining techniques to detect trade based manipulations and extends previous works available in the literature by incorporating the analysis of intraday trade prices in addition to closing prices. Furthermore, the analytical models described reinforce the results of previous market manipulation studies that are based on traditional statistical and econometrical methods, providing an alternative portfolio of methods and techniques originating from the data mining and knowledge discovery areas.	Jurisdiction: Single Customer: Regulator Provider: Venue Surveillance Team	Unsupervised anomalies detection algorithm and supervised learning algorithms (C5.0 and CR&T) used to describe patterns of manipulation in a way that is understandable by humans.
Case Study Two: 'insider trading' manipulations	Discusses a case scenario that considers the perspective of a regulating authority dealing with an insider trading manipulation case. The customer is also interested in relating the unusual trading activity with unusual traders' behaviour, such as an increase in the volume of insider trading, or links between coordinated traders performing last minute purchases.	Jurisdiction: Single Customer: Regulator Provider: Venue Surveillance Team	Unsupervised anomalies detection algorithm and supervised learning algorithms (C5.0 and CR&T) used to describe patterns of manipulation in a way that is understandable by humans. Use of OLAP, queries, and visualisation techniques.
Case Study Three: Social networking tools for 'cross-border' 'wash-sales' manipulations	Deals with a case of 'cross-border' manipulation where conspirators carried out a 'pump-and-dump' scheme using off-shore accounts. Specifically, a Canadian citizen coordinated a group of manipulators, which used accounts in the Bahamas and in the Turks and Caicos Islands to engage in fraudulent activities of microcap stocks. A combination of data mining and social networking analysis is used in order to analyse patterns of manipulations.	Jurisdiction: Multiple Customer: Regulator Provider: Venue Surveillance Team	Unsupervised anomalies detection algorithm and supervised learning algorithms (C5.0 and SVMs) used to describe patterns of manipulation in a way that is human understandable. Use of OLAP, queries, social networking and visualisation techniques.
Case Study Four: High-Frequency Trading Quote Stuffing	Presents a high frequency trading analysis of a particular Exchange Trade Fund (ETF) and discusses how stuff-quoting and other HFT strategies can affect the normal function of trading systems. Suggestions are put forward on how a real-time engine could have been used to detect consistent 'stuff quoting' and 'stub-quoting', and other behaviours.	Jurisdiction: Any Customer: Any Provider: Any	A full market simulation is used to model the behaviour of trading systems, and its effects on the display book, best bids, best asks, and trade prices. A combination of indicators and views are used in order to show efficient real time case rules and circuit breaker mechanisms. Discusses an integrated market monitoring and surveillance system conceptual prototype.

Each case study presented its own results and conclusions, and knowledge gained was abstracted away to be integrated later into this final 'conclusions and lessons learnt' sub-section following the *circumscription* and *operation and goal knowledge* loops. Moreover, each case study followed the design research methodology in order to create independent artefacts that solve the particular problems that were introduced in each of them.

In Case Study 1, the challenge of building market monitoring and surveillance components was addressed. These were built using real trading and textual data, and taking as a base real cases of manipulation. In particular, as described throughout the case, the emphasis on the data mining task was to build classification models whose rules are humanly comprehensible, especially with the aim of allowing the researcher to understand the patterns in data and to compare them with previously reported patterns (*circumscription* and *operation and goal knowledge* loops). In terms of the instantiation of the case as a service system, it was shown how a value proposition is presented from the SRO (the provider) to the Regulator (the customer), in terms of services of monitoring and surveillance that are performed by the exchange using the infrastructure and the decision trees that were described in the case details.

Case Study 2, addressed the challenge of building market monitoring and surveillance components, especially data mining components, that are working prototypes for the detection of insider trading manipulations. These were built using real trading and textual data, and taking as a base a real insider trading case. This case study highlighted that the difficulty of insider trading detection was the ability of the system to discriminate between this and other types of market manipulations, in particular trade-based manipulations, as these tend to have similar patterns, i.e. jumps in prices or volume. Another important aspect of insider trading is the possibility of manipulators who target both the derivative and the stock of the company for which they possess material information, mainly because of the great leverage and reduced capital requirements when trading in the derivative markets. This aspect, however, is not touched in this case and is left for further research. In terms of the instantiation of the case as a service system, the

case study considers a similar configuration of value networks and value propositions to that presented in Case Study 1.

Case Study 3 addresses the challenge of using social network analysis components for the detection of 'wash sales', 'match trades' and in general, any kind of manipulation in which the coordination of different traders is taking place, even in the case of cross-border or cross-jurisdiction manipulations. The terminology and terms introduced in the previous chapters helped nicely to define and describe a real case scenario of cross-jurisdiction manipulation, in which a group of traders coordinated activities from different countries to target and manipulate a particular set of stocks traded in the OTCBB and Pink Sheets markets located in the USA. Although the analysis architecture is very similar to that presented in Case Study 2, the difference lies in two aspects: firstly in the availability of fully functional social networking analysis components, such as the NodeXL; and secondly in the fact that the customer and provider of the detection engine is assumed to be the same agent, i.e. the regulator. In terms of the instantiation of the case as a service system, the case study considers a configuration of value networks and value propositions in which the provider and the customer of the service is assumed to be the Regulator. The case also presents an alternative view of the world which illustrates the interactions between two jurisdictions, thereby recognising the cross-jurisdictional aspects of the case.

The purpose of Case Study 4 is to study the consequences of HFT quote-stuffing practices in market systems. In order to do so, a market simulator was built and used to model trading systems and their behaviour when confronted with HFT quote-stuffing practices. The case presented some preliminary evidence on how HFT quote stuffing can increase the gap of best bid and ask prices between markets, contrary to the evidence that HFT has helped align prices across markets (Hendershott et al., 2011, Jovanovic and Menkveld, 2011). This is very important, as manipulators could profit by artificially creating latencies in trading data feeds, that would make arbitrage possible by taking advantage of the HFT induced price differences between markets.

Using these findings, a combination of indicators and views were used in order to show how efficient circuit breaker mechanisms and order control mechanisms could be implemented by an exchange in order to safeguard the appropriate behaviour of its trading system. Moreover, suggestions are put forward on how a real-time engine could have been used to detect potential HFT manipulative behaviour and other misconducts.

Using the data generated in simulations, the case showed views of some indicators that could be implemented to monitor the markets and illustrated what an ideal market monitoring and surveillance engine organisation should look like. It suggests that the 'Case Rules' component in the real time engine layer could be used to implement the order protection and trade protection rules by creating indicators that use complex event processors, for example, to calculate and deploy filtering rules, rejecting or accepting invalid orders before these are processed in the trading system. Similarly, a 'Circuit Breaker' component could be used to calculate, monitor, and take actions, again using complex event processors, that halt trading whenever price collars are exceeded. As in previous cases, the '2M-3S' framework worldviews and SVN views could be used to represent any configuration of customer-provider that satisfies the delivery of monitoring and surveillance services, either in an single jurisdiction or in a multiple jurisdiction setting.

Overall, the cases helped us to understand how market monitoring and surveillance systems can be used to study a set of market misbehaviours that are common or that have attracted attention for their novelty. It is clear that investors and regulators could benefit by understanding how each of these cases should be addressed individually and by understanding the similarities and expected functionalities of these systems. For instance, although all four cases dealt with different types of manipulations and different types of data, it is still possible to identify common processes and components of detection engines that need to be considered when defining these systems, such as, components for data pre-processing, analysis, reporting, as well as, knowledge management components.

Another lesson that can be deducted from the cases is the expected complexity of dealing with manipulations that involve cross-jurisdiction scenarios of non equity products. This is especially important considering the lack of consolidated data available for non equity markets, both at the pre and post trade level. This fact highlights the level of coordination, perhaps through Memorandum of Understanding, that is expected from authorities in order for cross-border market monitoring and surveillance systems to work properly. Such levels of coordination are far from being optimum, and as such, this is an area of research and policy work that would require further studies.

In terms of new high-frequency trading scenarios and learning, it was possible to highlight the importance of avoiding any kind of conducts that introduce latencies in the systems. The cases were also useful to propose ways in which this problem could be addressed through independent monitoring and/or clearing services.

Finally, it is possible to identify other aspects that were not covered by these case studies. Of great help in this task are the taxonomy instantiations that were given at the end of each case study. From them it is possible to observe that little work was presented with respect to the details of text mining components, and their functionalities. This is also true for future forms of high-frequency trading manipulations, or any other real time analysis capabilities. In addition, although some future conceptual functionalities were described, such as, multimedia data analysis and voice recognition, no actual work was dedicated to these areas.

7 Conclusions

This chapter summarises this thesis. It begins by restating the motivation, aims and objectives, the structure and contributions of the work, and a discussion of the limitations. Finally, policy implications are also discussed before ending with suggestions for a number of future research areas on the basis of the work presented in the thesis

7.1 Overview of the work

The aim of the thesis is the development of a framework for analysing market monitoring and surveillance systems in order to provide a common foundation for researchers and practitioners. In order to achieve this aim, the thesis examined the new requirements for the operation of financial markets, the recent consultations on the structure and governance of EU and US markets, as well as, future usage scenarios and emerging technologies. The usefulness of the proposed framework was evaluated through four critical case studies, which not only helped to understand with practical examples the way markets monitoring and surveillance systems work, but also investigated their weaknesses, potential evolution and ways to improve them.

The thesis is organised as follows: Chapter 1 discussed extensively how the adoption of new technologies and increased competition contributed to greater market integration and market fragmentation which, unexpectedly, also helped create the settings for bizarre episodes such as the 'flash crash'. Chapter 1 emphasised the need for societies to aim at reducing uncertainty, increasing transparency and guaranteeing accountability by means of monitoring, surveillance and appropriate audit trailing of financial systems in order to safeguard the important role financial markets plays in our societies. The chapter also discussed why markets need to be monitored not only to meet the objective of reducing market manipulations and fraud but also to level the playing field for all participants. Finally, it described how, when necessary conditions are not met, market efficiency and market integrity both deteriorate.

Helping to create strong *foundations* for the 2M-3S framework, chapter 2 presented a review of market manipulations and MMSS, which relates to the *knowledge based* aspects of the methodology. It began by presenting a brief overview of how the problem of market manipulation has been historically defined and investigated by the financial and economic community, and highlighted that manipulations do indeed exist, and that they can be very profitable for manipulators but detrimental to other market participants. The chapter also proposed a number of definitions and a corresponding taxonomy for different types of market manipulations. It also provided a description of the market monitoring and surveillance systems, components and processes, and the allocation of responsibility for monitoring and surveillance. The chapter discussed the technology used to perform the tasks of monitoring and surveillance, and provided examples of systems that have been deployed. The chapter concluded with a summary of the system requirements, limitations and gaps in the existing MMSS.

Chapter 3 provided a complete review of another important part of the *knowledge base*, describing and justifying in detail the research methodology and its philosophical perspective. More specifically, chapter 3 discussed the method that was used to research and design an innovative solution that addresses the gaps identified in the understanding and functionality of market monitoring and surveillance systems. The chapter started by providing a general definition of the research approach followed by a justification of why this methodology was appropriate for the task at hand. It continued with an in-depth description of how the methodology was implemented. Chapter 4 organised the definitions, properties and parts of the *concepts* that are used in this thesis by presenting a collection of taxonomies relevant to market monitoring and surveillance systems. In order to present and exemplify the taxonomies, different manipulation scenarios and the financial terminology used to describe them was discussed.

In chapter 5 *models* and *methods* were introduced. Specifically, principles of Service-Dominant logic (SD), Service Science (SS) and Systems Thinking (ST) were used to introduce the framework for Market Monitoring and Surveillance Service Systems or '2M-3S' framework. Drawing on concepts from SD logic, SS,

and ST, the current problems and characteristics of the financial markets were described as service systems which exist and interact in a service-oriented economy. Using the ST perspective, the concepts of archetype and leverage were used to propose solutions for the current problems through systems concepts and terminology. Chapter 5 also expanded previously defined concepts and terminology by presenting a model for a typical detection engine. Details of its sub-systems and relationships between them were provided.

In terms of *evaluation* of the utility of the framework, chapter 6 presented individual *instantiations* of the artefact by means of *case studies.* Evaluation was carried out on a case–by-case basis, judging the utility of the proposed framework in solving and increasing understanding of a part or the whole of the problem and the use of the models introduced previously. Each case study presented its own results and conclusions, and knowledge gained was abstracted away to be integrated later into overall conclusions and lessons learnt sub-section following the *circumscription* and *operation and goal knowledge* loops of the research methodology.

7.2 Contributions and limitations of the work

This thesis delivers several contributions to knowledge. In relation to research question 1, the thesis examines the context in which market monitoring and market surveillance systems are currently being used, presenting a set of definitions and requirements that address the gaps and expected functionalities of such systems. In addition, this thesis analysed their characteristics, processes and performance using several analysis perspectives, such as the systems thinking *about* and *from* reasoning perspectives. For example, the *limits to grow* and *shifting the burden* archetypes were used to talk about typical behaviour patterns that can be found in financial markets, proposing new ways of delivering market manipulations and surveillance systems through independent and/or commercial providers of such services.

In relation to research question 2, the thesis proposed a taxonomy of concepts which form the basis for a body of knowledge of market monitoring and surveillance systems. This contribution delivers a more comprehensive approach

for describing and organising the concepts in the domain. These taxonomies were also used to define the concepts involved in each of the case studies, and also to identify concepts available in the domain that were not covered in the thesis.

In relation to research question 3, the thesis adopted concepts and theories from service systems, service dominant logic and systems thinking in order to examine under a socio-technical perspective markets as service systems, where, monitoring and surveillance, consists of various services provided by different stakeholders including market operators, regulators, and third parties.

In relation research question 4, this thesis proposed and designed a new framework for market monitoring and surveillance systems. Figure 5-21 presents a schematic view of the components and underpinnings of the 2M-3S framework scope. In this framework the market monitoring and surveillance activities are described as a user or customer-driven service value network. Different SVN configurations were put forward, specifically for two types of market monitoring and surveillance service provision options: a 'close' and an 'open stake-holder' SVN. These represent two examples of service provision, one in which the regulators are empowered as customers and another where individual investors make use of independent commercial market monitoring and surveillance services.

These SVN configurations were deployed and expanded by means of their instantiation in each case study. This included different types of manipulation scenarios, as well as, different types of data, markets and analysis techniques. Together they helped us to identify generic processes and resources that are necessary for dealing with different market misbehaviours, as well as, to recognize weaknesses and future usage scenarios that would require new and improved functionalities and methods. Another important contribution of this work is the analysis of both cross border and high frequency trading manipulation cases. These manipulation scenarios have been dealt infrequently in the literature, but are expected to be of great importance in the near future. This expectation is based on the recent consultations on the structure and governance of EU and US markets and also on the ubiquitous adoption of information and communication technologies of financial markets around the world.

In relation to research question 5, the thesis discussed the implications of the proposed framework under each of the cases examined, including cross-border and HFT scenarios. It also discussed different kinds of possible business models that may be applicable, and finally, drew conclusions from a wide variety of perspectives across all the case studies. Implications and future work are further discussed on section 7.3.

Despite the extensive work reported in this thesis, there are still challenges that need to be addressed through further investigation. With respect to the construction of the framework, is still possible to enhance and expand its worth, being a limitation a formal description of its structures and properties. Future work will have to deal with the enhanced description of relationships and constraints of service systems, presenting details, for instance, about value propositions, novel service configurations and analytical and mathematical properties that could lead to a full formalization of the framework. In addition, this work shares several of the caveats of previous work regarding a selection bias towards poor manipulations (Aggarwal and Wu, 2006). This stems from the fact that only those cases in which at least a suspicion of manipulation is present, either by a formal SEC or market investigation, have been used. It is possible, therefore, that cases in which manipulations occurred but were not observed or reported may have been ignored, likewise cases in which the SEC did indeed carry out an investigation but this did not result in any action against the fraudsters due to lack of evidence or budget constraints.

Regarding the coverage of the cases and datasets, focusing on only a few case studies can potentially skew the outcomes and the rules of the models for detecting manipulations, since there could have been structural changes in the way the financial markets and fraudsters work that are not taken into consideration in the selected sample. However although the knowledge gained from the analysis of the cases is very valuable, this thesis pursues as a higher goal the creation of a common body of knowledge, and this can still be achieved even if particular nuggets of knowledge, such as the detection rules, are biased or become obsolete.

7.3 Implications and Future Work

Further research should address further evaluation of the framework, especially with respect to the different users and consider aspects such as completeness, usability, and usefulness. As mentioned previously, future work should also fully formalise the framework describing its analytical and mathematical properties. Future work should also look for possible extensions of the framework, for instance, a section of the framework dedicated exclusively to deal with text mining techniques could be of great help in order to complement and enhance the body of knowledge that was delivered in this thesis. In particular, a formal ontology of concepts, definitions and properties could help to enhance the detection capabilities and automation of the surveillance systems, but could also help to organise the knowledge in a way that is more easily accessible for humans and machines.

In terms of new instantiations of the framework, further research should address insider trading as applied to both derivatives and other non equity markets, and should also consider additional cross-border scenarios based on the definitions provided. An area for future work could be the consideration of different ways of studying the patterns of cross-border trading in real time data and linking them with off-line data. With respect to HFT manipulation scenarios, whole new avenues of research are still to be discovered. One path that could be followed is the enhancement of the HFT trading simulator that was presented in this thesis. This could imply the development of a trading simulator which allows the testing and checking of more types of trading rules and halts, as well as, to perform all kinds of stress tests to HFT trading algorithms and trading systems. This new simulator should be tested in new cases scenarios and should include functionalities for dealing with other products and securities, again with special emphasis on cross-border and non equity markets scenarios. As proposed in the thesis, this simulator could be implemented adopting a culture of 'many-to-many', open, transparent and global monitoring and surveillance services, thus, considering simulations that are built over a fully consolidated tape scenario.

In terms of policy implications, in Case Study 4 a question was posed in this respect: in the presence of HFT, what steps that are necessary to safeguard the

spirit of rules such as the Order Protection Rule, which were created to preserve market integrity, when advances in market efficiency are in direct conflict with such rules? In the researcher's opinion, it is first necessary to increase awareness of the unintended consequences of increasing the speed of trading, as well as to create mechanisms to guarantee that these consequences are managed and kept under a certain minimum. A plausible solution would be to enforce a policy or rule in which exchanges are obliged to provide retroactive compensation to investors for any losses caused by the desynchronization of prices. In order to achieve this, markets would need to collect and store all relevant information about, for instance, the order book and its depth, in order to make exchanges accountable for any price differences that they help create.

From the monitoring and surveillance point of view it is crucial that delays are not introduced to the systems deliberately, and thus it is essential that exchanges can filter out invalid and/or strange quotes before they are processed by the trading systems. With respect to the quote life, for instance, a minimum duration could be enforced. With respect to the price and quantity, collars could be introduced, rejecting, for example, any order which is too far from the last valid best bid or ask offer, which would additionally help to tackle any stub-quoting problems. Cancellation ratios could also be monitored, and all cancellations that exceed a certain limit should also be rejected.

Overall, this thesis delivered a novel artefact with working tools for the analysis and development of market monitoring and surveillance systems. It additionally presented an organised body of knowledge for the detection of different types of manipulation and market abuse, describing not only concepts and models, but also data and text mining tools that could be used to tackle information- and trade-based manipulations. These could be used by regulating authorities and in self- or non-self-regulated venues, as well as by other types of customers, such as individual investors, institutional investors and broker-dealer firms among others. These could also enable a new configurations and business models for market monitoring and surveillance, in which different types of customers could carry out peer and/or public monitoring and surveillance in the form of on-demand services.

8 References

ABIOYE, E. (2011) BUSINESS INTELLIGENCE – FINANCIAL FRAUD A CROSS-BORDER CASE. *School of Computer Science.* Manchester, The University of Manchester.

AGGARWAL, R. K. & WU, G. (2006) Stock Market Manipulations. *Journal of Business,* 79**,** 1915-1953.

AITKEN, M. J., HARRIS, F. H. & JI, S. (2010) Trade-Based Manipulation and Market Effciency after the Introduction of Real-Time Surveillance: A Cross-Market Comparison. *Working paper series available at SSRN eLibrary*.

ALDRIGE, I. (2010) *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems.,* New Jersey, John Willey & Sons.

ALLEN, F. & GALE, D. (1992) Stock-price manipulation. *Review of Financial Studies,* 5**,** 503-529.

ALLEN, F. & GORTON, G. (1992) Stock price manipulation, market microstructure and asymmetric information. *European Economic Review,* 36**,** 624-630.

BAGNOLI, M. & LIPMAN, B. L. (1996) Stock price manipulation through takeover bids. *RAND Journal of Economics,* 27**,** 124-147.

BALLS, E. (2007) Speech by the Economic Secretary to the Treasury. *FSA Regulation Conference.*

BARANDELA, R, SANCHEZ, J, S., GARCIA, V, RANGEL & E, G. (2003) Strategies for learning in class imbalance problems. *Pattern Recognition,* 36**,** 3.

BARILE, S. & POLESE, F. (2010) Smart Service Systems and Viable Service Systems: Applying Systems Theory to Service Science. *Service Science,* 2**,** 21-40.

BARNIER, M. (2010) Keynote Speech at the Public hearing on the “Review of the Markets in Financial Instruments Directive (MiFID)”. Brussels.

BENABOU, R. & LAROQUE, G. (1992) Using privileged information to manipulate markets: Insiders, gurus, and credibility. *Quarterly Journal of Economics,* 107**,** 921.

BOARDMAN, J. & SAUSER, B. (2008) *Systems thinking: coping with 21st century problems,* Boca Raton, CRC Press.

BODIE, Z., KANE, A. & MARCUS, A. (2008) *Essentials of Investments,* New York, McGraw-Hil.

BOLLERSLEV, T. (1986) Generalized autoregressive conditional heterocedastocity. *Journal of Econometrics,* 31**,** 307-327.

BROGAARD, J. (2010) High Frequency Trading and Its Impact on Market Quality. *Working Paper Series SSRN eLibrary.*

BUCKINGHAM SHUM, S. (1997) Negotiating the Construction and Reconstruction of Organisational Memories. *Journal of Universal Computer Science,* 3**,** 899--928.

CARHART, M. M., KANIEL, R. O. N., MUSTO, D. K. & REED, A. V. (2002) Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds. *Journal of Finance,* 57**,** 661-693.

CASTI, J. L. (1992) *The simply complex in reality rules,* New York, Wiley.
CHAMBERLAIN, T. W., CHEUNG, C. S. & KWAN, C. C. Y. (1989) Expiration-Day Effects of Index Futures and Options: Some Canadian Evidence. *Financial Analysts Journal,* 45**,** 67.
CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C. & WIRTH, R. (2000) CRISP-DM 1.0 Step-by-step data mining guide.
CHLISTALLA, M. (2011) High-frequency trading, better than its reputation? , Deutsche Bank Research.
CHRISTIANSEN, H. (2000) Cross-Border Trade in Financial Services: Economics and Regulation. *Financial Market Trends,* 2000**,** 216.
CHRISTIANSEN, H. & KOLDERTSOVA, A. (2009) The Role of Stock Exchanges in Corporate Governance. *Financial Market Trends,* 1**,** 1-32.
CHUNSHENG, Z., JIANPING, M. E. I. & GUOJUN, W. U. (2003) Behavior Based Manipulation: Theory and Prosecution Evidence. *Working Papers (Faculty) -- University of Michigan Business School.*
COMERTON-FORDE, C. & RYDGE, J. (2006) Market Integrity and Surveillance Effort. *Journal of Financial Services Research,* 29**,** 149-172.
CONCEICAO, C. (2006) Tackling cross-border market abuse. *Journal of Financial Regulation and Compliance,* 14**,** 29-36.
CUMMING, D. & JOHAN, S. (2008) Global Market Surveillance. *American Law and Economics Review,* 10**,** 454-506.
DONOHO, S. (2004) Early detection of insider trading in option markets. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.* Seattle, WA, USA, ACM.
DOW JONES FACTIVA (2009).
EDGE, M. (2007) Proactive Fraud Managemet using Financial Fraud Modelling Language and Signature Based Methods. *Informatics - Manchester Business School.* Manchester, University of Manchester.
EDGE, M. E., SAMPAIO, P. R. F. & CHOUDHARY, M. (2007) Towards a Proactive Fraud Management Framework for Financial Data Streams. *Dependable, Autonomic and Secure Computing, 2007. DASC 2007. Third IEEE International Symposium on.*
EFRON, B. (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics,* 7**,** 1-26.
EUROPEAN COMMISSION (2010) Consultation on the review of the Markets in Financial Instruments Directive (MiFID).
FAMA, E. F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance,* 25**,** 383-417.
FAN, H., HUANG, Y.-A., WANG, H. & YU, P. S. (2004) Active mining of data streams. *Proceedings of the Fourth SIAM International Conference on Data Mining.*
FAO (2011) *FAO Food Price Index.* Food and Agriculture Organization of the United Nations.
FAWCETT, T. & PROVOST, F. (1997) Combining data mining and machine learning for effective fraud detection. *AI Approaches to Fraud Detection and Risk Management.*

FAYYAD, U. & PIATETSKY-SHAPIRO, G. (1996) From data mining to knowledge discovery in databases. *AI Magazine,* 17**,** 37.

FELIXSON, K. & PELLI, A. (1999) Day end returns--stock price manipulation. *Journal of Multinational Financial Management,* 9**,** 95-127.

FINANCIAL SERVICES AUTHORITY (2007) Handbook.

FLYVBJERG, B. (2006) Five Misunderstandings about Case-Study Research. *Qualitative Inquiry,* 12**,** 219-245.

FRIEDER, L. & ZITTRAIN, J. (2007) *Spam Works: Evidence from Stock Touts and Corresponding Market Activity, Harvard University Berkman Center for Internet and Society*

Working paper series.

GERARD, B. & NANDA, V. (1993) Trading and Manipulation Around Seasoned Equity Offerings. *Journal of Finance.*

GOLDBERG, H., KIRKLAND, J., LEE, D., SHYR, P. & THAKKER, D. (2003) The NASD Securities Observation, News Analysis & Regulation System (SONAR). *Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence.* Acapulco, Mexico.

GREGOR, S. (2006) The Nature of Theory in Information Systems. *MIS Quarterly,* 30**,** 611-642.

GREGOR, S. & JONES, D. (2007) The Anatomy of a Design Theory. *Journal of the Association for Information Systems,* 8**,** 312-335.

GRÖNROOS, C. (2008) Adopting a service business logic in relational business-to-business marketing: value creation, interaction and joint value co-creation. *Otago Forum.*

GRUBER, T. (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition,* 5**,** 199-220.

HARRIS, L. (1989) A Day-end Transaction Price Anomaly. *Journal of Financial and Quantitative Analysis,* 24**,** 29-45.

HENDERSHOTT, T., JONES, C. M. & MENKVELD, A. J. (2011) Does Algorithmic Trading Improve Liquidity? *The Journal of Finance,* 66**,** 1-33.

HEVNER, A., MARCH, S., JINSOO, P. & RAM, S. (2004) Design Science in Information Systems Research. *MIS Quarterly,* 28**,** 75-105.

ITURRIETA, F. & BOADLE, A. (2011) Merged South American bourses off to slow start. Reuters U.S.

JARROW, R. A. (1992) Market Manipulation, Bubbles, Corners, and Short Squeezes. *Journal of Financial & Quantitative Analysis,* 27**,** 311-336.

JARROW, R. A. (1994) Derivative Security Markets, Market Manipulation, and Option Pricing Theory. *Journal of Financial & Quantitative Analysis,* 29**,** 241-261.

JENSEN, D. (1997) Prospective assessment of AI technologies for fraud detection: A case study. *AAAI Workshop on AI Approaches to Fraud Detection and Risk Management.*

JOVANOVIC, B. & MENKVELD, A. J. (2011) Middlemen in Limit-Order Markets. *Working Paper Series SSRN eLibrary.*

KATIKA, A. (2011) Investigating Financial Fraud in HFT. *School of Computer Science.* Manchester, The University of Manchester.

KHWAJA, A. I. & MIAN, A. (2005) Unchecked intermediaries: Price manipulation in an emerging stock market. *Journal of Financial Economics,* 78**,** 203-241.
KIRKLAND, J. D. & SENATOR, T. E. (1999) The NASD Regulation Advanced-Detection System. *AI Magazine,* 20**,** 55.
KNAUP, H., SCHIESSL, M. & SEITH, A. (2011) Speculating with Lives. How Global Investors Make Money Out of Hunger. Der Spiegel Online.
KOPANAKIS, I. & THEODOULIDIS, B. (2003) Visual data mining modeling techniques for the visualization of mining outcomes. *Journal of Visual Languages & Computing,* 14**,** 543-589.
KUBAT, M. & MATWIN, S. (1999) Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. *14th International Conference on Machine Learning.* Nashville, USA.
KUMAR, P. & SEPPI, D. (1992) Futures Manipulation with `Cash Settlement'. *Journal of Finance,* 47**,** 1485-1502.
KWAN, S. K. & MIN, J. H. (2008) An Evolutionary Framework of Service Systems. *Journal of Harbin Institute of Technology,* 15.
KWAN, S. K. & YUAN, S.-T. (2011) Customer-Driven Value Co-Creation in Service Networks. *The Science of Service Systems.* Springer.
LEINWEBER, D. J. (2001) Three Hundred Years of Stock Market Manipulations. *Journal of Investing,* 10**,** 7.
LEUNG, A., ZHUANG, Y. & FONG, S. (2004) On designing a flexible e-payment system with fraud detection capability. *e-Commerce Technology, 2004. CEC 2004. Proceedings. IEEE International Conference on.*
MADRIGAL, A. (2010) Market Data Firm Spots the Tracks of Bizarre Robot Traders. The Atlantic.
MAGLIO, P. & SPOHRER, J. (2008) Fundamentals of service science. *Journal of the Academy of Marketing Science,* 36**,** 18-20.
MAGLIO, P. & SPOHRER, J. (2010a) New York, Springer.
MAGLIO, P. & SPOHRER, J. (2010b) *Why a handbook?,* New York, Springer.
MAHONEY, P. G. (1999) The stock pools and the Securities Exchange Act. *Journal of Financial Economics,* 51**,** 343-369.
MARCH, S. T. & SMITH, G. F. (1995) Design and natural science research on information technology. *Decision Support Systems,* 15**,** 251-266.
MAYHEW, D. & ANDERSON, K. (2007) Whither Market Abuse (in a More Principles-Based Regulatory World)? *Journal of International Banking Law and Regulation,* 22**,** 515-531.
MERRICK, J. J., NAIK, N. Y. & YADAV, P. K. (2005) Strategic trading behavior and price distortion in a manipulated market: Anatomy of a squeeze. *Journal of Financial Economics,* 77**,** 171-218.
MONGKOLNAVIN, J. & TIRAPAT, S. (2009) Marking the Close analysis in the Thai Bond Market Survaillence using association rules. *Expert Systems with Applications,* 36**,** 8523-8527.
NANEX (2010) Analysis of the Flash Crash.
NANEX (2011) May 6'th 2010 Flash Crash Analysis - Continuing Developments and Research.
NASDAQ OMX GROUP (2011) NASDAQ data on demand services.

NEVILLE, J., JENSEN, D., FRIEDLAND, L. & HAY, M. (2003) Learning relational probability trees. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* Washington, D.C., ACM.

NEVILLE, J., ŞIMŞEKZG, Ö., JENSEN, D., KOMOROSKE, J., PALMER, K. & GOLDBERG, H. (2005) Using relational knowledge discovery to prevent securities fraud. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining.* Chicago, Illinois, USA, ACM.

OGUT, H., DOGANAY, M. & AKTAS, R. (2009) Detecting stock-price manipulation in an emerging market: The case of Turkey. *Expert Systems with Applications,* 36**,** 8523-8527.

PHUA, C., LEE, V., SMITH, K. & R., G. (2005) A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*.

PIRRONG, C. (2004) Detecting Manipulation in Futures Markets: The Ferruzzi Soybean Episode. *American Law and Economics Review,* 6**,** 28-71.

PIRRONG, S. C. (1995) The self-regulation of commodity exchanges: the case of market manipulation. *Journal of Law and Economics,* 38**,** 141-206.

POLANSKY, S., KULCZAK, M. & FITZPATRICK, L. (2004) NASD Market Surveillance Assessment and Recommendations. Final Report. *Achievement of Market Friendly Initiatives and Results Program (AMIR 2.0 Program).* National Association of Securities Dealers (NASD) and U.S. Agency of International Development.

PRITCHARD, A. C. (2003) Self-regulation and securities markets. *Regulation,* 26**,** 32-39.

PUTNINS, T. J. (2009) Closing price manipulation and integrity of stock exchanges. *Discipline of Finance, Faculty of Economics and Business.* Sydney, University of Sydney.

QI, R. G., FAN, Z., SHEN, H. & YU, M. (2007) Editorial: Towards Service Science, Engineering and Practice. *International Journal of Services Operations and Informatics (IJSOI),* 2**,** 103-113.

RITTEL, H. J. & WEBBER, M. M. (1984) *Planning problems are wicked problems,* New York, John Wiley & Sons.

SADOGHI, M., LABRECQUE, M., SINGH, H., SHUM, W. & JACOBSEN, H.-A. (2010) Efficient event processing through reconfigurable hardware for algorithmic trading. *Proc. VLDB Endow.,* 3**,** 1525-1528.

SALERNO, J., J. , CARDILLO, R., A. & ZHANG, Z., M. (2005) Comparing various algorithms for discovering social groups with uni-party data. IN BELUR, V. D. (Ed.). SPIE.

SCHAPIRO, M. L. (2010) Speech by SEC Chairman: "Strengthening Our Equity Market Structure". New York, U.S. Securities and Exchange Commission.

SENATOR, T. E. (2000) Ongoing management and application of discovered knowledge in a large regulatory organization: a case study of the use and impact of NASD Regulation's Advanced Detection System (RADS). *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.* Boston, Massachusetts, United States, ACM.

SENGE, P. M. (1994) *The Fifth Discipline: the art and practice of the learning organization,* New York, Currency Doubleday.
SHARPE, W. F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk *Journal of Finance,* 19**,** 425-442
SHON, D. A. (1993) *The reflective practitioner: How professionals think in action,* New York, Basic Books.
SILVER, M. S., MARKUS, M. L. & BEATH, C. M. (1995) The information technology interaction model: a foundation for the MBA core course. *MIS Q.,* 19**,** 361-390.
SIMON, H. A. (1996) *The science of the Artificial,* Cambridge, MA., MIT Press.
SMITH, M., MILIC-FRAYLING, N., SHNEIDERMAN, B., MENDES RODRIGUES, E., LESKOVEC, J. & DUNNE, C. (2010) NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010. Social Media Research Foundation, http://www.smrfoundation.org.
SPOHRER, J. C., ANDERSON, L., PASS, N. & AGER, T. (2008) Service Science eService Dominant Logic. *Otago Forum.*
SPOHRER, J. C., MAGLIO, P., BAILEY, J. & GRUHL, D. (2007) Steps Toward a Science of Service Systems. *IEEE Computer,* 40**,** 71-77.
STANDARD AND POOR'S COMPUSTAT RESOURCE CENTER (2009).
STOLL, H. R. & WHALEY, R. E. (1991) Expiration-Day Effects: What Has Changed? *Financial Analysts Journal,* 47**,** 58-72.
TAKEDA, H., VEERKAMP, P. & YOSHIKAWA, H. (1990) Modeling Design Process. *AI Magazine.*
TANIGUCHI, M. & HAFT, M. (1998) Fraud detection in communication networks using neural and probabilistic methods. *Proceedings of the IEEE International conference on Acoustics, Speech, and Signal Processing.*
THE WORLD BANK (2011) Food Price Watch.
THEODOULIDIS, B. (2011) Information Management: A Wicked Problem? *Working Paper Series.*
U.S. COMMODITY FUTURES TRADING COMMISSION & U.S. SECURITIES AND EXCHANGE COMMISSION (2010a) Findings Regarding the Market Events of May 6, 2010. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues.
U.S. COMMODITY FUTURES TRADING COMMISSION & U.S. SECURITIES AND EXCHANGE COMMISSION (2010b) Preliminary Findings Regarding the Market Events of May 6, 2010. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues.
U.S. SECURITIES AND EXCHANGE COMMISSION (2010a) Concept Release on Equity Market Structure. Washington DC.
U.S. SECURITIES AND EXCHANGE COMMISSION (2010b) SEC Approves New Rules Prohibiting Market Maker Stub Quotes. *Press Release.*
VAISHNAVI, V. & KUECHLER, W. (2004) Design Research in Information Systems. January 20, 2004, last updated August 16, 2009 ed.
VAN BOMMEL, J. (2003) Rumors. *Journal of Finance,* 58**,** 1499-1520
VARGO, S. L. & LUSCH, R. F. (2004) Evolving to a New Dominant Logic for Marketing. *Journal of Marketing.* American Marketing Association.

VARGO, S. L. & LUSCH, R. F. (2008) Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science.* Springer Science & Business Media B.V.

VARGO, S. L., MAGLIO, P. P. & AKAKA, M. A. (2008) On value and value co-creation: A service systems and service logic perspective. *European Management Journal.*

VILA, J.-L. (1989) Simple games of market manipulation. *Economics Letters,* 29**,** 21-26.

VITALE, P. (2000) Speculative Noise Trading and Manipulation in the Foreign Exchange Market. *Journal of International Money and Finance,* 19**,** 689-712.

WESTPHAL, C. & BLAXTON, T. (1998) *Data mining solutions: methods and tools for solving real-world problems,* New York, NY, US, John Wiley & Sons, Inc.

WHARTON RESEARCH DATA SERVICES (2009).

WILBY, J., MACAULAY, L. & THEODOULIDIS, B. (2011) Intentionally holistic knowledge intensive service systems. *International Journal of Services Technology and Management,* 16**,** 126-140.

WITTEN, I. & FRANK, E. (2005) *Data mining : practical machine learning tools and techniques,* San Francisco, Calif, London : Morgan Kaufmann.

XIA, M. (2007) Applying data mining in market abuse detection. *Faculty of Engineering.* Manchester, University of Manchester.

YAHOO! FINANCE (2009).

YUFENG, K., CHANG-TIEN, L., SIRWONGWATTANA, S. & YO-PING, H. (2004) Survey of fraud detection techniques. *Networking, Sensing and Control, 2004 IEEE International Conference on.*

ZAKI, M., DÍAZ, D. & B.THEODOULIDIS (2010) Using Text Mining To Analyze Quality Aspects Of Unstructured Data: A case Study For 'Stock-Touting' Spam Emails. *16th Americas Conference on Information Systems (AMCIS 2010).* Lima, Peru.

ZHONGFEI, Z., JOHN, J. S. & PHILIP, S. Y. (2003) Applying data mining in investigating money laundering crimes. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* Washington, D.C., ACM.

ZWEIG, J. (1997) Watch out for the year-end fund flimflam. *Money Magazine,* November**,** 130-133.

Appendix 1: Watch-list for suspicious block hours and hit ratio

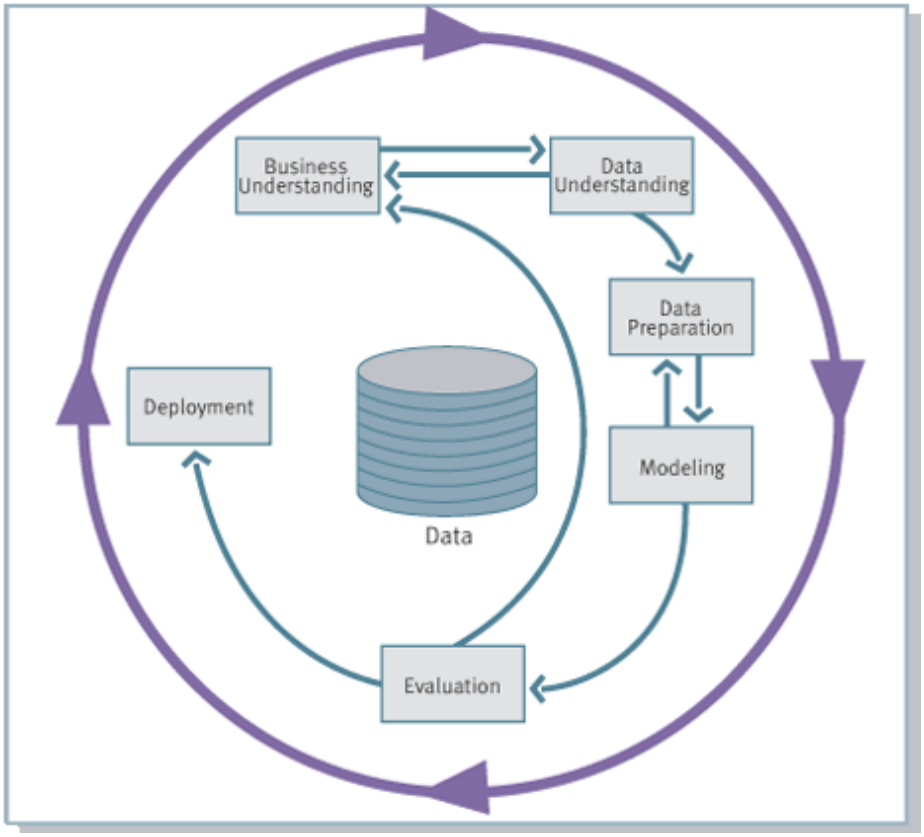
This appendix shows the values of indicators and the classification of the blocks included in the watch list according to the level 1 C5.0 Tree. The table at the bottom reports the *confusion matrix* resulting from the application of the C5.0 tree to the blocks mentioned.

SYMBOL	DATE	X$TI_Hour	StockPrc	Size	Nr_Trades	Ri	AR	ABS_Z01	ABS_ZAR1	ABS_ZS1	ABS_ZT1	ABS_ZV1	RO2	RAR2	RS2	RT2	RSD3	RV3	QEND	PRECOND	News_Dummy	HOUR14+_FLAG	T_F_BYHAND	C5.0_Tree	HIT?
AKSY	17/01/2003	16	5.24	100	1	-0.22	-0.61	0.51	1.12	0.44	1.02	0.22	-93.3	-5.37	0.01	0.07	0.93	1.03	0	4	0	T	T	F	0
AKSY	14/02/2003	16	5.83	100	1	3.33	2.47	13.73	5.93	0.51	0.80	0.16	-170.2	-115.16	0.01	0.06	1.05	1.28	0	4	0	T	T	T	**1**
AKSY	20/02/2003	16	5.89	20900	5	-0.01	0.09	0.01	0.14	0.31	0.89	0.42	-0.1	1.28	0.42	0.17	1.09	0.57	0	4	0	T	T	F	0
AKSY	05/03/2003	16	6.34	1500	1	4.46	4.23	15.37	12.36	0.86	0.82	1.04	-125.2	116.72	0.05	0.02	0.85	0.79	0	4	0	T	T	T	**1**
AKSY	13/05/2003	12	8.68	39200	139	4.40	4.20	9.32	7.51	0.52	2.78	2.47	24.1	41.69	1.54	3.64	1.91	3.98	0	2	0	F	T	T	**1**
AKSY	28/05/2003	10	9.24	74500	135	3.93	3.79	18.49	8.88	0.56	1.34	2.08	-137.5	-24.17	1.76	2.59	0.74	0.71	0	1	0	F	T	T	**1**
AKSY	11/06/2003	12	9.81	39300	80	2.27	2.19	5.63	4.37	0.51	1.18	1.66	19.8	13.09	1.77	2.13	1.35	1.77	0	2	0	F	T	T	**1**
AKSY	20/06/2003	12	10.95	18800	57	1.64	1.64	7.36	5.44	0.31	1.28	1.61	-575.9	28.44	1.33	2.23	0.56	0.39	0	2	0	F	T	T	**1**
AKSY	24/07/2003	11	15.05	149100	405	-4.01	-4.00	11.52	8.60	4.21	8.97	3.33	-37.2	-30.47	7.57	9.09	0.70	0.59	0	1	0	F	T	T	**1**
AKSY	25/07/2003	10	11.49	576700	1574	-23.40	-22.91	28.18	25.58	17.09	18.83	1.56	591.4	197845.01	25.87	28.90	1.42	1.88	0	1	0	F	T	T	**1**
BIF	31/03/2003	15	4.42	30000	6	1.32	1.25	3.77	3.87	5.18	2.90	1.90	-41.4	-16.92	7.82	5.45	0.61	0.55	1	5	0	T	T	T	**1**
BIF	30/06/2003	15	4.97	2900	1	0.15	0.28	0.43	0.78	0.15	0.56	1.49	3.9	9.53	0.76	0.45	0.92	0.79	1	5	0	T	T	F	0
BIF	30/06/2003	16	5.15	38000	1	3.62	3.69	11.56	11.44	5.54	0.52	2.73	244.7	365.14	9.94	0.48	0.83	0.71	1	6	0	T	T	T	**1**
BIF	30/09/2003	15	5.17	4800	6	0.20	0.41	0.44	0.90	0.00	1.60	0.18	18.3	-58.42	1.00	3.21	0.98	0.56	1	5	0	T	T	F	0
BIF	30/09/2003	16	5.30	41900	1	2.43	2.41	5.57	5.52	3.21	0.38	0.77	66.5	76.33	8.50	0.49	0.95	0.53	1	6	0	T	T	T	**1**
BIF	31/12/2003	15	5.90	1200	3	0.16	-0.01	0.64	0.04	0.39	0.35	1.64	-9.7	0.20	0.44	1.41	0.66	0.35	1	5	0	T	T	F	0
BIF	**31/12/2003**	**16**	**6.30**	**85300**	**4**	**6.84**	**6.89**	**27.99**	**28.60**	**21.62**	**0.78**	**1.61**	**2666.4**	**-179.03**	**31.55**	**1.90**	**0.65**	**0.35**	**1**	**6**	**0**	**T**	**T**	**T**	**1**
BTF	31/03/2003	15	11.93	8700	7	1.65	1.59	6.21	5.73	2.75	3.00	0.23	68.6	-80.22	4.44	5.00	1.03	1.24	1	5	0	T	T	T	**1**
BTF	30/06/2003	15	12.95	13500	20	-0.58	-0.52	3.04	2.63	3.51	5.08	1.15	14.3	8.25	4.49	6.59	1.08	1.19	1	5	0	T	T	T	**1**
BTF	30/06/2003	16	12.95	400	1	0.05	0.08	0.26	0.38	0.84	0.58	2.01	-1.2	-1.18	0.12	0.28	1.13	1.78	1	6	0	T	T	F	0
BTF	30/09/2003	15	13.48	1200	2	0.21	0.29	1.34	1.71	0.14	0.10	0.35	-28.6	-7.25	0.86	1.15	1.04	0.87	1	5	0	T	T	F	0
BTF	30/09/2003	16	13.53	16000	2	0.38	0.35	2.39	1.99	10.65	0.11	0.50	109.4	-15.35	11.65	1.18	1.05	0.94	1	6	0	T	T	T	**1**
BTF	31/12/2003	15	15.21	1600	3	0.11	-0.01	0.24	0.03	0.36	0.03	0.80	2.2	-1.62	0.46	1.05	1.37	0.93	1	5	0	T	T	F	0
BTF	31/12/2003	16	15.23	3000	1	0.09	0.10	0.20	0.23	0.09	0.37	0.74	1.7	13.47	0.87	0.35	1.32	0.85	1	6	0	T	T	F	0
ESPR	08/01/2003	17	6.16	5200	2	-17.11	-17.15	27.92	27.98	0.54	0.82	1.28	2383.0	317.04	0.48	0.10	1.13	1.21	0	4	0	T	T	T	**1**
ESPR	16/01/2003	10	6.98	3900	26	12.57	12.54	19.78	19.73	0.38	0.85	0.90	4470.1	-602.78	0.57	2.14	0.35	0.08	0	1	0	F	T	T	**1**
ESPR	31/01/2003	15	6.90	49400	123	1.87	1.96	8.08	8.22	7.86	10.09	0.86	408.3	-46.20	12.93	12.81	1.06	1.42	0	4	0	T	T	T	**1**
ESPR	31/01/2003	16	7.30	21900	7	6.54	6.37	15.90	14.59	1.63	0.28	4.89	103.8	306.44	4.01	0.51	1.83	5.50	0	5	0	T	T	T	**1**
ESPR	21/02/2003	17	7.35	5000	2	9.70	9.67	47.55	46.41	0.13	0.59	0.86	-2752.9	-263.37	0.81	0.19	1.09	1.14	0	4	0	T	T	T	**1**
ESPR	07/03/2003	12	8.30	66800	230	6.01	5.85	7.99	8.11	6.24	8.58	1.56	24.2	27.85	7.40	9.89	1.11	1.90	0	2	0	F	T	T	**1**
ESPR	07/03/2003	13	8.65	118600	201	5.22	5.22	4.04	4.16	7.70	3.82	4.86	11.6	12.87	10.61	6.58	1.85	6.19	0	2	0	F	T	T	**1**
ESPR	09/05/2003	15	11.62	160700	474	-3.46	-3.52	10.04	9.67	13.10	26.42	0.98	-195.8	220.97	15.05	19.83	0.76	0.64	0	3	0	T	T	T	**1**
ESPR	09/05/2003	16	12.02	5000	2	3.68	3.59	5.09	4.87	0.35	0.43	2.28	-37.6	-28.10	0.32	0.05	1.58	4.96	0	4	0	T	T	T	**1**
ESPR	06/06/2003	10	15.88	153700	288	6.49	6.34	14.75	14.46	6.11	10.37	1.60	39.1	63.69	7.93	9.74	0.72	0.81	0	1	0	F	T	T	**1**
ESPR	12/06/2003	13	16.23	117500	199	3.29	3.34	9.62	11.11	2.90	4.48	1.58	-219.2	-76.11	5.19	5.49	0.33	0.10	0	2	0	F	T	T	**1**
ESPR	26/06/2003	9	18.09	1712400	2316	1.42	1.39	1.86	1.87	29.44	32.85	0.48	-21.3	-28.38	24.63	24.50	1.09	0.73	0	1	0	F	T	T	**1**
ESPR	29/07/2003	12	17.31	721500	1500	-7.47	-7.84	10.27	11.59	6.54	9.06	3.13	131.8	137.07	10.27	13.53	1.70	3.78	0	2	0	F	T	T	**1**
FF	31/03/2003	15	13.98	1500	4	0.58	0.49	1.45	1.19	0.00	0.85	0.01	54.6	-16.79	1.00	2.18	1.00	0.93	1	5	0	T	T	F	0
FF	30/06/2003	15	15.35	3800	4	0.67	0.90	2.67	3.07	1.06	0.57	2.77	-29.1	-60.62	1.96	1.56	0.77	0.66	1	5	0	T	T	F	0
FF	30/09/2003	15	15.70	22600	35	0.00	0.18	0.01	0.68	2.39	5.23	2.29	-0.1	-1.84	3.63	5.83	1.33	1.79	1	5	0	T	T	F	0
FF	30/09/2003	16	15.73	400	1	0.16	0.13	0.54	0.54	0.88	0.78	1.45	-1.7	-1.28	0.06	0.14	1.21	1.59	1	6	0	T	T	F	0
FF	31/12/2003	15	17.68	8500	20	0.34	0.16	1.86	0.90	0.59	0.82	1.38	8.0	18.76	0.45	1.88	0.64	0.45	1	5	0	T	T	F	0
JHFT	31/03/2003	15	12.00	12000	20	2.66	2.56	9.45	8.45	9.18	5.96	1.33	-51.6	-28.46	9.47	13.95	0.56	0.57	1	5	0	T	T	T	**1**
JHFT	30/06/2003	15	14.65	8300	14	2.33	2.63	4.12	4.73	14.90	8.64	0.61	18.9	18.09	17.17	11.35	0.94	0.59	1	5	0	T	T	T	**1**
JHFT	30/06/2003	16	14.94	0	1	0.00	0.18	0.00	0.28	0.49	0.23	0.24	0.0	1.06	0.00	0.61	1.02	1.53	1	6	0	T	T	F	0
JHFT	30/09/2003	15	14.95	1200	3	0.43	0.77	1.07	1.63	1.49	1.07	1.46	-4.9	-8.57	2.86	2.65	0.86	0.70	1	5	0	T	T	F	0
JHFT	30/09/2003	16	15.24	900	2	3.53	3.52	8.54	7.04	0.83	0.44	1.15	-48.0	-62.48	1.99	1.62	0.89	0.72	1	6	0	T	T	T	**1**
JHFT	31/12/2003	15	18.00	19825	43	7.39	7.17	31.51	31.45	9.54	11.39	2.46	115.8	207.98	13.06	16.75	0.54	0.39	1	5	0	T	T	T	**1**
OME	10/06/2003	15	5.59	109900	138	7.24	7.10	12.94	11.97	17.06	22.84	1.46	121.4	138.85	20.09	20.50	1.19	1.35	0	3	0	T	T	T	**1**
ZAP	03/01/2003	10	30.72	200	2	0.07	0.06	10.58	0.41	0.51	1.40	0.49	28.0	-0.50	0.46	4.00	1.47	1.50	0	1	0	F	T	F	0
ZAP	03/01/2003	11	31.10	10000	3	1.20	1.36	52.53	8.80	21.65	2.09	2.39	118.3	-13.49	24.32	4.50	3.79	7.82	0	1	0	F	T	T	**1**
ZAP	03/03/2003	14	37.71	22400	58	4.74	5.17	23.45	16.20	69.28	53.18	0.95	260.9	254.79	48.00	66.92	0.81	0.81	0	4	0	T	T	T	**1**
ZAP	05/03/2003	10	38.90	75500	133	6.91	6.87	7.50	6.85	16.38	9.85	1.07	72.2	47.24	34.79	27.52	1.71	1.39	0	1	0	F	T	T	**1**
ZAP	10/06/2003	15	39.74	33400	82	9.54	9.35	186.02	48.28	225.04	322.94	1.03	1218.1	525.54	107.74	1230.00	0.61	0.53	0	3	0	T	T	T	**1**
ZAP	14/07/2003	10	58.75	25700	66	4.14	3.84	5.72	5.53	3.88	4.07	1.36	19.5	18.29	7.21	7.62	1.25	1.54	0	1	0	F	T	T	**1**

T_F_BYHAND vs C5.0_Tree		F	T	Total
T	Count	17	38	55
Hit Ratio	Row %	30.91	69.09	100

Appendix 2: Knowledge Discovery Process flow

According to CRISP-DM (Chapman et al., 2000)



Appendix 3: Abnormal returns calculation

Abnormal returns were calculated using the standard Capital Assets Pricing Model (CAPM) (Sharpe, 1964). Specifically, the abnormal returns were calculated as the difference between the observed return and the expected return:

$$AR^{h}{}_{i} = R^{h}{}_{i} - ER^{h}{}_{i}$$

Where:

$AR^{h}{}_{i}$ = Abnormal Return of asset *i* in block hour *h*
$R^{h}{}_{i}$ = Observed Return
$ER^{h}{}_{i}$ = Expected Return

The expected return was defined by the CAPM model:

with:

$$ER^{h}{}_{i} = r_{f}{}^{d-1} + (R^{h}{}_{m} - r_{f}) \cdot \beta_{i}^{m-1}$$

$ER^{h}{}_{i}$ = Expected Return of asset *i* in block hour *h*
$r_{f}{}^{d-1}$ = Risk free rate (10-year US treasury note return) of the *previous day*
$R^{h}{}_{m}$ = Market Return (SP&500 return) in block hour *h*
$\beta_{i}{}^{m-1}$ = Beta of asset *i* of the *previous month*

and the observed return as:

$$R^{h}{}_{i} = \frac{P^{h}{}_{i} - P^{h-1}{}_{i}}{P^{h-1}{}_{i}}$$

with:

$R^{h}{}_{i}$ = Observed Return
$P^{h}{}_{i}$ = Representative price of asset *i* in block hour *h*

As can be seen from the formulas, the use of the CAPM model which contains proxies and lagged information for risk-free returns and Betas could distort the precision of the intended abnormal return measure. In this sense, it is assumed that even if the CAPM is not the most accurate way to calculate abnormal returns, at least efforts are made to capture the effects of important market movements that could affect the stock prices, especially when the market is reacting to significant new information, such as Federal Reserve announcements, terrorist attacks, armed conflicts and other similar developments. The reader must bear in mind that the objective of the paper is to describe patterns in stock intraday trading, for which precision of one indicator can be secondary if the pattern can be better described only with states of the variables such as 'high', 'low', or 'average'. This is reason behind the use of the decision trees as a tool for analysis.

Appendix 4: Example of a news item

Released on 21st September 2000

"<?xml version=""1.0"" encoding=""UTF-8""?>"
 <ppsarticle>
  <article>
  <accessionNo>prn0000020010812dw9l03qe1</accessionNo>
  <reference>distdoc:archive/ArchiveDoc::Article/prn0000020010812dw9l03qe1</reference>
  <baseLanguage>EN</baseLanguage>
  <copyright>

</copyright>
  <headline>
"   <paragraph display=""Proportional"" truncation=""None"" lang=""EN"">"

Zomax Incorporated Comments on Expected 2000 Financial Results

</paragraph>
  </headline>
  <leadParagraph>
"   <paragraph display=""Proportional"" truncation=""None"">"

"MINNEAPOLIS, Sept. 21 /PRNewswire/ -- Zomax Incorporated (Nasdaq: ZOMX)"
today announced that its third quarter 2000 financial results will be lower
than current consensus estimates. The Company expects third quarter 2000
revenues of $57 to $59 million and diluted earnings of $.15 to $.17 per share.
"Revenues through the nine months ending September 29, 2000 are expected to be"
"$177 to $179 million, with earnings of $.57 to $.59 per share. Gross margins"
will continue strong at approximately 30% with operating income margins of
approximately 15%. Full year revenues are expected to be $237 to $245
"million, with earnings of $.75 to $.80 per share."

</paragraph>
"   <paragraph display=""Proportional"" truncation=""None"">"

The shortfall is primarily the result of general market softness and in
"particular the European market, a major customer modifying a third quarter"
"program that the Company was unable to replace with other business, an"
increase in polycarbonate prices due to increasing crude oil prices and
further weakening of European currencies.

</paragraph>
  </leadParagraph>
  <publicationDate>
   <dateTime>2000-09-21T20:46:00Z</dateTime>
  </publicationDate>
  <sourceName>PR Newswire</sourceName>
" <company code=""ZOMOPT"">"
   <name>

ZOMAX INC (USA)

</name>

Appendix 5: Difference of Means

T-Test for difference of means TotalGap and GapLife.

*Cells contain: Mean

Field One	Field Two	Mean One*	Mean Two*	Correlation	Mean Difference*	Importance
gap_life_Sim1	gap_life_Sim2	7609.689	4299.504	0.912 Strong	3310.185	1.000 Important
gap_life_Sim1	gap_life_Sim3	7609.689	6277.906	0.986 Strong	1331.783	1.000 Important
gap_life_Sim1	gap_life_Sim4	7609.689	10576.152	0.944 Strong	-2966.463	1.000 Important
gap_life_Sim2	gap_life_Sim3	4299.504	6277.906	0.934 Strong	-1978.403	1.000 Important
gap_life_Sim2	gap_life_Sim4	4299.504	10576.152	0.803 Strong	-6276.649	1.000 Important
gap_life_Sim3	gap_life_Sim4	6277.906	10576.152	0.918 Strong	-4298.246	1.000 Important

*Cells contain: Mean

Field One	Field Two	Mean One*	Mean Two*	Correlation	Mean Difference*	Importance
TotalGap_Sim1	TotalGap_Sim2	0.009	0.011	-0.031 Strong	-0.002	1.000 Important
TotalGap_Sim1	TotalGap_Sim3	0.009	0.012	0.198 Strong	-0.003	1.000 Important
TotalGap_Sim1	TotalGap_Sim4	0.009	0.019	-0.055 Strong	-0.011	1.000 Important
TotalGap_Sim2	TotalGap_Sim3	0.011	0.012	0.010 Strong	-0.001	1.000 Important
TotalGap_Sim2	TotalGap_Sim4	0.011	0.019	0.149 Strong	-0.008	1.000 Important
TotalGap_Sim3	TotalGap_Sim4	0.012	0.019	0.091 Strong	-0.007	1.000 Important