

Data-Driven Models & Mathematical Finance: Apposition or Opposition?



Babak MAHDAVI-DAMGHANI

Oriel College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Machine Learning

Trinity 2019

Acknowledgements

I would like to declare that this report is my own work unless otherwise specified. I would like to thank my supervisor, Pr. Stephen Roberts (University of Oxford) for offering me a full scholarship and a generous stipend. I would like to also thank him for his help and guidance during my doctoral years and during the writing of this technical report and the papers [1, 2, 3, 4, 5, 6, 7, 8]. I would like to thank Dr. Alexandre Kabla (University of Cambridge) for introducing me to a couple of critical ideas (the asymmetric cost of pattern recognition and cellular automaton models) which contributed in inspiring me years later, to develop some of the ideas behind UTOPE-ia [9] and the HFTE model [4] model. I would like to thank Pr. Chris Rogers (University of Cambridge) for guiding me through a project which years later inspired me to develop the Cointelation model [10, 11]. I would like to thank Pr. Cont (University of Oxford), Pr. El Karoui (Sorbonne University) and Pr. Denis Couvet (École Polytechnique), for respectively introducing me to the École Polytechnique admission process, offering me a full scholarship, introducing me to formal mathematical modelling and therefore inspiring me to interpret years later some of the findings behind the IVP model [12, 13]. I would like to thank Maryann McDonagh for giving a chance at attending an Ivy League school. I would like to also thank Pr. Iraj Zandi (University of Pennsylvania) for first introducing me to the scientific method, Dr. Patrick McSharry (University of Oxford) and Vasile Palade (University of Oxford) for guiding me during my MSc dissertation [14] which led to my DPhil admission. I would like to also thank Paul Wilmott for publishing my first few papers [10, 11, 9, 12, 13, 4] and making the very first of them the cover story of his magazine [9] which contributed towards starting my academic career. I would like to thank the thesis jury: Pr. Jean-Philippe Bouchaud (École Polytechnique) and Pr. Álvaro Cartea (University of Oxford) for their valuable input which served to improve the thesis greatly. Last but not the least, I would like to thank my parents for their support and generosity during all these years.

Abstract

The aftermath of the financial crisis of 2009 as well as the multiple Flash Crashes of the early 2010s resulted in social uproars in the general population and ethical malaises in the scientific community [15, 9, 11, 10] which triggered noticeable changes in Quantitative Finance (QF).

More specifically, QF was instructed to change [16, 17, 18] and become more realistic as opposed to more convenient. The concurrent rise of Big Data (BD) [19] and Data Science (DS) [20] contributed to facilitating these changes. More specifically, in terms of defining new models, we saw a significant increase in the use of Machine Learning (ML) overtaking traditional Mathematical Finance (MF) models. In this thesis we consider the impact of such data-driven modelling transition in finance. In order to illustrate these changes the thesis is divided into two parts, each consisting of three and four chapters.

The first part of the thesis consists of examples in which BD has been exposing the limitations of traditional Financial Mathematics assumptions. Specifically, we develop in that context the Cointelation [11, 10, 5], the IVP [12, 13, 6], the modified Heston [8] and the Responsible VaR [7] models, all data driven modifications of distinguished Financial Mathematics models. We also illustrate how the sum of traditional Financial Mathematics and ML methods can be larger than their individual parts. For instance, we expose how Deep Learning by constraints and Stochastic Calculus can, with the help of feature engineering, allow us to formalize useful dynamical strategies [5].

In the second part, we take a bottom-up approach to algorithmic trading and introduce the High Frequency Financial Trading Ecosystem (HFTE) [4] and illustrate some intriguing connections to the world of evolutionary dynamics. We introduce the concept of path of interaction [4, 3] as a way to test concepts such as strategy invasion. We then explore the challenges associated with properly regulating the algorithmic trading markets, in the era of flash crashes, by formalizing a particle filter methodology [3].

Part I

Introduction

Contents

Acknowledgements	3
Abstract	4
I Introduction	5
Contents	6
Lists	14
Figures	14
Tables	18
Algorithms	19
1 Opening	21
1.1 Context	21
1.1.1 The Rise of Big Data	21
1.1.1.1 Definition	21
1.1.1.2 How Big is Big Data?	22
1.1.1.3 Scope	23
1.1.2 A Market Changing Financial Crisis	23
1.1.2.1 The Subprime Crisis as a Triggering Effect	23
1.1.2.2 A Call for a Modelling Revolution	23
1.1.2.3 Exactitude vs Complexity	25
1.1.2.4 From Financial Mathematics to Machine Learning	26
1.2 Problem Formulation	27
1.2.1 Bottom-Up & Top-Down Approach for Trading	27
1.2.1.1 Agent-Based Intelligent System & Deep Learning	27
1.2.1.2 Adversary Model as a Key for Enhancement	30

1.2.1.3	Stability of the Market and Multi-Target Tracking	31
1.2.1.4	Socially Responsible & Consumer Finance	31
1.2.2	Big Data Triggering Change	32
1.2.2.1	Exposing the Limitations of the Wings	32
1.2.2.2	Dimensionality Handling and Proxying	32
1.2.2.3	The Problem of Normalizing Rolling Contracts	33
1.2.2.4	Relationship between Stochastic & Local Volatility	35
1.2.2.5	Portfolio Optimization for Cointelation	35
1.2.2.6	Clustering for Distribution Forecasting	36
1.3	Agenda	36
1.3.1	Models Assuming Data to Data Reassuming Models	36
1.3.2	A Bottom-Up Approach to Algorithmic Trading	37

II From Models Assuming Data to Data Reassuming the Models 38

2	Cointelation, Inferred Correlation & Portfolio Optimization	39
2.1	Measured Correlation is Misleading	39
2.1.1	Traditional Financial Mathematics Assumption	39
2.1.2	Data Suggests Dependence and Heteroscedasticity	41
2.1.2.1	Direct Observation	41
2.1.2.2	Indirect Observation	43
2.1.3	Consequences	44
2.2	Cointelation & Inferred Correlation	46
2.2.1	Rational	46
2.2.2	Mathematical Specification	47
2.2.3	Interesting Properties	47
2.2.3.1	Inferred Correlation	47
2.2.3.2	Number of Crosses	50
2.2.4	Parameter estimation	51
2.2.5	Proposed Test	52
2.3	Socially Responsible & Consumer Finance	53
2.3.1	Approved Persons	53
2.3.2	Application to the example of Oil and BP	53
2.3.2.1	Physical Reason	53
2.3.2.2	Correlation & timescales	53

2.3.2.3	Prediction & Biological Explanation	54
2.3.3	Marketing Material & Proper Market Conduct	56
2.4	Portfolio Optimization	58
2.4.1	Mathematical Definition	58
2.4.2	Markowitz Portfolio Theory Review	58
2.4.2.1	Foundations	58
2.4.2.2	Optimization methodology for a pair of assets	58
2.4.3	Ornstein-Uhlenbeck Theory Review	59
2.4.3.1	Foundations	59
2.4.3.2	Optimization Methodology for a pair of assets	60
2.4.4	Stochastic Portfolio Theory Review	61
2.4.4.1	Introduction	61
2.4.4.2	Problems	61
2.4.4.3	Framework	62
2.4.5	Mean-Variance Criterion	64
2.4.5.1	Expected Return and Variance of Portfolio	64
2.4.5.2	Optimal Investment Using the MVC	65
2.4.6	The Stochastic Control Solution	68
2.4.6.1	Power Utility Maximization Problem	68
2.4.6.2	Definition and Assumptions	68
2.4.6.3	Optimal Investment Strategy	70
2.4.7	Deep Learning for solving our PDE	71
2.4.7.1	General Idea	71
2.4.7.2	Neural Network Architecture	73
2.4.8	Dynamic Switching	74
2.4.8.1	Signal Decomposition	75
2.4.8.2	Finding the right phase	75
2.4.9	Results	76
3	Anomaly Detection & Volatility Surface de-Arbitraging	81
3.1	Vanilla Options Model	81
3.1.1	A Notorious History	81
3.1.2	Log-Normal Assumption	82
3.1.3	Normality Assumption	83
3.1.4	Garman-Kohlhagen Model	84
3.1.5	Implied Volatility from Options Prices	85

3.2	Normalizing Volatility Contracts	87
3.2.1	Context	87
3.2.2	Normalizing Listed Options Rolling Contracts	87
3.2.3	Implied Volatility Surface Edges Problem	88
3.3	Arbitrage Condition on the Strike	89
3.3.1	From First Principles	89
3.3.2	Necessary but not Sufficient Condition	90
3.4	Arbitrage Condition on the Tenor	93
3.5	De-arbitraging	96
3.5.1	The General Idea	96
3.5.2	The Case of the FX Asset Class	97
4	Big Data Changing the Vanilla Options Landscape	99
4.1	Volatility Parametrization	99
4.1.1	Motivation	99
4.1.2	Schonbucher's Model	100
4.1.3	The SABR Model	101
4.1.4	The Stochastic Volatility Inspired Model	102
4.1.4.1	The Raw SVI	102
4.1.4.2	The Natural SVI	103
4.2	Implied Volatility Parametrization	107
4.2.1	The Generalized SVI Model	107
4.2.1.1	Big Data Fueling an Industry Change	107
4.2.1.2	Academic Interpretation	108
4.2.1.3	Downside Transform	110
4.2.2	Implied Volatility Parametrization	111
4.2.2.1	Overall Formula	111
4.2.2.2	Modelling the Bid-Ask Wings Curvature	113
4.2.2.3	Modelling the Bid-Ask ATM Spread	114
4.2.2.4	Position Size	115
4.3	De-arbitraging with the IVP parametrization	116
4.4	Updating Volatility Data	117
4.4.1	Reclaiming Data Science's Intuitive Meaning	117
4.4.2	Updating Volatility from Listed Markets	117
4.4.3	Updating Volatility outside Listed Markets	118
4.4.4	The Midprice Smoothing Approach	118

4.4.4.1	The Midprice Constrained Smoothing Approach . . .	118
4.4.4.2	The Liquidity Adjustment Smoothing Approach . . .	119
4.4.4.3	The Forecasting Smoothing Approach	120
4.5	Proxying Volatility with Sparse Data	121
4.5.1	Implied Volatility and Risk Factor Decomposition	122
4.5.2	Proxying and Economic Factors	122
4.5.2.1	Asset Class	122
4.5.2.2	Economic sector	123
4.5.2.3	Product Diffusion Types	123
4.5.2.4	Geographical Location	123
4.5.2.5	Liquidity Profile & Size of the Company	123
4.5.2.6	Linear Interpolation & Parameters Surrounding Tenors	124
4.5.3	Sequential Bootstrapping	124
4.5.3.1	Definition	124
4.5.3.2	Proxying Missing IVP Factors & Priorities	124
4.5.3.3	Post Proxy Selection & Smoothing	125
4.5.3.4	Optimization by Constraint for Sparse Data	125
4.5.4	Tracking Volatility & Resampling Risk Factors	126
4.6	Statistical Arbitrage for Vanilla Options	130
4.6.1	Point-wise Approach	130
4.6.2	Simplified IVP & Closed Form Calibration	130
4.6.2.1	Formal Bayesian Set-up	132
5	Clustering for Distribution & Regime Change Forecasting	133
5.1	Introduction	133
5.1.1	Objectives	133
5.1.1.1	Leading vs Lagging Risk System	133
5.1.1.2	SDE vs Gaussian Processes	134
5.1.2	Mathematical formulation of Anticipative VaR	134
5.1.3	Agenda	134
5.2	The SDE Approach	135
5.2.1	Generalized Diffusion for Risk Factors Model	135
5.2.2	Calibration	136
5.2.2.1	Choice of Model Assumption	136
5.2.2.2	Calibration of the Primary Parameters	137
5.2.2.3	Secondary Parameters in the Anticipative VaR	138

5.2.3	Benefits	138
5.2.4	Drawbacks	139
5.3	The Machine Learning Approach	139
5.3.1	Mathematical Specifications	139
5.3.2	Calibration	143
5.3.2.1	The Bayesian set-up	143
5.3.2.2	A straightforward learning algorithm	143
5.3.3	Benefits	143
5.4	Application to Pairs Trading	144
5.5	Application To Risk Management	147
5.5.1	Margining: Classic & Anticipative Methodologies	147
5.5.2	Drift in Risk & Misconceptions	147
5.5.3	Margining in the context of Anticipative VaR	148
5.5.4	Results & Rolling Conditional Distribution	149
5.6	Reconciling Responsive & Stable VaR	150
5.6.1	Generating Risk Scenarios	150
5.6.2	Responsive VaR	150
5.6.2.1	The Exponential Weighting Approach	151
5.6.2.2	The Local to Historical Volatility Approach	151
5.6.3	Responsive vs. Stable VaR	151
5.6.4	Responsible VaR	153

III From a Top-Down to a Bottom-Up Approach in Financial Modelling 156

6	Agent-Based Intelligent System, from Shallow to Deep Learning	157
6.1	The Financial Game of Life	158
6.2	Generative Adversarial Networks	159
6.3	Electronic Trading	160
6.3.1	Description	160
6.3.2	Variable Definition	160
6.4	Neural Net Architecture & Learning Potential	160
6.4.1	A Brief Qualitative History	160
6.4.2	Kolmogorov-Arnold's Superposition Theorem	163
6.5	Intelligent Agents & Financial Strategies	164
6.5.1	The High Frequency Financial Funnel	164

6.5.2	The EWMA Architectures	166
6.5.2.1	Trend Following	166
6.5.2.2	Mean Reversion	167
6.5.3	Multi Linear Regression NN Format	169
6.5.4	Regularized NN Format & Lasso Regression	171
6.5.5	XOR NN Format	172
6.6	Network HFFF & Deep Learning Thoughts	175
6.6.1	Deep Learning Brief Recent History	175
6.6.2	Methodology	175
6.6.3	Our Hypothesis	176
7	Evolutionary Dynamics & Strategy Ecosystem	177
7.1	Review of Markov Chain Monte Carlo Models	177
7.1.1	Metropolis-Hastings Algorithm	178
7.1.2	Hamiltonian Monte Carlo	178
7.1.3	Gibbs Sampling	178
7.1.4	Ordered Over-relaxation	179
7.1.5	Slice Sampling	179
7.1.6	Multiple-try Metropolis	180
7.1.7	Reversible-Jump	180
7.2	Game Theory Review	181
7.2.1	Prisoner's Dilemma	181
7.2.2	Axelrod's Computer Tournament	182
7.2.3	Evolutionary Dynamics	183
7.2.4	Minority Game	184
	7.2.4.1 A Short Introduction	184
	7.2.4.2 Criticism	186
7.3	Theoretical Biology Review	187
7.4	Formalizing the Evolutionary Process	189
7.4.1	Survival & Birth Processes	191
7.4.2	Inheritance with Mutations	191
7.5	Evolutionary Dynamics Simulation	193
7.5.1	Observations	193
7.5.2	Philosophical Interpretation	195
7.5.3	Regulation	196
	7.5.3.1 Optimal Control Theory	196

7.5.3.2	Partial Differential equation Specification	196
7.5.3.3	Algorithm Trading Systemic Risk	197
7.6	Path of Interaction	198
7.6.1	HFTE Game	198
7.6.1.1	Definitions	198
7.6.2	Strategy Tournament	202
7.6.2.1	Foreword	202
7.6.2.2	Results	204
7.6.2.3	Few Interesting Hypotheses	207
8	Stability of Financial Systems and Multi-Target Tracking	210
8.1	Classic Methods in Multi-Target Tracking	210
8.1.1	Linear Methods	210
8.1.1.1	Kalman Filter	210
8.1.1.2	Extended Kalman Filter	212
8.1.2	Non-Linear Methods	213
8.1.2.1	Importance Sampling	213
8.1.3	Resampling Methods	215
8.1.3.1	Sequential Monte Carlo Methods	215
8.1.4	Scenario Tracking Algorithm	217
8.1.4.1	Introduction	217
8.1.4.2	Time-Varying Number of Targets Problem	219
8.1.4.3	Measurement-to-Target Association Problem	219
8.2	HFTE SMC Tracking Methodology	219
8.2.1	Direct Approach	220
8.2.2	Simplified Simulation	220
IV	Conclusion, Appendices & Bibliography	223
9	Conclusion	224
9.1	Covariance and Data Reassuring Models	224
9.1.1	Summary	224
9.1.2	Future Research	225
9.1.2.1	Cointelated Portfolio Performance measure	225
9.1.2.2	The n -Cointelated case	225
9.1.2.3	Application to Cryptocurrencies	226

9.1.2.4	Band-wise Gaussian mixture	226
9.1.2.5	Normalizing Options Contracts	227
9.1.2.6	Particle Filter for Implied Volatility MTT	227
9.1.2.7	Additional Liquidity issues for Implied Volatility	227
9.1.2.8	De-arbitraging the FX case	228
9.1.2.9	Harmonizing Stochastic & Local Volatility	228
9.2	Bottom-Up Approach to Trading	229
9.2.1	Summary	229
9.2.2	Current & Future Research	229
9.2.2.1	Classification Simplification	229
9.2.2.2	The State Space can be improved	230
9.2.2.3	Order-book Dynamics	230
9.2.2.4	Increased HFFF complexity does not equate to Invasion	230
9.2.2.5	Birth & Death Processes	231
9.2.2.6	Complex Food Webs	231
9.2.2.7	Diversity & Stability	231
9.2.2.8	Profitable & Unprofitable Ecosystem Asymmetry	232
9.2.2.9	Morality & HFTE Games	232
9.2.2.10	Regularization & Regulation	232
9.3	Closing Words	234
Appendices		235
	DGM NN Architecture	236
	Proof of Lemma 2.4.1	236
	Analytical solution for HJB equation Derivation	240
	Convergence of Heston to natural SVI	244
	Heston to natural SVI	244
	Proof for right and left wing sub-linearity	246
	Acronyms	248
	A-D	248
	E-M	248
	N-Z	250
Bibliography		250
Index		270

List of Figures

1.1	Natural Gas flash crash of 06/08/2011 [21]	22
1.2	Simple Implied Volatility & Standard Pillar Relation on a Rolling Day	34
1.3	Complex Implied Volatility & Standard Pillar Relation on a Rolling Day	34
2.1	Examples of drift departures with different measured correlations [11, 22].	40
2.2	Target & Walmart with measured correlation at different timescales[11].	42
2.3	Implied Correlation and Ratio [10, 11] for SP500 & Eurostoxx.	44
2.4	Simulations of equations (2.1) & (2.4) with $\rho = 0.9$	45
2.5	Correlation distributions with same risk under Equations (2.1) & (2.4)	46
2.6	Cointelation example touching the whole correlation spectrum [10]: on the top figure we have two Cointelated pairs S and S_g simulated with $\rho = -1$ as well as their measured correlation (on the bottom figure) at different timescale	48
2.7	Number of cross formula of equation (2.7) compared to simulated data	51
2.8	Visual representation for the sampling zones for θ [10].	52
2.9	Oil and BP and their measured and Inferred Correlation in the context of the FCA “Client Best Interest Rule” and “misleading statement and actions”	55
2.10	Inferred Correlation vs “Fake Correlation”. Simulation of S , S_l and S_r on the top figure and their measured correlation in the bottom figure (in red we have the inferred correlation, in blue the measured correlation of cointelated pairs and in green the max of the measured correlation of non mean reverting pairs). The chosen parameters are $\theta = 0.002$, $\rho = -0.5$, $\sigma = \sigma_l = 0.002$ and $\lambda = 1/2$	57
2.11	Merton [23] non-linear PDE using DGM & independent peer review [24].	74

2.12	(a) one simulated scenario based on cointelation model (2.5) with parameters: $\mu = 0.05, \sigma = 0.17, \sigma_l = 0.16, \theta = 0.1, \rho = -0.6$ and scaled spread: $\theta(S_t - S_{l,t})$; (b) portfolio return and optimal weight using Dynamic Switching approach; (c) portfolio return and optimal weight using Machine Learning approach.	77
2.13	Histogram of excess ($P\&L$) for ML_{LS} vs SC at terminal time T	77
2.14	Histogram of excess ($P\&L$) for ML vs FM at terminal time T	78
2.15	(a) Simulated Cointelated Pairs S_t and $S_{l,t}$ (left) and the combined cumulative $P\&L$ (right); (b) ρ -centered strategy signals (left) and strategy specific cumulative $P\&L$ (right); (c) θ -centered strategy signals (left) and strategy specific cumulative $P\&L$ (right).	79
2.16	(a) Simulated Cointelated Pairs S_t and $S_{l,t}$ (left) and the combined cumulative $P\&L$ (right); (b) ρ -centered strategy signals (left) and strategy specific cumulative $P\&L$ (right); (c) θ -centered strategy signals (left) and strategy specific cumulative $P\&L$ (right).	80
3.1	Arbitrage Free IVS with 15 standard tenors & 5 standard strikes [12]	86
3.2	Vogt's [25] total variance example verifying $b(1 + \rho) \leq \frac{4}{T}$ (left figure: with the x axis being the log-moneyness and the y axis being the implied variance) and the corresponding $\partial_{K,K}^2 BS(\sigma^2(K, T))$ approximating the (supposed) always positive pdf (right figure: with the x axis being the log-moneyness and the y axis being the non normalized pdf).	92
3.3	Geometrical representation for three FX pairs and their correlation. .	95
3.4	Visualization for the Core Simple de-Arbing Idea	96
3.5	FX Strike Space Misalignment of Figure 3.4	97
4.1	Impact of a change in the value of parameter a in the rSVI/gSVI/IVP	104
4.2	Impact of a change in the value of parameter b in the rSVI/gSVI/IVP	104
4.3	Impact of a change in ρ parameter in the rSVI/gSVI/IVP	105
4.4	Impact of a change in the value of parameter m in the rSVI/gSVI/IVP model	105
4.5	Impact of a change in the value of parameter σ in the rSVI/gSVI/IVP	106
4.6	Crude Oil Bid Ask Smile for the 1 Year Expiry as given by the CME	109
4.7	IVP & SVI with 1 year expiry S&P 500 index options [26] on 26/11/2016	110
4.8	Impact of a change in the value of parameter β in the gSVI/IVP model	111
4.9	Impact of a change in the value of parameter ψ in the IVP model . .	113

4.10	Impact of a change in the value of parameter α in the IVP model . .	114
4.11	Impact of a change in the value of parameter p in the IVP model . .	115
4.12	Impact of a change in the value of parameter η_ψ in the IVP model . .	115
4.13	Impact of a change in the value of parameter η_α in the IVP model . .	116
4.14	Sets of Fixed Points Modeled by Different Sets of IVP Parameters . .	132
5.1	Gaussian distribution mimicking approximately Figure 2.8. The top figure, represents a zone in which our forecast by the mean reverting SDE is significantly above its historical mean (therefore the forecasted distribution is biased on the downside). The bottom figure represents the zone in which our mean reverting SDE is significantly below its historical (therefore the forecasted distribution is biased on the upside). In red we have the forecasted distribution which is centered around zero because we are at the historical long term mean	141
5.2	Exact same situation as in figure 5.1 but with two additional zones (in green)	142
5.3	Example of forecasted distributions, in green under the Anticipative VaR hypothesis. In zones 1 and 5, the ATM is about the historical mean and therefore the forecasted distribution is biased on the downside. In zones 3 and 4, we have the exact opposite case and finally in zone 2 the distribution is symmetrical because the ATM is equal to the long term mean μ	149
5.4	Youngman's [27] example of One-Day 99% VaR BBB Corporate Bonds.	152
5.5	Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.000$	154
5.6	Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.990$	154
5.7	Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.999$	155
5.8	Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 1.000$	155
6.1	Conway's Game of Life rules illustrated	158
6.2	Three snapshots of a simulation of Conway's Game of Life	158
6.3	Order-book visual representation	161
6.4	Simple Neural Network Modelling a Linear Regression	162
6.5	Feedfoward ANN with 1 Hidden Layer & the XOR Function	162
6.6	The High Frequency Financial Funnel	166
6.7	The EWMA Strategy in HFFF format	168
6.8	MACD Strategies (or a difference of EWMA's) in HFFF format . . .	169
6.9	The MLR strategy in HFFF format	170

6.10	Another MLR strategy in HFFF format	170
6.11	A Lasso regression in HFFF format	172
6.12	The XOR strategy in HFFF format	174
6.13	Another XOR strategy in HFFF format	174
6.14	Illustration for intuitive strategy invasion: “denser NN” lead to in- creased potential for invasion.	176
7.1	Some classic Game theory Representations [28, 29, 30, 31]	182
7.2	Minority Game (right) with $m = 3$ & possible strategy Table (left) [32]	184
7.3	$A(t)$ with $m = 15, 7, 2$ with visible seasonality [32].	185
7.4	Relation between σ and m $N = 101, 201, 301, 501, 701$ [32].	186
7.5	Stable 3-species Lotka-Volterra Simulation	189
7.6	Illustration for the Death and Birth processes in our EP.	192
7.7	Hypothesis of how an HFTE Simulation should perform.	194
7.8	Unstable 3-Species Lotka-Volterra Simulation	198
7.9	Illustration for a hypothetical Strategy Invasion Map [4, 3]	202
7.10	Instability increases with an additional strategy.	207
8.1	Stratified Sampling illustration	216
8.2	Particle Filter on market scenarios on milestones [2, 3, 5, 11, 23, 47] . Each of the 6 rows of the 15 ecosystems corresponds to the probability of ecosystem S_i at milestone m	222
9.1	Under-fitting, just right and over-fitting examples [33].	233
2	Bird’s-eye view of our DGM [23]	236
3	Detailed view of our DGM [23]	236
4	Four scenarios of σ and θ from high to low	242
5	Four scenarios of ρ and μ from high to low	243

List of Tables

1.1	Few Video Presentation Links & References (NICtV)	37
2.1	Johansson maximum eigenvalue test for Walmart and Target [11] . .	41
4.1	Proposed Proxying Missing Factor to Proxying priorities hash table .	125
4.2	Implied Volatility Grid and Formula	130
5.1	Generalized SDE to known SDEs	137
5.2	Example of full revaluation table under classic Risk engines	148
5.3	Example of full revaluation table under anticipative Risk engines . . .	148
6.1	XOR Relationship Between Open Interest, Price & Signal	173
7.1	Evolutionary Dynamics Related Strategies	183
7.2	Path of Interaction for 2 TF Strategies with $\uparrow\uparrow$ Seeds and Full OB . .	201
7.3	P&L in Path of Interaction for 2 Strategies with $\uparrow\uparrow$ Seeds and Full OB	205
7.4	P&L in Path of Interaction for 3 Strategies with $\uparrow\uparrow$ Seeds and Full OB	206

List of Algorithms

1	Deep Galerkin Method (DGM)	73
2	Update Implied Volatility	119
3	HASH FUNCTION $\mathcal{H}(\lambda)$	128
4	Band-Wise Gaussian Mixture	144
5	GMforCointelatedPairs(P, h)	146
6	Backpropagation	167
7	Gibbs Sampling	179
8	Simplified TF Strategy	200
9	Simplified MLR Strategies	203
10	Simplified XOR Strategies	203
11	Kalman Filter	211
12	Extended Kalman Filter	213
13	Resample	215
14	Sequential Monte Carlo	218
15	Particle Filter on Simplified HFTE Strategies State Space	221
16	Scenarios Hash Table H_I^s	221

Chapter 1

Opening

1.1 Context

1.1.1 The Rise of Big Data

1.1.1.1 Definition

The multiple industrial applications arising from the concurrent rise in information retrieval and computer storage capabilities has opened up Big Data (BD) in a spectacular fashion [19, 34, 35, 20, 36]. BD's arrival is unique since the scope is both deep and far reaching. But what really is Big Data? Though used sometimes loosely partly because of a lack of formal definition, the interpretation that seems to best describe Big Data is the one associated with large body of information that we could not comprehend when used only in *smaller amounts*¹ [19]. This characterization seems to indicate that the realm of the definition goes fundamentally beyond simply reducing the confidence interval of a parameter whose estimation would benefit from an increase of the sample size. This latter intuition is the natural statistician point of view. In fact the term “datafication” has recently been introduced in order to replace the misleading term that is Big Data in order to make sure readers research the term instead of guessing its meaning [19]. A good way to illustrate this point would be for instance to examine Figure 1.1. This *new*² data, at the high-frequency domain, allows us to reexamine the market, an old problem, with a new angle. This new angle is what we call the Bottom-Up approach³ rather than assume the Top-Down⁴. Big

¹This is a direct quote [19] which we expand on next.

²To be understood here in the context HFT data not being available (therefore new) in the past (for example during the time Bachelier wrote his influential work [37] which pushed scientist to abandon the bottom-up approach for the benefits of the top down instead).

³E.g., strategies interacting explain the (perceived random) dynamics of the market

⁴E.g., the market, assumed random, explains the designed dynamics of the strategies (This assumption makes rational a business model constructed around the idea of dynamical strategies such

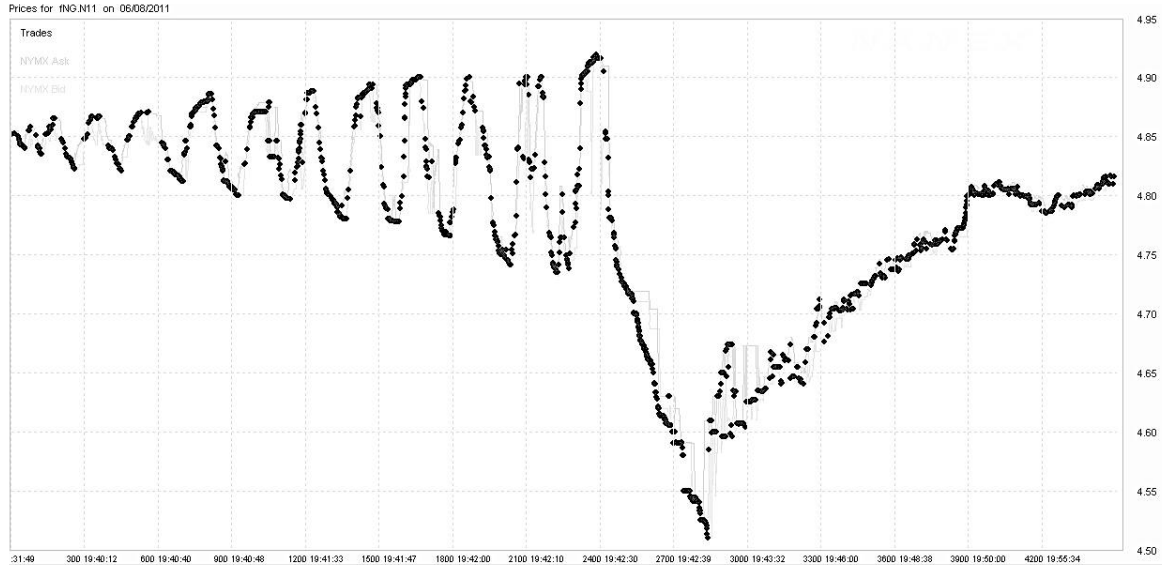


Figure 1.1: Natural Gas flash crash of 06/08/2011 [21]

Data suggests real innovation as opposed to merely improvement of the status quo. Though the part of the definition that infers increase in size is partly indicative of the definition, it only tells half of the story. In that sense “Innovative⁵ Data” would have been a more intuitive term, though perhaps arguably less marketing friendly. Taking its literal sense though, we can legitimately ask how big is Big Data?

1.1.1.2 How Big is Big Data?

There exist many anecdotal claims illustrating the size of Big Data. For instance it's been suggested that if we were to take as reference the time when information was not stored digitally (for instance during the third century BC), where it was believed that the Library of Alexandria housed the sum of all human knowledge, then today, there are arguably 320 times the number of inhabitants worth of data available. More specifically if all this data was placed on CDs and these latter CDs were stacked up, the CDs would form five separate piles that would all reach to the moon [19]. Another interesting fact reported is that as much as 90% of current data was created in the last couple of years [20]. Though these figures are often the most cited by researchers, there are legitimate questions around the quality and the usefulness of the data being stored. For instance Facebook likes, which may have been

as hedging for instance).

⁵The idea that there is increase in the available data (the “Big” in Big Data) is implied in the latter formulation but on top of that the intuition that it also brings change is encompassed with this proposed terminology as well.

bought or censured, constitute a source of data equal in value to perhaps Geophysics data. There are also small disagreements with respect to how fast, the size of Data is growing, but it is estimated to roughly double every three years [19, 35].

1.1.1.3 Scope

The Internet and Cloud Computing fast growth have led to the exponential growth of data in most (if not all) industries. This hot topic attracted the attention of all segments of the population going from government, academia, and the industry with far reaching opportunities but also grand challenges such as computational and system complexity⁶. No matter how one looks at these figures the rise of Big Data is real, its size and scope are changing our lives and our civilization at a very rapid speed. The industry has many applications of BD. The one we have chosen to explore is the financial industry as it is at the heart of our economy, comes under much scrutiny and can create systemic risk [38] with far reaching impacts.

1.1.2 A Market Changing Financial Crisis

We can go as far back as few centuries for the construction of the financial system with perhaps its first serious mathematization attempt occurring about a century ago [37]. However, the event that is most relevant to this thesis happened about 10 years ago with additional posterior signs which served to remind us that the impact of this event was not over⁷ [1].

1.1.2.1 The Subprime Crisis as a Triggering Effect

The financial crisis of 2009 and the resulting social uproar in the general population, induced an ethical malaise in the scientific community [15, 9, 11, 10], which changed the market in many ways. More specifically, after the subprime crisis governments strongly pushed the regulators to develop more efficient risk monitoring systems⁸ and to review the current modelling pillars so as to avoid similar crises in the future.

1.1.2.2 A Call for a Modelling Revolution

The new candidate sector under inspection quickly became the one of algorithmic systematic trading. For instance, the flash crash of May 6, 2010, in which the Dow

⁶with already proposed solutions [34].

⁷for example the multiple flash crashes.

⁸In this context risk is viewed as a mixture of Market and Reputation.

Jones Industrial Average lost almost 10% of its value in matter of minutes, exacerbated the scrutiny. However, the current state of the art risk models, are the ones inspired by the last subprime crisis and are essentially models of networks in which each node can be impacted by the connected nodes through contagion [38] and is better suited to lower frequency models. Indeed, on 06/08/2011 a seemingly relatively unnoticed event occurred on the natural gas commodities market. We say “relatively unnoticed” simply because the monetary impact was limited and finance is unfortunately an industry in which warning signs are usually dismissed until it is too late. We can see from Figure 1.1 that clearly something unusual is occurring. This feeling is exacerbated by the strong intuition that only interacting agents falling into some sort of quagmire could yield such series of patterned oscillations followed by a crash in value. Indeed, commodities has historically been seen as a physical market. This means that the price dynamics is supposed to be driven by supply and demand of commodities which can be transformed, consumed, stored and/or produced. This particular point is a unique feature compared to the other markets (Equities, FX, or Rate). Also Figure 1.1 suggests that the common, though perhaps a bit lazy view, that crashes occur through, totally unpredictable [39] events may not be true for algorithmic trading. In fact, the science of market impact has gathered noticeable steam in the recent past [40]. These few examples amongst others have led the scientific community to encourage revolutionary changes to occur, possibly in the form of agent-based modelling [16, 17, 18] in lieu of traditional financial mathematics models. It is in this fundamental opposition of views that part of the title of this thesis must be understood. Indeed, traditional financial mathematics programs focused on derivatives⁹ were chastised and rethought [41]. This decline popularized Machine Learning (ML) and more specifically Gaussian processes (GP) within them because they provided a flexible non-parametric framework to which, one could incorporate growing data. The latter academic scheme is already making good progress [42] at modelling the options market for example but it seems there are outstanding issues especially when coherence as defined by arbitrage constraints is taken into consideration. On the Risk side, VaR must now take into account the procyclicality of the market. More specifically, data suggests that a big market move is likely to follow another big market move [43, 38], risk models must adapt quickly, incorporate Bayesian statistics, adjust to the sudden increase of volatility of the market and therefore be

⁹In which highest likelihood and mathematical convenience prevailed over data supported by the market.

responsive. Paradoxically, in order to eliminate the risk associated to liquidity shortage and the resulting systemic risk, VaR should, on top of being Responsive, remain (paradoxically) as stable as possible [44]¹⁰.

1.1.2.3 Exactitude vs Complexity

Convolved financial products with high volatility or/and low liquidity or/and without any societal need, other than as speculative tool, such as exotic products were chastised [45]. Many desks were closed as a result. Indeed, the last crisis led to a modelling revolution. Fueled by the lessons learnt from trading these products, based on wrong but mathematically convenient assumptions¹¹, we saw a concurrent:

- tactical step back¹² in the complexity of the products in order to gain a more focused momentum for the future Quantitative models,
- compulsory step forward in liquidity modelling since the subprime crisis was arguably a liquidity activated crisis.

The product class that took the niche of exotics became simpler vanilla products, which hedging property has still utilitarian value¹³, more liquid, less volatile and therefore more in-line with the role of derivatives at their inception. Generally speaking liquidity modelling became of central focus like never before especially through government led initiatives [44] such as for example the Fundamental Review of the Trading Book (FRTB). As the risk models increased in sophistication, questions around coherence of scenarios also became of central importance [12]. The Capital Requirement of each financial institution is now linked to its VaR¹⁴ and the latter must be calculated with historical data. Finally, *Profit and Loss*¹⁵ (P&L) associated to trading should be mapped to appropriate risk factors.

¹⁰We reconcile these discordant instructions later in the document.

¹¹The correlation model assumption was misleading[11] when it came to pricing complex subprime products.

¹²Coming back to the basis of the derivatives markets which is to provide insurance against big market moves as opposed to create these big market moves.

¹³For example a farmer would use a put options in order to hedge himself against the prices of its crop going down few months before maturity.

¹⁴Later corrected to Expected Shortfall, but this change is irrelevant in the context of this paper as going from one to the other when one has the simulated scenarios is relatively easy.

¹⁵This is the most basic way to look at a strategy performance: it is usually calculated as the cumulative sum of the returns.

1.1.2.4 From Financial Mathematics to Machine Learning

Going all the way back to the early stages of the 20th century and Louis Bachelier pillar contribution [37] to Mathematical Finance, the world of Quantitative Finance has been gradually enhanced by the other STEM fields. Probability Theory, Computer Science, Physics, and Statistical Mechanics, among other fields, have steadily but surely brought Quantitative Finance amongst the most challenging STEM fields as its interdisciplinary nature (Mathematics, Computer Science, Physics, Economics & Finance) makes it increasingly difficult to master in full. This has given rise to further granular specialization¹⁶. Added to these challenging historical enhancements, today the other STEM fields¹⁷ are accelerating these changes by contributing themselves to the complexity. More specifically Bayesian Statistics, Signal Processing Statistics [3], Game Theory [4] and above all Machine Learning [19] are increasingly contributing to the interdisciplinary complexity. With the rise of Big Data [19] and Data Science [20], the Bottom-Up approach of agent based modelling, is in total opposition to the Top-Down approach used by traditional Financial Mathematics (in which Brownian Motion is used everywhere there is uncertainty). Added to this algorithmic trading antagonistic change in the markets' point of view, we have been witnessing an interesting re-balancing shift between models and data. More specifically, we are moving from an approach in which models assumed data towards one in which data is reassuming the models [11, 10, 12, 13] and slowly making distinguished Financial Mathematics models obsolete. To some extent the subprime crisis can be seen as the triggering effect which has seen the rise of Machine Learning and the coinciding decline of traditional Financial Mathematics models within the world of Quantitative Finance. However, as we will see more in details the opposition between these two fields can sometimes be turned into an apposition¹⁸.

Remark Mathematics is without a doubt the best tool we have to analyse and derive meaning from a specific hypothesis. In the course of this thesis we do not criticize the wonderful tool that is mathematics but rather the convenient assumptions behind some of the financial mathematics models. Also it is only fair to note that Financial Mathematics is itself an evolving field and many of the novel Data Driven techniques

¹⁶We now have “Pricing Quants”, “Algorithmic Trading Quants”, “Risk Methodology Quants”, “Structurers”, “Model Validation Quants”, “Quant Developpers”, “Quantitative Traders”, Data Scientists etc.

¹⁷As we will see in part III.

¹⁸This wording needs to be understood the following way: the whole of the two fields can be more than their individual sums.

were, are or will be part of Financial Mathematics in different geographical locations and labs. The Financial Mathematics that we are referring to in this thesis is the one which accepted the convenient mathematical assumptions that drove the spectacular growth of Financial Derivatives starting in the 90s and perhaps culminating with the Subprime Crisis.

1.2 Problem Formulation

We formalize here few problems in Quantitative Finance which we hope will illustrate the modelling revolution suggested by some of the highest authorities in the field [16, 17]. More specifically, we take a bottom-up approach to algorithmic trading and show how remarkable the results of this antithetical approach to the status quo can reveal about the real complexity of the markets. We also expose through this mean the shift in the field from models assuming data to one in which data re-assumes the models.

1.2.1 Bottom-Up & Top-Down Approach for Trading

1.2.1.1 Agent-Based Intelligent System & Deep Learning

We learn about the bottom-up vs the top-down approach in introductory systems engineering classes at the undergraduate level. However, by the time one gets into the most advanced postgraduate financial mathematics classes, the models have become dogma. Indeed at these more advance stages, it becomes much more important to be able to derive or infer meaning via these believes rather than understand the limitations of these core modelling assumptions and improve the models from inception. In fact, these beliefs are so much anchored in our common academic psyches that wrong¹⁹ models get Nobel Prizes²⁰ and few lead to market crashes²¹.

Remark The latter award is sometimes abusively called the “Nobel Prize in Economics”, the same way the Fields Medal is sometimes called the “Nobel Prize in Mathematics”. The Nobel Prize in Economics does not, in fact, exist. The same way Alfred Nobel left Mathematics out of his will, he also left Economics out of his will. This latter fact is less known as the wording of the Economics Prize is much closer to the wording of the other Nobel Prizes. The exact wording of the Economics Prize is

¹⁹A good opportunity to remind us that all *models are wrong but some are useful*.

²⁰See: Black-Scholes model and Long-Term Capital Management history [46].

²¹The subprime crisis is arguably due to an implementation around the wrong assumptions (correlation vs domino effect [38]).

the “Nobel *Memorial* Prize in Economic Sciences” which was awarded for the BSM in 1973. The Prize was in fact launched by the Sveriges Riksbank (Sweden’s Central Bank) in *memory* of Alfred Nobel.

In fact the embarrassment of the repeated market crashes has led the highest Quantitative Finance experts²² to call for a modelling revolution [16] in the shape and form of an agent-based intelligent system point of view. Also, note that Quantitative Finance is often criticized as being more a social science because many²³ of its assumptions are wrong²⁴. This peculiarity is not exclusive to Quantitative Finance. Indeed, other STEM fields share some of this embarrassing fact. For example, in Physics questions around the gravitational force are still outstanding and Newton theory only works within the confine of our planet and not beyond [47]. In biology the individual centered view of evolution, though would explain a great deal of our surrounding was ultimately gently put aside when the gene centered view of evolution appeared [48]. These two theories had at inception the highest ever recorded academic impact of their respective fields but were ultimately “incomplete”²⁵. In any case, how is this relevant to the mentioned strategy of information processing? The current modelling approach in Quantitative Finance is the top-down²⁶ approach and the one we are suggesting is the bottom-up approach²⁷. Indeed the current modelling format takes as view that financial underliers follow a random walk like process and that the latter converges towards the Wiener process²⁸. Formally, in the top-down approach we assume that in the “limits” a change to some price process S_t follows a log-normal diffusion process. Recall that the log-normal assumption arises from the Wiener process itself resulting in the assumption of the random walk:

Definition (Wiener Process): W_t has four main properties: $W_0 = 0$ a.s. $\forall t > 0$, $W_{t+s} - W_t$ is independent of W_s where $s < t$, $W_{t+s} - W_t \sim \mathcal{N}(0, s)$ and W_t is continuous in t .

²²Jean-Philippe Bouchaud was awarded the very prestigious “Quant of the Year” award the year this thesis was written.

²³I am being a little politically correct here as I would rather replace “many” by “all”.

²⁴Though Mathematics still remains the mother of all sciences especially when it comes to deriving meaning from these assumptions.

²⁵I personally feel less inclined to label these two contributions as “wrong”, a terminology I find easier to use in the context of the historical Financial Mathematics models. I prefer the terminology “incomplete” instead. Indeed, the degree of in-exactitude seems to be on different scales.

²⁶One in which we surrender in front of the complexity of the market and assume it is random or close to random.

²⁷Which has been assumed to be impossible ever since Bachelier’s doctoral dissertation [37] more than a century ago.

²⁸Small caveat here: not all models in Financial Mathematics assume a pure Brownian motion obviously but the majority incorporate stochasticity which the bottom-up approach does not assume.

Definition (Random Walk to Wiener Process): Let ξ_1, ξ_2, \dots be i.i.d. random variables with mean 0 and variance 1. For each n , we define $W_n(t) = \frac{1}{\sqrt{n}} \sum_{1 \leq k \leq [nt]} \xi_k$ where $t \in [0, 1]$. By the central limit theorem (and more rigorously Donsker theorem) $\lim_{n \rightarrow \infty} W_n(t) - W_n(s) \sim N(0, t - s)$.

However, this top-down approach with, for instance, the assumption of the increments being i.i.d.²⁹ has been criticized both in the low frequency domain [11] as well as in the higher frequency domain (see analysis associated to Figure 1.1) so much so that “overlay”³⁰ models [51] have been incorporated to the BS model to account for the mis-pricing induced by the log-normal assumption³¹.

Remark Note that the random walk assumption is still a central useful pillar of the buy side. This is because the business model is constructed taking the risk neutral approach (more specifically the price of an option is the price of its replication).

These later models where in turn challenged with the arrival of Big Data³² which exposed new limitations [12] of these overlay models but these latter models where in turn also shown to be incomplete when the question of liquidity came into play [13]. You may take two approaches in analyzing the consistent failure of these models:

- either you accept that all these models are *too*³³ incomplete because the core assumptions which we use to derive them, are too far from reality. Therefore, the approach consisting of waiting for the next crash (to incorporate the solution of the newly perceived limitation with yet another overlay model) is unacceptable,
- or you could choose to simply assume that this is the natural course of the scientific method and for a lack of a better alternative we can accept to work with this limiting assumptions,

The repetitive market crashes, the subsequent punitive sanctions taken by the regulators (where instructions to work with new models that directly contradict these root mathematical assumptions³⁴) suggests that the timing is now right to take a step

²⁹To make the analysis more convenient as opposed to more exact.

³⁰E.g., Dupire’s local volatility model [49, 50].

³¹This is the whole rationale behind the implied volatility surface.

³²There is no consensus on whether this claim is true.

³³We stress this word as all models are incomplete by definition but some are useful and some others were useful in the past. The world of Quantitative model may gradually evolve or can change drastically as a result of new information.

³⁴The regulators have recently advised the banks that their model should take into account the procyclicality of the markets [44], directly contradicting the mathematical assumption around the returns being i.i.d..

back and reexamine the bottom-up approach for a more robust scientific method in the future. Though this may seem a little difficult to imagine, taking the bottom-up approach in the scientific method is more likely to allow us to conquer the complexity of Financial modelling the same way Physics and the other STEM fields conquered their topics.

Remark An interesting analogy can be made with respect to how the gene centered view of evolution³⁵ completely re-shuffled our understanding of natural selection and gave the opportunity to explain altruism better. By analogy, we are trying to communicate the idea that the change of perspective from the market centered view (Top-Down) of the financial systems is the wrong way to understand the fluctuation of the market and that the strategy centered view (Bottom-Up) of the financial system provides an opportunity to explain the fluctuations of the market more truthfully despite being more challenging.

However, this bottom-up approach at the intelligent agent level presents a great deal of challenges. The first one to take into account is to recall that small simple increments in information processing are the basis of any viable complex biological system. For instance what created the complexity of the eye in evolution was a slow process which went from the simple photoreceptor to the folded area (cavity) and finally a complex eye³⁶.

1.2.1.2 Adversary Model as a Key for Enhancement

The creation of the eye was for survival purposes. It was developed in both predators and preys. Indeed, the interaction between them favored the enhancement on the complexity of the eye. The key word here to note in the scientific process is the one of *interaction*. This critical element of the scientific process was perhaps best exposed by Conway's work on the Game of Life [52, 53]. The scientific process in the approach is to use simple rules at the microscopic level in order to explain the complexity at the macroscopic level. DeepMind's AlphaGo was produced with a similar idea: Adversary Models³⁷ competing in order to create data, later used for training. A question naturally arises here. How can we apply this methodology to Quantitative Finance and more specifically market microstructure? What would be a simple mathematical structure (in the form of a DNA) that would both allow

³⁵To be understood as in opposition to the individual centered view of evolution

³⁶We took a bit of jump here as there are many more steps including but not limited to the evolution of the cornea.

³⁷Or better known as Generative (GAN) [54] in the Neural Network literature.

simple recognizable strategies to arise from a random swarm of strategies with the environmental pressure being the profit that they make and how these mathematical structures, through interaction with competing strategies, can create fluctuations in the market as well as create a pressure for these strategies to adapt and improve their models? How can we make sure we do not incorporate the idea of foresight in our design? Can we make a parallel with other known biological systems such as ecosystems? These tasks are perhaps overly ambitious. However, can we come up with models and ideas that would inspire research in a new direction? In this thesis, these questions will fall under the part in which Data Driven Models forge a clear opposition to classic Financial Mathematics. However, as we will illustrate with few examples these two fields can also enhance each other and their relationship can therefore be classified, in some circumstances, as a symbiotic enhancement (the contrast of opposition and apposition was however more easily conveyed in the title through assonance).

1.2.1.3 Stability of the Market and Multi-Target Tracking

The electronic trading market is increasingly regarded as the source of the next big market crash [4, 16]. More specifically the fact that this type of trading is characterized by interacting algorithms, at the high-frequency domain, make these potential crashes especially dangerous because they come very quickly (on top of unexpectedly). It becomes of central importance to be able to decipher market movement as a result of these interacting algorithms. However the specifications of these algorithms are always hidden for proprietary reasons which makes the task of the regulators seemingly impossible. However, we can legitimately ask oneself this question. Can we come up with a solution to this regulatory quagmire? More specifically, the construction of scenario based particle filter may offer us a hope in at least laying down the foundations of finding a new potential path towards handling these problems. What are these first steps towards finding a solution?

1.2.1.4 Socially Responsible & Consumer Finance

The social uproar that led the last financial crisis not only exposed the limitations of the current mathematical models but also exposed limitations in the systems that allowed us to determine early enough potential future problems. Can we come up with what these issues might be at the individual level? What about at the population level? Are they induced by the Financial Mathematics culture? Can the Machine

Learning culture help us understand our own limitations? Can we come up with recommendations that may improve the status quo?

1.2.2 Big Data Triggering Change

BD has changed the options market a significant way. More specifically, given that the last subprime crisis was a derivatives led market crash, the regulators have put enhanced scrutiny on these derivatives. Can we have examples of how BD has changed the way we model these derivatives? Remember that BD suggest a change of modelling paradigm and not merely an improvement of estimation. First though, how can we gather BD from the observed prices? How can we proxy³⁸ this data and reconcile it with what is observable? Can we give examples of these methodologies in the different banking functions, going from trading to risk? How can we handle dimensionality in this new enhanced data context?

1.2.2.1 Exposing the Limitations of the Wings

Though the limitations of the Black-Scholes model had already been partially exposed in the past, especially with respect to the Implied Volatility Surface construction [49, 50], the arrival of BD exposed new limitations in the pricing models. More specifically the way the wings should behave is not currently making consensus especially when it comes to data [13, 55, 56, 12, 51, 57]. In what way were wings mis-priced in some circumstances? How does that change the current pricing models? Can we expand on these old obsolete models to incorporate these BD induced changes? Can we incorporate regulator driven changes especially around incorporating liquidity in the models themselves [44]?

1.2.2.2 Dimensionality Handling and Proxying

How can we handle dimensionality when we deal with the implied volatility (IV) surface? More specifically since the IV is a surface and therefore of infinite dimension, how can we decompose the risk factors of the latter in a useful manner? How can we handle the dimensionality of this structure in general? Can we use the decomposition of the IV into risk factor allow us to come up with proxying methodologies which aim would be to reconstruct closely related IV which data is sparse?

³⁸The method consisting of replacing unavailable data with the closest alternative is called interpolation or extrapolation in a situation in which we have data of the same underlier (though at different time, strikes or expiry) available but in situation in which none of this is available we use different types of data (for example of a closely related underlier or other similar methods) usually referred to as a proxy.

1.2.2.3 The Problem of Normalizing Rolling Contracts

Most listed derivatives markets typically offer new contracts once a month with a two year expiry on a fixed date. This means that once two years have elapsed from the first issuance of the listed contracts, we have every months the contracts which were issues 2 years back that expires. The days in which a contract is issued, we have a new set of data points conveying information in the relevant implied volatility surface with an Over Night (ON) to a two year expiry window at different strikes. More specifically, we have a new information about short dated options (which expired the day of issuance) and information about the surface every month in between these two time-lapses. It is usually agreed that there exists 9 important standardized pillars in which linear interpolation in variance space gives reasonable results³⁹. These pillars are ON, 1 Week (1W), 2W, 1M, 2M, 3M, 6M, 1 Year (1Y) and 2Y. Figure 1.2 gives an illustration of these pillars the instant in which a simultaneous issuance of the longest expiry options with the expiry of the shortest expiry contracts occurs. Figure 1.3 illustrates the case in which the last expiry was more than a day away associated with the challenges in estimating these invisible points. Note that in these two figures F represents the forward price and K the strike. The x axis represents therefore the moneyness⁴⁰ of the option. Another problem is the one associated to the question of marked implied volatility surface (IVS) update in cases of non listed volatilities. How can we diffuse this newly arrived data onto the other points of the IVS? These few changes and their mix will have to be adequately addressed in our proposed methodologies at the distribution level (to reflect the various scenarios of IVS changes). Also this specific point ties in with the one of arbitrage with a trading perspective: anomaly detection for IVS has increasingly become of central importance post subprime crisis. There are however, interesting questions. What makes an IVS coherent? How can one detect and turn an IV into a coherent scenario which takes into consideration the essence of what changed its shape (for example in the context of scenario creation). Are the anomalies associated to the IVS the same for every asset class or are there idiosyncratic differences?

³⁹For example on the FX markets.

⁴⁰Note here that different markets have different expressions of moneyness with for instance, equities being expressed in log-moneyness, rates in moneyness and FX in delta space. We will expand on this point in Section 3.1.

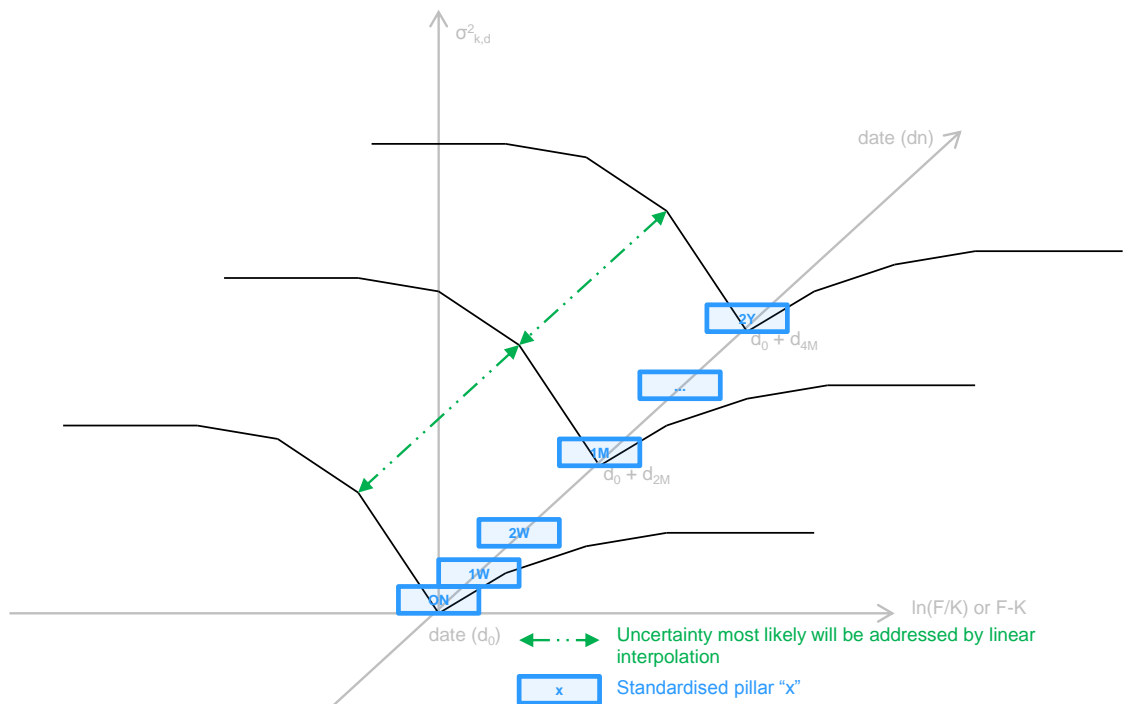


Figure 1.2: Simple Implied Volatility & Standard Pillar Relation on a Rolling Day

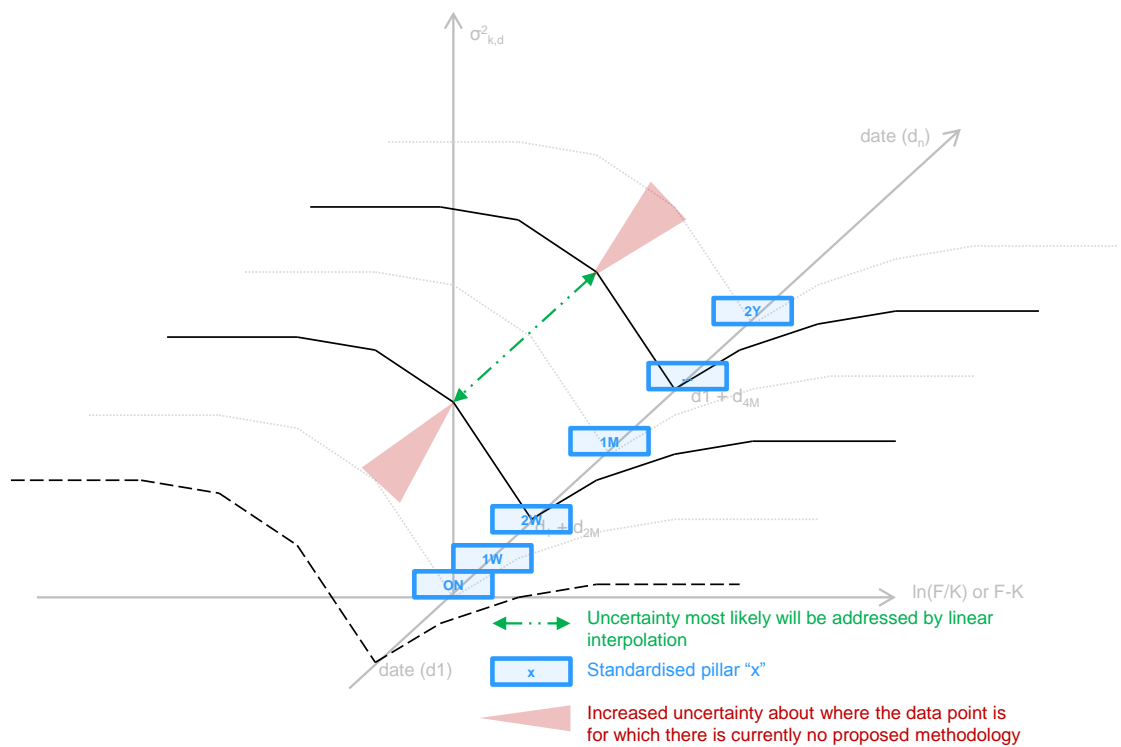


Figure 1.3: Complex Implied Volatility & Standard Pillar Relation on a Rolling Day

1.2.2.4 Relationship between Stochastic & Local Volatility

Gatheral [51] introduced the stochastic volatility inspired (SVI) parametrization at Merrill Lynch in 1999. The parametrization was claimed to have two key properties that have led to its subsequent popularity with practitioners. First, for a fixed time to expiry T , the implied Black-Scholes variance $\sigma_{BS}^2(k, T)$ is linear in the log-strike k as $|k| \rightarrow \infty$ consistent with Roger Lee's moment formula [58]. Second the parametrization had connections to the Heston model [59], arguably the most popular stochastic volatility model. However, at the same time, we may observe [8] instances in which the wings are sub-linear (e.g. Figure 4.7). This latter point led to the SVI being decommissioned at Merrill Lynch (now Bank of America Merrill Lynch) and discussions, seemingly independent⁴¹, of the Heston model have exposed the limitation of the latter in modelling the smile. Both the Heston [60] and the SVI [51] are popular models within the Quantitative Finance community so much so that their link was recently proposed [59] but their industry implementation does not work well with observed data⁴². More specifically, we can ask ourselves the following related questions. If the Heston model is incomplete, how can we enhance it rather than replace it? Similarly if the SVI is limiting, what model can we replace it with in such a way that the skeleton of the model is not violated and the fixes are, something that would resemble an add-on. How can modify these two models such that their connection is still maintained [59]? Mahdavi-Damghani [11, 10] recently introduced the inferred correlation⁴³ concept which is meant to represent the closest correlation related concept which also captures some of the limitations of the assumed classic models. We will see if it can bring about this add-on to the Heston model to help us capture the smile effect better within a stochastic volatility setting. We will see how BD has triggered here interesting enhancements to the world of QF in general.

1.2.2.5 Portfolio Optimization for Cointelation

The cointelation model [11, 10] recently introduced was first applied to regulatory and risk purposes. Its more direct application of pairs trading was not implemented. How can we calibrate the cointelation model in order to extract useful trading signals? More specifically, can we decompose the model in such a way so as to capitalize on

⁴¹We will show how the SVI being decommissioned because of its inability to model the wings is related to the Heston model being unable to accurately model the smile.

⁴²We develop this point later in the thesis.

⁴³Though we will see that we are discussing here the concept of "Assumed Correlation" and not Inferred Correlation.

known models and perhaps formulate the problem in a simpler form? Can we use Deep Learning if we get stuck with traditional Financial Mathematics?

1.2.2.6 Clustering for Distribution Forecasting

The current risk models available to practitioners are at best Responsive and therefore lagging with respect to regime changes. This means that at least one risk breaches is needed for the mathematical model to be able to adjust to the changing market conditions. How can we come up with mathematical specifications for risk systems that would be leading as opposed to lagging? It seems quite intuitive that a VaR model cannot be Stable and Responsive at the same time. Can we reconcile these conflicting risk concepts⁴⁴?

1.3 Agenda

We have organized the thesis in two parts each containing three to four chapters, trying to stress the fundamental “Opposition” in the way the scientific approach is organized between these two closely related fields but also show that, the detailed resolutions of the problem suggests, from time to time, an “Apposition”.

1.3.1 Models Assuming Data to Data Reassuming Models

More specifically, in the first part, labeled as II in the table of contents, consists of exposing how the triggering effect of the multiple crises has brought us into an interesting era in which we are going from models assuming data to one in which the data is re-assuming the models. More specifically we introduce the Cointelation model in order to expose how by taking a more descriptive approach [11] of the Black Scholes Log-Normal diffusion model [37, 46], we may measure a correlation of -1 where in fact the real long term correlation is $+1$. We introduce in that effect the concept of Inferred correlation [10] which can be understood as the more realistic estimation of the real risk associated in holding two Cointelated pairs. We also show how Deep Learning and Clustering can help us in the context of Cointelated pairs [5]. More specifically we see how a pure Financial Mathematics approach or a pure Machine Learning approach have their own limitations. But as we will the elegant combined use of these two closely related fields can yield superior results. In the spirit of exposing additional examples of models re-assuming the data, we introduce the gSVI

⁴⁴which are, interestingly, equally desirable model features for risk managers, traders and financial mathematics practitioners despite their apparent discordant properties

model [12] which is a data driven change in the traditional Financial Mathematics model known as the SVI [51]. We enhance the latter model by introducing the IVP model [13, 6] which is a modification of the latter to include a liquidity overlay. Since the SVI converges towards the Heston model [59], we also show how these two data driven changes create consequences in the assumptions in the Heston model [61] itself by introducing the modified Heston model [8]. We also introduce other secondary concepts such as the one of Responsible VaR [7], a way to reconcile discordant risk measures.

1.3.2 A Bottom-Up Approach to Algorithmic Trading

The second part of the thesis, labelled as III in the table of contents, we take a bottom-up approach to algorithmic trading instead of the top down approach and introduce in this effect the HFTE [4] and illustrate some of the interesting connections to the world of evolutionary dynamics. In that effect we also introduce the concept of path of interaction [4, 3] as a way to test the burden of proof pertaining to concepts such as strategy invasion. We then explore the challenges associated to properly regulating the algorithmic trading markets in the era of flash crashes by formalizing a particle filter methodology [3].

Remark Note that the below table refers to few of the recorded introductory presentations I had about some, but not all of the topics discussed in this thesis. The tables in page 248 have been added to help the reader with potentially unfamiliar acronyms.








Research Links	HFTE	IVP	UTOPE	Cointelation
Video 1				
Video 2			N/A	
Papers	[4, 3, 1, 2]	[13, 12, 6, 8]	[9]	[10, 11, 5, 7]

Table 1.1: Few Video Presentation Links & References (NICTv)

Part II

From Models Assuming Data to Data Reassuming the Models

Chapter 2

Cointelation, Inferred Correlation & Portfolio Optimization

In this chapter, in the context of the aftermath of the last crisis, we re-examine couple of the simplest (conveniently chosen) assumptions in Quantitative Finance and examine their real economical consequences. More specifically, in Section 2.1, we examine the consequences of assuming constant volatility and returns being i.i.d. [46, 39] in the context of modelling relationship between assets. We do this in order to expose the dangers of these assumptions. In this context we introduce the Cointelation model [11, 10] in Section 2.2, which we propose to be a more realistic enhancement to the Black-Scholes log-normal assumption. We expose how critical it is to have a more descriptive approach to the financial markets by stressing applications associated to the challenging regulatory environment. More specifically, we discuss an application in socially responsible and consumer finance in Section 2.3. We also develop concepts related to the Cointelation mode such as the Inferred Correlation and the Number of Crosses formulas. Finally, we revisit classic portfolio optimization techniques and propose a new methodology at the crossroad between classic financial mathematics and a pure machine learning solution involving the modelling of a PDE with a Deep Neural Network in Section 2.4.

2.1 Measured Correlation is Misleading

2.1.1 Traditional Financial Mathematics Assumption

Definition (Correlated Log-normal Pairs): Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ where \mathbb{P} is the historical probability measure under which the discounted price of the underlier, S , is not necessarily a martingale. We define in

this set-up the set of two SDE's given by equation (2.1) which we arbitrarily name the Correlation model

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \quad (2.1a)$$

$$\frac{dS_{r,t}}{S_{r,t}} = \mu dt + \sigma dW_t^r, \quad (2.1b)$$

$$d\langle W_t, W_t^r \rangle = \rho dt, \quad (2.1c)$$

where ρ is the correlation coefficient between the returns of assets S and S_r . In order to abide by the format of the Cointelation model we introduce later in this chapter, the process $(S)_{t \geq 0}$ is arbitrarily called the leading process, $(S)_{r,t \geq 0}$ the lagging process, μ the drift and σ the volatility.

Remark In the context of this chapter we use Pearson's correlation coefficient [62, 63] of equation (2.2) in the context of measured correlation.

$$\rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}. \quad (2.2)$$

It is worth noting that in order to use the estimation from equation (2.2) in a meaningful way, certain assumptions must be met. More specifically the two key elements central to our argument are that the variance must remain constant (homoscedasticity) and the returns i.i.d.¹ If these two assumptions are not met, we can obtain a very misleading measured correlations, as illustrated in Figure 2.1. The latter is split in

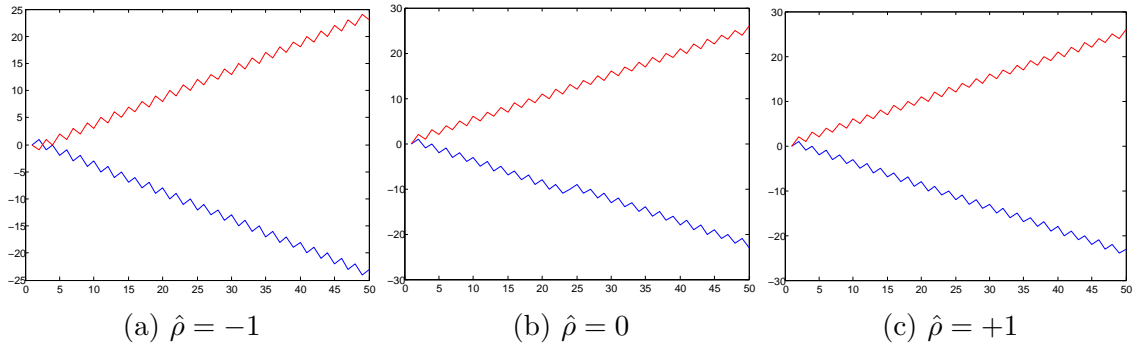


Figure 2.1: Examples of drift departures with different measured correlations [11, 22].

three columns of sub-figures. The y axis corresponds to a hypothetical unit of return (for example normal) and the x axis represents a hypothetical unit of time. What is interesting about these three figures is that despite having the same drift, which can be seen as the long term drift, their immediate infinitesimal correlation hits the

¹The stocks to return process must remain independent.

three symbolic edges of the correlation spectrum. Indeed, the one on the left reveals a measured correlation of -1 (arguably intuitive), the one in the middle reveals a measured correlation of 0 (halfway through the stochastic process, the way the two lines fluctuate changes phase), and finally the one on the right reveals a measured correlation of $+1$ because the volatility is not constant.

2.1.2 Data Suggests Dependence and Heteroscedasticity

Though the problem of heteroscedasticity had been raised decades before the sub-prime crisis, its occurrence has revived interest in the topic. Assets do not seem to follow the mentioned set of system of differential Equations (2.1) and this can be seen both directly or indirectly.

2.1.2.1 Direct Observation

For instance, if we take a direct approach, the examples of Walmart and Target² (Figure 2.2) or Oil and BP (Figure 2.9), we can see that although the measured correlation on a daily basis is significantly lower than 1, it changes at different time horizons and during big market moves (e.g. during a crisis), the underliers crash at the same time. For these observations, the traditional correlation model seem to be a poorer fit than the Cointegration assumption. Table 2.1 gives the statistical results of the Cointegration test performed on Walmart and Target [11]. More specifically we can see that the test statistics rejects the null hypothesis that there is no Cointegration with a p-value of 1% with and without a constant. The data therefore suggests that the correlation model is not the right choice when it comes to model such pairs.

Cointegration test	Test statistic	Critical values		
		10%	5%	1%
with constant	15.98	13.75	15.67	20.2
without constant	20.92	16.85	18.96	23.7

Table 2.1: Johansson maximum eigenvalue test for Walmart and Target [11]

Remark For instance, within the commodities market, it costs a certain amount of money to turn Crude Oil to Brent and therefore if one of these assets becomes

²Two competing companies in the same sector and geographical location.

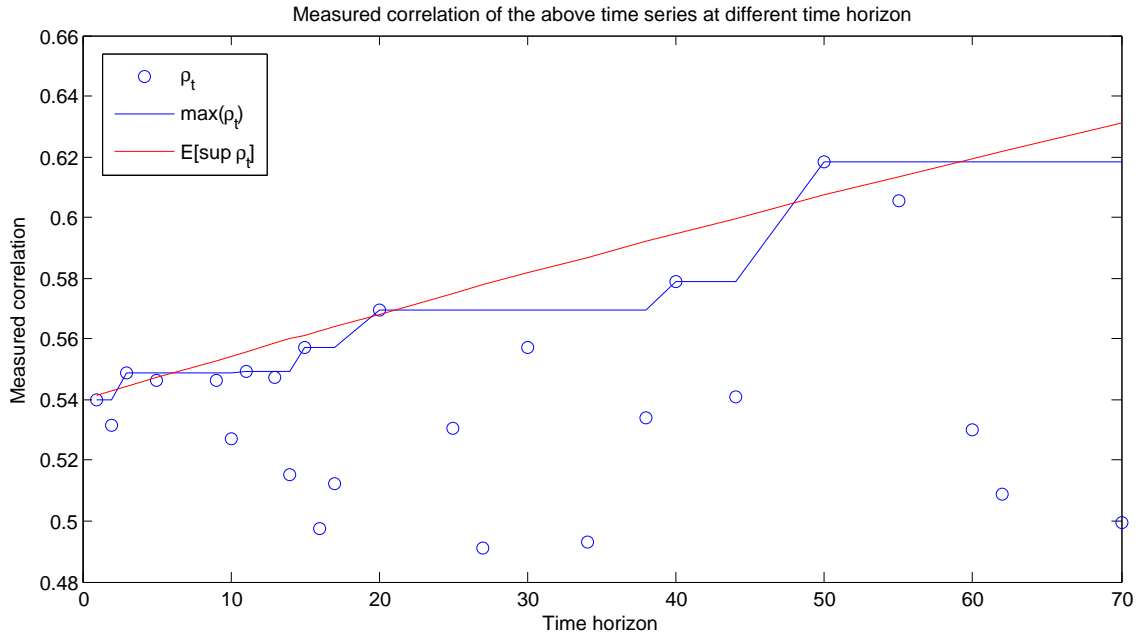
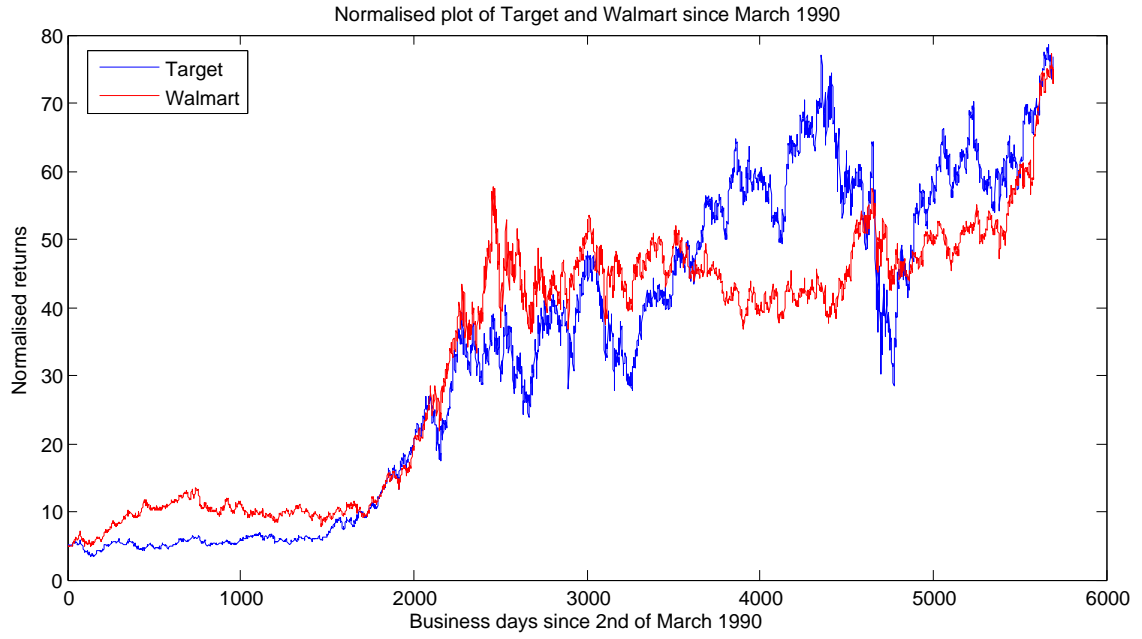


Figure 2.2: Target & Walmart with measured correlation at different timescales[11].

relatively expensive, it becomes economic to buy the other and pay for its transformation. Similarly, within one sector (e.g., Target and Walmart from Figure 2.2³ with

³We will explore more this in detail when we discuss the regulatory consequences in Subsection 2.3.3.

its classic mean reversion test results listed in Table 2.1) competition within rival assets may create similar mean reversion forces. Generally speaking, if one product becomes expensive, it makes other products, sometimes, seemingly orthogonal, more attractive. The short-term risk associated with the departure from this model is insured by both random events and high frequency index traders who hedge themselves on the market while structurers sell the relevant products.

2.1.2.2 Indirect Observation

In the indirect approach, traders understand that measured correlation does not accurately represent the relationship between assets in the long term as it fails to capture the long term drift of the underliers. As such, traders understand that in situations where the perceived equilibrium between assets is away from its historical mean then the expectation of future realized correlation should be smaller than its historical mean correlation. A way to represent this concept mathematically is via the equation of a semi circle. Indeed let Ψ be the set of points (x, y) available for marketed correlation then the semi-circle that will best fit these market correlation can be retrieved using

$$\{\hat{x}_c, \hat{y}_c, \hat{r}\} = \arg \min_{x_c, y_c, r} \sum_{i=1}^N [(x + x_c)^2 + (y + y_c)^2 - r^2], \quad (x, y) \in \Psi. \quad (2.3)$$

Note that in Figure 2.3, the red dotted line represent the best fit as per the sum of squares measure of the semi-circle of equation (2.3). The blue circles represent “marked” correlation. The Figure supports the idea that traders believe that underliers have the same drift in the long run. Indeed this figure suggests that there is a long term relative mean (e.g. $\frac{SP500}{Eurostoxx} \approx \mu$) between the SP500 and the Eurostoxx. This implies that if this relative ratio is away from its economical equilibrium, then this is reflected in the price of exotic options. More specifically, traders on the sell side store what they believe implied correlation ought to be in order to have P&L for basket options⁴. What this figure suggests is that, the further away this ratio is from around 2.1, the lower the implied correlation (because traders anticipate that the two underliers should mean revert). The fitted curve from Figure 2.3 is significant as it does suggest that practitioners have accepted that mean reversion and therefore the assumptions around the returns of assets being i.i.d is not true.

⁴E.g., “worst off”.

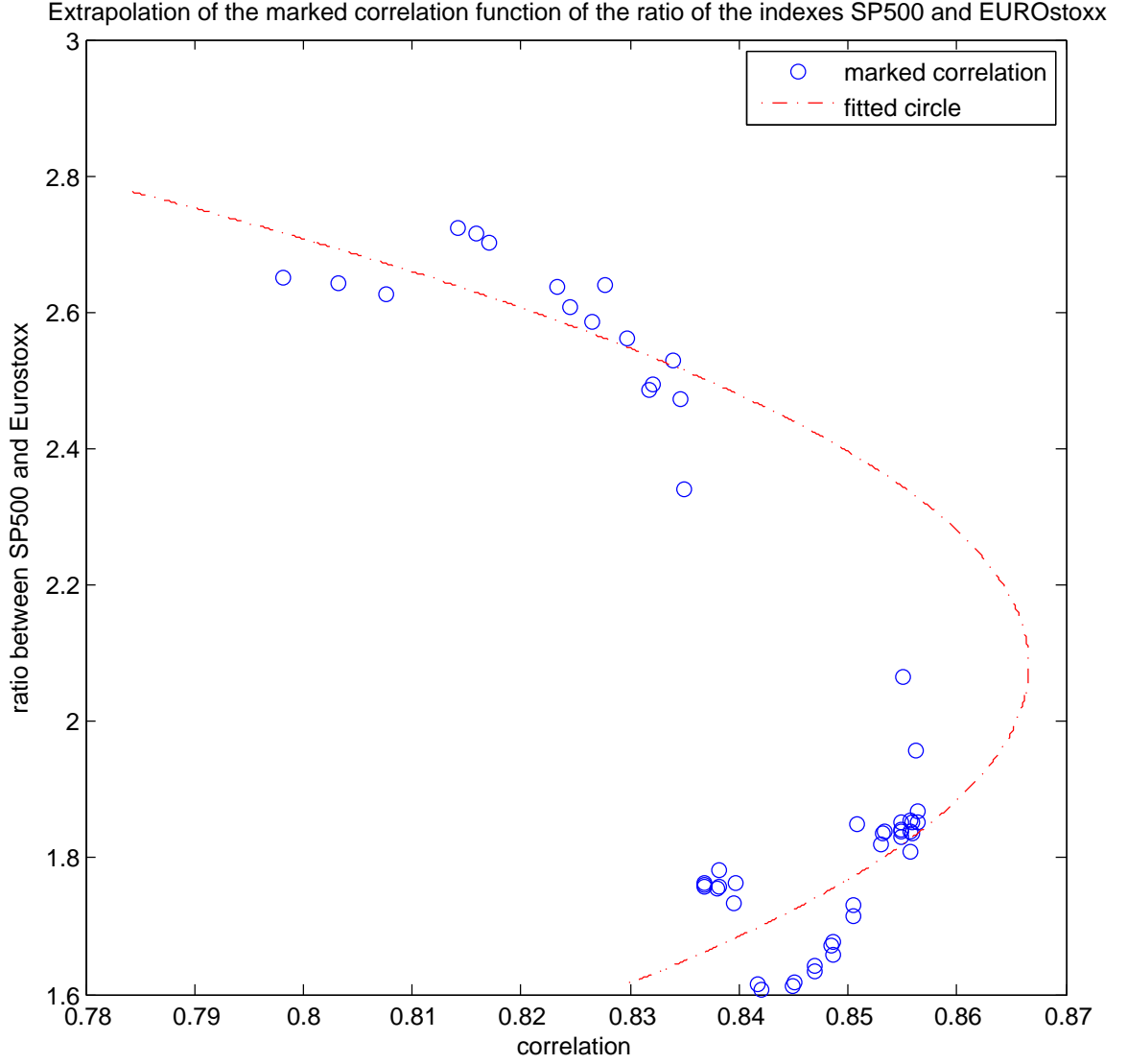


Figure 2.3: Implied Correlation and Ratio [10, 11] for SP500 & Eurostoxx.

2.1.3 Consequences

It becomes quite clear that choosing the model of equation (2.1) instead of one which really captures the mean reverting effect can have extremely misleading effects [11]. We try next to take a simplistic alternative to our Correlation model by replacing it with a known model to reflect these market observations.

Definition (Cointegration Model): Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ where \mathbb{P} is the historical probability measure under which the discounted price of the underlier, S , is not necessarily a martingale. Our bespoke Cointegration

model is defined by the set of two SDE's given by:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t, \quad (2.4a)$$

$$dS_{g,t} = \theta(S_t - S_{g,t})dt + \sigma_l S_{g,t} dW_t^\perp, \quad (2.4b)$$

where the process $(S)_{t \geq 0}$ is called the leading process, $(S)_{g,t \geq 0}$ the lagging process, r the drift, the σ the volatility and θ the speed of mean reversion and W_t^\perp such that $d\langle W_t, W_t^\perp \rangle = 0$.

Remark As a caveat the term “cointegration” is a slight abuse of terminology as the traditional terminology refers to the econometrics literature usually associated to Engle and Granger [64, 65, 66]. However, the term cointegration in this chapter means “mean reverting”. For the interest of using an easy to remember term that was meant to be close enough to correlation we picked cointegration like in our original paper [11, 10].

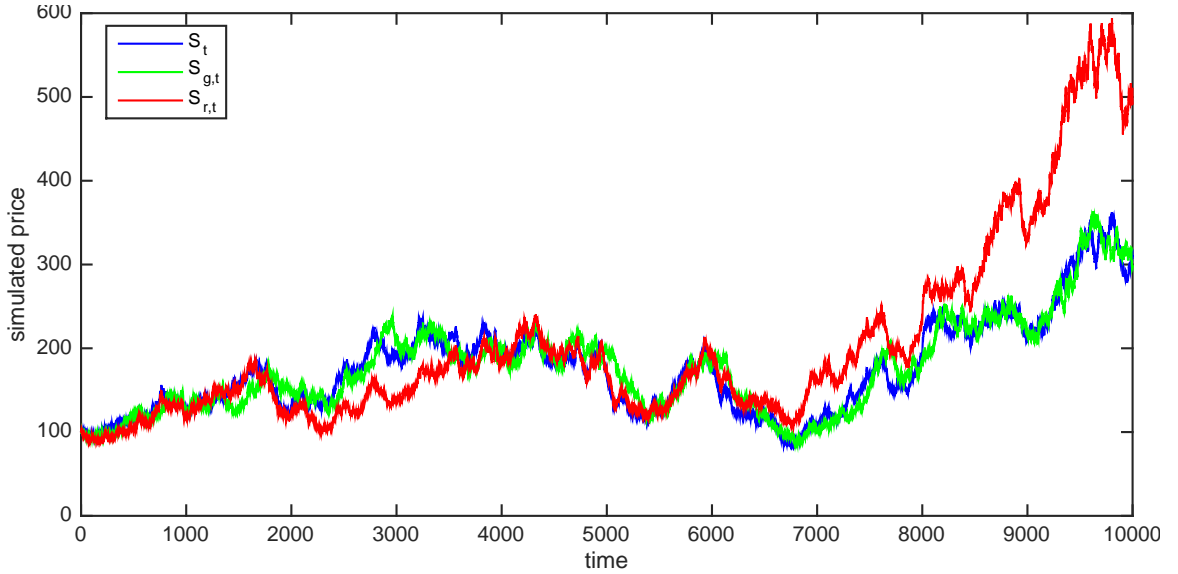


Figure 2.4: Simulations of equations (2.1) & (2.4) with $\rho = 0.9$

Figure 2.4 contain three stochastic processes S_t , $S_{g,t}$ and $S_{r,t}$ introduced earlier in this chapter. The one in blue, S_t , represents the leading process in equations (2.1a) and (2.4a), the one in red, $S_{r,t}$ represents equation (2.1b) and the one in green, $S_{g,t}$, the one of equation (2.4b). The y axis corresponds to a hypothetical unit of return (for example normal) and the x axis represents a hypothetical unit of time. The measured correlation between the returns of the blue and green processes is around 0.2 and the measured correlation between the blue with the red is 0.9. The point we would like

to make here is that these value are very much un-intuitive when you observe the long term drifts as the green and blue line seem to be more “similar”, yet have a low correlation. The implication of model selection⁵ because quite critical when it comes to constructing an optimal portfolio and measuring its risk. For instance, we have performed the following experiment. We have selected the parameters of our models in such a way so as to equate their risk level but then have isolated the measured correlation under each model and plotted them on Figure 2.5. Here again we can see that error in model selection can mean that people interpret very related assets as totally unrelated. In reality models of equations (2.1) and (2.4), though

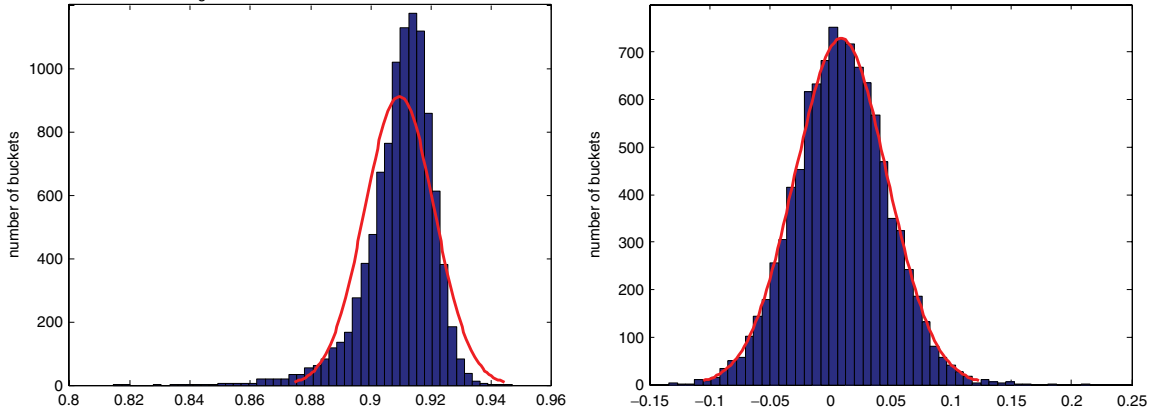


Figure 2.5: Correlation distributions with same risk under Equations (2.1) & (2.4)

individually incomplete, when merged into a more general form, create a powerful tool: the Cointelation model.

2.2 Cointelation & Inferred Correlation

2.2.1 Rational

The relationship of any two asset can be roughly decomposed of a short term and long term risk. The rational for the long term risk is that during the time of rare market crashes all assets tank. However, during the more “normal” periods a shorter term risk/relationship drives the co-movement of assets. These co-movements are accompanied with small but still existent mean reversion forces from one asset to the other for which a substitution from one to the other exists [10, 11].

⁵Equation (2.1) instead of the model of equation (2.4).

2.2.2 Mathematical Specification

Cointelation [10] is a portmanteau neologism in finance, designed to signify a hybrid method between the **cointegration** and the **correlation** models.

Definition (Cointelation Model): Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ where \mathbb{P} is the historical probability measure under which the discounted price of the underlier, S , is not necessarily a martingale. The Cointelation model is defined by the set of two SDE's given by:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \quad (2.5a)$$

$$dS_{l,t} = \theta(S_t - S_{l,t})dt + \sigma_l S_{l,t} dW_t^l, \quad (2.5b)$$

$$d\langle W_t, W_t^l \rangle = \rho dt, \quad (2.5c)$$

where ρ is the correlation between the two Brownian Motions. The process $(S)_{t \geq 0}$ is called the leading process, $(S)_{l,t \geq 0}$ the lagging process, μ the drift of S , σ is the volatility of S and θ the speed of mean reversion of S_l .

Remark It should be noted that Multiscale Autoregressive Models (MAR) [67] may to some extent model some aspects of the Cointelation model. More specifically by properly adjusting the parameters of the model, one can capture both the long term and short term characteristic of the process [67] which is something the MAR models share with the Cointelation model. However the MAR model applications is more natural to wavelets [68] and although the method associated to expressing the correlation of a stationary process on the dyadic tree [69] is certainly intellectually inspiring, it feels like an overkill when applied to financial applications. Also, the context of the AR process and its spectral factorization, an exponentially growing number of coefficients are involved with an exponentially growing number of polynomial constraints. This, prompted the authors to abandon this naive representation [69]. On the other hand the Cointelation model allows for a minimal adjustment to a mainstream practitioners friendly model with only one additional parameter (θ) which feels more appropriate in this context than the MAR models.

2.2.3 Interesting Properties

2.2.3.1 Inferred Correlation

One important element of our proposed model is that, although it models mean reversion in the long run (therefore a long term positive correlation), it can also

model negative correlation in the short run so much so that it can model the whole correlation spectrum $[-1, 1]$. Figure 2.6 illustrates the idea that if one chooses $\rho = -1$ in equation (2.5c), the whole correlation spectrum of $[-1, 1]$ can be reached as a function of the timescale. Therefore all the possible traditional range associated to factoring risk is encompassed in a simple 2 factor model. The reason why we observe

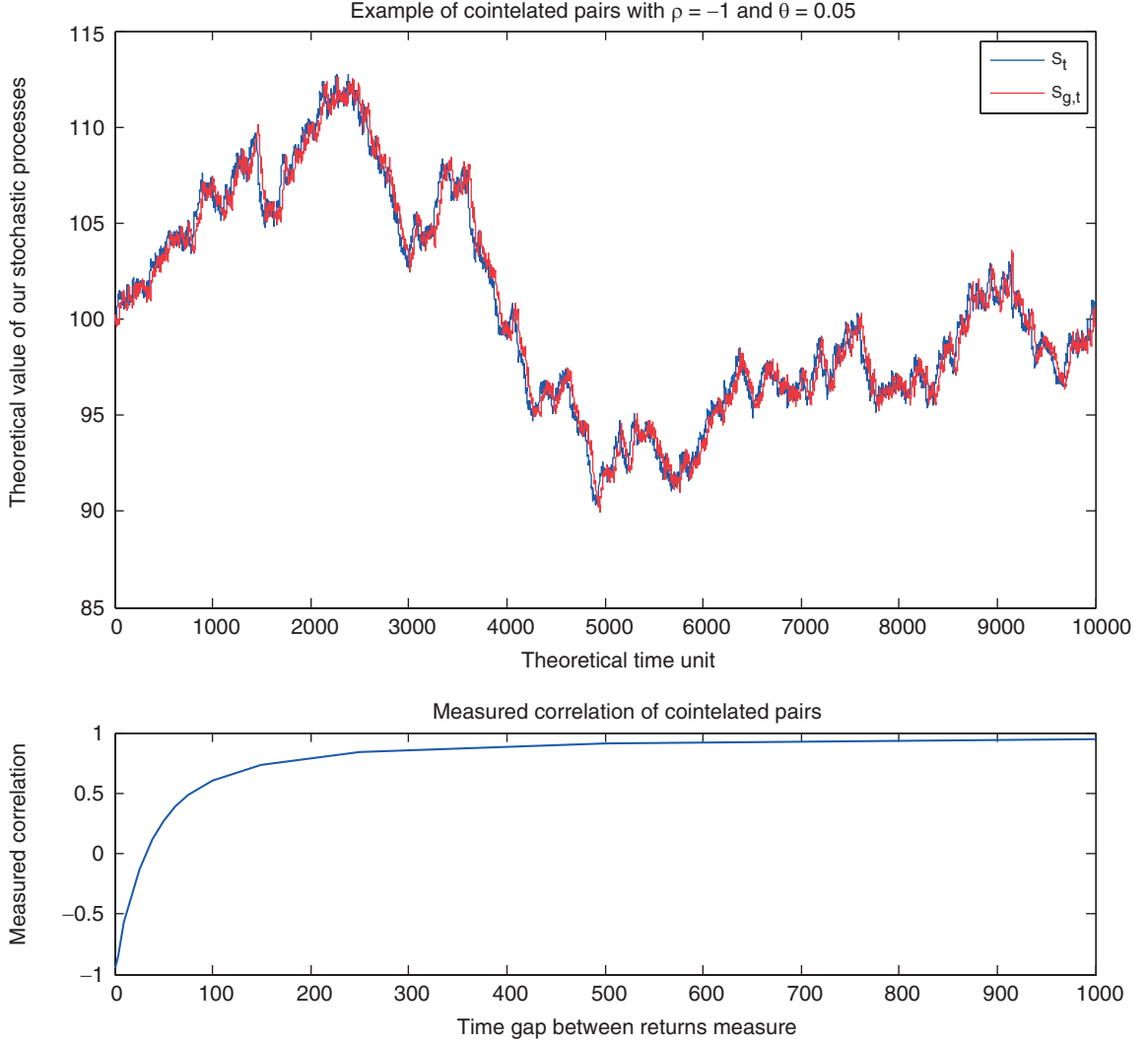


Figure 2.6: Cointelation example touching the whole correlation spectrum [10]: on the top figure we have two Cointelated pairs S and S_g simulated with $\rho = -1$ as well as their measured correlation (on the bottom figure) at different timescale

this interesting phenomenon is because $S_{l,t}$ tries to “catch up” with S_t and the latter phenomenon is dominated in the short run by the stochastic part of the SDE in which the instantaneous correlation is -1 . When the spread between the two SDEs becomes significant, we observe a mean reversion. This contributes towards measuring a higher

correlation between the two SDEs as the timescale increases.

Remark The Inferred Correlation concept was introduced in the context of the Cointelation test but also as a mean to translate the relationship between cointelated pairs into a concept practitioners are more familiar with: Pearson’s correlation coefficient [62, 63] of equation (2.2). The idea of this empirical formula is that at the smallest timescales the estimation for the measured correlation should be equal to ρ from equation (2.5c), should asymptotically converge towards 1 as the timescale increase and the speed at which it converges towards 1 is controlled by θ (the closer θ gets to 1, the faster the measured correlation converges to 1).

Inferred Correlation Empirical Formula: We consider a stock price process $(S_t)_{t \geq 0}$ with natural filtration $(\mathcal{F}_t)_{t \geq 0}$, and we define the forward price process $(F_t)_{t \geq 0}$ by $F_t := \mathbb{E}(S_t | \mathcal{F}_0)$ then considering the dynamics of equation (2.5) we have:

$$\rho_\tau^* \approx \rho + (1 - \rho) [1 - \exp(-\theta\lambda(\tau - 1))] \quad (2.6)$$

where $\rho_\tau^* = \mathbb{E}[\sup_{0 < t \leq \tau} \rho_t]$, $\tau \in \mathbb{Z}^*$, $\theta \in [0, 1]$.

Remark We initially set $\lambda \approx 1.75$ for “regular financial data” [10]. In reality λ is itself a function of the other parameters, though far less sensitive. Improving this approximation is currently an open problem with the Cointelation test. More recent additional simulations have shown that λ is in fact closer to 1/2 than 1.75.

Note that this last proposition formalizes mathematically the idea that during market crashes, even what appears to be the most anti-correlated assets can share the same risk factor in the longer timescales. Figure 2.2 provides an actual time series that is proposed to be Cointelated [11].

Remark In equation (2.6) if we replace the converging constant 1 to x (your relaxed variable of convergence), equation (2.6) becomes $\rho_\tau^* \approx \rho + (x - \rho) [1 - \exp(-\theta\lambda(\tau - 1))]$. This can be a simple way to modify the Inferred Correlation formula as to leave flexibility on ρ_τ^* when $\tau \rightarrow \infty$.

Inferred Correlation is a (some may argue risk averse) formula for the correlation term structure for assets in which the correlation term structure would otherwise be impossible to calculate with available data.

2.2.3.2 Number of Crosses

The second step of the Cointelation test [10] introduces the concept of *Number of Crosses* for which we have an approximate empirical formula. Its intuitive rationale is that, compared to the number of time purely correlated SDEs (e.g., without the mean reversion component⁶) the number of times the discrete version of the cointelated SDEs cross is more than if they were random and the bigger is the value of θ the more often the discretized SDEs *cross paths* per unit of time⁷. At $\theta = 1$, the spread is reabsorb completely and the departure becomes a martingale (e.g., equal chances of going up or down) which makes the expected number of crosses to be exactly half of the data's length. Now that we have the boundaries, we need to find a good interpolation technique. As we can see from figure 2.7 the data suggest that the formula should not be linear. After researching several paths we found that the beta distribution allowed a good fit with the data parameter used (any ρ , $\sigma = 1\%$ and a data's length equal to 1000000).

Number of Crosses Empirical formula: If we discretize equation (2.5), then we can approximate the number of time the two stochastic process, $x = S_{i \in [1,2,\dots,L]}$ and $y = S_{l,i \in [1,2,\dots,L]}$, cross paths, by equation (2.7)

$$\mathbb{E}[\Gamma_{x,y}(\theta, \rho, \sigma, \sigma_l, L)] \approx L \times \frac{1}{2} \int_0^\theta \mathcal{B}(x, a, b) dx \quad (2.7)$$

with L , the length of the data, $\theta \in [0, 1]$, $\int_0^\theta \mathcal{B}(x, a, b) dx$ the cumulative density function of $\mathcal{B}(\theta, a, b)$ introduced earlier in equation (7.8) with $a = 0.57$ and $b = 1$.

Remark We make three remarks. First, the number of cross formula can easily be approximated using simple packages such as Excel's BETA.DIST's built in function ($=L*(\text{BETA.DIST}(\theta, 0.57, 1, \text{TRUE}))*0.5$). Also note that this approximation is an improvement of our first attempt⁸ at modelling this idea [10]. More specifically some of the boundaries of this formula were criticized during the viva [70]. The reworked formula incorporates the remarks made. Also note that the formula is invariant with respect to either ρ , σ or σ_l . Finally it may be argued that the concept of Number of Crosses is somewhat related to the concept of Number of Turning Points [71] in Gaussian kernels. Like in the Number of Crosses, the number of turning points is controlled by the lengthscale⁹ but remains random [71].

⁶ $\theta = 0$.

⁷the mirror concept in continuous time can be thought of a version of local time.

⁸Where $\mathbb{E}[\Gamma_{x,y}(\theta, L)] \approx L \left[\gamma(1 - \theta) + \frac{1}{2}\sqrt{\theta} \right]$.

⁹Corresponding to L in equation (2.7).

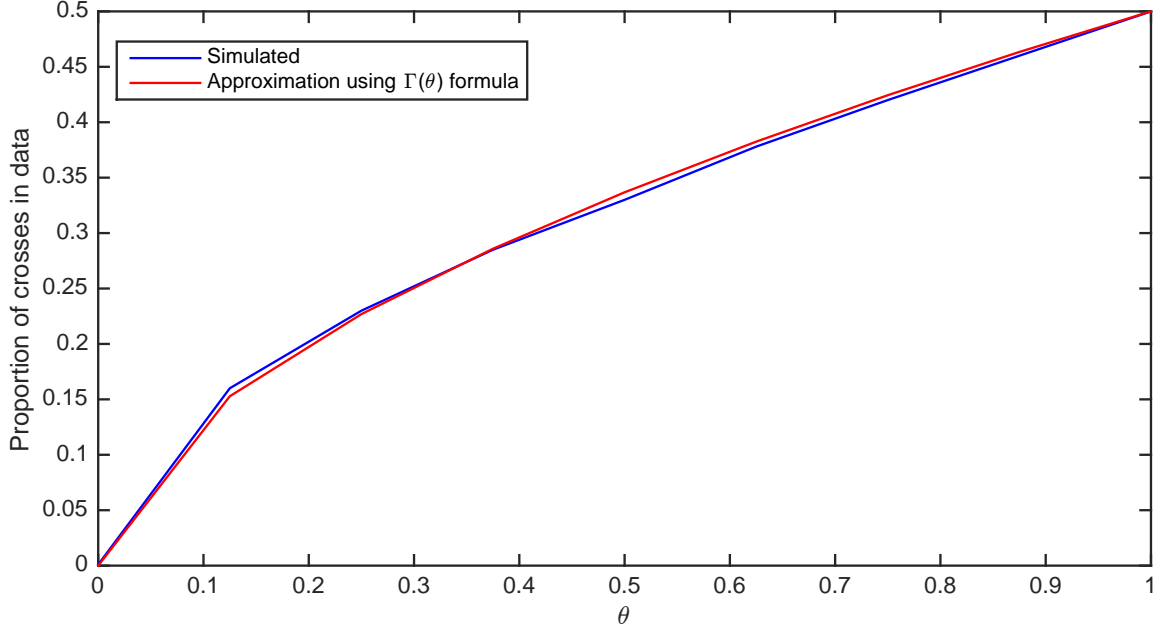


Figure 2.7: Number of cross formula of equation (2.7) compared to simulated data

2.2.4 Parameter estimation

Using the methodology of the original paper [10], we can use sequential estimation in order to estimate the key parameters, ρ , θ and σ as per Equations (2.8a), (2.8b) and (2.8c) respectively.

$$\hat{\rho}_1 = \mathbb{E}[\rho_1], \quad (2.8a)$$

$$\hat{\theta} = \mathbb{E} \left[\frac{dS_{l,t}}{(S_t - S_{l,t})dt + \sigma S_{l,t} \left(\hat{\rho}_1 dW_t + \sqrt{1 - \hat{\rho}_1^2} dW_t^\perp \right)} \right], \quad (2.8b)$$

$$\hat{\sigma} = \mathbb{E} \left[\frac{dS_{l,t} - \hat{\theta}(S_t - S_{l,t})dt}{\hat{\rho}_1 dW_t + \sqrt{1 - \hat{\rho}_1^2} dW_t^\perp} \right]. \quad (2.8c)$$

Similar to the variance reduction methodology described by [10], we define $B_+ = \left| \frac{\max(S_t - S_{l,t}, t \in [0, T])}{2} \right|$ and $B_- = \left| \frac{\inf(S_t - S_{l,t}, t \in [0, T])}{2} \right|$. We note that the estimation of θ has a higher variance when $Z_\sigma = B_+ > |S_t - S_{l,t}| > B_-$ where σ , on the other hand has quality samples. The reverse is true when $Z_\theta = |S_t - S_{l,t}| > B_+ \cup |S_t - S_{l,t}| < B_-$. We can therefore sample θ in Z_θ and σ in Z_σ . Figure 2.8 gives a representation of these sampling zones.

Remark Note that estimating θ can be done via other different ways such as inverting the number of crosses formula (2.7) like it is shown by equation (2.9) or with a

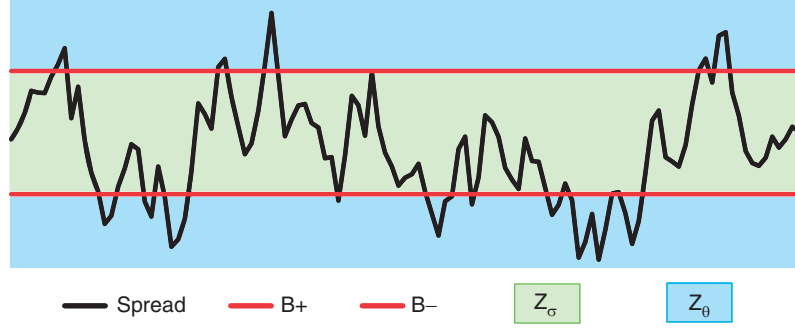


Figure 2.8: Visual representation for the sampling zones for θ [10].

combination of methods [10].

$$\theta \approx \lambda \left(\frac{\hat{\Gamma}_{x,y}(\theta, l)}{l} - \gamma \right)^2. \quad (2.9)$$

2.2.5 Proposed Test

Definition (Cointelation Test): Two stochastic processes representing financial data are Cointelated if the following four conditions are verified. The “Inferred Correlation” hypothesis and the “Number of Crosses” hypothesis must be verified. Also the underlying assets must have reasonable physical connection that would suggest that their spread should mean revert. Note that in the instances where the first two points are not verified exactly, the correlation model cannot possibly be a substitute as correlation is a special case of cointelation (where $\theta = 0$) model.

Remark To some extent few of the concepts that we are developing here could be approximated by a copula-adjusted correlation model or preferably a mean-adjusted model. More specifically, using [72]’s notation we can define a general Gaussian based multivariate density f in equation (2.10):

$$f_{(p)}(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \prod_{i=1}^p \sigma_i |R|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} u^T R^{-1} u \right), \quad (2.10)$$

where $u = (u_1, \dots, u_p)^T$, $u_i = \frac{x_i - \mu_i}{\sigma_i}$, R the correlation matrix and p the number of marginal distributions. In the case of the cointelation model’s translation into a mean-adjusted model bivariate copula, we set $p = 2$ and $u_2 = \frac{S_{l,t} - S_t}{\sigma_l}$ in equation (2.10). As for the copula-adjusted correlation’s translation into the cointelation model, interesting work has been developed in that context on the premise that correlation

between financial objects becomes stronger as the market is going down, especially during market crashes [73]. The work is centered around the formalization of the *local correlation* concept which is itself an enhancement of a conditional correlation model developed as a response to the last subprime crisis [73, 74, 75, 76, 77]. The main limitations of these models are mostly due to the somewhat arbitrary regions for the conditional correlation as well as the inclusion of indicative functions. From the admission of the authors themselves the local correlation concept has computation and estimation that are far more difficult than the corresponding computation and estimation of the conditional correlation (which we have seen have other issues as well) [73].

2.3 Socially Responsible & Consumer Finance

2.3.1 Approved Persons

The legal duties of an approved person are given in Section 2 of the Financial Services and Markets Act (FSMA 2000) [78]. This also spells out the purpose of regulation by specifying the Financial Services Authority's (FSA) four statutory objectives. The first one of these statutory duty is to maintain confidence in the financial system, the second is to promote public understanding, the third is to protect consumers, and finally, the last one consists of reducing financial crimes. Approved persons follow the following principles¹⁰: 1) integrity; 2) skill, care, and diligence; 3) proper standard of market conduct; 4) dealing with the regulator in an open way; 5) proper organization of business; 6) skill, care, and diligence in management; and 7) finally compliance with regulatory requirements [78].

2.3.2 Application to the example of Oil and BP

2.3.2.1 Physical Reason

Because BP is a company focusing on the production of Oil it is normal that its long term performance is strongly linked to the price of Oil. Thus, the relationship between Oil and BP is a good candidate for the cointelation model.

2.3.2.2 Correlation & timescales

If one examines the correlation between Oil and BP (Figure 2.9) we realize that the measured correlation between Oil and BP is a function of the gaps in the measurement

¹⁰Similar objectives are set out by the SEC.

of the returns. For example the correlation coefficient of the returns between Oil and BP every 12 days is 0.37, whereas if we compute it every 70 days we obtain a correlation coefficient of 0.56 (see Figure 2.9). How do we explain this discrepancy? The difference is due to the fact that in the long run Oil and BP have the similar trend but in the short run we obtain a weaker perceived relationship due to the numerous noises that can only impact one of these assets and not the other (example the very temporary impact of the Gulf of Mexico accident on the spread between Oil and BP in Figure 2.9). Because of the physical reasons we have just described the relationship between Oil and BP is a good candidate for the cointelation model. In fact with the current financial mathematics literature this duality between differences in the short term relationship and the long term relationship can only be explained in a simple way by the Cointelation model. Using the variance reduction technique used in [10] and [12], one can estimate θ (in the last 5 years, for daily returns $\hat{\theta} \approx 1\%$) and then plug in θ into equation (2.6) and we can obtain a pretty accurate representation for $\mathbb{E}[\max \hat{\rho}_\tau]$ (Figure 2.9). Figure 2.2, we looked at earlier in this chapter shows that a similar study can be done with Target and Walmart.

Remark Another point could be made on whether 0.37 is very different from 0.56. The point here is not really whether these numbers are statistically significant but rather how significant is the way they were fetched and how an unsophisticated investor interprets these numbers: 0.37 is the correlation out of 70 possible timescales in this example and this is obviously significant on its own. However, the fact that it comes from an unnatural timescale¹¹ (12 days) adds to the significance of the misleading practices used in order to fetched this number.

2.3.2.3 Prediction & Biological Explanation

Because there is not enough data to measure the longer term correlation one can use equation (2.6) to forecast the yearly correlation estimate which gives $99\% = 0.47 + (1 - 0.47)(1 - e^{-1.75(1\%)(254-1)})$ a very strong long term relationship. Note that this value may appear high. The reason for this perception is twofold. First this gives an estimate of $\mathbb{E}[\max \hat{\rho}_\tau]$ which should be equivalent to $\mathbb{E}[\hat{\rho}_\tau]$ provided the right phase is chosen. The probability that the right phase is chosen decreases as τ increases. However, the longer the timescale, the smaller the penalization of being “out of phase”. The second reason is more qualitative and inspired by the UTOPE concept [9] which is essentially a terminology used in order to express the idea of a

¹¹Neither Daily, Weekly or Monthly returns.

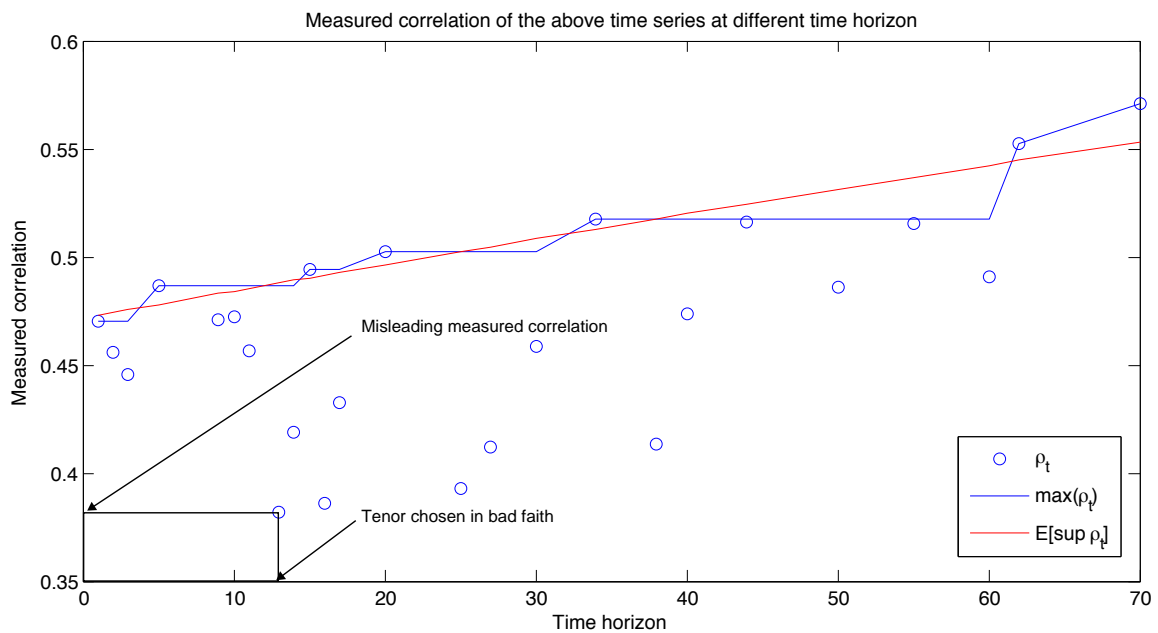
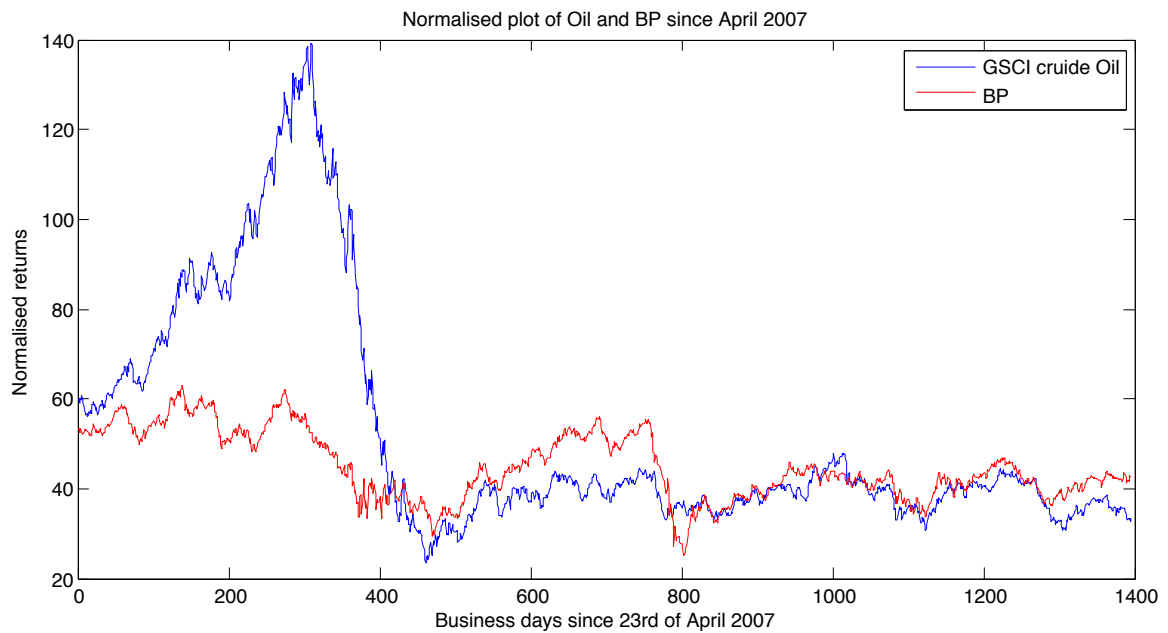


Figure 2.9: Oil and BP and their measured and Inferred Correlation in the context of the FCA “Client Best Interest Rule” and “misleading statement and actions”

human brain “over fitting” a relationship¹². Indeed because the benefit of seeing a true pattern far outweighs the cost of being mistaken with respect to false patterns

¹²Seeing patterns where there are none such as apophenia.

our minds are automatically set up to see patterns sometimes when there are none. The same way it was difficult for our ancestors to grasp that evolution occurs since we cannot physically “see” (with the naked eye) evolution occurring we might not “see” why correlation should increase that much. This is because the time scale in which we have evolved is the product of environmental pressures that did not select for these “long term” qualities that would make us grasp these differences of time scale. We do not live long enough. You cannot “see” evolution with the naked eyes, you notice it through evidence long after it ocured. Similarly seeing the pattern that the correlation is low on a time scale we are comfortable with makes it difficult to extrapolation in longer timescales. The careers in Finance rarely last more than 40 years. Had the careers lasted 400 years, these phenomenons would be much more intuitive. It might not be very intuitive at first to see why the long term correlation which we have never really seen is really that high¹³ but the mathematical evidence suggest the opposite [11].

Remark Long time scales do imply fake correlations with a pure Brownian motion. The induced inter point correlations reduces the number of effective degrees of freedom so much that even correlation close to unity becomes statistically insignificant. In Figure 2.10, we have in green the speed at which the “max” correlation (for pure orthogonal Brownian motions) increases as a function of time and we can see in red how much faster the inferred correlation increases through time. Both do increase with time which is expected, but the latter does it faster.

2.3.3 Marketing Material & Proper Market Conduct

We have seen that in the example of Oil and BP the correlation was very dependent of the sampling frequency τ . It could therefore be tempting for a sales person to try to capitalize on the lowest value of this measured correlation. For example a commodities sales could advertise the lowest measured correlation available to try to attract long term investors who usually invest in the equities market and who seek diversification in the long term. These investors do not necessarily understand that the correlation estimates are an increasing function of τ in cointelated pairs. These investors who then adjust their portfolio based on the theory of Markowitz [79] would get an overall risk in their portfolio which would not go towards their diversification strategy. These practices are in violation of the Client Best Interest Rule and the rule on Misleading Statement and Actions laid out by the FCA.

¹³Or rather we see it during the crisis in which the correlation of everything goes to 1 but prefer to interpret it differently.

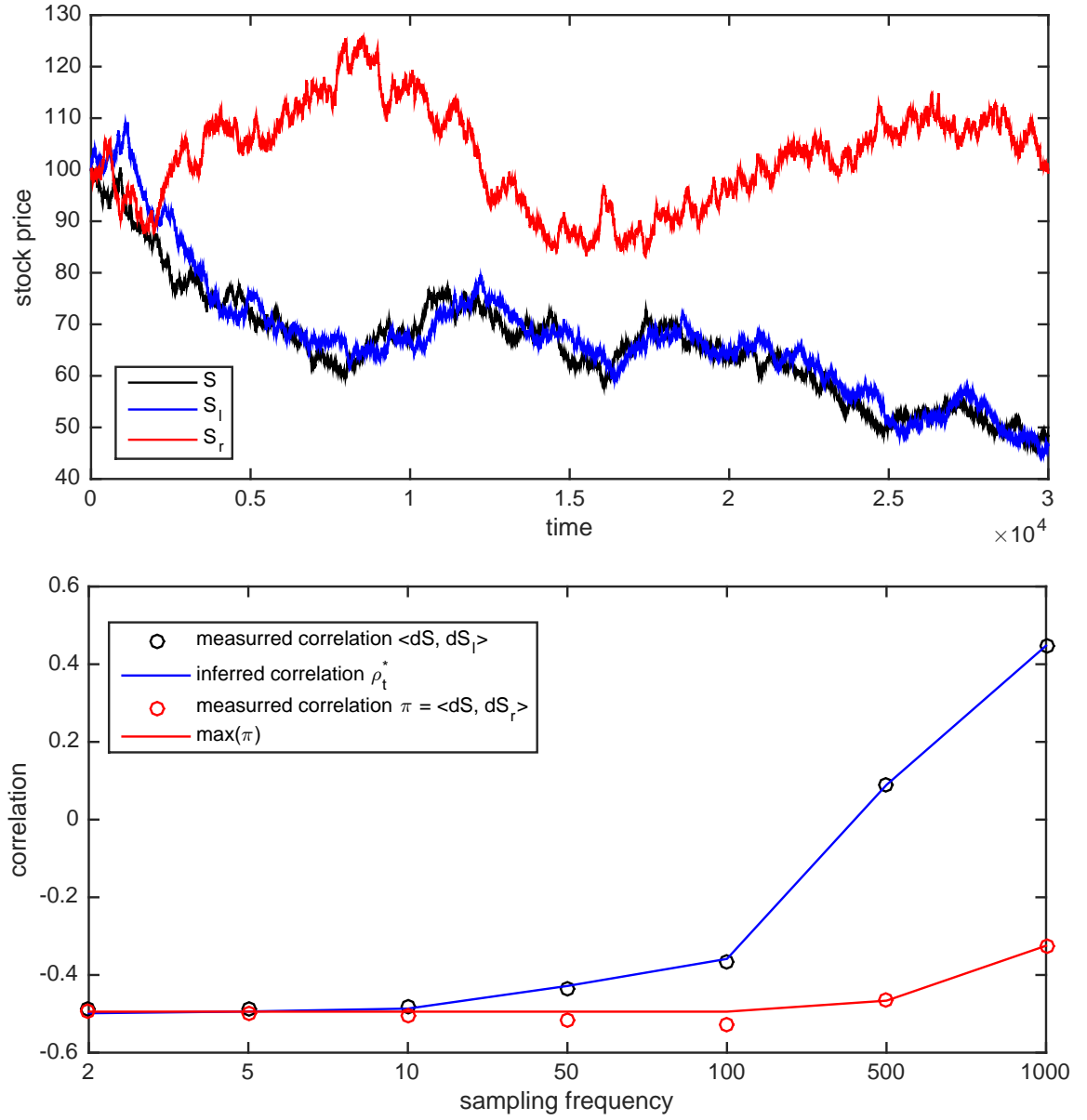


Figure 2.10: Inferred Correlation vs “Fake Correlation”. Simulation of S , S_l and S_r on the top figure and their measured correlation in the bottom figure (in red we have the inferred correlation, in blue the measured correlation of cointelated pairs and in green the max of the measured correlation of non mean reverting pairs). The chosen parameters are $\theta = 0.002$, $\rho = -0.5$, $\sigma = \sigma_l = 0.002$ and $\lambda = 1/2$

Remark Note that some of the rest of the chapter is a joint work [5] (for which I am the first author).

2.4 Portfolio Optimization

2.4.1 Mathematical Definition

We propose in this section an optimal portfolio strategy, leveraging on available methodologies [80, 79].

2.4.2 Markowitz Portfolio Theory Review

2.4.2.1 Foundations

The foundation of modern portfolio theory (MPT) was established by Harry Markowitz in 1952 with his seminal paper [79] in which he proposed expected return and variance to be the criterion for the portfolio selection. More specifically, the problem of an agent who wishes to build a portfolio with the maximum possible level of expected return, given a limit of variance is, considered with a focal point being the portfolio efficientness. Concepts such as “efficient frontier”, or set of efficient mean-variance combinations, were introduced subsequently. Besides introducing the concept, the paper also describes the methodology in detail. Markowitz paved the way for studying theoretically the optimal portfolio choice of risk-averse agents. Based on the ideas developed by Markowitz, Tobin published his famous research work on agents’ liquidity preferences and the separation theorem [81]. Later, the Capital Asset Pricing Model (CAPM) was introduced independently by Sharpe [82] and Lintner [83].

2.4.2.2 Optimization methodology for a pair of assets

Consider a probability space $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, with \mathbb{P} our historical probability measure, and a filtration, $(\mathcal{F}_t)_{(t \geq 0)}$, generated by two dimensional Brownian motion: (W^a, W^b) . Let $r_a S_a$, $r_b S_b$ be our discounted prices. Given a set of parameters $(r_a, r_b, \sigma_a, \sigma_b, \rho)$, we can define the set of Stochastic Differential Equations’ (SDE):

$$\frac{dS_t^a}{S_t^a} = r_a dt + \sigma_a dW_t^a, \quad (2.11a)$$

$$\frac{dS_t^b}{S_t^b} = r_b dt + \sigma_b dW_t^b, \quad (2.11b)$$

$$d\langle W_t^a, W_t^b \rangle = \rho dt. \quad (2.11c)$$

We assume the dynamic of risk-free asset $M(t)$ with continuously compounded risk free rate r satisfies equation

$$dM(t) = rM(t)dt. \quad (2.12)$$

We then define an optimal strategy using Markowitz methodology. More specifically we consider the classic mean-variance optimization problem where we initially define the two important measures for our portfolio return and variance given by:

$$r_p = hr_a + (1 - h)r_b, \quad (2.13a)$$

$$\sigma_p^2 = [h\sigma_a + (1 - h)\sigma_b]^2 = h^2\sigma_a^2 + 2h(1 - h)\sigma_a\sigma_b\rho + (1 - h)^2\sigma_b^2. \quad (2.13b)$$

where h is the weight of asset a , $(1 - h)$ the weight of asset b , σ_a the volatility of asset a , σ_b the one of asset b , r_a the returns of a , σ_p the volatility of the portfolio, r_b the returns of b and r_p the returns of the portfolio.

Definition (Sharpe Ratio): The *Sharpe Ratio* (SR) is given by

$$\text{SR} = \frac{\mathbb{E}[r_p] - \mathbb{E}[r]}{\sigma_p}$$

where r is the risk free return. SR can be divided in two components. First, we have the expected returns $\mathbb{E}[r_p] = \sum_i h_i \mathbb{E}(R_i)$ where r_p is the return on the portfolio, r_i is the return on asset i and h_i is the weighting the proportion of i -asset in the portfolio. Second we have $\sigma_p^2 = \sum_i h_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} h_i h_j \sigma_i \sigma_j \rho_{ij}$, is our portfolio return variance.

Our problem is to maximize the Sharpe Ratio (SR). We achieve this by finding the optimal weight h^{**} , more formally given by equation (2.14b) using the methodology of equation (2.14a).

$$h^{**} = \arg \max_h \frac{\mathbb{E}[r_p - r]}{\sigma_p}, \quad h \in (0, 1), \quad (2.14a)$$

$$h^{**} = \frac{\sigma_b^2 - \sigma_a \sigma_b \rho}{\sigma_a^2 - 2\rho \sigma_a \sigma_b + \sigma_b^2}. \quad (2.14b)$$

2.4.3 Ornstein-Uhlenbeck Theory Review

2.4.3.1 Foundations

A stochastic control approach to the problem of pairs trading was proposed by Mudchanatongsuk, Primbs and Wong [80].

Remark The notation used by Mudchanatongsuk, Primbs and Wong [80] is slightly different from the way we have defined the Cointelation model. We strove to respect their name in the following paragraph in order to abide by practices associated to literature review. A notable difference is the θ parameter which they use as long term mean, which we use as speed of mean reversion.

By modelling the spread of a stock prices as an Ornstein-Uhlenbeck (OU) process [84] a portfolio optimization based stochastic control problem is formulated. The optimal position to this control problem in closed form is computed by solving the corresponding Hamilton-Jacobi-Bellman equation [85]. The dynamics of the risk free rate r is given by equation (2.12). Denote by $A(t)$ and $B(t)$ the prices of two assets at time t with $B(t)$ following the geometric Brownian motion,

$$dB(t) = \mu B(t)dt + \sigma B(t)dZ(t) \quad (2.15)$$

with drift μ , and volatility σ . Here $Z(t)$ is a standard Brownian motion. The spread between the two relevant assets at time t is denoted by $X(t)$ as

$$X(t) = \ln[A(t)] - \ln[B(t)], \quad (2.16)$$

and assumed to follow the mean-reverting process

$$dX(t) = \kappa[\theta - X(t)]dt + \eta dW(t), \quad (2.17)$$

where θ is the long-term equilibrium to which the spread reverts, κ the rate of mean reversion¹⁴ and η is the volatility of the spread. Let ρ denote the instantaneous correlation coefficient between Z and W , therefore $\langle dW(t), dZ(t) \rangle = \rho dt$.

2.4.3.2 Optimization Methodology for a pair of assets

Under the above assumptions and by means of Ito's lemma, the dynamics of $A(t)$ are,

$$\frac{dA(t)}{A(t)} = \left(\mu + \kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta \right) dt + \sigma dZ(t) + \eta dW(t). \quad (2.18)$$

The wealth dynamic of the self-financing portfolio $V^h(t)$ is then described by:

$$dV^h(t) = V^h(t) \left[h(t) \frac{dA(t)}{A(t)} + \tilde{h}(t) \frac{dB(t)}{B(t)} + \frac{dM(t)}{M(t)} \right], \quad (2.19)$$

which can be rewritten as

$$\frac{dV^h(t)}{V^h(t)} = \left[h(t) \left(\kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta \right) + r \right] dt + \frac{1}{V^h(t)} \eta dW(t),$$

where $h(t)$ is the portfolio weight of stock A and $\tilde{h}(t) = -h(t)$ is the portfolio weight of asset B at time t . Assuming that an investors' preference can be represented by

¹⁴with $\kappa > 0$.

the utility function $U(x) = \frac{1}{\gamma}x^\gamma$, with $x \geq 0$ and $\gamma < 1$, the stochastic control problem is of the form

$$h^*(t, x) = \sup_{h(t)} \mathbb{E} \left[\frac{1}{\gamma} (V^h(t))^\gamma \right], \quad (2.20)$$

subject to $V(0) = v_0$, $X(0) = x_0$, $dX(t) = \kappa[\theta - X(t)]dt + \eta dW(t)$ and $dV^h(t) = V^h(t) \left[h(t)(\kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta) + r \right] dt + \eta dW(t)$. The optimal weight $h^*(t, x)$ is given by

$$h^*(t, x) = \frac{1}{1 - \gamma} \left[\beta(t) + 2x\alpha(t) - \frac{\kappa(x - \theta)}{\eta^2} + \frac{\rho\sigma}{\eta} + \frac{1}{2} \right],$$

where $\alpha(t) = \frac{\kappa(1 - \sqrt{1 - \gamma})}{2\eta^2} \left(1 + \frac{2\sqrt{1 - \gamma}}{1 - \sqrt{1 - \gamma} - (1 + \sqrt{1 - \gamma}) \exp\left(\frac{2\kappa(T - t)}{\sqrt{1 - \gamma}}\right)} \right)$ and with complicated $\beta(t) = \frac{1}{2\eta^2[(1 - \sqrt{1 - \gamma}) - (1 + \sqrt{1 - \gamma}) \exp\left(\frac{2\kappa(T - t)}{\sqrt{1 - \gamma}}\right)]} \times \left(\gamma\sqrt{1 - \gamma}(\eta^2 + 2\rho\sigma\eta) \left[1 - \exp\left(\frac{2\kappa(T - t)}{\sqrt{1 - \gamma}}\right) \right]^2 - \gamma(\eta^2 + 2\rho\sigma\eta + 2\kappa\theta) \left[1 - \exp\left(\frac{2\kappa(T - t)}{\sqrt{1 - \gamma}}\right) \right] \right)$.

Proof For the derivation, we refer to Mudchanatongsuk, Primbs & Wong [80].

2.4.4 Stochastic Portfolio Theory Review

2.4.4.1 Introduction

Samo and Vervuurt [86] consider the inverse problem of Stochastic Portfolio Theory (SPT): learning from data an optimal investment strategy, based on any given set of trading characteristics. Initially introduced and developed by Robert Fernholz [87], SPT is a mathematical theory for analyzing the stock market structure and portfolio behavior. Samo and Vervuurt [86] labeled the methodology as “descriptive” as opposed to “normative”, as well as consistent with the observed behavior of actual markets. Normative methodologies, the basis for earlier theories like modern MPT [79] and the CAPM [88], are absent from SPT. For instance SPT provides the rule-based mathematical weaponry to explain under what conditions it becomes possible to outperform a cap-weighted benchmark index.

2.4.4.2 Problems

SPT has, however, several problems and limitations. First finding relative arbitrages is difficult since they are inverse problems. Second, the exclusion of possibilities of bankruptcies seems quite unrealistic for practitioners. Third, the SPT set-up is developed only for investment strategies that are driven by market capitalizations.

These limitations were more or less addressed in [86] by adopting a Bayesian non-parametric approach. The broad range of investment strategies is considered driven by a function defined on an arbitrary space of trading characteristics, on which a Gaussian process prior is placed. One of the aims of SPT is to construct portfolios which outperform an index, or benchmark portfolio, over a given time-horizon with probability one, whenever this might be possible. One such investment strategy is the so-called diversity-weighted portfolio. The idea of the latter is to re-calibrate the weights of the market portfolio, by raising them all to some given power $p \in (0, 1)$ and then re-normalising¹⁵.

2.4.4.3 Framework

We recall in this subsection some of the mathematical formalization for SPT. Fernholz [87] assumes the dynamics of n stock capitalization processes $X_i(\cdot)$ given by (2.21).

$$dX_i(t) = X_i(t) \left(b_i(t)dt + \sum_{\nu=1}^d \sigma_{i\nu}(t)dW_{\nu(t)} \right), \quad (2.21)$$

where $i \in [1, n]$, $t \geq 0$, $W_1(\cdot), \dots, W_d(\cdot)$ are independent standard Brownian motions with $d \geq n$, and $X_i(0) > 0$ are the initial capitalizations. All processes are assumed to be defined on a probability space (Ω, \mathcal{F}, P) , and adapted to a filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{(0 \leq t < \infty)}$ that satisfies the usual conditions and contains the filtration generated by the “driving” Brownian motions. The processes of rates of return $b_i(\cdot)$ of volatilities $\sigma(\cdot) = \sigma_{i\nu}(\cdot)_{(1 \leq i \leq n, 1 \leq \nu \leq d)}$, are \mathbb{F} -progressively measurable and assumed to satisfy the integrability condition:

$$\sum_{i=1}^n \int_0^T \left(|b_i(t)| + \sum_{\nu=1}^d (\sigma_{i\nu}(t))^2 \right) dt < \infty, \quad (2.22)$$

\mathbb{P} -a.s., $\forall T \in (0, \infty)$, as well as the non-degeneracy (ND) condition:

$$\exists \epsilon > 0 \text{ such that: } \zeta' \sigma(t) \zeta \geq \epsilon \|\zeta\|^2, \quad (2.23a)$$

$$\forall \zeta \in \mathbb{R}^n, t \geq 0; \quad \mathbb{P}\text{-a.s.} \quad (2.23b)$$

For the diversity-weighted portfolio with negative parameter p the following no-failure (NF) condition is imposed:

$$\exists \phi \in (0, 1/n) \text{ such that: } P(\mu_{(n)}(t) > \phi \forall t \in [0, T]) = 1. \quad (2.24)$$

¹⁵More information on the methodology can be found in [89].

Remark It is interesting to note that, because of the non realistic NF conditions enhancements are constantly developed, notably a negative-parameter variant ($p < 0$) [90] of the diversity-weighted portfolio [89]. The strategy is claimed to outperform the market, with probability one, over sufficiently long time-horizons, under ND assumption on the volatility structure and under the assumption that the market weights admit a positive lower bound. Several modifications of this portfolio are put forward, which outperform the market under milder versions of the latter NF condition.

A popular implementation is the diversity-weighted portfolio with parameter $p \in \mathbb{R}^*$, defined as in [89]:

$$h_i^{(p)}(t) := \frac{[\mu_i(t)]^p}{\sum_j^n [\mu_j(t)]^p}, \quad i = 1, \dots, n, \quad (2.25)$$

with

$$h_{(1)}^{(p)}(t) \geq h_{(2)}^{(p)}(t) \geq \dots \geq h_{(n)}^{(p)}(t), \quad (2.26)$$

where $\max_{i=1}^n h_i^{(p)}(t) = h_{(1)}^{(p)}(t)$, $h_{(n)}^{(p)}(t) = \min_{i=1}^n h_i^{(p)}(t)$ and

$$\mu_i(t) := \frac{X_i(t)}{\sum_j^n X_j(t)}, \quad i = 1, \dots, n. \quad (2.27)$$

The method for how to optimize p ¹⁶ was not mentioned. However it can be noticed that, while preserving the rankings of all stocks (more weights on stocks with higher returns), decreasing p in (2.25) decreases allows the portofolio to get closer to an equally weighted basket.

Remark Our current observation for the SPT is that it is a popular theory gaining momentum in academia with nevertheless assumptions that are not accepted by practitioners. Besides few mathematical necessary assumptions which are obviously not verified in an economical point of view (e.g. no bankruptcies are allowed)¹⁷, there are more theoretical assumptions not necessarily useful in practice. This has given birth to competing practitioners driven models such as the risk parity theory, for example used by hedge funds or funds of hedge funds¹⁸. More specifically, in practice the differences in the $\mu_i(t)$'s of equation (2.27) are more explained by the stochastic part of equation (2.21) than its deterministic part so the optimization is more explained by statistics of order than by optimal allocation based on value. Lastly the theory

¹⁶though tables have been provided in an empirical study [89].

¹⁷but are nevertheless not necessarily show-stoppers in terms of practical implementations.

¹⁸Aquila Capital, Northwater, Wellington, Invesco, First Quadrant, Putnam Investments, ATP, PanAgora Asset Management, BlackRock, 1741 Asset Management, Neuberger Berman, Alliance Bernstein, AQR Capital Management, Clifton Group, Salient Partners, Schroders, Natixis Asset Management and Allianz Global Investors.

is lagging as opposed to leading and makes it sub-optimal in relative value arbitrage despite quality attempts done in that regard [86, 89].

2.4.5 Mean-Variance Criterion

As we have seen in Section 2.4.2.2, Markowitz's efficient portfolios can be represented graphically to form efficient frontiers [79]. In this section we use the Mean-Variance Criterion to optimize the portfolio of two stocks which follow the Cointelation model [10] and one risk-free asset.

2.4.5.1 Expected Return and Variance of Portfolio

A portfolio considers a combination of n potential assets, with an initial capital/wealth of $V(0)$ and weights h_1, h_2, \dots, h_n , such that $\sum_i^n h_i = 1$. Let $h_i V(0)$ be the amount invested in security i for $i = 1, 2, \dots, n$. Given the weights h_1, \dots, h_n , the number of shares to invest in security i is

$$n_i = \frac{h_i V(0)}{S_i(0)}. \quad (2.28)$$

The value of the portfolio at time t is $V(t) = \sum_{i=1}^N n_i S_i(t)$. Thus the percentage of the portfolio invested in asset i is

$$h_i(t) = \frac{n_i S_i(t)}{\sum_{i=1}^N n_i S_i(t)}, \quad (2.29)$$

with $\sum_{i=1}^N h_i(t) = 1$. The rate of return of asset i over $[t - \Delta t, t]$ is given by

$$R_i(t) = \frac{S_i(t) - S_i(t - \Delta t)}{S_i(t - \Delta t)} = \frac{S_i(t)}{S_i(t - \Delta t)} - 1. \quad (2.30)$$

The rate of return of portfolio, $R_p(t)$, over $[t - \Delta t, t]$ is defined as

$$R_p(t) = \frac{V(t) - V(t - \Delta t)}{V(t - \Delta t)}. \quad (2.31)$$

The return of the portfolio is a linear combination of the returns of individual assets as follows

$$\begin{aligned} R_p(t) &= \frac{V(t)}{V(t - \Delta t)} - 1 = \sum_{i=1}^N \frac{n_i S_i(t)}{\sum_{j=1}^N n_i S_i(t - \Delta t)} - 1 \\ &= \sum_{i=1}^N \frac{n_i S_i(t - \Delta t) S_i(t)}{\sum_{j=1}^N n_i S_i(t - \Delta t) S_i(t - \Delta t)} - 1 \\ &= \sum_{i=1}^N h_i(t) (R_i(t) + 1) - 1 = \sum_{i=1}^N h_i(t) R_i(t). \end{aligned}$$

Remark The literature sometimes uses log returns, defined to be:

$$r_i(t) = \ln \left(\frac{S_i(t)}{S_i(t - \Delta t)} \right). \quad (2.32)$$

This is because for a short period of time the log return is approximately equal to the rate of return

$$r_i(t) = \ln \left(\frac{S_i(t)}{S_i(t - \Delta t)} \right) = \ln(R_i(t) + 1) \approx R_i(t). \quad (2.33)$$

The return of portfolio in this case is $r_p = \sum_{i=1}^N h_i r_i$. The expected return of portfolio, $E(r_p)$, and the variance of the return of portfolio, $\sigma^2(r_p)$, are give by $E(r_p) = E \left(\sum_{i=1}^N h_i r_i \right)$ and $\sigma^2(r_p) = Var \left(\sum_{i=1}^N h_i r_i \right)$. Using the properties listed in [91] we obtain the following:

$$E(r_p) = \sum_{i=1}^N h_i E(r_i), \quad (2.34)$$

$$\sigma^2(r_p) = \sum_{i=1}^N \sum_{j=1}^N h_i h_j \sigma(r_i, r_j) = \sum_{i=1}^N h_i^2 \sigma^2(r_i) + \sum_{i=1}^N \sum_{j \neq i}^N h_i h_j \sigma(r_i, r_j), \quad (2.35)$$

where $E(r_i)$ and $\sigma^2(r_i)$ are the expectation and the variance of returns of asset i respectively and $\sigma(r_i, r_j)$ is the covariance of returns of i and j .

2.4.5.2 Optimal Investment Using the MVC

Let the probability space (Ω, \mathcal{F}, P) with filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by two-dimensional Brownian motions. We assume the dynamic of the risk-free asset $M(t)$ with continuously compounded risk free rate $r \geq 0$ satisfying

$$dM(t) = rM(t)dt. \quad (2.36)$$

Denote by $S(t)$ and $S_l(t)$ the prices of two assets at time t , which dynamics follows the Cointelation model (2.5). We assume an initial wealth $w_0 > 0$ at time $t = 0$. The investment behavior is modelled by an investment strategy $h = (h_1, h_2, h_3)$. Here, $h_i \in [0, 1]$, $i = 1, 2, 3$, denotes the percentage of total wealth invested in i -th asset. Let h_1, h_2, h_3 denote respectively the portfolio weights for stocks S , S_l and risk-free asset M at time t . The weight are constant throughout the investment horizon $[0, T]$. We restrict our considerations to self-financing strategies. Denote by V_t^h the value of portfolio at time t given by:

$$V^h(t) = \frac{h_1 V^h(t)}{S(t)} S(t) + \frac{h_2 V^h(t)}{S_l(t)} S_l(t) + \frac{h_3 V^h(t)}{M(t)} M(t), \quad t > t_0,$$

with $V(t_0) = v_0$. The dynamic of the wealth process is given by:

$$dV^h(t) = V^h(t) \left[h_1 \frac{dS(t)}{S(t)} + h_2 \frac{dS_l(t)}{S_l(t)} + h_3 \frac{dM(t)}{M(t)} \right]. \quad (2.37)$$

Definition (Admissible strategies) Let $\mathcal{A}(v_0)$ denote the set of all admissible strategies corresponding to the initial condition $v_0 > 0$. We say that a trading strategy $h = (h_1, h_2, h_3)$ is *admissible* and write $h \in \mathcal{A}(v_0)$ if

(i) Given $v_0 > 0$ the wealth process $V^{v_0, h}(\cdot)$ corresponding to v_0, h satisfies

$$V^{v_0, h}(t) \geq 0, \quad 0 \leq t \leq T, \quad (2.38)$$

(ii) $h_i \geq 0$ for all $i = 1, 2, 3$,

(iii) $\sum_{i=1}^3 h_i = 1$.

Definition (Optimal Investment Strategy): An investment strategy is called optimal if

$$\begin{cases} E(r_p(h)) \geq E(r_p(\tilde{h})), & \text{for all } \tilde{h} \in \mathcal{A} \\ \text{Var}(r(h)) \leq \text{Var}(r(\tilde{h})), & \text{for all } \tilde{h} \in \mathcal{A}. \end{cases}$$

Definition (Mean Variance Criterion): An optimal strategy for the Mean Variance Criterion (MVC) is equivalent to an optimal strategy for maximizing the following utility function $U(T, h) = 2\tau E[r_p] - \sigma^2[r_p]$, where $\tau \geq 0$ is risk tolerance.

Thus, the portfolio problem becomes

$$\max_h U(T, h) \quad (2.39)$$

with constraints $\sum_{i=1}^N h_i = 1$ and $h_i \geq 0 \quad \forall i$. From equation (2.32) we have that the rate of return of our portfolio over $[0, t]$ is $R_p(t) = \frac{V^h(t) - V^h(t - \Delta t)}{V^h(t - \Delta t)} = \sum_{i=1}^3 h_i(t) R_i(t)$ and the log return of our portfolio, r_p is given by

$$r_p(t) = h_1 r_1(t) + h_2 r_2(t) + h_3 r_3(t), \quad (2.40)$$

where $r_i(t) \approx R_i(t)$, as we showed in equation (2.33).

Lemma 2.4.1 Denote by $V^{v_0, h}(t)$ the value of the portfolio corresponding to the admissible strategy $h \in \mathcal{A}_0$ and initial wealth $v_0 > 0$. Then

(i) The expectation of portfolio return (2.18) over $[0, T]$ is

$$\begin{aligned}
E(r_p(T)) &= h_1\nu_1(T) + h_2\nu_2(T) + h_3r(T), \\
\nu_1(T) &= \left(\mu - \frac{\sigma^2}{2}\right)T, \\
\nu_2(T) &= \left[\ln(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T}) - \ln(S_{l,0}) + \frac{1}{2} \right. \\
&\quad \left. - \frac{ce^{(2\mu+\sigma^2)T} + de^{(\mu-\theta+\sigma\sigma_l\rho)T} + (S_{l,0}^2 - c - d)e^{(\sigma_l^2-2\theta)T}}{2(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} \right], \\
\sigma_1^2(T) &= \sigma^2T, \\
\sigma_2^2(T) &= \frac{ce^{(2\mu+\sigma^2)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} + \frac{de^{(\mu+\sigma\sigma_l\rho-\theta)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} + \\
&\quad \frac{(S_{l,0}^2 - c - d)e^{(\sigma_l^2-2\theta)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} - 1, \\
\sigma_{12}(T) &= \ln\left(\frac{be^{(\mu+\sigma^2)T} + (S_0S_{l,0} - b)e^{(\sigma\sigma_l\rho-\theta)T}}{aS_0e^{\mu T} + (S_{l,0}S_0 - aS_0)e^{-\theta T}}\right).
\end{aligned} \tag{2.41}$$

(ii) The variance of portfolio return (2.13a) over $[0, T]$ is

$$\sigma^2(r_p(T)) = h_1^2\sigma_1^2(T) + 2h_1h_2\sigma_{12}(T) + h_2^2\sigma_2^2(T), \tag{2.42}$$

where $\nu_1(T)$ is the expected return of stock S over investment horizon $[0, T]$, $\nu_2(T)$ is the expected return of stock S_l over investment horizon $[0, T]$, $\sigma_1^2(T)$ is the variance of returns of stock S over investment horizon $[0, T]$, $\sigma_2^2(T)$ is the variance of return of stock S_l over investment horizon $[0, T]$, and $\sigma_{12}(T)$ is the covariance of returns of two stocks S and S_l over investment horizon $[0, T]$.

Proof See Appendix 9.3.

With expectation of equation (2.41) and the variance of equation (2.4.1) of the portfolio returns, we can obtain optimal strategies of equation (2.39).

Proposition 2.4.2 The optimal solution for the problem of equation (2.39) applied to the Cointelation model of equation (2.5) is:

$$h^* = \frac{1}{e'\Sigma^{-1}e}\Sigma^{-1}e + \tau\left[\Sigma^{-1}M - \frac{e'\Sigma^{-1}M}{e'\Sigma^{-1}e}\Sigma^{-1}e\right], \tag{2.43}$$

where $e' = [1, 1, 1]$, $M = [E(r_S), E(r_{S_l}), E(r_B)]$ and variance-covariance matrix is

$$\begin{bmatrix}
\text{Var}(r_S) & \text{Cov}(r_S, r_{S_l}) & 0 \\
\text{Cov}(r_S, r_{S_l}) & \text{Var}(r_{S_l}) & 0 \\
0 & 0 & 0
\end{bmatrix}.$$

Note that the expressions for $E(r_S)$, $E(r_{S_l})$, $E(r_B)$ and $Var(r_S)$, $Var(r_{S_l})$ are given in the Appendix 2.4.1.

Proof For the proof we refer to [92].

2.4.6 The Stochastic Control Solution

2.4.6.1 Power Utility Maximization Problem

A stochastic control approach to the problem of pairs trading was proposed in [80]. In this section we mainly follow [80], though assume slightly different dynamics for stock prices. More specifically, Mudchanatongsuk, Primbs and Wong [80] assume the price dynamics of one of the stocks to be a geometric Brownian motion and model the log-difference of the stock prices as an Ornstein-Uhlenbeck process. We however, assume the dynamics of stock prices are governed by the Cointelation model of equation (2.5), where one of the stocks follow the geometric Brownian motion and the second stock mean reverts around the first one.

2.4.6.2 Definition and Assumptions

Let (Ω, \mathcal{F}, P) be a complete probability space with a Brownian filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by two-dimensional Brownian motion, $(\tilde{W}(t), W(t))_{t \geq 0}$. We consider the same set-up as in Section 2.4.5.2: a portfolio of two stocks and one risk-free asset. The stocks dynamics follows the Cointelation model of equation (2.5) and the dynamic of risk-free asset is given in equation (2.36). We assume an initial wealth $v_0 > 0$ at time $t = 0$. The positions of the different assets composing the portfolio are allowed to be adjusted continuously up to a fixed horizon T . The investment behavior is modelled by an investment strategy $h = (h_1, h_2, h_3)$ where, $h_i(t)$, $i = 1, 2, 3$, denote the percentage of total wealth invested in asset i at time t . Let $h_1(t)$, $h_2(t)$ denote the portfolio weights for stocks S and S_l , respectively at time t . In addition, we restrict our considerations to self-financing strategies. We use the definition of admissible control as in [93] for the definition of admissible control and controlled process.

Definition (Control Processes): Given a subset U of \mathbb{R}^3 , we denote by \mathcal{U}_0 the set of all progressively measurable processes $h = \{h_t, t \geq 0\}$ valued in U . The elements of \mathcal{U}_0 are called control processes.

Denote by $V^h(t)$ the value of portfolio corresponding to strategy h at time t , which is given by

$$V^h(t) = \frac{h_1(t)V^h(t)}{S_t}S_t + \frac{h_2(t)V^h(t)}{S_l(t)}S_l(t) + \frac{h_3(t)V^h(t)}{M(t)}M(t).$$

The dynamic of the portfolio value is given by equation (2.44).

$$dV^h(t) = V^h(t) \left[h_1(t) \frac{dS_t}{S_t} + h_2(t) \frac{dS_{l,t}}{S_{l,t}} + h_3(t) \frac{dM(t)}{M(t)} \right]. \quad (2.44)$$

Plugging into (2.44) the dynamics for $S(t)$ and $S_l(t)$ from equation (2.5a) and (2.5b) respectively we get:

$$dV^h(t) = V^h(t) \left[h_1(\mu dt + \sigma dW(t)) - h_1 \left(\theta \left(\frac{S_t}{S_{l,t}} - 1 \right) dt + \sigma_l d\tilde{W}(t) \right) + r dt \right]$$

Lemma 2.4.3 *Let $Z(t) := S_t/S_{l,t}$. The dynamics of $Z(t)$ are given by*

$$dZ(t) = [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(Z(t) - 1)]Z(t)dt + Z(t)(\sigma dW(t) + \sigma_l d\tilde{W}(t)) \quad (2.45)$$

Proof By Ito's quotient rule:

$$d \left(\frac{S_t}{S_{l,t}} \right) = \frac{dS_t}{S_t} \frac{S_t}{S_{l,t}} - \frac{dS_{l,t}}{S_{l,t}} \frac{S_t}{S_{l,t}} + \frac{d\langle S_l, S_l \rangle_t}{S_{l,t}^2} \frac{S_t}{S_{l,t}} - \frac{d\langle S, S_l \rangle_t}{S_t S_{l,t}} \frac{S_t}{S_{l,t}} \quad (2.46)$$

Then

$$\begin{aligned} dZ(t) &= \mu Z(t)dt + \sigma Z(t)dW_t - \theta(Z(t) - 1)Z(t)dt \\ &\quad - \sigma_l Z(t)d\tilde{W}(t) + \frac{\sigma_l^2 S_l^2(t)}{S_l^2(t)} Z(t)dt - \sigma\sigma_l\rho Z(t)dt \\ &= [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(Z(t) - 1)]Z(t)dt + (\sigma dW(t) - \sigma_l d\tilde{W}(t))Z(t). \end{aligned}$$

■

For each control process $h \in \mathcal{U}_0$ we rewrite the dynamics of the two-dimensional state process $P = (V, Z)$ as follows

$$dP(t) = a(t, P(t), h(t))dt + b(t, P(t), h(t))dA(t) \quad (2.47)$$

with initial value of $P(t_0) = p_0$ and $A(t) = (W(t), \tilde{W}(t))$ is the two-dimensional Brownian motion. The process P^h is called the controlled process. Let $[t_0, T]$ with $0 \leq t_0 < T < \infty$ be the relevant time interval and define $Q := [t_0, T] \times \mathbb{R}^2$. The coefficient functions $a : Q \times U \rightarrow \mathbb{R}^2$ and $b : \times U \rightarrow \mathbb{R}^{2 \times 2}$ are all continuous. Further, for all $h \in U$ let $a(\cdot, \cdot, h)$ and $b(\cdot, \cdot, h)$ be in $C^1(Q)$. We then define:

Definition (Admissible Control): Denoting $\mathcal{A}(t_0; p_0)$ the set of all admissible controls corresponding to the initial condition $(t_0; p_0) \in Q$, we say a control $\{h(t), \mathcal{F}_t\}_{t \in [t_0, t_1]}$ will be called *admissible* if the following conditions hold

- (i) for all $p_0 \in \mathbb{R}$ the corresponding controlled SDE (2.47) with initial condition $P(t_0) = p_0$ admits a pathwise unique solution $\{P^h(t)\}_{t \in [t_0, t_1]}$,
- (ii) $\forall k \in \mathbb{N}$ the integrability condition is satisfied: $E \left(\int_{t_0}^{t_1} |h(s)|^k ds \right) < \infty$,
- (iii) the corresponding state process P^h satisfies: $E^{t_0, p_0} \left(\sup_{t \in [t_0, t_1]} |P^h(t)|^k \right) < \infty$,
- (iv) only pairs trading is allowed: short one asset and long the other: $h_1 = -h_2$.

Since we consider the self-financing portfolio and due to the last point, the dynamic of wealth process become

$$\begin{aligned} dV^h(t) &= V^h(t)[(h_1[\mu - \theta(Z(t) - 1)] + r)dt + h_1[\sigma dW(t) + \sigma_l d\tilde{W}(t)]], V(0) = v_0 \\ dZ(t) &= [\mu + \sigma_l^2 - \sigma \sigma_l \rho - \theta(Z(t) - 1)]Z(t)dt + Z(t)[\sigma dW(t) - \sigma_l d\tilde{W}(t)], Z(0) = z_0 \end{aligned}$$

2.4.6.3 Optimal Investment Strategy

We assume that an investor's preference can be represented by the power utility function $U(x) = \frac{1}{\gamma}x^\gamma$, with $x \geq 0$ and $\gamma < 1$. It is our objective to maximize the objective (or utility) functional J over all admissible controls, i.e. determine an admissible control $h(\cdot)$ such that for each initial value (t_0, v_0) the utility functional below is maximized:

$$J(t_0, v_0; h) := \mathbb{E} \left[\frac{1}{\gamma} (V^h(t))^\gamma | V_{t_0}^h = v_0, Z_{t_0} = z_0 \right]. \quad (2.48)$$

The value function of the utility maximization problem is defined by

$$\mathcal{W}(t, v) := \sup_{h(\cdot) \in \mathcal{A}(t, v)} J(t, v, h). \quad (2.49)$$

Now we can formulate the optimization problem:

$$\sup_{h(\cdot) \in \mathcal{A}(0, v_0)} \mathbb{E} \left[\frac{1}{\gamma} (V^h(t))^\gamma \right] \quad (2.50)$$

Consider the function $G(t, v, z)$ such that $G \in C^{1,2}(Q)$. The Hamilton-Jacobi-Bellman (HJB) equation corresponding to the stochastic control problem (2.50) is

$$\frac{\partial G}{\partial t}(t, v, z) + \sup_{h \in \mathcal{A}} \mathcal{L}^h G(t, v, z) = 0, \quad (2.51)$$

subject to terminal condition

$$G(T, v, z) = v^\gamma. \quad (2.52)$$

The infinitesimal generator, $\mathcal{L}^h G(t, v, z)$ in (2.51), of the two dimensional state process $P = (V, Z)$, is given by

$$\begin{aligned}\mathcal{L}^h G(t, v, z) = & \frac{1}{2}[h_1^2(\sigma^2 - 2\sigma\sigma_l\rho + \sigma_l^2)v^2G_{vv} + 2h_1(\sigma^2 - 2\sigma\sigma_l\rho + \sigma_l^2)vzG_{vz} \\ & + (\sigma^2 - 2\sigma\sigma_l\rho + \sigma_l^2)z^2G_{zz}] + [h_1[\mu - \theta(z - 1)] + r]vG_v \\ & + [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(z - 1)]zG_z.\end{aligned}\tag{2.53}$$

Theorem 2.4.4 *If there exists an optimal control $h^*(\cdot)$ then G coincides with the value function:*

$$G(t, v, s) = \mathcal{W}(t, v, z) = J(t, v; h^*).$$

We cannot solve the PDE in equation (2.51) in closed form. We will however see in section 2.4.7 that we can find an approximation for the solution with Deep Learning. In the meantime we examine a pure Machine Learning approach.

2.4.7 Deep Learning for solving our PDE

A hybrid methodology between classic Financial Mathematics and what people consider pure Machine Learning is presented in this section. More specifically, we try to solve our nonlinear partial differential equation of equation (2.53) with a deep learning algorithm¹⁹. To some extent this section illustrates how the whole of two fields can outperform their individual parts²⁰.

2.4.7.1 General Idea

In the general case, we consider a PDE with d spatial dimensions:

$$\begin{aligned}\frac{\partial u}{\partial t}(t, x; \theta) + \mathcal{L}u(t, x) &= 0, \quad (t, x) \in [0, T] \times \Omega, \\ u(t, x) &= g(t, x), \quad x \in \partial\Omega, \\ u(t = 0, x) &= u_0(x),\end{aligned}\tag{2.54}$$

where $x \in \Omega \subset \mathbb{R}^d$ and \mathcal{L} is the infinitesimal generator, a function of all the other partial derivatives. Note that Ω represents the domain of our function and $\partial\Omega$ represents the domain at the boundaries. We implement here the Deep Learning method recently introduced in [24]. More specifically, the goal is to approximate the function

¹⁹Using the DGM [23] presented later but which architecture can be seen by Figures 2 and 3.

²⁰If you are a pure mathematician, please take this Systems Engineering jargon to mean that the fields enhance each other.

$U(t, x)$ with the deep neural network $f(t, x; \theta)$. Here $\theta \in \mathbb{R}^d$ is the set of neural network parameters. The objective function associated to our problem consists of three parts:

1. A measure of how well the approximation satisfies the differential operator:

$$\left\| \frac{\partial f}{\partial t}(t, x; \theta) - \mathcal{L}f(t, x; \theta) \right\|_{[0, T] \times \Omega, \nu_1}^2, \quad (2.55)$$

2. A measure of how well the approximation satisfies the boundary condition:

$$\left\| \frac{\partial f}{\partial t}(t, x; \theta) - g(t, x) \right\|_{[0, T] \times \partial\Omega, \nu_2}^2, \quad (2.56)$$

3. A measure of how well the approximation satisfies the initial condition:

$$\left\| \frac{\partial f}{\partial t}(0, x; \theta) - u(0, x) \right\|_{\Omega, \nu_3}^2. \quad (2.57)$$

Remark Here all three errors are measured in terms of L^2 -norm, i.e. $\|f(y)\|_{\mathcal{Y}, \nu}^2 = \int_{\mathcal{Y}} |f(y)|^2 \nu(y) dy$ with $\nu(y)$ being a density on region \mathcal{Y} .

The sum of all three terms above gives us the objective function associated with training the neural network:

$$\begin{aligned} J(f) = & \left\| \frac{\partial f}{\partial t}(t, x; \theta) - \mathcal{L}f(t, x; \theta) \right\|_{[0, T] \times \Omega, \nu_1}^2 + \left\| \frac{\partial f}{\partial t}(t, x; \theta) - g(t, x) \right\|_{[0, T] \times \partial\Omega, \nu_2}^2 + \\ & \left\| \frac{\partial f}{\partial t}(0, x; \theta) - u(0, x) \right\|_{\Omega, \nu_3}^2. \end{aligned} \quad (2.58)$$

Thus, the goal is to find a set of parameters θ such that the function $f(t, x; \theta)$ minimizes the error $J(f)$. When the dimension d is large, estimating θ by directly minimizing $J(f)$ is infeasible. Therefore, one can minimize the error $J(f)$ using a stochastic gradient decent on a Deep Neural Network, where we use a sequence of time and space points drawn randomly. We describe the DGM in algorithm (1).

Remark Note that our learning rate α_n must decrease with n [24] and a simple way to do this is by using an exponential weighted method where $\alpha_n \leftarrow \alpha_{n-1} * \lambda$ with $\lambda \in (0, 1)$.

Algorithm 1: Deep Galerkin Method (DGM)

Input: The untrained DGM NN architecture

Output: The trained NN approximating the function f

▷ Set Arbitrary Stopping Criteria

$\text{maxIteration} \leftarrow 10000, i \leftarrow 0, e \leftarrow 10^{-3}, e_t \leftarrow 1$

for $(i \leq \text{maxIteration})$ **AND** $(e_t \leq e)$ **do**

▷ Generate random points

$(t_n, x_n) \leftarrow \mathcal{U} \sim [0, 1]^2, (\tau_n, z_n) \leftarrow \mathcal{U} \sim [0, 1]^2, w_n \leftarrow \mathcal{U} \sim [0, 1]$

$s_n \leftarrow \{(t_n, x_n), (\tau_n, z_n), w_n\}$

▷ Calculate the squared error

$L_n^1 \leftarrow \left(\frac{\partial f}{\partial t}(t_n, x_n; \theta_n) - \mathcal{L}f(t_n, x_n; \theta_n) \right)^2$

$L_n^2 \leftarrow \left(\frac{\partial f}{\partial t}(\tau_n, z_n; \theta_n) - g(\tau_n, z_n) \right)^2$

$L_n^3 \leftarrow \left(\frac{\partial f}{\partial t}(0, x_n; \theta_n) - u(0, w_n) \right)^2$

$G(\theta_n, s_n) \leftarrow L_n^1 + L_n^2 + L_n^3$

▷ Take a descent step at the random points

$-\arg \max_{\theta_n} G(\theta_n, s_n), \alpha_n \leftarrow \alpha_{n-1} * \lambda,$

$\theta_{n+1} \leftarrow \theta_n \alpha_n \nabla_{\theta} G(\theta_n, s_n)$

▷ Calculate Error

$i \leftarrow i + 1, e_t \leftarrow G(\theta_n, s_n)$

2.4.7.2 Neural Network Architecture

The neural network (NN) architecture used is like a long short-term memory networks (LSTMs) though with small differences [24]. We describe the architecture of this NN.

$$\begin{aligned}
S^1 &= \sigma(\mathbf{w}^1 \cdot \mathbf{x} + b^1) \\
Z^l &= \sigma(\mathbf{u}^{z,l} \cdot \mathbf{x} + \mathbf{w}^{z,l} \cdot S^l + b^{z,l}), \\
G^l &= \sigma(\mathbf{u}^{g,l} \cdot \mathbf{x} + \mathbf{w}^{g,l} \cdot S^l + b^{g,l}), \\
R^l &= \sigma(\mathbf{u}^{r,l} \cdot \mathbf{x} + \mathbf{w}^{r,l} \cdot S^l + b^{r,l}), \\
H^l &= \sigma(\mathbf{u}^{h,l} \cdot \mathbf{x} + \mathbf{w}^{h,l} \cdot (S^l \odot R^l) + b^{h,l}), \\
S^{l+1} &= (1 - G^l) \odot H^l + Z^l \odot S^l \\
f(t, \mathbf{x}, \theta) &= \mathbf{w} \cdot S^{L+1} + \mathbf{b}.
\end{aligned}$$

where $l = 1, \dots, L$, \odot denotes Hadamard multiplication, L number of layers and σ the activation function. The rest of the subscript refer to the specific neurons for our NN architecture of Figures 2 and 3 in the Appendix²¹. The method was tested with several non linear PDEs independently [23] including non linear PDEs such as the Merton problem that is referred to in [23, 24]. The analytical solution and its NN

²¹Click link 9.3.

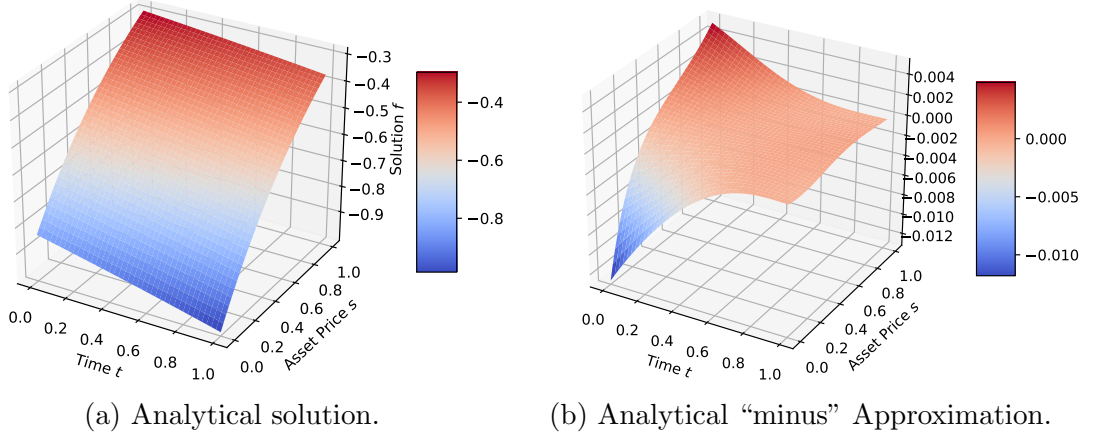


Figure 2.11: Merton [23] non-linear PDE using DGM & independent peer review [24].

approximation error are plotted in that of 2.11a and 2.11b respectively. We refer to the formal definition of the Merton PDE and its resolution in [23, 23, 24].

Remark Though close, the solution does not do as well around $t = 0$ which corroborates with the findings in [23].

We use the DGM described above to solve the reduced²² PDE

$$\begin{aligned}
& \tilde{\sigma}(\gamma - 1)ff_t - \frac{1}{2}(\tilde{\sigma}^2\gamma z^2 f_z^2 + \gamma[\mu - \theta(z - 1)]^2 f - \tilde{\sigma}(\gamma - 1)z^2 f f_{zz}) = \\
& -\tilde{\sigma}\gamma[\mu - \theta(z - 1)]z f f_z - \tilde{\sigma}(\gamma - 1)[\mu + \sigma_l^2 - \sigma\eta\rho - \theta(z - 1)]f f_z - \tilde{\sigma}\gamma(\gamma - 1)r f^2, \\
& f(T, z) = 1,
\end{aligned} \tag{2.60}$$

with $(t, z) \in [0, T] \times \mathbb{R}$ and $\forall z \in \mathbb{R}$ and no boundary or initial conditions.

2.4.8 Dynamic Switching

Now that we have decomposed the Cointelation model into two simpler distinctive strategies and solved them independently, we need to recompose them into a single strategy to recover the first problem. We achieve this by using a dynamic switching methodology. Assuming that an investor's preference can be represented by the utility functions $U(x) = \frac{1}{\gamma}x^\gamma$, with $x \geq 0$ and $\gamma < 1$ and $U(t, h) = 2\tau E[r_p] - \sigma^2[r_p]$ where τ is the risk aversion coefficient, we formulate the following portfolio optimization:

$$w_t = \sup_{h(t) \in \mathcal{A}(0, v_0)} \mathbb{E} \left[\frac{1}{\gamma} (V^h(t))^\gamma \right] 1_{\left\{ \left| \theta \left(\frac{s_t - s_{l,t}}{s_{l,t}} \right) \right| > \mu \right\}} + \max_{h(t)} [2\tau E(r_p) - \sigma^2(r_p)] 1_{\left\{ \left| \theta \left(\frac{s_t - s_{l,t}}{s_{l,t}} \right) \right| \leq \mu \right\}}. \tag{2.61}$$

²²Going from a 3-dimenssional PDE to a 2-dimenssional PDE using an anzast described in 9.3.

The optimization process consists of a triggering function $1_{\left\{\left|\theta\left(\frac{S_t-S_{l,t}}{S_{l,t}}\right)\right|>\mu\right\}}$ which switches between the MVC optimal strategy and the power utility optimal strategy. In the case where $\left|\theta\left(\frac{S_t-S_{l,t}}{S_{l,t}}\right)\right|$, we are in a situation in which the spread strategy is more interesting than the MVC approach and therefore we switch to the former, otherwise we switch back to MVC.

Proposition 2.4.5 *Dynamic switching achieves better results than mean-variance or power utility optimal strategies implemented separately.*

Proof This should be obvious as the dynamic switching allows to do at least as well as the best of the two strategies taken separately. See Figure 2.15 for illustration.

2.4.8.1 Signal Decomposition

We have the set of decomposed signals summarized as

$$\begin{aligned} h_1^{SC}(t) + h_2^{SC}(t) &= 0 \quad \text{and} \quad h_3(t) = 1, \\ h_1^{MVC}(t) + h_2^{MVC}(t) + h_3(t) &= 1, \end{aligned}$$

where $h_1(t) = -h_2(t)$ are the optimal weights for assets S and S_l correspondingly, which were computed in closed form in Section 2.4.5.2 and $h_1(t)$, $h_2(t)$, $h_3(t)$ are weights for assets S , S_l and risk free asset B obtained as a solution to classic MVC in Section 2.4.5.2. These signals are recomposed as below

$$h^S(t) = h_1^{MVC}(t) + h_1^{SC}(t), \tag{2.62}$$

$$h^{S_l}(t) = h_2^{MVC}(t) + h_2^{SC}(t). \tag{2.63}$$

We would like to use some of the results introduced in Section 2.4.6 but this Section uses a different jargon than the one associated to the cointelation model introduced in Section 2.4. We therefore need to reconcile these two sections through a proper transform function.

2.4.8.2 Finding the right phase

One other important aspect to note is that when dealing with cointelated pairs, after estimating θ , ρ , σ , an understanding of phase is still required. The latter point is discussed in this subsection. We have seen in equation (2.7) that the expectation of the number of crosses $\mathbb{E}[\Gamma_{S,S_l}]$ can be approximated by $L \times \int_0^\theta \mathcal{B}(x, a, b) dx$.

Definition (Beginning of Time): We call the *Beginning of Time* τ_0 , the inception time of the discrete version of two Cointelated pairs in which, $\Delta S_t = S_t - S_{t-1}$ and $\Delta S_{l,t} = S_{l,t} - S_{l,t-1}$ with arbitrary, $t > \tau_0$, $\Delta S_{\tau_0} := 0$, $\tau_0 := 0$.

Definition (Equilibrium & Disequilibrium): A Cointelated pair $(S_t, S_{l,t})_{t>0}$, is said to be in *Disequilibrium* at time τ if $|S_\tau - S_{l,\tau}| > 0$. Likewise, the same pair is said to be in *Equilibrium* if $S_\tau - S_{l,\tau} = 0$.

In the case in which our cointelated pair's τ_0 is unknown or two cointelated pairs, $(S_t, S_{l,t})_{t>0}$ started in *disequilibrium* (e.g. $|S_{\tau_0} - S_{l,\tau_0}| > 0$), then it is primordial to start the strategy in a case where they are in *equilibrium*.

Definition (Cointelated Pairs' Phase Corrector): The *Phase Corrector* of two cointelated pairs is a constant $c \in \mathbb{R}$ which maximize the number of times, n^* , the cointelated pairs cross paths:

$$n^* = \arg \max_c \hat{\Gamma}_{S_{\tau+c}, S_{l,\tau+c}}(\theta, L), \quad (2.64)$$

where $\hat{\Gamma}_{S_{\tau+c}, S_{l,\tau+c}}(\theta, L)$ is the number of times S and S_l cross paths.

2.4.9 Results

We have performed few simulations including the one in figures 2.12. More specifically in figure 2.12(a) we have simulated two cointelated paths. We have also plotted the cumulative P&L from both the Financial Mathematics methodology in 2.12(b) and the clustering methodology in 2.12(c). We can see that the ML methodology does better in that one example (e.g., 1737 vs 1888). However, because one example does not really establish superiority, we have performed several simulations and have gathered the results in figure 2.13. The histogram seem to indicate that the ML methodology does marginally better than the FM method, though the latter methodology, when doing better does so more significantly (e.g., the left tail of the distribution in figure 2.13 is fatter than the right tail). From histogram of performance in Figures 2.13 and 2.14 we have concluded that we have the following rankings for the approaches for our Cointelated pairs²³

$$SC < ML_{LS} < FM < ML.$$

We have also performed additional simulation which objective it to show how the signal decomposition works. Indeed figures 2.16 and 2.15 illustrate that the addition

²³For parameters $\mu = 0.05, \sigma = 0.17, \sigma_l = 0.16, \theta = 0.1, \rho = -0.6$ of cointelation model (2.5).

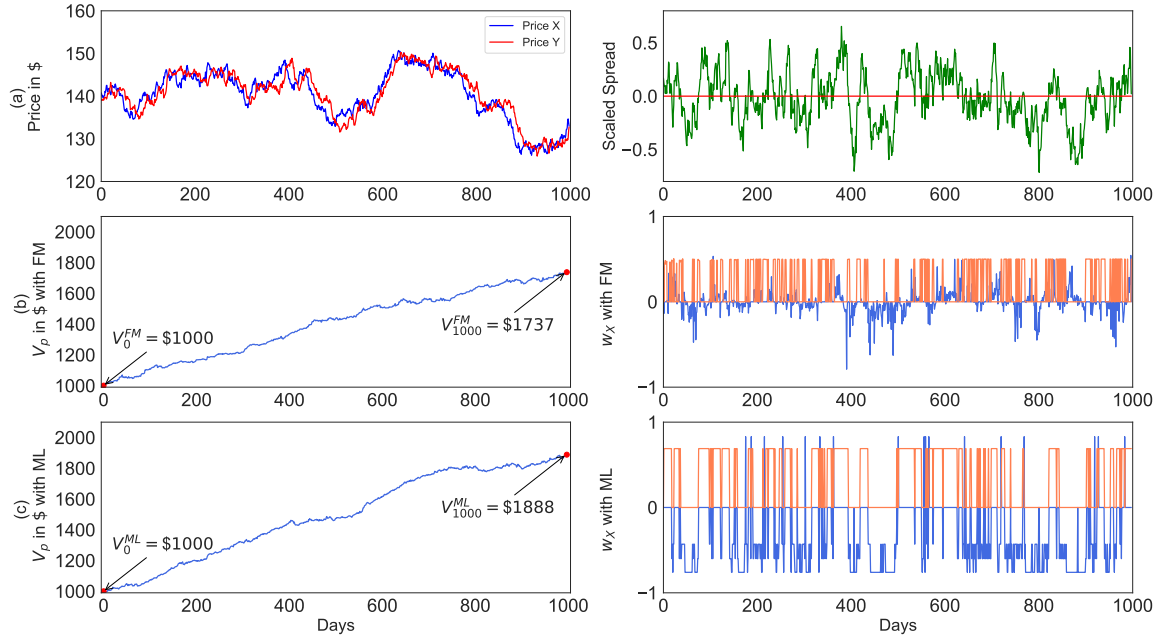


Figure 2.12: (a) one simulated scenario based on cointelation model (2.5) with parameters: $\mu = 0.05, \sigma = 0.17, \sigma_l = 0.16, \theta = 0.1, \rho = -0.6$ and scaled spread: $\theta(S_t - S_{l,t})$; (b) portfolio return and optimal weight using Dynamic Switching approach; (c) portfolio return and optimal weight using Machine Learning approach.

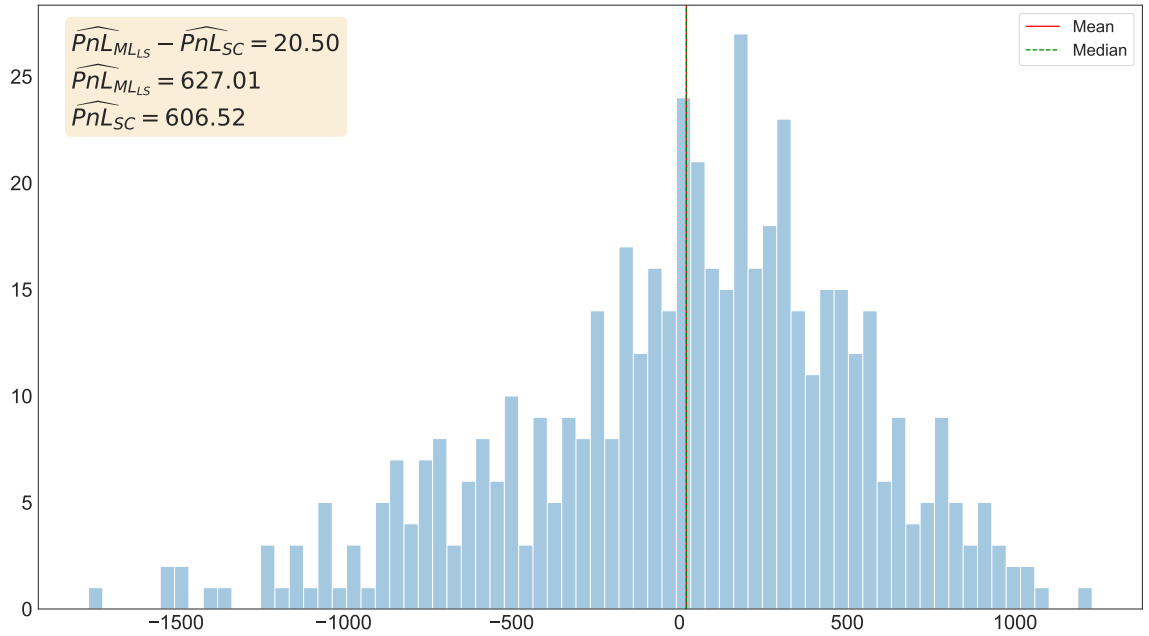


Figure 2.13: Histogram of excess (P&L) for ML_{LS} vs SC at terminal time T .

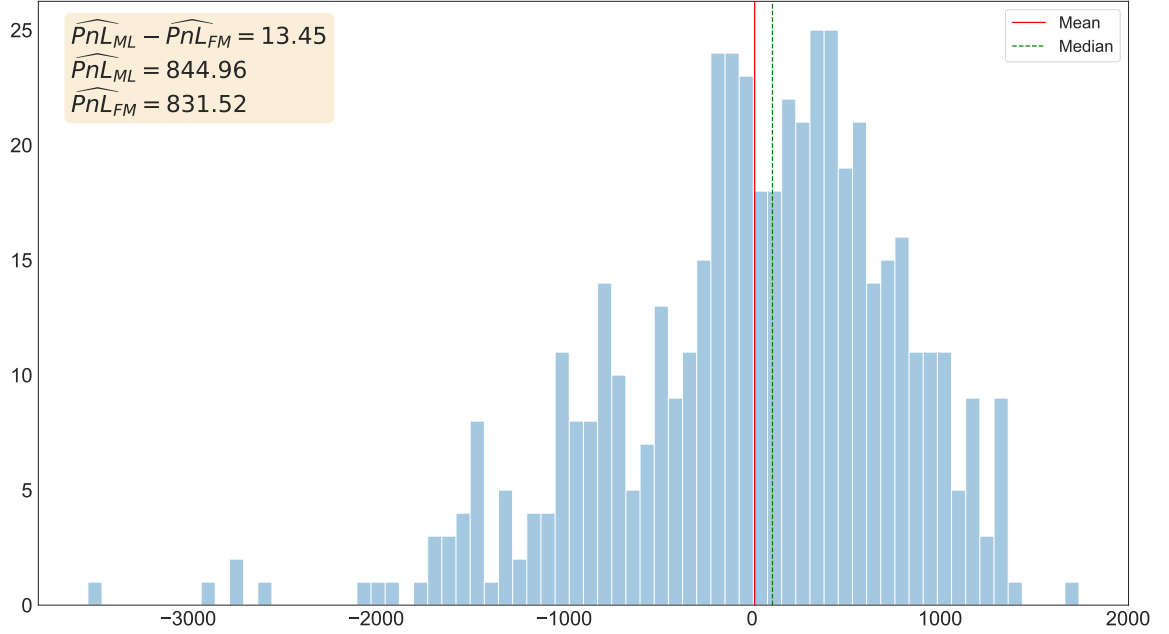


Figure 2.14: Histogram of excess ($P\&L$) for ML vs FM at terminal time T .

of the optimal strategy corresponding to stochastic control problem and the optimal strategy corresponding to Markowitz problem enhances our portfolio performance. More specifically, for figure 2.15 we see that the $SR = 1.98$ for the mean reverting centered strategy. We see that the $SR = 1.90$ for the Markowitz centered strategy. But we see that the overall strategy $SR = 2.71$. We have studied the cointelation model in the context of a dual portfolio optimization problem using three approaches centered around the idea of switching the weights of our portfolio according to a couple of strategies evolving in parallel. We labelled these strategies the Markowitz and OU centered strategies with a performance measure ranging from the MVC to power utility maximization. We first implemented what is considered a classic Financial Mathematics methodology using optimal control. The second methodology implemented was one using what is considered a Machine Learning approach using clustering. We found that these two methodologies lack efficiency in two very different ways. The First methodology is hampered in its resolution because the PDE of equation (2.51) does not have a closed form solution²⁴. On the other hand, the machine learning approach is easier to implement since we can bypass the complex SDE calibration issues and it also provides flexibility to regime change. However

²⁴It was noted during the viva [94] that there exists numerical methods in low dimension, usually associated to the domain of Financial Mathematics, that can approximate the PDE in question.

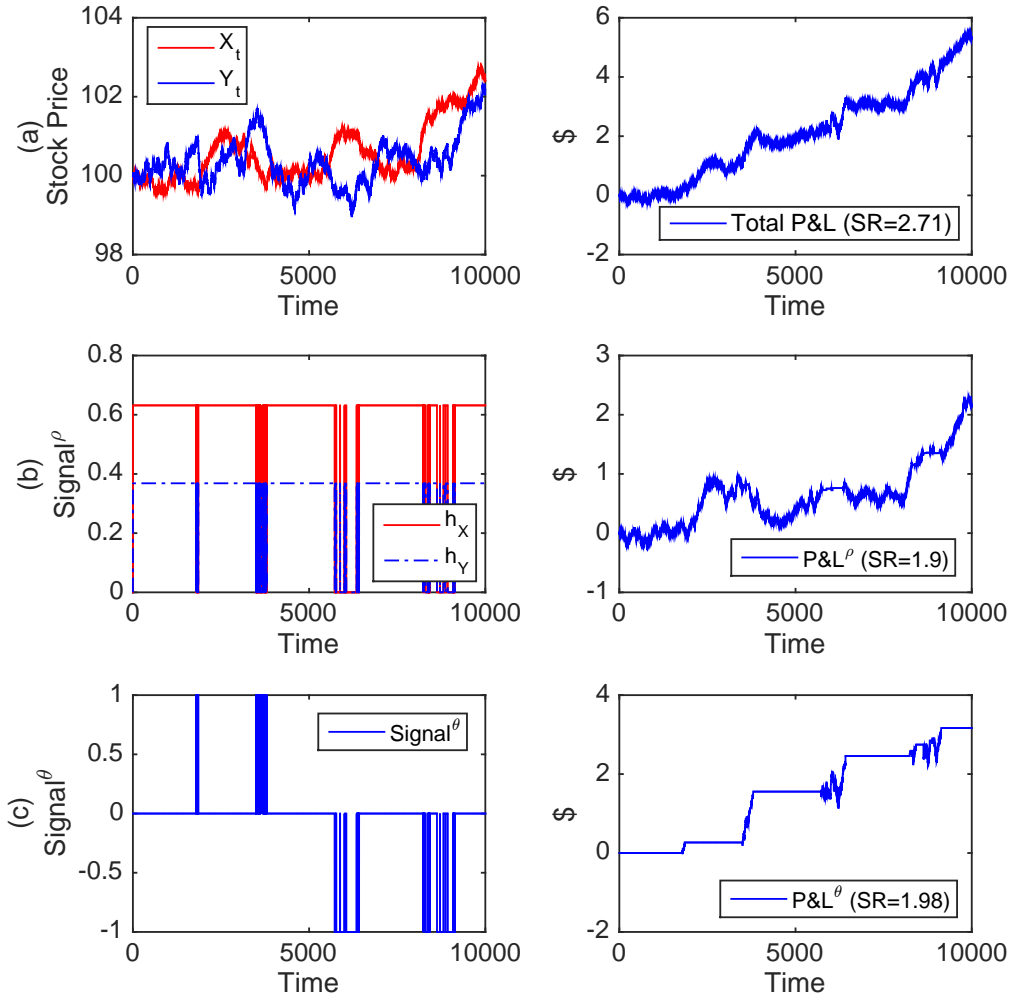


Figure 2.15: (a) Simulated Cointelated Pairs S_t and $S_{l,t}$ (left) and the combined cumulative P&L (right); (b) ρ -centered strategy signals (left) and strategy specific cumulative P&L (right); (c) θ -centered strategy signals (left) and strategy specific cumulative P&L (right).

one small issue with this Clustering methodology is its inability to handle boundary restrictions²⁵. As a result of the limiting aspects of these two methodologies, a third (hybrid) one was implemented. More specifically we started deploying this third methodology the same way we started the first one but instead implementing the DGM method to approximate a solution to our PDE. More specifically through a series of mathematical transforms (see Appendix page 240) where we go from a PDE in three dimension to one with only two, and though we may lose the financial intuition, we can find a solution for our PDE.

²⁵Though a sampling rejection method could theoretically be enforced.

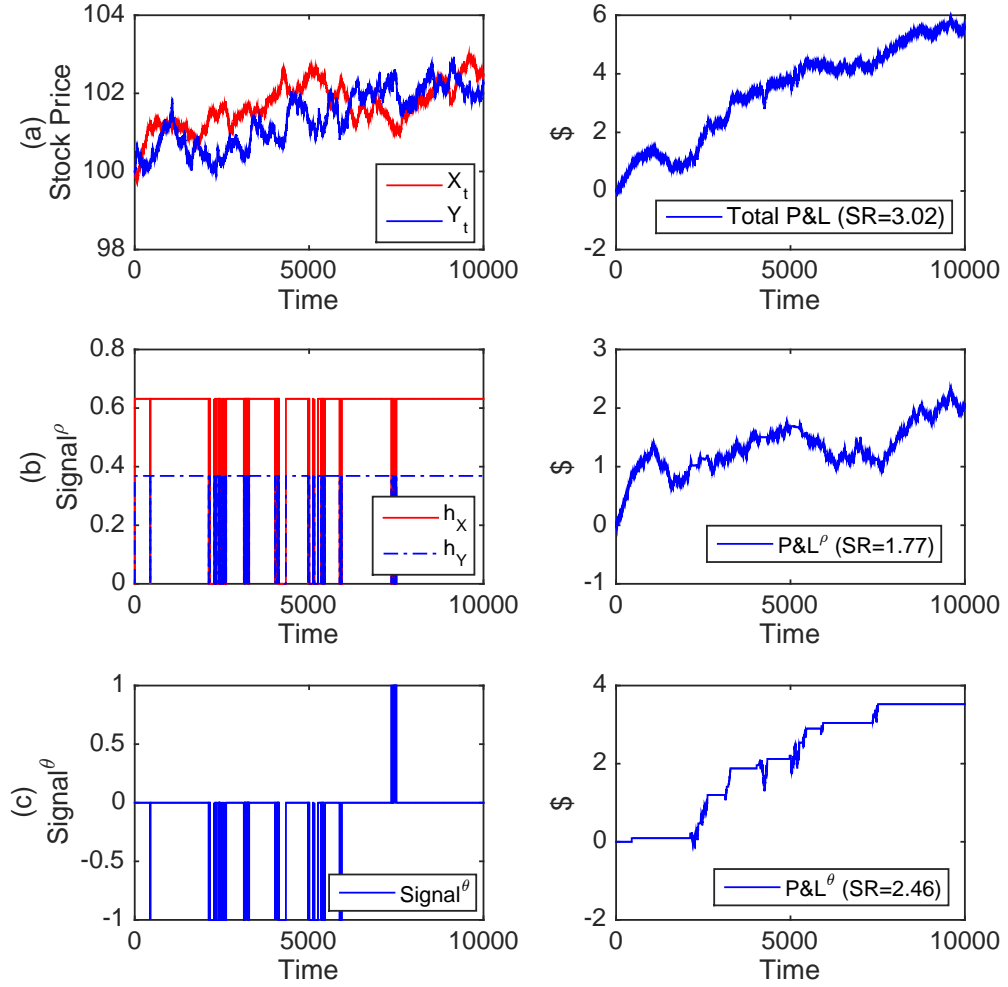


Figure 2.16: (a) Simulated Cointelated Pairs S_t and $S_{l,t}$ (left) and the combined cumulative P&L (right); (b) ρ -centered strategy signals (left) and strategy specific cumulative P&L (right); (c) θ -centered strategy signals (left) and strategy specific cumulative P&L (right).

Remark We know that covariance and correlation are related to each other through the following simple relationship: $\rho_{XY} = \sigma_{XY}/(\sigma_X\sigma_Y)$. In this chapter we introduced the concept of inferred correlation. In the spirit of adding levels of abstraction incrementally, and more specifically generalizing the study of “perceived correlation”, we propose to study aspects of implied covariance in the next Chapter.

Chapter 3

Anomaly Detection & Volatility Surface de-Arbitraging

This chapter explores issues around the Black-Scholes-Merton model, one of the most celebrated models in Financial Mathematics. We first explore some of its history in Section 3.1 as well as closely related models. We also show how the Implied Volatility Surface (IVS) was constructed as an add on to “fix” the model. We then move onto how to fetch and normalize the IVS from the normalized contracts in Section 3.2. We also explore some issues around arbitrage conditions in Sections 3.3 and 3.4. More specifically we study the history of the misunderstanding which led some of the literature to misinterpret a shortcut associated to arbitrage conditions on the strike space. We propose a simple solution in Section 3.5 to make the model more robust in situations where data is asynchronous. This chapter serves as an introduction to our next chapter.

3.1 Vanilla Options Model

3.1.1 A Notorious History

Vanilla options are the building blocks of the derivative markets. For instance, the celebrated Black-Scholes-Merton (BSM) [46], designed for the equities market, provides the mathematical weaponry to price European and American-style options. Proposed by Fischer Black and Myron Scholes, and then further developed by Robert Merton, it is one of the few Financial Mathematics model that was awarded the most prestigious award in Economics. The model was in fact already partially developed by Harold Bierman in the previous decade [95], though it was perhaps incomplete and less rigorous in its pricing but the (seed) idea was already there in 1967. Ever since the idea, the model has steadily improved [49, 50, 12] though still referred to as having

ambiguity today [96]. Nevertheless, starting with the bankruptcy of their own hedge fund¹, the BSM had significant impacts on the economy as it allowed leveraging². More specifically a call option, for example, allows, against relatively small fees, to gain exposure to substantial positive gains. This is arguably one of the reasons why Long-Term Capital Portfolio collapsed in the late 1990s and also why we had the latest subprime crisis³. Despite a promising start (21%, 43% and 41% return, respectively in the first three years 1995-1997), in 1998 LTC lost \$4.6 billion in less than four months following the 1997 Asian financial crisis and the 1998 Russian financial crisis. There are multiple reasons as to why the fund did not work as anticipated. The one that is most relevant to this chapter is associated with the assumption that volatility ought to be constant in the BSM formula. This led to the development of the implied volatility surface model [49, 50, 51]. The BSM can take three main forms depending on the underlying diffusion. First, the Log-Normal diffusion [46], secondly the Normal Assumption [97] (which became fashionable in the rates market in 2016 due to interest rates going negative) and finally the Garman-Kohlhagen model for the FX market [98], which formalizes the log-normal diffusion as a ratio of log-normal diffusions. We summarize in the next three subsections these variations of the model and explain in Section 3.1.5 how to fetch⁴ information on the implied volatility surface (IVS) from the observed prices process.

3.1.2 Log-Normal Assumption

Axiom 3.1.1 *Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ and \mathbb{Q} , our risk neutral probability measure, under which the stock price process $(S_t)_{t \geq 0}$ is considered a martingale. We consider this process to be log-normal if:*

$$dS_t/S_t = (r - q)dt + \sigma dW_t, \quad (3.1)$$

where r is the interest rate, q the dividend yield, σ the volatility and $W_t \sim N(0, t)$.

¹Except Fischer Black who was battling cancer during the same time, the remaining two authors own equity in “Long-Term Capital Management”.

²The concept of leveraging was arguably started with the creation of loans but options leveraging model is at another scale.

³Few believe that the last crisis was a liquidity crisis, or a correlation crisis: liquidity, correlation and leveraging, despite describing different QF concepts, all however share the fact that they are based on wrong and dangerous assumptions made on the mathematical models.

⁴Fetching is usually a jargon in Quantitative finance which means in this context transforming the price domain of a vanilla option into the implied volatility domain.

Lemma 3.1.1 *Following axiom 3.1.1, the process $(S_t)_{t \geq 0}$ under the risk neutral probability measure \mathbb{Q} is given by:*

$$S_T = S_t e^{\left(r - q + \frac{\sigma^2}{2}\right)(T-t) + \sigma W_{T-t}}. \quad (3.2)$$

Proof A straight forward application of Ito's lemma to $S_T = S_t e^{\left(r - q + \frac{\sigma^2}{2}\right)(T-t) + \sigma W_{T-t}}$ we obtain $dS_t/S_t = (r - q)dt + \sigma dW_t$. ■

Theorem 3.1.2 *Following axiom 3.1.1 and lemma 3.1.1, the price of a European Call options is given by*

$$C(S, t) = e^{-r(T-t)} [F_t N(d_1) - K N(d_2)], \quad (3.3a)$$

$$d_1 = \frac{1}{\sigma \sqrt{T-t}} \left[\ln \left(\frac{S_t}{K} \right) + \left(r - q + \frac{1}{2} \sigma^2 \right) (T-t) \right], \quad (3.3b)$$

$$d_2 = d_1 - \sigma \sqrt{T-t}, \quad (3.3c)$$

where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution, $T - t$ the time to maturity, S_t the spot price of the underlying asset, F_t the forward price, K the strike price, r the risk free rate, q the dividend yield and σ the volatility of returns of the underlying asset.

Proof A Call is the right but not the obligation to buy an asset S in the future that is $\max(S_t - K, 0) = (S_t - K)^+$. Therefore the price of a European Call is given by $C(S, t) = e^{-r(T-t)} \mathbb{E}^Q[S_T - K]^+ = e^{-r(T-t)} \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (S_T - K) 1_{\{S_T > K\}} e^{\frac{-x^2}{2\sigma^2}} dx$ where \mathbb{E}^Q is the expectation under the risk neutral probability and 1 the indicator function. We note that $F_t = S_t e^{-r(T-t)}$. We can also note that $S_T = S_t e^{\left(r - q + \frac{\sigma^2}{2}\right)(T-t) + \sigma \sqrt{T-t}x}$ where $x \sim N(0, 1)$. This implies that $S_T > K \Leftrightarrow x < \frac{1}{\sigma \sqrt{T-t}} \left[\ln \left(\frac{S_t}{K} \right) + \left(r - q + \frac{1}{2} \sigma^2 \right) (T-t) \right] = -d_2$. We can remove the indicator function and adjust the bounds of the integral function to obtain $C(S, t) = e^{-r(T-t)} \frac{1}{\sqrt{2\pi\sigma}} \int_{-d_2}^{\infty} (S_T - K) e^{\frac{-x^2}{2\sigma^2}} dx$. We set $\tilde{x} = x - \sigma \sqrt{t}$, we get $d\tilde{x} = dx$, $\tilde{x} = -d_2 - \sigma \sqrt{t} = -d_1$, also noticing that $1 - N(-d_2) = N(d_2)$ we obtain equation (3.3). ■

3.1.3 Normality Assumption

Usually, the least known of the three related models introduced in this section, and only popularized recently because of the arrival of negative rates on many markets, the pricing methodology using normal diffusion is given equation (3.6). Formally we have

Axiom 3.1.2 Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ and \mathbb{Q} , our risk neutral probability measure, under which the stock price process $(S_t)_{t \geq 0}$ is considered a martingale. We consider this process to be log-normal if:

$$dS_t = rdt + \sigma dW_t \quad (3.4)$$

where σ is the volatility and $W_t \sim N(0, t)$.

Lemma 3.1.3 Following axiom 3.1.2, the process $(S_t)_{t \geq 0}$ under the risk neutral probability measure \mathbb{Q} is given by:

$$S_t = S_0 + rt + \sigma \int_0^t dW_u \quad (3.5)$$

Proof $dS_t = rdt + \sigma dW_t \Leftrightarrow \int_0^t dS_u = \int_0^t rdu + \int_0^t \sigma dW_u \Leftrightarrow S_t - S_0 = r(t - 0) + \sigma \int_0^t dW_u$. This leads to $S_t = S_0 + rt + \sigma\sqrt{t}x$ where $x \sim N(0, t)$. ■

Theorem 3.1.4 Following axiom 3.1.2 and lemma 3.1.3, the price of a Call options is given by equation (3.3).

$$C(S, t) = e^{-r(T-t)}[(F_t - K)N(d) + \sigma\sqrt{T-t}N'(d)], \quad (3.6a)$$

$$d = \frac{F_t - K}{\sigma\sqrt{T-t}}, \quad (3.6b)$$

where the variables are the same as in proof of Subsection 3.1.2 and where $N'(\cdot)$ is the derivative of $N(\cdot)$.

Proof $C(S, t) = e^{-r(T-t)}E^Q[S_T - K]^+ = e^{-r(T-t)}\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}(S_T - K)1_{\{S_T > K\}}e^{\frac{-x^2}{2\sigma}}dx$ where E^Q is the expectation under the risk neutral probability and 1 the indicator function. We can also note that that $F_t = S_0 + rt$ so $S_t - K > 0 \Leftrightarrow S_0 + rt - K + \sigma\sqrt{t}x > 0$. By the symmetry principle $\Leftrightarrow \frac{F_t - K}{\sigma\sqrt{t}} < x$ we obtain equation (3.6). ■

3.1.4 Garman-Kohlhagen Model

The Garman-Kohlhagen (GK) model [98] is commonly used for the FX market.

Axiom 3.1.3 Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ and \mathbb{Q} , our risk neutral probability measure, under which the stock price process $(S_t)_{t \geq 0}$ is considered a martingale. We consider this process to be log-normal if:

$$dS_t/S_t = (r_d - r_f)dt + \sigma dW_t, \quad (3.7)$$

where r_d and r_f are the domestic and foreign interest rate, σ the volatility and $W_t \sim N(0, t)$.

Lemma 3.1.5 *Following axiom 3.1.3, the process $(S_t)_{t \geq 0}$ under the risk neutral probability measure \mathbb{Q} is given by:*

$$S_T = S_t e^{\left(r_d - r_f + \frac{\sigma^2}{2}\right)(T-t) + \sigma W_{T-t}} \quad (3.8)$$

Proof A straight forward application of Ito's lemma to $S_T = S_t e^{\left(r_d - r_f + \frac{\sigma^2}{2}\right)(T-t) + \sigma W_{T-t}}$ we obtain $dS_t/S_t = (r_d - r_f)dt + \sigma dW_t$. ■

Theorem 3.1.6 *Following axiom 3.1.3 and lemma 3.1.5, the price of a European Call options is given by*

$$C = S_0 e^{-r_f T} N(d_1) - K e^{-r_d T} N(d_2), \quad (3.9a)$$

$$d_1 = \frac{\ln(S_0/K) + (r_d - r_f + \sigma^2/2)T}{\sigma\sqrt{T}}, \quad (3.9b)$$

$$d_2 = d_1 - \sigma\sqrt{T}, \quad (3.9c)$$

where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution, $T - t$ the time to maturity, S_t the spot price of the underlying asset, K the strike, r_d and r_f the domestic and the foreign free rates, and σ the volatility of returns of the underlying asset.

Proof See proof for theorem 3.1.2 and replace the parameters accordingly. ■

3.1.5 Implied Volatility from Options Prices

Because vanilla options models suppose constant⁵ volatility, an assumption that is mathematically convenient but ultimately not verified by data, these models need a correction [49, 50] to complete their limitations. This overlay is what is commonly known as the Implied Volatility Surface (IVS).

Remark There are many reasons why the BSM is still used but for the sake of making the reasons brief, we can put forward the argument of the Greeks being critical across all main banking functions: in Front Office⁶, Risk Management⁷ and Product Control⁸.

⁵Note that the issue around volatility not being constant was known decades before [94] the BSM was invented.

⁶Hedging is used on daily basis on options' desks.

⁷VaR methodologies using sensitivities are quite common.

⁸Clearing methodologies require live P&L and often sensitivities are used.

To reconcile the BSM equation with market prices, the only non-observable value is the volatility input value. We call implied volatility the geometrical 3D structure (e.g. figure 3.1 is an example⁹) which takes as input a tenor and a moneyness¹⁰ and returns the volatility value which reconciles the BSM equation to the market observable price. There exists several well known numerical methods for fetching the implied volatility surface from the options market prices. For instance the Bisection method, the Newton-Raphson method, the Secant method or the Brent algorithm (their ensemble algorithm) are few of these methods¹¹.

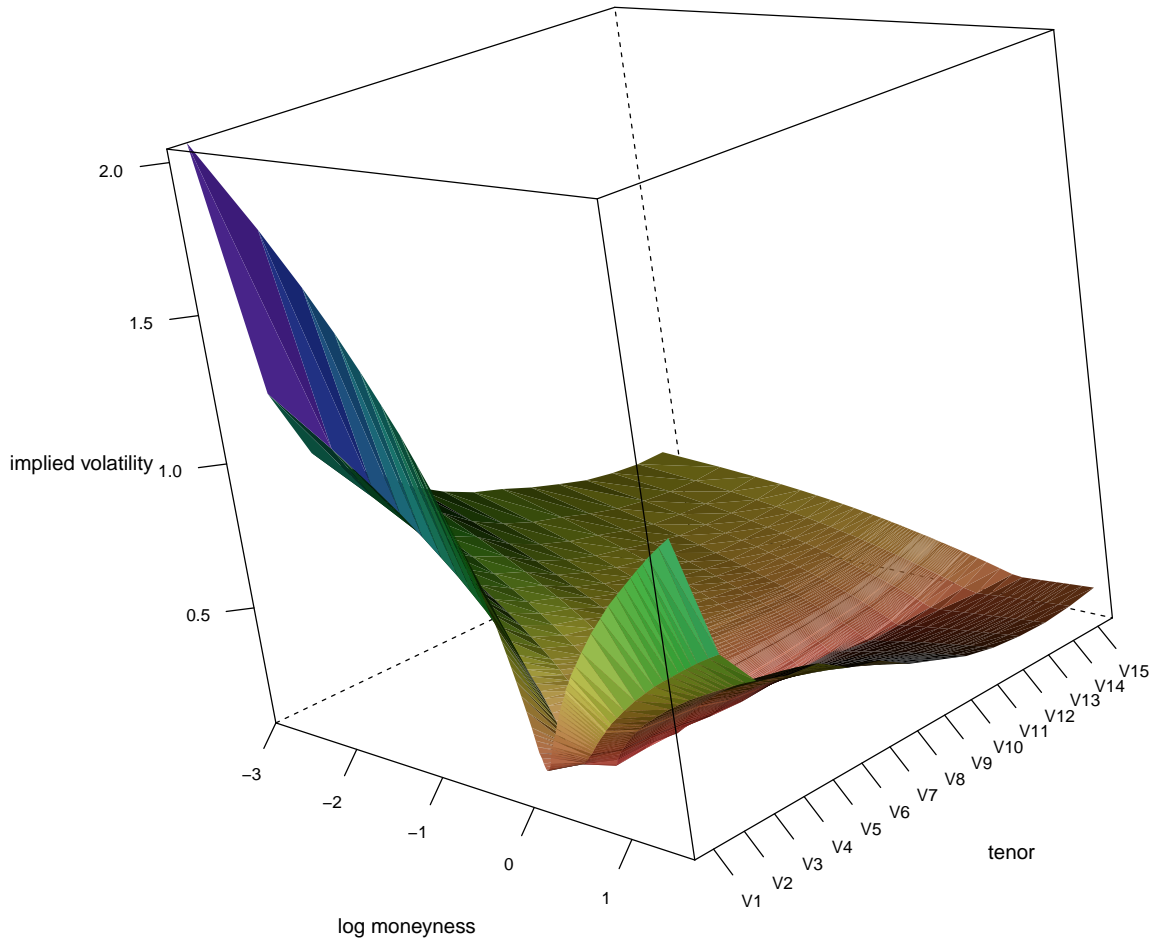


Figure 3.1: Arbitrage Free IVS with 15 standard tenors & 5 standard strikes [12]

⁹The data was taken from the equities market and the volatility was then stressed using the gSVI model to make it unique [12].

¹⁰or log-moneyness or delta space depending on which form of the BSM and which asset class we are dealing with.

¹¹We refer the reader to the classic literature [6] in that regard.

3.2 Normalizing Volatility Contracts

3.2.1 Context

The sub-prime crisis of 2008 had profound impacts on the financial industry overall. The resulting challenging regulatory environment pushed central counter party (CCP) clearing houses to come up with more sophisticated risk systems that do capture risk in a more intelligent fashion while providing capital requirement relief through cross margining at the same time. The Fundamental Review of the Trading Book (FRTB) requires the market participants to have trading systems in which liquidity is incorporated directly into the model. Also, now, the P&L explained must be more dynamic in order to address the enhanced sophistication and speed requirements of risk system. We have seen in the introduction the problem of normalizing rolling contracts and those associated with updating the Implied Volatility Surface (IVS).

3.2.2 Normalizing Listed Options Rolling Contracts

Before we discuss the solution of normalizing rolling contracts, we provide, a few definitions to make the Data Science easier.

Definition (Standardized Strikes): We call Standardized Strikes the set of industry accepted moneynesses which are used for the sake of storing data or for visualization. We denote these strikes by C^k (e.g. $-3, -2, -1, 0, 1$, which are arbitrary units in Figure 3.1).

Remark Different asset classes have different measures of moneyness. For example C^k is expressed in log-moneyness in equities, moneyness in rates and in delta space in FX. In the FX market, we have $C^k = \{10, 25, 50, 75, 90\}$ because the relevant exchanges list the options in this format.

Definition (Standardized Tenors): We call Standardized Tenors the set of industry accepted tenors which are used for the sake of storing data or for visualization. We denote these tenors by C^τ (e.g. v_1, v_2, \dots, v_{15} which are arbitrary units in Figure 3.1).

Remark Though industry standards may change, it is accepted currently, especially in the FX markets that there exist 10 standards tenors: overnight (ON), one week (1W), two weeks (2W), 1 month (1M), two months (2M), three months (3M), six months (6M), one year (1Y), one year and a half (18M) and two years (2Y).

Definition (Rolling Contracts): We call Rolling Contracts the sequence of predefined contracts that come in the market every month on a set day and which can be traded up to expiry.

Remark Let C^τ be the set of standardized pillars, C^k the set of standardized strikes and C^d the set of live contract expires. We denote by $\sigma_t(C_i^\tau, C_j^k)$ the implied volatility of the i th observed element of C^τ where $1 < i < |C^\tau|$ and j th element of C^k of C^d where $1 < j < |C^d|$.

Definition (Incomplete Standardized Strikes Data Set): Let us call $\tilde{C}^\tau \in C^\tau$ and $\tilde{C}^k \in C^k$ the set of incomplete data taken within the standardized strikes.

Definition (Interpolation vs Observable Data): We denote by $\hat{\sigma}_t(\vartheta, K)$ the linear interpolation in variance space of the implied volatility which is given by

$$\hat{\sigma}_t^2(\vartheta, C_j^k) = \frac{(C_{i+1}^\tau - \vartheta)\sigma_t^2(C_{i+1}^\tau, C_j^k) + (\vartheta - C_i^\tau)\sigma_t^2(C_i^\tau, C_j^k)}{C_{i+1}^\tau - C_i^\tau}, \quad (3.10)$$

where $\vartheta \in [C_i^\tau, C_{i+1}^\tau]$ and $1 < i < |C^\tau|$.

Remark The above definition does not include a definition of the edges of our implied volatility surface and also assumes that a perfect interpolation and extrapolation methodology already exist on the strike space.

3.2.3 Implied Volatility Surface Edges Problem

Figure 1.3 exposes the problem associated between the discrepancy that may occur on days when there is no roll between standardized pillars and the rolling contracts. More specifically let Υ be the time at which the contract rolls, as defined by equation (3.11a), then the longest tenor proxy can be better approximated by equation (3.11b) and the shortest by equation (3.11c).

$$\Upsilon = \inf\{t | \sigma_t^2(C_1^\tau, C_j^k) = \sigma_t^2(C_1^d, C_j^k)\}, \quad (3.11a)$$

$$\hat{\sigma}_t^2(\vartheta, C_j^k) = \frac{\sigma_\Upsilon^2(C_{i+1}^\tau, C_j^k)}{\sigma_\Upsilon^2(C_i^\tau, C_j^k)} \hat{\sigma}_t^2(C_i^\tau, C_j^k), \quad (3.11b)$$

$$\hat{\sigma}_t^2(\vartheta, C_j^k) = \frac{\sigma_\Upsilon^2(C_i^\tau, C_j^k)}{\sigma_\Upsilon^2(C_{i+1}^\tau, C_j^k)} \hat{\sigma}_t^2(C_{i+1}^\tau, C_j^k), \quad (3.11c)$$

with $i = |C^\tau| - 1$ for equation (3.11b) and $i = 1$ for equation (3.11c).

Definition (Normalized tenors) Without loss of generality we refrain from using the various symbols defined in this subsection and assume, unless otherwise specified, that throughout the thesis the implied volatility tenors are those obtained from the normalized tenors as opposed those from specific contracts.

The problem of updating implied volatility for the OTC market would require several different methodologies. More specifically, these different methodologies are designed for listed markets or over-the-counter (OTC).

3.3 Arbitrage Condition on the Strike

Axiom 3.3.1 *Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ and \mathbb{Q} , our risk neutral probability measure, under which the stock price process $(S_t)_{t \geq 0}$ is considered a martingale. We consider this process to be log-normal if:*

$$dS_t/S_t = \mu dt + \sigma dW_t, \quad (3.12)$$

where μ is the interest rate, σ the volatility and $W_t \sim N(0, t)$.

Lemma 3.3.1 *Following axiom 3.3.1, the process $(S_t)_{t \geq 0}$ under the risk neutral probability measure \mathbb{Q} is given by:*

$$S_T = S_t e^{(\mu + \frac{\sigma^2}{2})(T-t) + \sigma W_{T-t}}. \quad (3.13)$$

Proof A straight forward application of Ito's lemma to $S_T = S_t e^{(\mu + \frac{\sigma^2}{2})(T-t) + \sigma W_{T-t}}$ we obtain $dS_t/S_t = \mu dt + \sigma dW_t$. ■

In order to prevent arbitrages on the volatility surface we will start from basic principles and derive the constraints relevant to the strike and tenor.

3.3.1 From First Principles

Theorem 3.3.2 *Following the axiom 3.3.1 and lemma 3.3.1 a Call $C(\cdot)$ will be arbitrage free on the strike axis if and only if*

$$\frac{\partial^2 C}{\partial K^2} = \phi(S_T, T) > 0 \quad (3.14)$$

Proof Using Dupire's work [49, 50], we can write the price of a call according to equation $C(S_0, K, T) = e^{-rT} \mathbb{E}^{\mathbb{Q}}[S_T - K]^+ = e^{-rT} \int_K^{+\infty} (S_T - K) \phi(S_T, T) dS_T$ where $\phi(S_T, T)$ is the probability density of the call at time T . Let us differentiate $C(\cdot)$

twice with respect to K . We know $C(S_0, K, T) = e^{-rT} \mathbb{E}^\mathbb{Q}[S_T - K]^+ = e^{-rT} \int_K^{+\infty} (S_T - K) \phi(S_T, T) dS_T$. Now $\frac{\partial C}{\partial K} = -e^{-rT} \int_K^{+\infty} \phi(S_T, T) dS_T = -e^{-rT} \mathbb{E}(S_T > K)$. Also, we know that $0 \leq -e^{-rT} \frac{\partial C}{\partial K} \leq 1$. Differentiating a second time and setting $r = 0$ we find $\phi(S_T, T) = \frac{\partial^2 C}{\partial K^2}$. ■

Note that using numerical approximations we obtain equation (3.15)

$$\forall \Delta, C(K - \Delta) - 2C(K) + C(K + \Delta) > 0, \quad (3.15)$$

which is known in the industry as the arbitrage constraint of the positivity of the butterfly spread [99]. The way to see this is to note that given that the probability density must be positive we have $\frac{\partial^2 C}{\partial K^2} \geq 0$, using numerical approximation, we obtain $\frac{\partial^2 C}{\partial K^2} = \lim_{\Delta \rightarrow 0} \frac{[C(K-\Delta) - C(K)] - [C(K) - C(K+\Delta)]}{\Delta^2} = \lim_{\Delta \rightarrow 0} \frac{C(K-\Delta) - 2C(K) + C(K+\Delta)}{\Delta^2}$, therefore $C(K - \Delta) - 2C(K) + C(K + \Delta) \geq 0$.

Remark Gatheral and Jacquier [100] proved that the positivity of the butterfly condition comes back to making sure that the function $g(\cdot)$ from equation (3.16) is strictly positive.

$$g(k) := \left(1 - \frac{Kw'(k)}{2w(k)}\right)^2 - \frac{w'(k)^2}{4} \left(\frac{1}{w(k)} + \frac{1}{4} + \frac{w''(k)}{2}\right). \quad (3.16)$$

The idea from where equation (3.16) comes from is related to equation (3.14): $\frac{\partial^2 C}{\partial K^2} = \phi(\cdot)$. More specifically, applying this formula to the Black-Scholes equation leads to, for a given tenor

$$\phi(k) = \frac{g(k)}{\sqrt{2\pi w(k)}} \exp\left(-\frac{d_2(k)^2}{2}\right), \quad (3.17)$$

where $w(k, t) = \sigma_{BS}^2(k, t)t$ is the implied volatility at strike K and where $d_2(k) := \frac{-k}{\sqrt{w(k)}} - \sqrt{w(k)}$.

3.3.2 Necessary but not Sufficient Condition

The literature around the butterfly condition is not confined to equation (3.14). More specifically we have the below necessary but **not sufficient** condition which has created confusion in the literature. We explore this point next.

Axiom 3.3.2 Assuming $r = 0$, let us define the Black-Scholes call function $f : \mathbb{R} \times [0, \infty) \rightarrow [0, 1)$ in terms of the tail of the standard Gaussian distribution $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp(-\frac{y^2}{2}) dy$ and given by:

$$f(k, \nu) = \begin{cases} \Phi\left(\frac{k}{\sqrt{\nu}} - \frac{\sqrt{\nu}}{2}\right) - e^k \Phi\left(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}\right), & \text{if } \nu > 0, \\ (1 + e^k)^+, & \text{if } \nu = 0. \end{cases}$$

Note that the assumption, the proof and its choice of boundaries¹² is attributed to Rogers and Tehranchi [57]. Please refer to the original paper [57] if clarifications are required.

Proposition 3.3.3 *Following axiom 3.3.2 we have*

$$|T\partial_K\sigma^2(K, T)| \leq 4, \forall K, \forall T, \quad (3.18)$$

Proof Let us call $V_t(k, \tau)$ the implied variance at time $t \geq 0$ for log-moneyness k and time to maturity $\tau \geq 0$. Let us now label our Kappa¹³ and Vega¹⁴, with the convention that $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$:

$$f_k(k, \nu) = -e^k \Phi\left(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}\right), \quad (3.19a)$$

$$f_\nu(k, \nu) = \phi\left(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}\right) / 2\sqrt{\nu} \quad (3.19b)$$

Now define the function $I : \{(k, c) \in \mathbb{R} \times [0, \infty) : (1 + e^k)^+ \leq c < 1\} \longrightarrow [0, 1)$ implicitly by the formula: $f(k, I(k, c)) = c$. Calculus gives $I_c = \frac{1}{f_\nu}$ and $I_k = -\frac{f_k}{f_\nu}$, from here using the chain rule, designating $\partial_{k+}V$ as the right derivative. We have $\partial_{k+}V = I_k + I_c \partial_k \mathbb{E}[(S_\tau - e^k)^+] = -\frac{f_k}{f_\nu} - \frac{\mathbb{P}(S_\tau > e^k)}{f_\nu} < -\frac{f_k}{f_\nu} = 2\sqrt{\nu} \frac{\Phi\left(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}\right)}{\phi\left(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}\right)}$. Now using the bounds of the Mills ratio $0 \leq 1 - \frac{x\Phi(x)}{\phi(x)} \equiv \varepsilon(x) \leq \frac{1}{1+x^2}$, we have: $\partial_{k+}V \leq \frac{4}{k/V+1} < 4$. Similarly, we can show [57] that $\partial_{k-}V > -4$, therefore we have $|\partial_k V| < 4$. ■

Though elegant and certainly potentially very useful, had it been a necessary **and sufficient** condition for the volatility surface to be arbitrage-free on the strike axis, equation (3.18) was ultimately proven incomplete, having in the meantime guided the literature for some time with famous researchers using the results [51] in their research.

Counter Example We explore below a counter example.

Axiom 3.3.3 *Let us call $\chi_R = \{a, b, \rho, m, \sigma\}$ a given parameter set for the function $f : \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^{+,*}$ given by*

$$f(k, \chi_R) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}] \quad (3.20)$$

¹²E.g.: $f : \mathbb{R} \times [0, \infty) \longrightarrow [0, 1)$.

¹³Derivative with respect to the strike: equation (3.19a).

¹⁴Derivative with respect to the volatility: equation (3.19b).

Lemma 3.3.4 *Following axiom 3.3.3*

$$b(1 + |\rho|) \leq 4/T, \forall K, \forall T, \quad (3.21)$$

Proof Simply take $\partial_K f(k, \chi_R) = b(1 + |\rho|)$. Now applying proposition 3.3.3 to $b(1 + |\rho|)$ we obtain equation (3.21). ■

A counter example on the SVI¹⁵ inequality of equation (3.21) was provided by Axel Vogt [25] in equation (3.22) which prompted a re-examination of the pillar assumptions [13].

$$(a, b, m, \rho, \sigma) = (0.0410, 0.1331, 0.3586, 0.3060, 0.4153). \quad (3.22)$$

For instance, figure 3.2 represents the counter example of $|T\partial_K\sigma^2(K, T)| \leq 4$ applied to the Raw SVI parametrization¹⁶. More specifically on the left hand side of

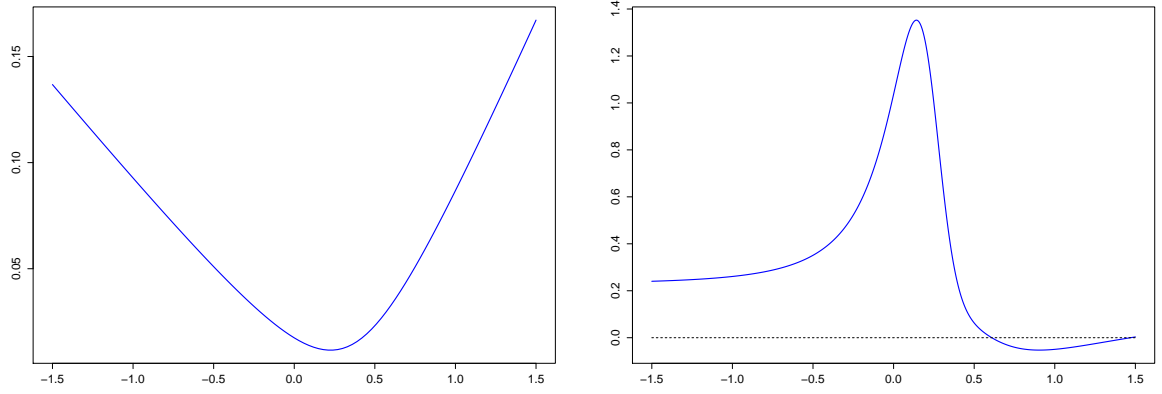


Figure 3.2: Vogt's [25] total variance example verifying $b(1 + |\rho|) \leq \frac{4}{T}$ (left figure: with the x axis being the log-moneyness and the y axis being the implied variance) and the corresponding $\partial_{K,K}^2 BS(\sigma^2(K, T))$ approximating the (supposed) always positive pdf (right figure: with the x axis being the log-moneyness and the y axis being the non normalized pdf).

Figure 3.2, we can observe a standard implied volatility plotted using the raw SVI equation (3.20) with Vogt's example [25]. The latter example verifies the constraint $|T\partial_K\sigma^2(K, T)| \leq 4$ yet, on the right hand side the PDF yields a negative value inferring an arbitrage. More specifically, this example when it is cross validated using first principles ($\partial_{K,K}^2 BS(\sigma^2(K, T)) > 0$, numerically approximated by equation (3.15)) which yields a negative value in a small portion of the unnormalized numerical PDF. Indeed, it can be clearly seen around moneyness of [0.6, 1.5] that the curve goes below

¹⁵Note that we introduce more formally this parametrization in section 4.1.4.1.

¹⁶Yielding $b(1 + |\rho|) \leq \frac{4}{T}$.

the $y = 0$ line yielding a butterfly arbitrage (sometimes a financial product that is understood to be an “unnormalized density”)¹⁷.

Remark If this is not clear why this is important, CCPs create “stress tests” based on historical scenarios. The latter scenarios need to be also coherent at the same time¹⁸.

3.4 Arbitrage Condition on the Tenor

Theorem 3.4.1 *Following the axiom 3.3.1 and lemma 3.3.1 a Call $C(\cdot)$ will be arbitrage free on the tenor axis if and only if*

$$C(K, T + \Delta) - C(Ke^{-r\Delta}, T) \geq 0. \quad (3.23)$$

Proof One application of Dupire’s formula [49, 50] is that the pseudo-probability density must satisfy the Fokker-Planck [101, 102] equation. This proof is taken from El Karoui [103]. Let us apply Itô to the semi-martingale. This is formally done by introducing the local time Λ_T^K :

$$\begin{aligned} e^{-r(T+\varepsilon)} (S_{T+\varepsilon} - K)^+ - e^{-rT} (S_T - K)^+ &= \int_T^{T+\varepsilon} re^{-ru} (S_u - K)^+ du \\ &\quad + \int_T^{T+\varepsilon} e^{-ru} 1_{\{S_u \geq K\}} dS_u + \frac{1}{2} \int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K. \end{aligned}$$

Local times are introduced in mathematics when the integrand is not smooth enough. Here the call price is not smooth enough around the strike level at expiry. Now we have: $E(e^{-ru} 1_{\{S_u \geq K\}} S_u) = C(u, K) + Ke^{-ru} P(S_u \geq K) = C(u, K) - K \frac{\partial C}{\partial K}(u, K)$. The term of the form $E\left(\int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K\right)$ is found due to the formula of local times, that is:

$$\begin{aligned} E\left(\int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K\right) &= \int_T^{T+\varepsilon} e^{-ru} du E(\Lambda_u^K) = \int_T^{T+\varepsilon} e^{-ru} du \sigma^2(u, K) K^2 \phi(u, K) \\ &= \int_T^{T+\varepsilon} \sigma^2(u, K) K^2 \frac{\partial^2 C}{\partial K^2}(u, K) du. \end{aligned}$$

Plugging these results back into the first equation we obtain:

$$\begin{aligned} C(T + \varepsilon, K) &= C(T, K) - \int_T^{T+\varepsilon} rC(u, K) du + (r - q) \int_T^{T+\varepsilon} \left(C(u, K) - K \frac{\partial C}{\partial K}(u, K)\right) du \\ &\quad + \frac{1}{2} \int_T^{T+\varepsilon} \sigma^2(u, K) K^2 \frac{\partial^2 C}{\partial K^2}(u, K) du. \end{aligned}$$

¹⁷If this is still not clear, please see section 3.3.

¹⁸Through de-arbitraging which we see in the next Section.

If we want to give a PDE point of view of this problem we can notice that $\phi(T, K) = e^{-rT} \frac{\partial^2 C}{\partial K^2}(T, K)$ verifies the dual forward equation:

$$\phi'_T(T, K) = \frac{1}{2} \frac{\partial^2 (\sigma^2(T, K) K^2 \phi(T, K))}{\partial K^2} - \frac{\partial^2 ((r - q) K \phi(T, K))}{\partial K}.$$

Integrating twice by parts, we find:

$$\frac{\partial e^{-rT} C(T, K)}{\partial T} = \frac{1}{2} \sigma^2(T, K) K^2 e^{rT} \frac{\partial^2 C(T, K)}{\partial K^2} - \int_K^{+\infty} (r - q) K e^{rT} \frac{\partial^2 C(u, K)}{\partial K^2} \partial K(T, K) du.$$

Now integrating by part again and setting dividends to 0 we find the generally admitted relationship:

$$\frac{\partial C}{\partial t} = \frac{\sigma^2}{2} K^2 \frac{\partial^2 C}{\partial K^2} - rK \frac{\partial C}{\partial K}$$

and therefore we have

$$\sigma = \sqrt{2 \frac{\frac{\partial C}{\partial t} + rK \frac{\partial C}{\partial K}}{K^2 \frac{\partial^2 C}{\partial K^2}}}.$$

From this formula and from the positivity constraint on Equation (3.14) we find that

$$\frac{\partial C}{\partial t} + rK \frac{\partial C}{\partial K} \geq 0.$$

Note that for very small Δ :

$$C(K e^{-r\Delta}, T) \approx C(K - Kr\Delta, T)$$

Using a Taylor expansion:

$$C(K - Kr\Delta, T) = C(K, T) - Kr\Delta \frac{\partial C}{\partial K} + \dots,$$

therefore

$$rK \frac{\partial C}{\partial K} \approx \frac{C(K, T) - C(K e^{-r\Delta}, T)}{\Delta}.$$

Using a forward difference approximation we also have:

$$\frac{\partial C}{\partial K} = \frac{C(K, T + \Delta) - C(K, T)}{\Delta}$$

and from the Fokker-Planck equation we have $\frac{\partial C}{\partial t} + rK \frac{\partial C}{\partial K} \geq 0$. Substituting, we obtain:

$$\frac{C(K, T + \Delta) - C(K, T)}{\Delta} + \frac{C(K, T) - C(K e^{-r\Delta}, T)}{\Delta} \geq 0.$$

Simplifying we find $C(K, T + \Delta) - C(K e^{-r\Delta}, T) \geq 0$. ■

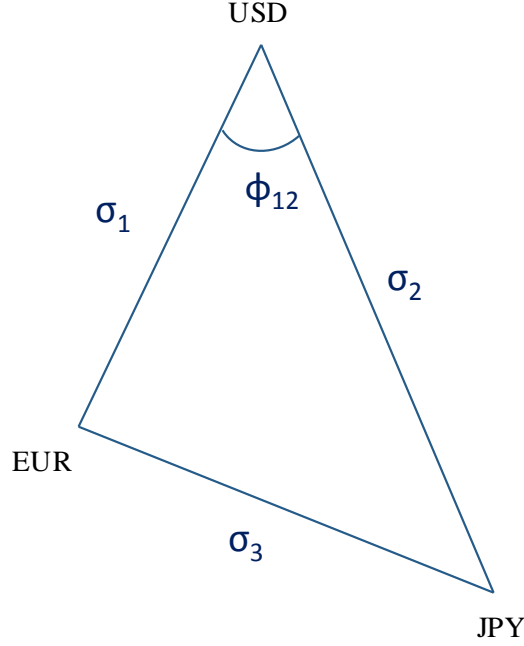


Figure 3.3: Geometrical representation for three FX pairs and their correlation.

Remark If you work in the FX market then there are some idiosyncratic properties of this specific market one can take advantage of [104]. For instance, let $S_{t,1}$ be the exchange rate of EUR/USD pair and σ_1 its implied volatility, $S_{t,2}$ the exchange rate of USD/JPY pair and σ_2 its implied volatility $S_{t,3}$ the exchange rate of EUR/JPY pair and σ_3 its implied volatility (see figure 3.3). Now notice that $S_{t,3}$ must equal to $S_{t,1} \times S_{t,2}$ else there are arbitrage opportunities induced. Geometrically we have

$$\sigma_1 + \sigma_2 + \sigma_3 > 2 \max(\sigma_1 + \sigma_2 + \sigma_3). \quad (3.24)$$

Therefore $\ln(S_{t,3}) = \ln(S_{t,1} \times S_{t,2})$. Taking the variance on each side we obtain the non-arbitrage condition on the volatility and the implied correlation given by

$$\sigma_3^2 = \sigma_2^2 + \sigma_1^2 + 2\rho_{1,2}\sigma_1\sigma_2.$$

By rearranging, the implied correlation can be isolated and given by

$$\rho_{1,2} = \frac{\sigma_3^2 - \sigma_2^2 - \sigma_1^2}{2\sigma_1\sigma_2} = \cos \phi_{1,2}, \quad (3.25a)$$

$$\text{or, } \arccos \rho_{1,2} = \phi_{1,2}, \quad (3.25b)$$

align Figures (3.3) and equation (3.24) shows a visual representation of this non-arbitrage constraint with $\phi_{1,2}$ representing the angle between σ_1 and σ_2 . The relationship between $\rho_{1,2}$ and $\phi_{1,2}$ is given by equation (3.25b).

3.5 De-arbitraging

As we have seen in Section 3.1.5, there exist several pricing methodologies. We denote by $C(\cdot)$ the relevant pricing methodology. But independently from the underlying diffusion model, for our de-arbitraging methodology, we make sure that for every pillar tenor and every pillar strikes the relevant points are mutually arbitrage free.

3.5.1 The General Idea

How do we algorithmically insure that a volatility surface is arbitrage free? We can simply do that with a nested for loop. In Figure 3.4 the x axis represents the strike space and the y axis the implied volatility space. A discretized version of the IVS de-arbitraging methodology would then consist of taking the initial volatility (for example the IV which has been stressed by one of the diffusions defined in Section 5) and loop over all points while adjusting each adjacent point that induce an arbitrage.

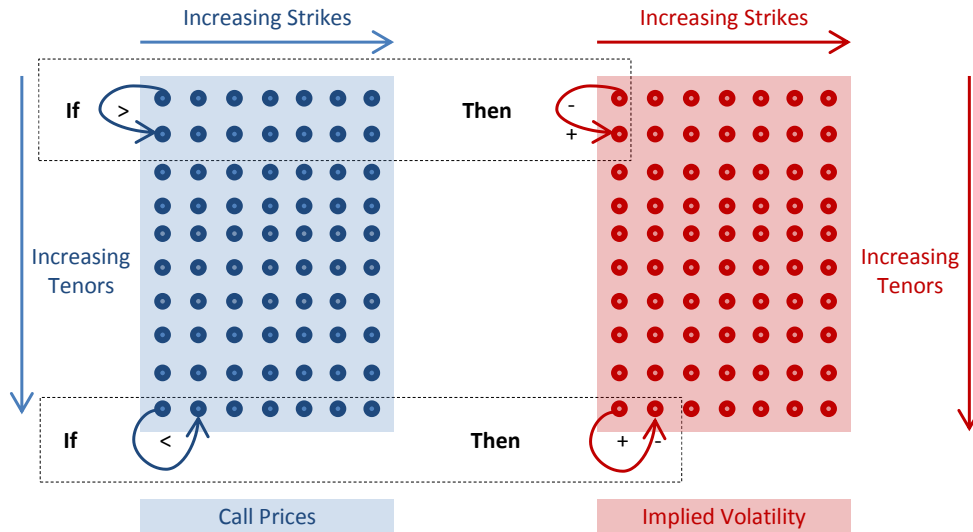


Figure 3.4: Visualization for the Core Simple de-Arbing Idea

Remark We are yet to find a volatility surface that is difficult to de-arbitrage. We have performed many simulations but we did not feel that presenting these simulations would bring any interesting insight.

3.5.2 The Case of the FX Asset Class

This intuitive representation of Figure 3.4 no longer works with the FX pillars like it is shown by Figure 3.5. Indeed, though the tenor pillars (ON, 1W, 2W, 1M, 2M, 3M, 6M, 1Y, 18M, 2Y) are not really providing additional challenges in the methodology, the fact that, on the strike axis, the market data is listed on the delta space (please see Subsection 3.1.4) creates additional complications. The classic de-arbing algorithms assumes that the data is conveniently aligned in log-moneyness space. We, however, know that $\Delta_f = \phi e^{-rft} N(\phi \frac{1}{2} \sigma \sqrt{t})$ so there is a way to turn the data in delta space of the FX market into log-moneyness but the delta to log-moneyness conversion creates increasing mis-alignments as the tenor increases and the convenient falling variance check no longer works. We take this opportunity to make a couple of remarks here.

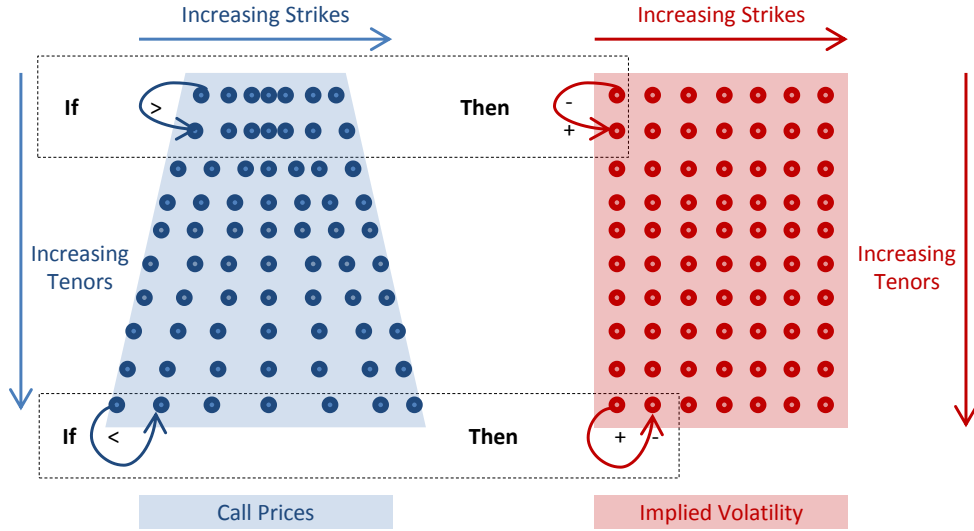


Figure 3.5: FX Strike Space Misalignment of Figure 3.4

Remark These Data Science related issues are of utmost importance. Great data on an average model will give you mediocre results, but bad data on an exceptional model will give you catastrophic results. The problem of normalization onto pillars and the process of making the result volatility surfaces arbitrage free is critical in all aspects of trading going from opportunity detection to risk management.

Remark Though, they have been recently studied formally [105] using Sinkhorn's algorithm, non-parametric de-arbitraging methodology are quite simplistic but stability and convergence are sometimes a problem [12]. This is because as one moves away from the arbitrage zone, one may recreate arbitrages on adjacent points.

We have explored, in this chapter, the three main forms of the vanilla options pricing models. We also went over the arbitrage conditions that make the IVS coherent. Though the main function of this chapter was to prepare the reader with respect to the next chapters, and therefore relatively poor in terms of original contribution, we believe to have clarified few historic mistakes that have created confusion in the literature. More specifically though we knew how to derive the strike and the tenor conditions from first principles, elegant shortcuts had emerged in the literature that were necessary but not sufficient conditions (which the literature had taken as sufficient). We also stressed the importance of using standardized pillars and have given a simple methodology to achieve this standardization. Finally, we made an intuitive introduction to de-arbitraging methodologies stressing the importance of making sure the de-arbitraging is performed on data which is of good quality and standardized appropriately. Generally speaking, because the volatility surface is a mathematical concept of infinite dimension, and because questions may arise in terms of Interpolation and Extrapolation techniques, parametrization approaches have been created to handle dimensionality reduction. As it happens, reducing the dimensionality, also helps the de-arbitraging algorithms to converge. We explore some of these methodologies in the next Chapter.

Chapter 4

Big Data Changing the Vanilla Options Landscape

In this chapter we go over few of the noticeable changes that have occurred in the area of implied volatility parametrization. First we review the literature in Section 4.1. More specifically we focus on the discrepancy between data as observed in the market and assumptions associated to the wings sub-linearity. We take this opportunity to introduce our own model which is a data and regulatory driven change from known models in Section 4.2. We propose our de-arbitraging methodology in Section 4.3. We spend the next two Sections (4.4 and 4.5) exploring data sciences related issues such as updating and proxying volatility data in a hostile environment. Finally, in Section 4.6 we discuss a few statistical arbitrage ideas.

4.1 Volatility Parametrization

4.1.1 Motivation

Volatility parametrization, a way to reduce the dimensionality of surfaces¹ may have very different motivations depending on ones function within a bank. For instance, if pricing related, methods associated to Partial Differential Equations (PDE) require convergence between the parametrization and the Heston model [59]. Within the same sell side, parametrization can be used for proxying (e.g. Section 4.6). On the buy side, parametrization can be used for statistical arbitrage as well as risk management. If the objective is more IT related, then parametrization can help in optimization. For instance, a volatility surface is free of arbitrage when Equations (3.15) and (3.23) are satisfied. If using a brute force method, for example moving

¹A mathematical concept of infinite dimension.

implied volatility points towards the arbitrage free frontier, can create arbitrages in other places of the volatility surface. Lagnado and Osher [106], Crepey [107] and other authors have proposed relevant minimization algorithms. However, the difficulties that can occur in handling de-arbitraging and dimensionality of a surface with (infinite) points create a natural need to limit the risk factors to a manageable level. Using volatility parametrization instead of a grid can reduce the degree of freedom and help in that optimization. These are among other aspects some of the motivations behind parametrizing the volatility surface.

4.1.2 Schonbucher's Model

In 1999 Schonbucher [108] introduced his parameterized version of the volatility surface. The main advantage of the Schonbucher's model is that it can be derived from Heston [61], a stochastic volatility model used by many financial institutions.

Axiom 4.1.1 *Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{Q})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ and \mathbb{Q} , our risk neutral probability measure, under which the stock price process $(S_t)_{t \geq 0}$ is considered a martingale. Formally we have*

$$\begin{cases} dS_t &= rS_t dt + \sigma_t S_t dW_t \\ d\sigma_{k,t} &= u_t dt + \gamma_t dW_t + \nu_t dW_t^\perp \\ \gamma_t^2 + \nu_t^2 &= 1 \end{cases}$$

where dW_t and dW_t^\perp are independent of each other. The implied stock volatility function $\sigma_{k,t}$ is yet to be specified, to simplify the notations let us redefine this implied volatility variable as I_t and let γ_t be the correlation between the instantaneous volatility and the spot price and ν_t chosen such that $\gamma_t^2 + \nu_t^2 = 1$ [108].

Schonbucher asserts that the implied volatility surface should be modeled through Equation (4.1).

Theorem 4.1.1 *Following axiom 4.1.1, the implied variance, $I^2(\cdot)$ converges asymptotically to*

$$I^2(x) = a + bx + \sqrt{\frac{1}{4}\sigma^4 + x^2\nu^2} \quad (4.1)$$

where x denotes the log moneyness².

²Log moneyness of an underlier S with a forward price F with respect to a strike K is $\ln \frac{K}{F}$.

Proof Apply Itô's lemma to the call price to obtain the drift restriction $rC^{BS} = \frac{\partial C^{BS}}{\partial t} + rS \frac{\partial C^{BS}}{\partial S} + u \frac{\partial C^{BS}}{\partial I} + \gamma \sigma S \frac{\partial^2 C^{BS}}{\partial I \partial S} + \frac{1}{2} \left(\sigma^2 S^2 \frac{\partial^2 C^{BS}}{\partial S^2} + \frac{\partial^2 C^{BS}}{\partial I^2} \right)$. Using the Black-Scholes formula for the call and its derivatives this reduces to a joint restriction on the implied and instantaneous volatility of equation (4.2).

$$Iu = \frac{1}{2(T-t)}(I^2 - \sigma^2) - \frac{1}{2}d_1d_2\nu^2 + \frac{d_2}{\sqrt{T-t}}\sigma\gamma, \quad (4.2a)$$

$$d_1 = \frac{x}{I\sqrt{T-t}} + \frac{1}{2}I\sqrt{T-t}, \quad (4.2b)$$

$$d_2 = d_1 - I\sqrt{T-t}, \quad (4.2c)$$

where we have used the standard definitions for d_1 and d_2 given by Equations (4.2b) and (4.2c). Equation (4.2a) blows up as $T-t$ goes to zero. This imposes the condition that $(I^2 - \sigma^2) - d_1d_2\nu^2(T-t) + d_2\sigma\gamma\sqrt{T-t} = O(T-t)$ and thus in the limit, noting that $\lim_{t \rightarrow T} d_1\sqrt{T-t} = \lim_{t \rightarrow T} d_2\sqrt{T-t} = \frac{1}{I} \ln(S/K)$ and by setting a zero correlation between spot and implied volatility³ we obtain $I^2(x) = \frac{1}{2}\sigma^2 + \sqrt{\frac{\sigma^4}{4} + x^2\nu^2}$. Note that there is no at the money (ATM) skew but this is easily remedied by adding an extra linear⁴ term which gives $I^2(x) = a + bx + \sqrt{\frac{\sigma^4}{4} + x^2\nu^2}$. We therefore find the original equation (4.1). ■

4.1.3 The SABR Model

The “Stochastic Alpha, Beta, Rho”, a stochastic version of the Constant Elasticity of Variance (CEV) model [109], commonly known as “SABR” model was developed in 2002 [110] by Patrick S. Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward. The SABR model's dynamics are given by

$$dF_t = \sigma_t F_t^\beta dW_t, \quad (4.3a)$$

$$d\sigma_t = \alpha \sigma_t dZ_t, \quad (4.3b)$$

with F being a forward⁵ on S , σ_t its volatility, $\langle dW_t, dZ_t \rangle = \rho dt$, $-1 < \rho < 1$, $0 \leq \beta \leq 1$ and $\alpha \geq 0$. Though advertised to be cross asset, the SABR model ended up being exclusively used on the rates market and unfortunately was made obsolete due to the rates market going negative in 2015. This latter fact was addressed recently [111].

³a simplification done by Schonbucher but not necessarily very realistic, specifically on the commodities market where for physical reasons the stochastic processes driving the commodities are more driven by mean reversion than classic correlation [10].

⁴The linearity of the wings was at that time justified [56] even-though later challenged [12]. We will develop this in details in Section 4.2.1.2.

⁵E.g., LIBOR forward rate, or forward swap rate, or a forward stock price.

Remark Interest rates being unable to go negative was an assumption, places like LCH⁶, assumed before 2015. Interest rates turning negative created crashes in LCH's initial margins pricers which was subsequently addressed by taking the normal assumption instead.

For the special cases of $\alpha = 0$ and $\beta = 1$, no closed-form solution for the implied volatility can be obtained. In the more general case, we can express the solution in terms of the implied volatility of the option. Indeed using the Black formula, and forcing it onto the SABR model, the implied volatility is approximately given by:

$$\sigma_{\text{impl}}^n = \alpha \frac{F_0 - K}{D(\zeta)} \left\{ 1 + \left[\frac{2\gamma_2 - \gamma_1^2}{24} \left(\frac{\sigma_0 C(F_{\text{mid}})}{\alpha} \right)^2 + \frac{\rho\gamma_1}{4} \frac{\sigma_0 C(F_{\text{mid}})}{\alpha} + \frac{2 - 3\rho^2}{24} \right] \varepsilon \right\},$$

where we let $C(F) = F^\beta$, $F_{\text{mid}} = (F_0 + K)/2$, $\zeta = \frac{\alpha}{\sigma_0} \int_K^{F_0} \frac{dx}{C(x)} = \frac{\alpha}{\sigma_0(1-\beta)} \times (F_0^{1-\beta} - K^{1-\beta})$, $\gamma_1 = \frac{C'(F_{\text{mid}})}{C(F_{\text{mid}})} = \frac{\beta}{F_{\text{mid}}}$, $\gamma_2 = \frac{C''(F_{\text{mid}})}{C(F_{\text{mid}})} = -\frac{\beta(1-\beta)}{F_{\text{mid}}^2}$, and $D(\zeta) = \ln \left(\frac{\sqrt{1-2\rho\zeta+\zeta^2}+\zeta-\rho}{1-\rho} \right)$.

Remark Note that with the log-normal assumption, we obtain

$$\sigma_{\text{impl}} = \left\{ 1 + \left[\frac{2\gamma_2 - \gamma_1^2 + 1/F_{\text{mid}}^2}{24} \left(\frac{\sigma_0 C(F_{\text{mid}})}{\alpha} \right)^2 + \frac{\rho\gamma_1}{4} \frac{\sigma_0 C(F_{\text{mid}})}{\alpha} + \frac{2 - 3\rho^2}{24} \right] \varepsilon \right\} \times \alpha \frac{\ln(F_0/K)}{D(\zeta)},$$

4.1.4 The Stochastic Volatility Inspired Model

4.1.4.1 The Raw SVI

Like the Schonbucher and the SABR models [108, 110], practitioners perceived that the advantage of the Raw Stochastic Volatility Inspired (rSVI) model, sometimes referred to as simply SVI [51], is that it can be derived from the Heston model [61, 59], though not advertised as such at inception [51]. Above and beyond an arguable point, the advantage of the SVI over the other parametrizations mentioned so far is its simplicity. However, like the other models it came with linear wings (e.g. Figures 4.6 and 4.7), no embedded Bid Ask and did not come with non-arbitrage constraints or at least not with correct arbitrage conditions [51, 13, 8]. This latter mistake was addressed in 2012 [100]. For each time to expiry, Gatheral proposes

$$\sigma_{BS}^2(k) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}], \quad (4.4)$$

⁶The London Clearing House is the most important CCP in the world.

where k is the log-moneyness, a adjusts the vertical displacement of the smile, b adjusts the angle between the left and right asymptotes, σ adjusts the smoothness of the vertex, ρ adjusts the orientation of the graph and m is the horizontal displacement of the smile. The advantage of Gatheral's model was that it was a parametric model that was easy to use, yet had enough complexity to model most of the volatility surface and its dynamics⁷ (or at least to the same extent Schonbucher's model does). Note that Schonbucher's market model has one fewer parameter than the SVI: the parameter m which centers the volatility surface around its minimum strike per tenor⁸. Other than this, the two models are equivalent. But at the same time it was simple enough that a solution could be found using simple optimization by constraint algorithms. Figure 4.3 illustrates the change in the ρ parameter (the skew risk), Figure 4.2 illustrates the change in the b parameter (the vol of vol risk), Figure 4.1 illustrates the change in the a parameter (the general volatility level risk), Figure 4.5 illustrates the change in the σ parameter (the ATM volatility risk) and finally, Figure 4.4 illustrates the change in the parameter m (the horizontal displacement risk).

4.1.4.2 The Natural SVI

There exist couple of additional equivalent forms of the SVI developed by Gatheral and Jacquier [100] which are closely linked to the Raw SVI of Section 4.1.4.1. First, the natural SVI parametrization of implied variance for each time to maturity is given by

$$\sigma_{BS}^2(k) = \Delta + \frac{w}{2} \left\{ 1 + \zeta \rho (k - \mu) + \sqrt{(\zeta(k - \mu) + \rho)^2 + (1 - \rho^2)} \right\}, \quad (4.5)$$

where $w \geq 0$, $\Delta \in \mathbb{R}$, $|\rho| < 1$ and $\zeta > 0$. Though, this equation looks different from the raw SVI model; we show next that they are equivalent. These different forms of the SVI are useful for different functions within a financial institutions. The natural SVI is seemingly more used in pricing and the raw SVI for risk management and statistical arbitrage strategies (the parameters being more orthogonal). The transform functions mentioned are useful when it comes to making sure different functions in a financial institution can communicate efficiently.

Remark We take this opportunity to introduce lemmas 4.1.2 and 4.1.3 which we will use in theorem 4.1.4.

⁷we will see its main limitation when we explore the gSVI.

⁸Please refer to page 89 for the definition of Normalized Tenors.

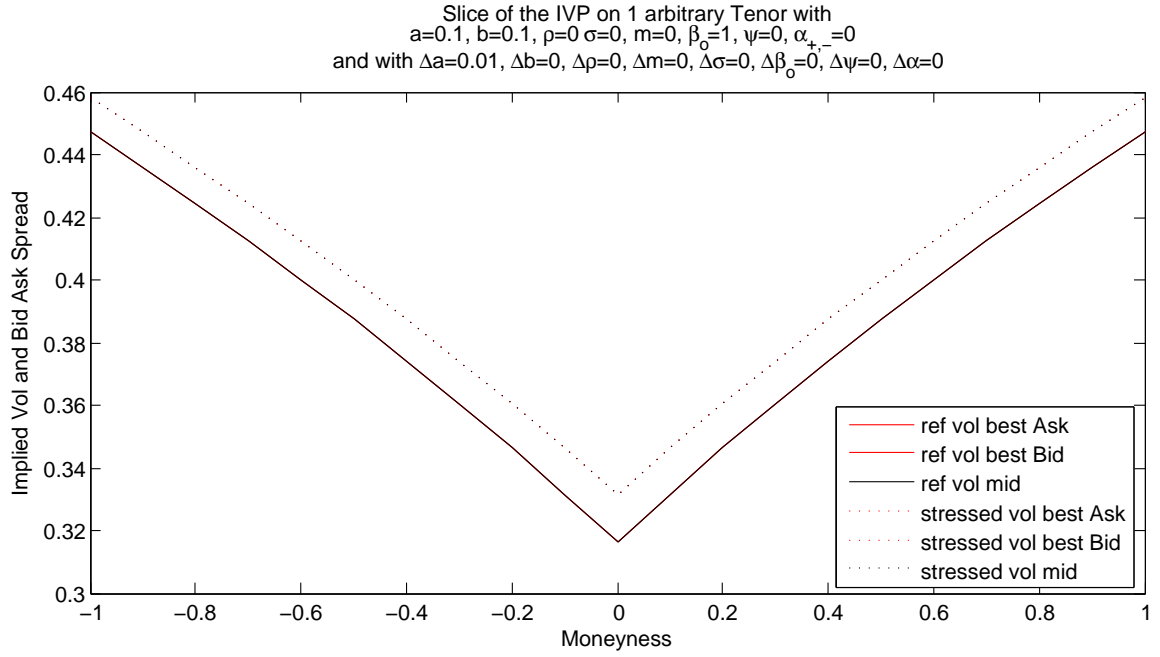


Figure 4.1: Impact of a change in the value of parameter a in the rSVI/gSVI/IVP

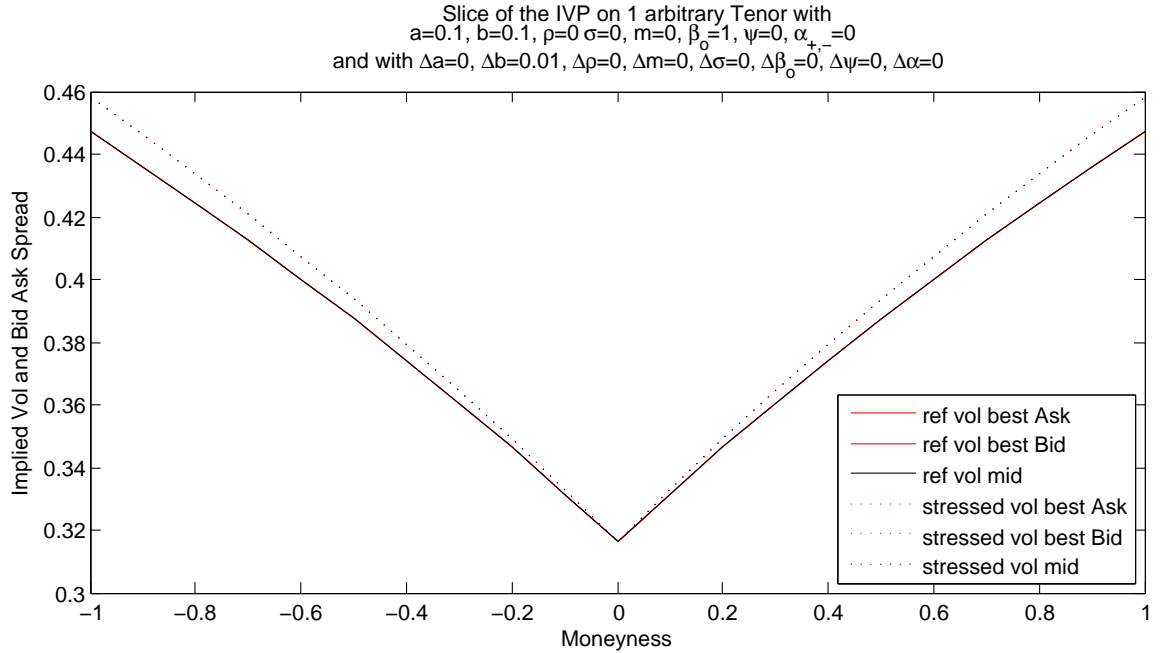


Figure 4.2: Impact of a change in the value of parameter b in the rSVI/gSVI/IVP

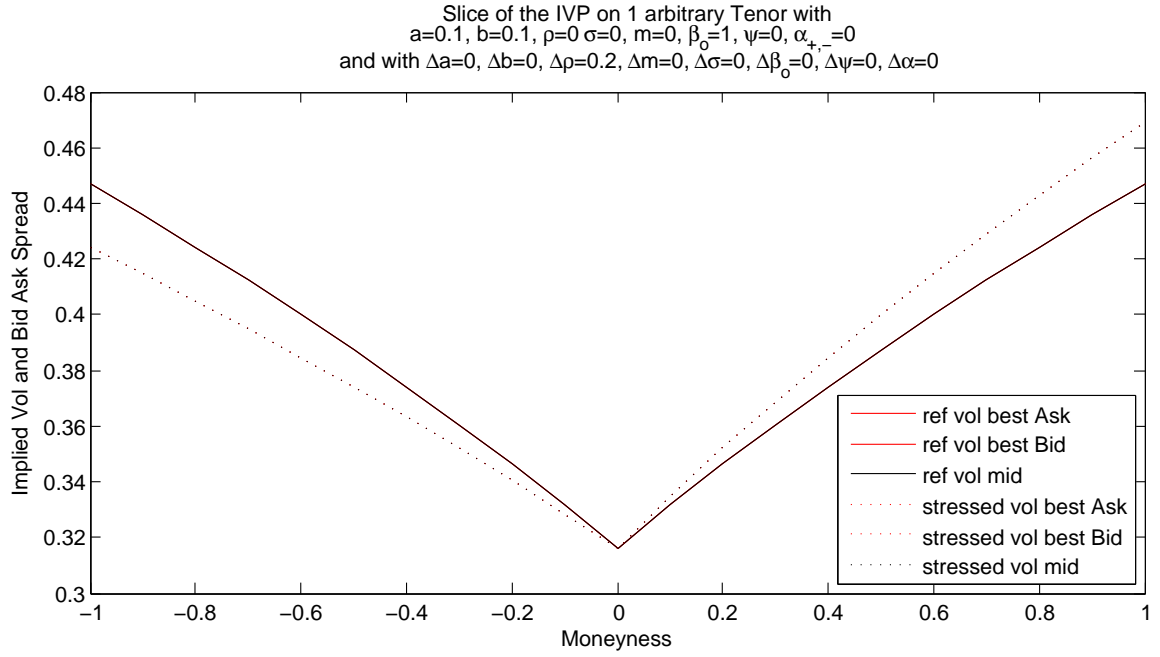


Figure 4.3: Impact of a change in ρ parameter in the rSVI/gSVI/IVP

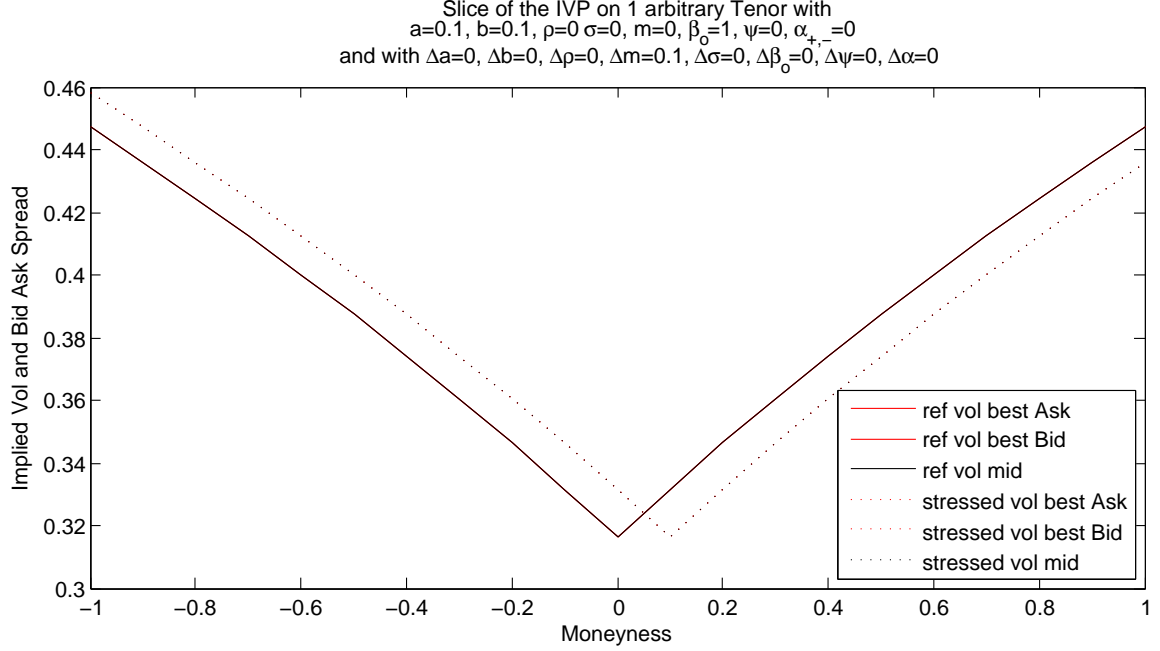


Figure 4.4: Impact of a change in the value of parameter m in the rSVI/gSVI/IVP model

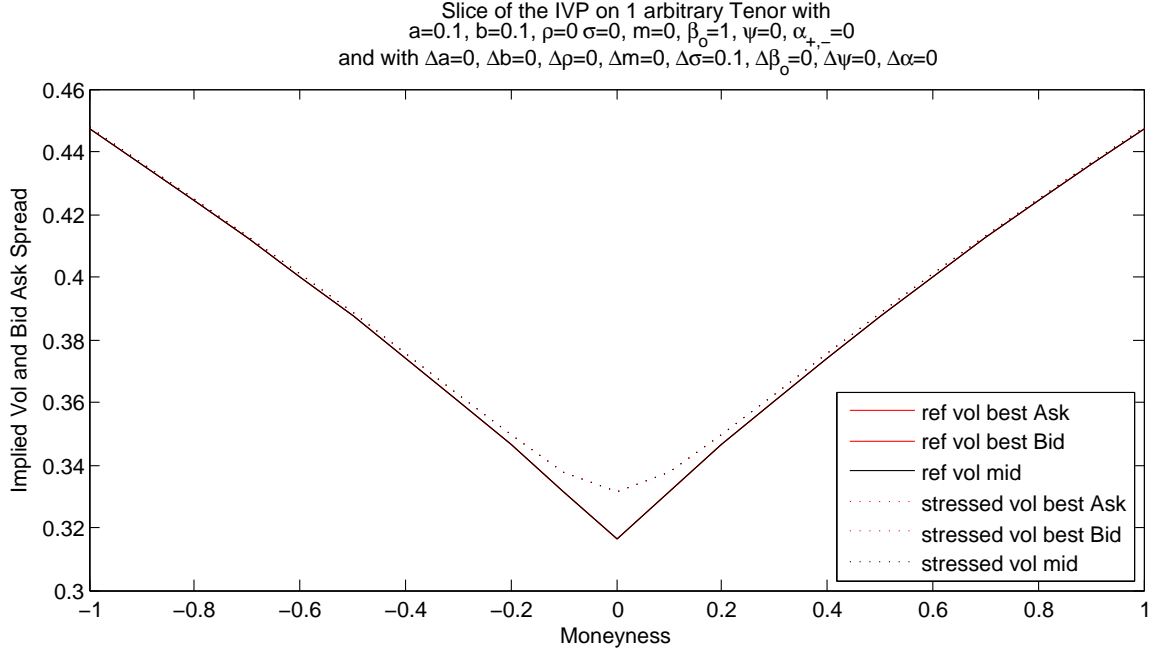


Figure 4.5: Impact of a change in the value of parameter σ in the rSVI/gSVI/IVP

Lemma 4.1.2 (Raw to Natural SVI Transform) Let $\chi_R = \{a, b, \rho, m, \sigma\}$ be a given parameter set for the function $f: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^{+,*}$ given by $f(k, \chi_R) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}]$, and let $\chi_N = \{\Delta, \mu, \rho, w, \zeta\}$ be a given parameter set for the function $g: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^{+,*}$ given by $g(k, \chi_N) = \Delta + \frac{w}{2} \left\{ 1 + \zeta \rho(k - \mu) + \sqrt{(\zeta(k - \mu) + \rho)^2 + (1 - \rho^2)} \right\}$. Then we define a set of 5 functions $\cup_{i=1}^5 T_i$ mapping χ_R onto χ_N the following way:

$$(a, b, \rho, m, \sigma) = \left(\Delta + \frac{w}{2}(1 - \rho^2), \frac{w\zeta}{2}, \rho, \mu - \frac{\rho}{\zeta}, \frac{\sqrt{1 - \rho^2}}{\zeta} \right) \quad (4.6)$$

Proof Using equation (4.5) and rearranging it in the form of equation (4.4), we note that the ρ of χ_N is the same as χ_R , we can then solve the 4 other parameters sequentially noticing that $a = \Delta + \frac{w}{2}(1 - \rho^2)$, $b = \frac{w\zeta}{2}$, $m = \mu - \rho/\zeta$ and $\sigma = \sqrt{1 - \rho^2}/\zeta$. ■

Lemma 4.1.3 (Natural to Raw SVI Transform) Let $\chi_R = \{a, b, \rho, m, \sigma\}$ be a given parameter set for the function $f: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^{+,*}$ given by $f(k, \chi_R) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}]$, and let $\chi_N = \{\Delta, \mu, \rho, w, \zeta\}$ be a given parameter set for the function $g: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^{+,*}$ given by

$g(k, \chi_N) = \Delta + \frac{w}{2} \left\{ 1 + \zeta \rho(k - \mu) + \sqrt{(\zeta(k - \mu) + \rho)^2 + (1 - \rho^2)} \right\}$. Then we define a set of 5 functions $\cup_{i=1}^5 T_i$ mapping χ_N onto χ_R in the following way

$$(\Delta, \mu, \rho, w, \zeta) = \left(a - b\sigma(1 - \rho)^{3/2}, m + \frac{\rho\sigma}{\sqrt{1 - \rho^2}}, \rho, \frac{2b\sigma}{\sqrt{1 - \rho^2}}, \frac{\sqrt{1 - \rho^2}}{\sigma} \right). \quad (4.7)$$

Proof Using equation (4.4) and rearranging it in the form of equation (4.5), we note that the ρ in χ_R is the same as χ_N , we can then solve the 4 other parameters sequentially noticing that $\Delta = a - b\sigma(1 - \rho)^{3/2}$, $\mu = m + \rho\sigma/\sqrt{1 - \rho^2}$, $w = 2b\sigma/\sqrt{1 - \rho^2}$ and $\zeta = \sqrt{1 - \rho^2}/\sigma$. ■

Theorem 4.1.4 Let $\chi_R = \{a, b, \rho, m, \sigma\}$ and $\chi_N = \{\Delta, \mu, \rho, w, \zeta\}$ be a given set of parameters defining respectively the Raw SVI given by $\sigma_R^2: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^{+,*}$ given by $\sigma_R^2(k, \chi_R) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}]$, and $\sigma_R^2(k, \chi_N) = \Delta + \frac{w}{2} \left\{ 1 + \zeta \rho(k - \mu) + \sqrt{(\zeta(k - \mu) + \rho)^2 + (1 - \rho^2)} \right\}$ respectively then $\sigma_R^2(k, \chi_R)$ and $\sigma_R^2(k, \chi_N)$ are equivalent functions.

Proof Using proof of equivalence, we can use Lemma 4.1.2 to prove that $\sigma_R^2(k, \chi_N) \Rightarrow \sigma_N^2(k, \chi_R)$ and $\sigma_R^2(k, \chi_N) \Leftarrow \sigma_N^2(k, \chi_R)$ using Lemma 4.1.3, and therefore $\sigma_R^2(k, \chi_N) \Leftrightarrow \sigma_N^2(k, \chi_R)$. ■

Remark Note that there exist other versions of the SVI, notably the SVI Jump-Wings (SVI-JW) parametrization which do not offer any advantage in the context of this document. We refer the motivated reader to the original paper [100] for more information.

4.2 Implied Volatility Parametrization

4.2.1 The Generalized SVI Model

4.2.1.1 Big Data Fueling an Industry Change

We have seen in Section 1.1.1 that the definition that best describes Big Data is the one associated to large body of information that we could not comprehend when used only in smaller amounts [19]. In that context, Jim Gatheral developed the SVI model at Merrill Lynch in 1999 and implemented in 2005. The SVI was subsequently decommissioned in 2010 because of its limitations in accurately pricing out of the money variance swaps (for example short maturity Var Swaps on the Eurostoxx are overpriced when using the SVI). This is because the wings of the SVI are linear and have a tendency to overestimate the out of the money (OTM) variance swaps. Figures 4.7 and 4.6 illustrate that latter point.

4.2.1.2 Academic Interpretation

Benaïm, Friz and Lee [56] gave a mathematical justification for this market observation. Their paper suggests that the implied volatility cannot grow asymptotically any faster than \sqrt{k} but may grow slower than \sqrt{k} when the distribution of the underlier has finite moments (e.g., does not have heavy tails).

Remark Note that this latter point is another necessary but not sufficient condition to make the implied volatility arbitrage free. Please see the previous chapter to get clarification.

Gatheral and Jacquier explained that the SVI model is consistent with Roger Lee’s moment formula [58]. However Roger Lee’s exact claims are that “the moment formula has implications for skew extrapolation: it rejects functions that grow faster than $|x|^{\frac{1}{2}}$, and **unless S_T has finite moments of all orders, it rejects those that grow slower than $|x|^{\frac{1}{2}}$** ”. Gatheral’s claim forewent the latter point. We will show here first that the higher moments of the log-normal distribution are actually finite even-though the distribution itself is considered heavy-tailed, and therefore mathematically the assumption on the linearity of the wings is rejected by that claim alone⁹.

Definition (Heavy-tailed distribution) The distribution of random variable X with cumulative distribution function (cdf) F is said to have heavy-tails if:

$$\lim_{x \rightarrow \infty} e^{kx} \Pr[X > x] = \infty \quad \text{for all } k > 0. \quad (4.8)$$

Remark Gatheral and Jacquier [100] misquoted Lee’s statement on tail-wing behavior of implied variance, which original statement mentioned that implied variance cannot grow faster than $|k|$ but can grow slower than $|k|$ when the underlying asset price has finite moments of all orders. This latter statement was misquoted by claiming that the Black-Scholes implied variance was exclusively linear. Later, Benaïm and Fitz [55] improved Roger Lee’s moment formula and concluded: “In models without moment explosion (Black-Scholes, Merton’s jump diffusion model, FMLS with $\beta = 1...$) the moment formula indicates sublinear behaviour of the implied variance”.

Lemma 4.2.1 (Moments of log-normal distribution) *The log-normal distribution does not has finite moments of all orders.*

⁹The log-normal distribution is wrongly considered part of the heavy tail distribution family due to the conception that the tails are heavier (fatter) than the normal distribution but does not fulfill the formal definition of equation (4.8). It’s possible Gatheral in [59] made this common mistake.

Proof If X is log-normal, then $Y = \ln X$ is normal. Consider $E[X^k] = E[e^{kY}] = \int_{y=-\infty}^{\infty} e^{ky} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy$. Now observe that, $ky - \frac{(y-\mu)^2}{2\sigma^2} = -\frac{2k\sigma^2 y + y^2 - 2\mu y + \mu^2}{2\sigma^2} = -\frac{1}{2\sigma^2} (y^2 - 2(\mu + k\sigma^2)y + (\mu + k\sigma^2)^2 + \mu^2 - (\mu + k\sigma^2)^2) = -\frac{(y - (\mu + k\sigma^2))^2}{2\sigma^2} + \frac{k(2\mu + k\sigma^2)}{2}$. Thus the k^{th} moment is simply $E[X^k] = e^{k(2\mu + k\sigma^2)/2} \int_{y=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y - \mu')^2/(2\sigma^2)} dy$ where $\mu' = \mu + k\sigma^2$. But this latter integral is equal to 1, being the integral of a normal density with mean μ' and variance σ^2 . So $E[X^k] = e^{k(2\mu + k\sigma^2)/2}$.

Remark The log-normal distribution is wrongly considered part of the heavy tail distribution family due to the conception that the tails are heavier (fatter) than the normal distribution but does not fulfill the formal definition of equation (4.8).

In any case this suggests that the linear wings of the SVI model may overvalue really deep OTM options, as one can observe in the markets. Figures 4.7 and 4.6 illustrate, for a couple of different asset classes, how the linearity of the wings assumption fails when far from the at the money (ATM).

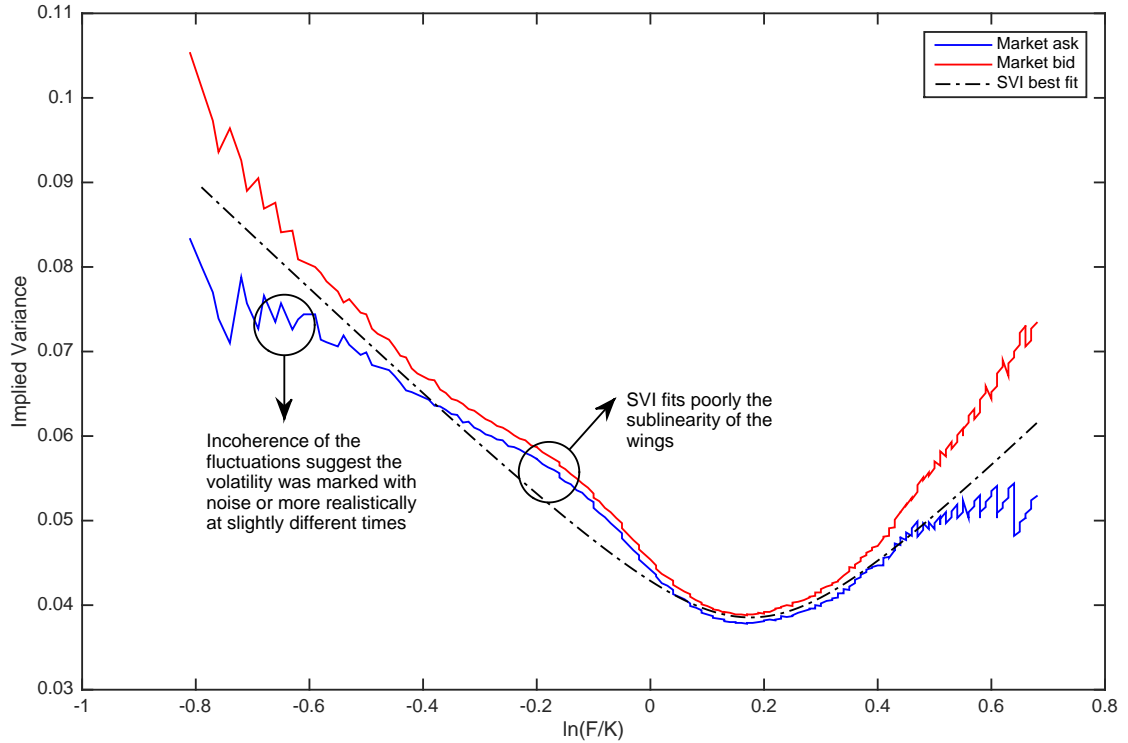


Figure 4.6: Crude Oil Bid Ask Smile for the 1 Year Expiry as given by the CME

Remark In figure 4.6, the data has been downloaded by Farah Haddadin and taken from the CME - live market data. The time stamp is 2018-09-10, the expiration date is 2018-09-17 and the future price is \$68.355.

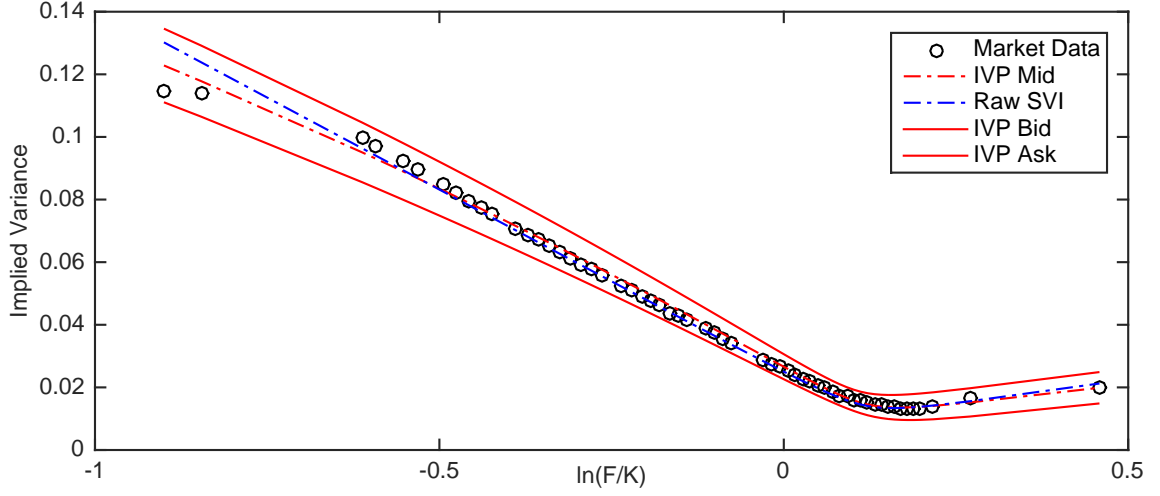


Figure 4.7: IVP & SVI with 1 year expiry S&P 500 index options [26] on 26/11/2016

4.2.1.3 Downside Transform

The downside transform in the gSVI was given by $z = \frac{k}{\beta^{|k-m|}}$, $1 \leq \beta \leq 1.4$ to address the mentioned wing limitations [12]. The resulting model becomes equation (2.2.3.1).

$$\sigma_{gSVI}^2(k) = a + b \left[\rho(z - m) + \sqrt{(z - m)^2 + \sigma^2} \right],$$

$$z = \frac{k}{\beta^{|k-m|}}, 1 \leq \beta \leq 1.4.$$

There are many ways of defining the concept of downside transform. One general approach would be to define μ and η such that equation (4.1a) defines the change of variable from strike space to modified strike space. The idea being that the further away the option is from the ATM, the bigger the necessary adjustment on the wings. Few examples of such downside transform are given by Equations (4.1a), (4.1b) and (4.1c).

$$z = \frac{k}{\beta^{\mu + \sigma_l |k-m|}}, \quad (4.1a)$$

$$z = e^{-\beta |k-m|} (k - m), \quad (4.1b)$$

$$z = \ln(\beta |k - m|). \quad (4.1c)$$

We can, for example, choose $\mu = 1$ and $\eta = 4$ and have the transformation in the form $z = \frac{k-m}{\beta^{1+4|k-m|}}$ because it yielded interesting results on the FX markets [13] and also because it relaxes the constraint on β since we incorporate a Bid Ask layer. Figure

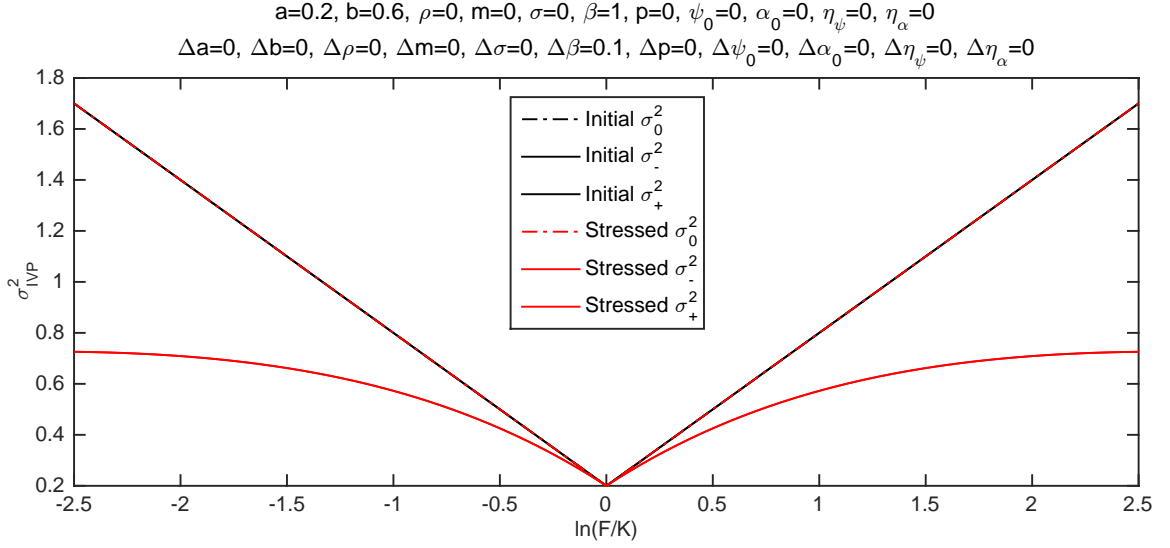


Figure 4.8: Impact of a change in the value of parameter β in the gSVI/IVP model

4.8 represents the impact of the β parameter in the gSVI. The geometric properties of the gSVI, more specifically its ability to model smile, skew and inverse skew, while at the same time correcting the linear wings of the SVI makes it applicable to all markets (FX, Commodities, equities and rates).

Remark Section 3 is dedicated to the discussion associated to anomaly detection in the context of arbitrage handling. However, we thought of making the following point here. There exist two constraints that make the gSVI “often” (as illustrated by Figure 3.2) arbitrage free. The first condition on the falling variance in equation (3.18) does not change. However the butterfly condition changes. Using some of the market conventions around what was initially thought to be an equivalent condition to the butterfly condition [58], namely equation (3.18), we obtain equation (4.2), which happens to be a necessary but not sufficient condition.

$$\left| T \frac{1 + |k - m| \ln \beta}{\beta^{|k - m|}} \left(b\rho + \frac{(\frac{k}{\beta^{|k - m|}} - m)}{\sqrt{(\frac{k}{\beta^{|k - m|}} - m)^2 + \sigma^2}} \right) \right| \leq 4. \quad (4.2)$$

4.2.2 Implied Volatility Parametrization

4.2.2.1 Overall Formula

By incorporating the information on the gSVI, the ATM Bid Ask spread and the curvature adjustment of the wings we define [12, 13] the Implied Volatility surface

Parametrization (IVP) split with its mid in equation (4.3) below¹⁰

$$\sigma_{IVP,o,\tau}^2(k) = a_\tau + b_\tau \left[\rho_\tau(z_{o,\tau} - m_\tau) + \sqrt{(z_{o,\tau} - m_\tau)^2 + \sigma_\tau^2} \right], \quad (4.3a)$$

$$z_{o,\tau} = \frac{k}{\beta_\tau^{1+4|k-m|}}, \quad (4.3b)$$

which is essentially the gSVI [12] model, introduced in the previous section, as well as its liquidity parameters in equation (4.4).

$$\sigma_{IVP,+, \tau}^2(k) = a_\tau + b_\tau \left[\rho_\tau(z_{+, \tau} - m_\tau) + \sqrt{(z_{+, \tau} - m_\tau)^2 + \sigma_\tau^2} \right] + \alpha_\tau(p), \quad (4.4a)$$

$$z_{+, \tau} = z_{o, \tau} [1 + \psi_\tau(p)], \quad (4.4b)$$

$$\sigma_{IVP, -, \tau}^2(k) = a_\tau + b_\tau \left[\rho_\tau(z_{-, \tau} - m_\tau) + \sqrt{(z_{-, \tau} - m_\tau)^2 + \sigma_\tau^2} \right] - \alpha_\tau(p), \quad (4.4c)$$

$$z_{-, \tau} = z_{o, \tau} [1 - \psi_\tau(p)], \quad (4.4d)$$

$$\alpha_\tau(p) = \alpha_{0, \tau} + (a_\tau - \alpha_{0, \tau})(1 - e^{-\eta_{\alpha\tau} p}), \quad (4.4e)$$

$$\psi_\tau(p) = \psi_{0, \tau} + (1 - \psi_{0, \tau})(1 - e^{-\eta_{\psi\tau} p}), \quad (4.4f)$$

where a , b , ρ , m and σ represent the same risk factors as that is raw SVI which we introduced in Subsection 4.1.4.1 and β is the same parameter as the gSVI in Subsection 4.2.1. The ATM liquidity function is represented by equation (4.4e), where p represents the position size, α_0 the smallest volatility increment associated to the smallest position size that one can trade and η_α , the price elasticity. The curvature liquidity function is represented by equation (4.4f), where p represents the position size, ψ_0 the smallest volatility increment associated to the smallest position size that one can trade and η_ψ , the price elasticity. Next, the latter liquidity parameters are being discussed in more details.

Remark Note that once the Bid Ask spread is incorporated, we care less about the midprice in the context of vanilla options market making. Though the mid price may have arbitrage at the portfolio level, the Bid-Ask relaxes these feasible arbitrage constraints given by all possible butterfly, and calendar spread contracts. These two adjustments are

$$\forall \Delta, \forall T, C(K, T + \Delta, \sigma_{IVP,+,t}(k)) - C(Ke^{-r\Delta}, T, \sigma_{IVP,-,t}(k)) \geq 0, \quad (4.5a)$$

$$C(K - \Delta, \sigma_{IVP,+,t}(k)) - 2C(K, \sigma_{IVP,-,t}(k)) + C(K + \Delta, \sigma_{IVP,+,t}(k)) > 0. \quad (4.5b)$$

¹⁰With flexibility on the way the downside transform is implemented.

4.2.2.2 Modelling the Bid-Ask Wings Curvature

One of the innovations of the gSVI [12] when compared to the SVI [51] is the adjustment of the Wings using a change of variable or downside transform: see equation (2.2.3.1). Let us call the parameter which aim is to adjust the wings of the midprice by $\beta_{o,\tau}$. Someone wanting to sell an option would want to sell it at a higher price than the midprice so the dampening effect of the bid ($\beta_{+,\tau}$) should be smaller than the one of the mid therefore $\beta_{+,\tau} < \beta_{o,\tau}$. Using the same logic, the dampening of the ask price should be $\beta_{o,\tau} < \beta_{-,\tau}$. The constraints on the β 's is given by

$$1 < \beta_{+,\tau} < \beta_{o,\tau} < \beta_{-,\tau}. \quad (4.6)$$

In order to control the addition of these new parameters we incorporate the ψ parameter, mentioned in equation (4.7), to account for the symmetry for this Bid Ask adjustment. Figure 4.9 illustrates how the variable ψ affects the β parameters and hence the Bid Ask price curvatures.

$$\begin{aligned} \beta_{+,\tau} &= (1 - \psi_\tau)\beta_{o,\tau}, \\ \beta_{-,\tau} &= (1 + \psi_\tau)\beta_{o,\tau}. \end{aligned} \quad (4.7)$$

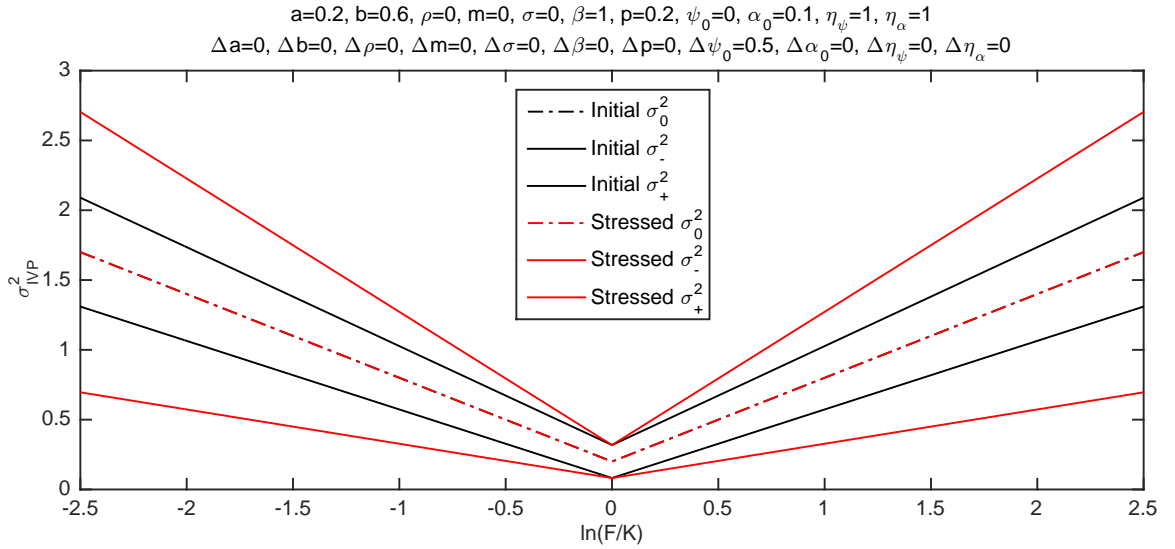


Figure 4.9: Impact of a change in the value of parameter ψ in the IVP model

Remark Note that the main reason why this discrepancy exists between the ATM Bid-Ask spread and the rest of the implied volatility surface is mainly due to the supply and demand. Indeed, usually exotic options trader try to hedge against the

different risk factors. The delta¹¹ of an option consists of a linear position of the underlier but the vega¹² of an option is only accessed through vanilla options which liquidity is the highest around the ATM and the further you get from this ATM, the less the economical need for these partial points.

4.2.2.3 Modelling the Bid-Ask ATM Spread

The curvature adjustment via the β parameters models the idea that the further away the option is from the ATM, the bigger the Bid Ask spread, however this change of variable yield a Bid Ask spread of 0 ATM. It is therefore necessary to adjust for this issue by adding an ATM bid ask factor that will take the form of $\min(a_\tau/2, \alpha_\tau)$, where α_τ is the tentative “ATM Bid Ask half spread” but adjusted if its value is such that it will be higher than the lowest point of the implied vol. Equation (4.4e) models this concepts and Figure 4.10 illustrates how the variable α adjusts the ATM Bid Ask spread.

Remark There are ways to improve this idea, especially if we assume that the implied volatility behaves like a log-normal distribution around 0 but the benefit to complexity ratio was optimal this way.

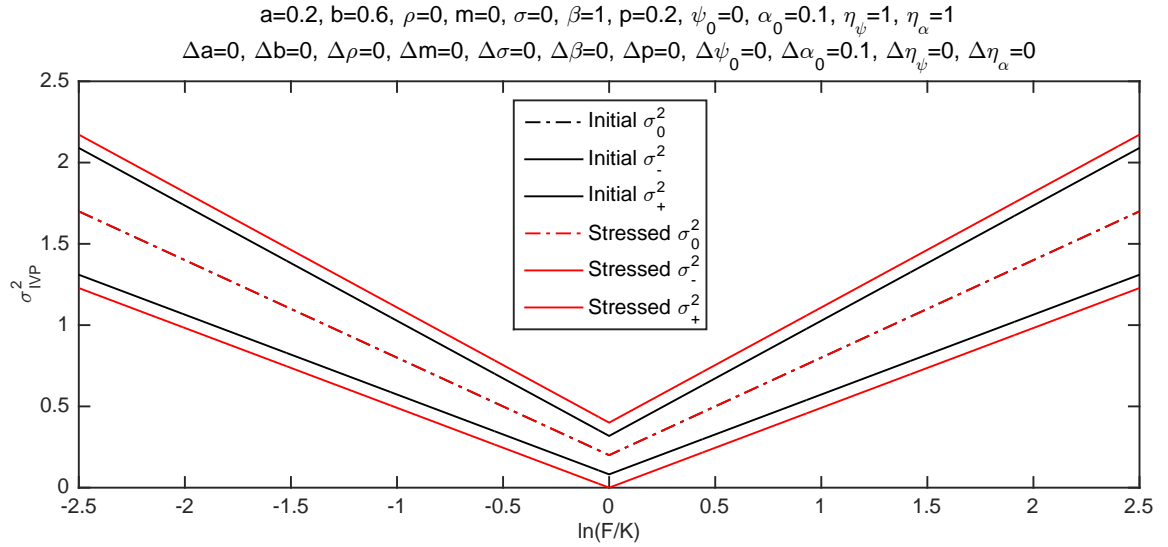


Figure 4.10: Impact of a change in the value of parameter α in the IVP model

¹¹ $\partial V/\partial x$.

¹² $\partial V/\partial \sigma$.

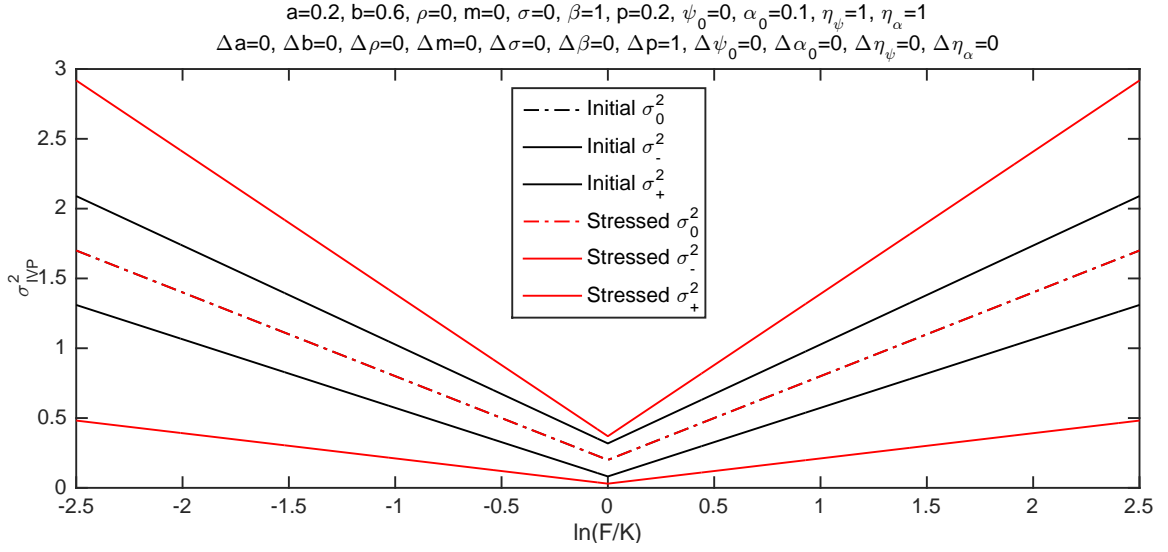


Figure 4.11: Impact of a change in the value of parameter p in the IVP model

4.2.2.4 Position Size

The functions $\alpha(p)$ (Figure 4.10) and $\psi(p)$ (Figure 4.9) model the ATM and wing curvature of the Bid-Ask keeping in mind the idea that the bigger the position size the bigger the market impact and hence the wider the Bid-Ask. This market impact parameter is controlled by p (Figure 4.11). Finally, a couple of additional parameters model the elasticity of the liquidity: η_ψ (Figure 4.12) and η_α (Figure 4.13).

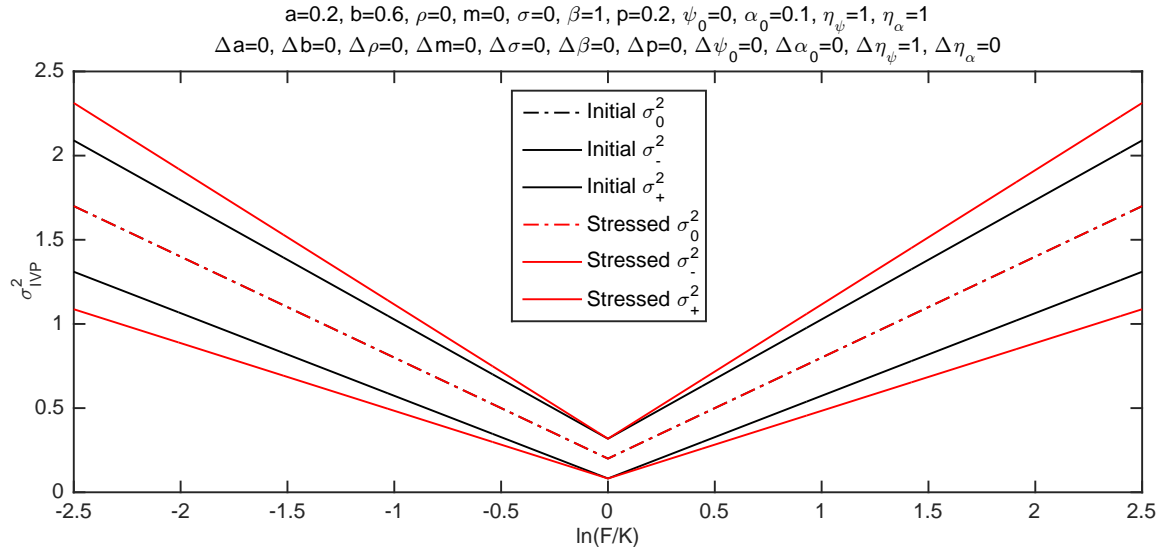


Figure 4.12: Impact of a change in the value of parameter η_ψ in the IVP model

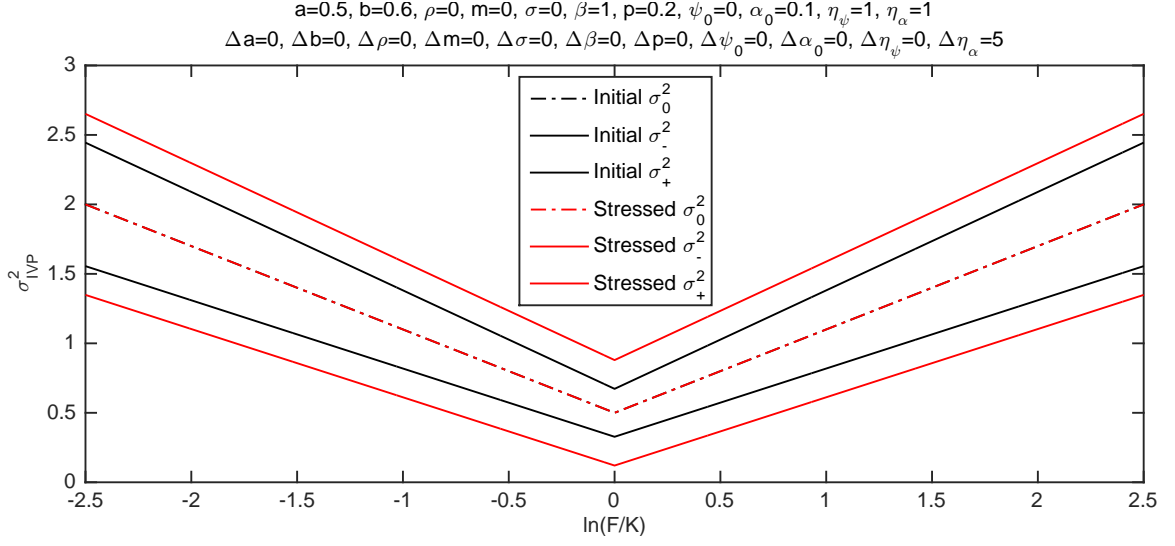


Figure 4.13: Impact of a change in the value of parameter η_α in the IVP model

Remark Note that during the writing of the thesis, the objective was to improve the status quo while not losing any elements of it. Mainly the original point of our model (gSVI and especially IVP) is its ability to model the data supported sublinearity of the wings as well as add a liquidity element. None of the current models incorporates these two elements. On top of this argument, there is nothing that the SVI can model that the IVP cannot and this by construction.

4.3 De-arbitraging with the IVP parametrization

We finished chapter 3 with Figure 3.4 which we explained to be an intuitive approach to de-arbitraging. Equation (4.8) formalize the idea more mathematically. Namely, we try to find the shortest distances between the input volatility grids and its closest arbitrage free mirror, $\tilde{\sigma}_t(\tau, d)$, spanned by the IVP parameters subject to the Call spread and Calendar spread Conditions of equations (4.8b) and (4.8c) respectively.

$$\text{solve: } \hat{\sigma}_t(\tau, d) = \underbrace{\arg \min}_{\tilde{\sigma}_t(\tau, d)} \sum_{\tau} \sum_d \left[C\left(\sigma_{i,t}(\tau, d)\right) - C\left(\tilde{\sigma}_t(\tau, d)\right) \right]^2 \quad (4.8a)$$

subject to: $\forall \tau$ and $\forall K$

$$C\left(K - \Delta, \sigma_0(K - \Delta, \tau)\right) - C\left(K, \sigma_0(K, \tau)\right) \geq 0 \quad (4.8b)$$

$$C\left(K, \tau + \Delta, \sigma_0(K, \tau + \Delta)\right) - C\left(K e^{-r\Delta}, \sigma_0(K e^{-r\Delta}, \tau)\right) \geq 0 \quad (4.8c)$$

The constraints of this optimization problem can be lifted by incorporating the constraints in the objective function directly. More specifically, by setting

$$\hat{\sigma}_t(\tau, d) = \arg \min_{\tau} \sum_d [\sigma_{i,t}(\tau, d) - \tilde{\sigma}_t(\tau, d)]^2 + \mathcal{K}(CS_1(K, \tau) + CS_2(K, \tau)) \quad (4.9)$$

where \mathcal{K} is the constraint scalar¹³. We need to incorporate the constraints into the objective function, and therefore have for the butterfly (or call spread) constraint

$$CS_1(K, \tau) = \left| C(K - \Delta, \sigma_0(K - \Delta, \tau)) - C(K, \sigma_0(K, \tau)) \right| \times 1_{\mathcal{B}}, \quad (4.10)$$

where the function $1_{\mathcal{B}} = 1$ if $C(K - \Delta, \sigma_0(K - \Delta, \tau)) - C(K, \sigma_0(K, \tau)) > 0$ and 0 otherwise. For the calendar spread constraint, we have

$$CS_2(K, \tau) = \left| C(K, \tau + \Delta, \sigma_0(K, \tau + \Delta)) - C(K e^{-r\Delta}, \sigma_0(K e^{-r\Delta}, \tau)) \right| \times 1_{\mathcal{C}}. \quad (4.11)$$

where $1_{\mathcal{C}} = 1$ if $C(K, \tau + \Delta, \sigma_0(K, \tau + \Delta)) - C(K e^{-r\Delta}, \sigma_0(K e^{-r\Delta}, \tau)) > 1$ and 0 otherwise. These indicator function enforce directly, in the objective function, the relevant constraints.

4.4 Updating Volatility Data

4.4.1 Reclaiming Data Science's Intuitive Meaning

Data Science is more known for the methodologies associated to understanding data [112] than extracting information from structured data [113]. However, we propose to add few methodologies associated to restructuring unstructured data in this section. The importance of the latter is critical in any mathematical analysis and the data is useless on its own (without the classic tools in Mathematical Finance).

4.4.2 Updating Volatility from Listed Markets

In the of presence of a listed market¹⁴, in which all the log moneynesses are visible across rolling contracts, proxying the unknown points is done in two steps. We first use the falling variance formula¹⁵ to provide us with a possible volatility smile and we pick the relevant moneyness on that smile. However, the latter needs to be checked for coherence and the surface needs to go through a de-arbitraging methodology, provided by equation (4.8), in order to reach the closest arbitrage free volatility.

¹³A large enough number to make sure the constraints are respected, but not too large to create numerical instabilities.

¹⁴Note that this is to contrast with over the counter (OTC) markets.

¹⁵Equation (3.23).

4.4.3 Updating Volatility outside Listed Markets

There are few ways to see how the updating function should be defined in the absence of a listed market. If we take a naive lagging approach we can certainly update the volatility surface point-wise. For instance, let Σ_t be the normalized matrix as explained from Subsection 1.2.2.3 and let $\sigma_t^{i,j}$ be the j th strike of the i th tenor of Σ_t , then the isolated arrival $\sigma_{t+\epsilon}^{i,j}$ overrides Σ_t by copying all but the $\sigma_t^{i,j}$'s element while substituting the latter with $\sigma_{t+\epsilon}^{i,j}$. This methodology is commonly used in the market but the lack of liquidity in uncommonly traded moneynesses and the asynchronicity of the trades in these moneynesses means that the volatility surface that is recorded and stored by the end of the day is one made of apples and oranges. This is best seen with fluctuations in the wings that are not coherent if the data had arrived at the same time. Figure 4.6 illustrates this latter point¹⁶.

4.4.4 The Midprice Smoothing Approach

Though not perfect, this search and replace process can be enhanced by a smoothing stage which consists of making the resulting implied volatility more realistic. The first option available is to apply a simple de-arbitraging methodology on the mid price. In algorithm (2) the function *dearbMid* achieves this objective. This widely used methodology may in fact re-position the newly arrived data $\sigma_{t+\epsilon}(i, j)$ back to its old position $\sigma_t(i, j)$. The process can be enhanced by adding a time kernel to put more weight on the points that are most recent.

4.4.4.1 The Midprice Constrained Smoothing Approach

One way to address the limitation described by the *dearbMid* algorithm (2) is to add a constraint to equation (4.8a) to get equation (4.12)

$$\hat{\Sigma}_{mid,t} = \underbrace{\arg \min}_{\Sigma_t} \sum_{q=1}^{|C^\tau|} \sum_{p=1}^{|C^k|} [\sigma_t(p, q) - \tilde{\sigma}_t(p, q)]^2 + \mathcal{K} (CS_1 + CS_2), \quad (4.12)$$

where $\hat{\Sigma}_{mid,t}$ represents the collections of modified implied volatility midprice money-ness and tenors in C^τ and C^k . CS_1 and CS_2 are as before, the calendar and butterfly conditions as defined by Equations (3.23) and (3.15) and finally \mathcal{K} is a scalar. This penalizes potential moves from the newly arrived point $\sigma_t(i, j)$. In algorithm (2) the function *dearbAroundPointAndPropagade* achieves this objective.

¹⁶It also illustrates a point about the SVI and the sub-linearity of the wings which we will delve on more in section 4.2.1

Algorithm 2: Update Implied Volatility

Input: Current implied volatility Σ_t and new arrived data $\sigma_{t+\epsilon}^{i,j}$.

Output: Updated implied volatility $\Sigma_{t+\epsilon}$.

▷ Make a copy implied vol

$$\Sigma_{t+\epsilon} \leftarrow \Sigma_t$$

▷ Search and replace

for $p = 1$ **to** $|C^\tau|$ **do**

for $q = 1$ **to** $|C^k|$ **do**

if $(i = p) \vee (j = q)$ **then**

$\sigma_{t+\epsilon}^{p,q} \leftarrow \sigma_{t+\epsilon}^{i,j}$

▷ Smoothing stage

if $\mathbb{F} = \text{"dearbMid"}$ **then**

We de-arb the mid implied volatility disregarding liquidity;

$\Sigma_{t+\epsilon} \leftarrow \text{dearbMid}(\Sigma_{t+\epsilon})$

else if $\mathbb{F} = \text{"dearbLiquidity"}$ **then**

We de-arb the mid implied volatility focusing on the liquidity parameters;

$\Sigma_{t+\epsilon} \leftarrow \text{dearbLiquidity}(\Sigma_{t+\epsilon})$

else if $\mathbb{F} = \text{"dearbAroundPointAndPropagade"}$ **then**

We de-arb while making sure $\sigma_{t+\epsilon}^{i,j}$ is not modified during smoothing;

$\Sigma_{t+\epsilon} \leftarrow \text{dearbAroundPointAndPropagade}(\Sigma_{t+\epsilon})$

else if $\mathbb{F} = \text{"EducatedGuessForecast"}$ **then**

We use Generalized Bumping methodology to modify the implied vol;

$\Sigma_{t+\epsilon} \leftarrow \text{EducatedGuessForecast}(\Sigma_{t+\epsilon})$

else

Default setting: no smoothing;

$\Sigma_{t+\epsilon} \leftarrow \Sigma_{t+\epsilon}$

▷ Return state

return $\Sigma_{t+\epsilon}$

4.4.4.2 The Liquidity Adjustment Smoothing Approach

It can however be questionable to temper with data points that have been observed just recently on the basis of one new point. One conservative way to adjust the implied volatility as a market maker is to include the liquidity model introduced in Section 4 and only adjust the liquidity components in such a way as to comfortably envelop all the points, like it is the case with the constraints of equation (4.13)

$$\hat{\Sigma}_{liq,t} = \underbrace{\arg \min}_{\alpha_q, \psi_q} \sum_{q=1}^{|C^\tau|} \sum_{p=1}^{|C^k|} [\sigma_t(p, q) - \tilde{\sigma}_t(p, q)]^2 + \mathcal{K}_1(CS_1 + CS_2), \quad (4.13)$$

where $\hat{\Sigma}_{liq,t}$ represents the collections of modified implied volatility liquidity parameters of all tenors with the corresponding liquidity adjusted arbitrage conditions (Equa-

tions (4.5b) and (4.5a). In algorithm (2) the function *dearbLiquidity* achieves this objective

Remark We recommend reading chapter 3 concurrently.

4.4.4.3 The Forecasting Smoothing Approach

Let us recall briefly the Principal Component Analysis (PCA) methodology. Let $M = (X_{ij}) \in \mathbb{R}^{m \times n}$ be a matrix of explanatory variables, \bar{M} is its centered version and \tilde{M} is its reduced and centered version such that

$$\bar{M} = \begin{bmatrix} X_{1,1} - \bar{X}_1 & \cdots & X_{1,N} - \bar{X}_N \\ \vdots & \ddots & \vdots \\ X_{K,1} - \bar{X}_1 & \cdots & X_{K,N} - \bar{X}_N \end{bmatrix} \text{ and } \tilde{M} = \begin{bmatrix} \frac{X_{1,1} - \bar{X}_1}{\sigma(X_1)} & \cdots & \frac{X_{1,N} - \bar{X}_N}{\sigma(X_N)} \\ \vdots & \ddots & \vdots \\ \frac{X_{K,1} - \bar{X}_1}{\sigma(X_1)} & \cdots & \frac{X_{K,N} - \bar{X}_N}{\sigma(X_N)} \end{bmatrix}.$$

We look for the vector u such that the projection of the cloud to u has maximal variance. This projection of X on u is denoted by $\pi_u(\tilde{M}) = \tilde{M} \cdot u$ and its variance $\pi_u(\tilde{M})^T \cdot \pi_u(\tilde{M}) = u^T \cdot \underbrace{\tilde{M}^T \cdot \tilde{M}}_C \cdot u$. We also have C diagonalizable. We call P the change of basis and $\Delta = \text{Diag}(\lambda_1, \dots, \lambda_N)$ the diagonal matrix made of its spectrum

$$\pi_u(\tilde{M})^T \cdot \pi_u(\tilde{M}) = u^T P^T \Delta P u = (Pu)^T \underbrace{\Delta}_{v} (Pu).$$

The values $(\lambda_1, \dots, \lambda_N)$ of Δ 's diagonal are classified by decreasing order. The unit vector u which maximizes $v^T \Delta v$ is an eigenvector of C associated to the eigenvalue λ_1 , we then have $v^T \cdot \Delta \cdot v = \lambda_1$. The proof is done by setting up the optimization by constraint method¹⁷. The same methodology is used in order to find the other eigenvalues such that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. This latter observation allows us to quantify in terms of proportion the combined explanatory power of the first i th components $\frac{\lambda_1 + \lambda_2 + \dots + \lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_N}$. The PCA suffers from several limitations such as the fact that it relies on linear assumptions, it relies on orthogonal transformations of the original variables, directions with largest variance are assumed to be the most important, and finally PCA is not scale invariant.

Remark In Section 4 we will see methods surrounding the parametrization of the implied volatility surface and the reader may stop here and read the relevant chapter before proceeding.

¹⁷in which we maximize the variance projected onto u (under the constraint that its norm should be 1 and via the Lagrange multiplier α : $L(u, \alpha) = u^T \cdot C \cdot u - \alpha(u^T u - 1)$).

Consistent with the results of the SABR model [110], a PCA analysis on the risk factors (per tenors) of the implied volatility reveals that the three most important components explaining the dynamics of the implied volatility are, in decreasing importance, the “ATM Vol”, the “Vol of Vol” and the “Skew”. The most important factor that explains the shape of the implied volatility is therefore the overall volatility level. This is the main rationale behind the proposed methodology. indeed the last proposed adjustment is to take simply the first element in explaining the dynamics of the implied volatility by applying a proportional move equal to $\frac{\sigma_{t+\epsilon}(i,j)}{\sigma_t(i,j)}$ to the remainder of the point. In algorithm (2), \mathbb{F} represents the switching flag between the methodologies described in this section.

Remark Considering the benefits to complexity ratio, one may choose to use different smoothing algorithms depending on whether one is working within a trading desk, a risk methodology desk in a bank or within a CCP as quant researcher dealing with margin calls. Indeed, within a trading desk, it is important that the combination of products sold to the counter-party is such that, overall the trader is not arbitrageable. With this in mind “dearbLiquidity” will be the smoothing algorithm used in this case. Within a risk methodology desk, the quants usually work with the mid prices and therefore a “dearbMid” smoothing algorithm is recommended.

4.5 Proxying Volatility with Sparse Data

One of the consequences of Big Data’s arrival is that it helped showcase and highlight the limitations of some the classic financial mathematics models (e.g. Figure 4.7).

Remark Note that, to some extent, limiting assumptions in the branch of financial mathematics were already well documented. We can point here [94] that Fama [114], Samuelson [115] and Mandelbrot [116] had all raised concerns around related concepts (such as fat tails) more than half a century ago. The whole development of the Levy Process is centered around the limitations of Bachelier’s assumptions [37]. However, the data associated to algorithmic trading (e.g. Figure 4.7) is new. Fama, Samuelson, Mandelbrot or Paul Levy all had access to low frequency data which have certainly guided their research on a path that would perhaps be different had they gotten access to the kind of interactions that are taking place in figure 4.7 or on the options market.

The arrival of BD, paradoxically, exposed the lack of data in certain circumstances such as for example in certain strikes and moneynesses of illiquid option contracts and therefore generated the need to develop the data science branch called proxying.

We will take a look at a proposed methodology for proxying illiquid vanilla options in this section.

4.5.1 Implied Volatility and Risk Factor Decomposition

First, let us note that proxying only makes sense if there is no reliable data. But proxying carries the risk of being wrong. Decomposing a closely related products into risk factors is a necessary first step. The way we propose to decompose the volatility surface is through the IVP¹⁸ model presented by equations (4.3) and (4.4). Each of the resulting risk factors will then be used to fill in the missing information appropriately.

4.5.2 Proxying and Economic Factors

When choosing a proxy for missing data there are several aspects one must take into consideration.

4.5.2.1 Asset Class

First, it is important to note that the asset class (e.g. equities, commodities, rates, FX, etc) is an important driver for the shape of the risk factors. For example in commodities you have the skew parameter which is opposite to the one of the equities market¹⁹. Indeed people in commodities are afraid of prices going up (for example, because of a drought in the agricultural markets) but in equities people are more afraid of bankruptcies. For this reason calls are bought at a premium compared to puts in commodities and the reverse is true in equities. The choice of the proxy, when it comes to the ρ parameter²⁰, can therefore be a critical factor when it comes to proxying a product within a specific asset class. For example looking forward proxying skew for an illiquid single stocks in equities with commodities data will be a bad idea. For an illiquid single stock (for which ATM is available), it would be better to proxy skew with the ρ parameter of the S&P500 than the ρ parameter of the Goldman Sachs Commodities index.

¹⁸The formal introduction to the model is done in Section 4.

¹⁹With the exception of oil which skew behaves like in the equities market.

²⁰Captures the imbalance in price between the put compared to the call prices: used in the IVP model introduced in section 4.2.

4.5.2.2 Economic sector

Whether you are working within equities or commodities you can always find ways to fine tune how close the products that needs to be proxied is to the subgroups composing this asset class. For example in equities you could group the underliers into pharmaceuticals or electronics groups. In commodities you could further group the asset class into economic sector like energy, agricultural or metals. For the latter two, the sector can be further decomposed into subgroups²¹. You can always find ways to target your proxy when you have the choice. These different markets have very different volatility of volatilities. For example commodities are famous for being more volatile than the equities markets. Parameters like the b or the a in the IVP model²² will be very different in these two asset classes. The process of choose the right proxy is as much associated to economics than it is with mathematics.

4.5.2.3 Product Diffusion Types

We have shown in Section 3.1.5 that the derivative markets can be priced following different diffusion-based processes. Recall that the log-normal diffusion is assumed from the equities Market, the normal diffusion for the rates market and the Garman-Kohlhagen for the FX markets. Depending on the asset class, the prospect of the long term mean reversion of the underlier may change. For instance, even though traders on the rates market understand the limitations of the Black-Scholes model, the linearity of the wings can become increasingly a limiting factor for OTM options. In these markets the control, and therefore the proxy, for the β parameter in the IVP model becomes increasingly important.

4.5.2.4 Geographical Location

Geographical location is also another component to keep in mind. For instance, emerging markets have a very different risk profile than the market of developed countries.

4.5.2.5 Liquidity Profile & Size of the Company

For liquidity related parameters we suggest to take a look at the volume traded and depth of similar products. For example; indices are traded often with a deep order book but single stocks less.

²¹E.g., metals can be divided into precious and industrial metals, and agricultural can be divided into grains livestock etc.

²²From section 4.2.

4.5.2.6 Linear Interpolation & Parameters Surrounding Tenors

One other way of proxying parameters of an unobserved tenors is to take the linear interpolation of these specific tenors.

Remark The de-arbitraging methodology still needs to be implemented after proxying.

4.5.3 Sequential Bootstrapping

4.5.3.1 Definition

In this subsection we propose a sequential bootstrapping solution. First let us give a definition and go through couple of remarks to help us understand the methodology.

Definition (Sequential bootstrapping): let $\chi_t = \{a_t, b_t, \rho_t, m_t, \sigma_t, \beta_t\} \cup \{\psi_{0,t}, \alpha_{0,t}, \eta_{\psi,t}, \eta_{\alpha,t}\}$ where $t \in C^\tau$ the set of target IVS induced by the incomplete original point data $\tilde{\sigma}(p, q)$ (where $q \in \tilde{C}^k \in C^k, p \in \tilde{C}^\tau \in C^\tau$). Similarly let us call the set of proxies $\chi_{p,t} = \{a_{p,t}, b_{p,t}, \rho_{p,t}, m_{p,t}, \sigma_{p,t}, \beta_{p,t}\} \cup \{\psi_{0,p,t}, \alpha_{0,p,t}, \eta_{\psi,p,t}, \eta_{\alpha,p,t}\}$. We construct the optimization by constraint with the aim to construct a volatility surface which would go through every known $\tilde{\sigma}(p, q)$ and recover the remaining point using $\chi_{p,t}$ for the remaining points and subject to the usual arbitrage constraints of Equations (3.15) and (3.23).

Remark The set of proxies $\chi_{p,t}$ does not necessarily need to come from the same underlier. For example, assume we have three stocks A, B and C. Stock C is missing both $b_{o,t}$ and $\rho_{o,t}$ in its set $\chi_{o,t}$. Stock C is very close in terms of market fundamentals to B (e.g. stocks with the same sector and geographical location) and is close to C but further away in terms of these same fundamentals (e.g. stocks with the same sector but maybe different geographical location). However stock B has also a missing $\rho_{o,t}$ but it has $b_{o,t}$. The completed set $\chi_{p,t}$ for C will consist of the $b_{o,t}$ of B and the $\rho_{o,t}$ from A. For the sake of keeping the mathematical specifications under control, p in $\chi_{p,t}$, refers to the closest available proxied risk factor.

4.5.3.2 Proxying Missing IVP Factors & Priorities

There are no set methodologies that we are aware off when it comes to proxying. We tried to summaries the main points developed in Section 4.5.2 in the Table 4.1, bearing in mind that these are qualitative suggestions.

Missing IVP Risk Factor	Proxying priorities
a	usually given with the parameters σ and m
b	AOTSPST, GL, PTES, AC
ρ	AOTSPST, PT, AC, ES
m	usually given with the parameters a and σ
σ	usually given with the parameters σ and m
β	AOTSPST, PDT
α_0	LPSOTC, AOTSPST
ψ_0	LPSOTC, AOTSPST
η_α	LPSOTC, AOTSPST
η_ψ	LPSOTC, AOTSPST

Table 4.1: Proposed Proxying Missing Factor to Proxying priorities hash table

4.5.3.3 Post Proxy Selection & Smoothing

Once the proxies have been selected, one still needs to make sure that the resulting implied volatility surface goes through the points where data is available. We also need to ensure that the newly constructed IV is arbitrage free. We will refer the reader here to Section 3 for the theory and description behind our de-arbitraging methodology. The points addressed in that Section 3 will be assumed as understood in the following paragraphs.

4.5.3.4 Optimization by Constraint for Sparse Data

We will now formalize the optimization problem where the aim is to calibrate the set χ_t with $\chi_{p,t}$ and $\tilde{\sigma}_t(p, q)$ by minimizing the loss of current information. Namely, if we set $\Sigma_t = \cup_{q=1}^{|C^k|} \cup_{p=1}^{|C^\tau|} \sigma_t(p, q)$ the set of implied volatility points generated by the parametrization χ_t , its calibration will be given by equation (4.14). Note that equation (4.14) is the same as the classic de-arbitraging methodology using the IVP parameters but it is really the addition of the last constraint controlled by \mathcal{K}_3 that

allows for the appropriate calibration.

$$\hat{\Sigma}_t = \underbrace{\arg \min}_{\Sigma_t} \sum_{q=1}^{|C^\tau|} \sum_{p=1}^{|C^k|} [\sigma_t(p, q) - \sigma_{p,t}(p, q)]^2 \quad (4.14a)$$

$$+ \mathcal{K}_1(CS_1 + CS_2) + \mathcal{K}_2[\sigma_t(i, j) - \tilde{\sigma}_t(i, j)]^2 + \mathcal{K}_3 \sum_{i=1}^{10} [\chi_{p,t}(i) - \chi_t(i)]^2 \quad (4.14b)$$

with $\mathcal{K}_1 \approx \mathcal{K}_2 > \mathcal{K}_3 > 0$ meant to represent the hierarchy in the optimization process. As we have seen previously, we first make sure that the result IVP does not have any arbitrage. Then we make sure that the IVP goes through all the points observed and finally we try to be as close as possible to the selected proxies. The reason why we set $\mathcal{K}_1 \approx \mathcal{K}_2$ is that it is assumed that the points observed are mutually arbitrage free and that, therefore, we will never be faced with the case in which we need to select either an arbitrage free IV or one that goes through all the observed points.

Remark Assume the degree of certainty or comfort in the choice of proxies is different (e.g. you work in equities and you are definitively comfortable that the ρ parameter should be taken from an equity index but you are less sure of the value of parameter β). How would you modify the constraints in the sub-equation (4.14b) in such a way so as to engineer preferences on the choice of the proxies? Note that in sub-equation (4.14b) the 3rd term $\mathcal{K}_3 \sum_{i=1}^{10} [\chi_{p,t}(i) - \chi_t(i)]^2$ is the term engineering the preferences in the proxies. An additional layer of granularity can be engineered is the choice of the constants, namely $\mathcal{K}_3 \sum_{i=1}^{10} [\chi_{p,t}(i) - \chi_t(i)]^2$ would become in sub-equation (4.14b) $\tilde{\mathcal{K}}_3 \sum_{i=1}^{10} \mathcal{K}_{3,i} [\chi_{p,t}(i) - \chi_t(i)]^2$ with $\mathcal{K}_3 \geq \tilde{\mathcal{K}}_3 \mathcal{K}_{3i}$.

4.5.4 Tracking Volatility & Resampling Risk Factors

Provided the arrival of implied volatility data is always incomplete, we can interpret this arrival in terms of a change in one of the risk factors. This make the IVP an ideal candidate to track the whole surface using a particle filter²³ on the different risk factor scenarios. With this objective in mind, we need to define the scenario's likelihood function. We must first define the arrival and the weighting process associated to the arrival time. Let the set of IVP parameters which best fit our implied volatility surface be

$$\chi_t^i = \{\chi_t^{M,i} \cup \chi_t^{L,i}\}, \quad (4.15)$$

²³Please see chapter 8 for a literature review.

where $i \in C^\tau$, $\chi_t^i = \{a_t^i, b_t^i, \rho_t^i, m_t^i, \sigma_t^i, \beta_t^i\}$ and $\chi_t^{M,i} = \{\psi_{0,t}^i, \alpha_{0,t}^i, \eta_{\psi,t}^i, \eta_{\alpha,t}^i\}$ represent the mid-IV and liquidity parameters of each pillar tensors. In addition, denote by $\dot{\sigma}(p, q)$ (where $q \in C^k$ and $p \in C^\tau$) the most recent data and the array of weights. We must also rank the risk factors from most likely to least likely. There are several methods that can be used to perform the latter task, we recommend a stepwise regression [117] based on historical data as it is simple enough but recognize that other methods could be potentially better [118].

Definition (Risk Factor Importance): We define λ_0 , the r^2 ²⁴ contribution of a pointwise change of the IVS, $\lambda^p = \{\lambda_a, \lambda_b, \lambda_\rho, \lambda_m, \lambda_\sigma, \lambda_\beta\}$ the set of ranked r-squared contribution of each of the 6 parameters $\in \chi_t^i = \{a_t^i, b_t^i, \rho_t^i, m_t^i, \sigma_t^i, \beta_t^i\}$ where $i \in C^\tau$ representing the mid-IV and liquidity parameters of each pillar tensors. We denote $\lambda^s = \{\lambda_{(a,b)}, \lambda_{(a,\rho)}, \dots, \lambda_{(m,\beta)}\}$ the set of unique pairwise parameter changes. Note that there really is one set of pairwise parameters that interests us (the combine change in skew and vol of vol). We define the subset of $\lambda^p \cup \lambda_{(b,\rho)}$ of all accepted scenarios:

$$\lambda = \{\lambda_0, \lambda_a, \lambda_b, \lambda_\rho, \lambda_m, \lambda_\sigma, \lambda_\beta, \lambda_{(b,\rho)}\}. \quad (4.16)$$

We define, in algorithm (3), $\mathcal{H}(\cdot)$, the function that takes as input a uniform random variable $u \sim U[0, 1]$ and returns the set of parameter(s) associated to λ .

Finally, we denote by N_p^t the number of particles at time t . It always mean reverts towards N_p so we choose N_p instead for convenience sake. When information about a specific point, $\sigma_{d,\tau}$, of the IVS has arrived we can assume that this change may be isolated *but this specific point propagates arbitrage free constraints on the IVS which would otherwise remain constant* (BPC)²⁵. It could also correspond to a change in “vol of vol” (the b parameter in the IVP model: Figure 4.2) BPC. It could correspond to a change in “skew” (the ρ parameter in the IVP model: Figure 4.3) BPC. It could correspond to an individual change of any other of IVP parameters. It could also correspond to a multiple change of any of the IVP parameters (more specifically “spot-vol”). The reason why mapping onto the IVP parameter is a good idea is because the parameters of the IVP not only fit better than any parameterized model the market observed prices but also its parameters map to easily understandable economical risk factor like we have seen before. The sampling methodology will consist of changing each parameters of the 3 sets of types of sampling. The first type of sampling is one

²⁴Alternatively called the coefficient of determination, “r-squared” represents the proportion of the variance explained by the model.

²⁵We substitute this sentence in italic by BPC from now on

Algorithm 3: HASH FUNCTION $\mathcal{H}(\lambda)$

Input: λ
Output: a set of parameter(s) is returned
 $u \sim \mathcal{U} [0, \lambda_0 + \lambda_a + \lambda_b + \lambda_\rho + \lambda_m + \lambda_\sigma + \lambda_\beta + \lambda_{(b,\rho)}]$
if $0 \leq u < \lambda_0$ **then**
 | **return** (0)
else if $\lambda_0 \leq u < \lambda_a$ **then**
 | **return** (a)
else if $\lambda_a \leq u < \lambda_b$ **then**
 | **return** (b)
else if $\lambda_b \leq u < \lambda_\rho$ **then**
 | **return** (ρ)
else if $\lambda_\rho \leq u < \lambda_m$ **then**
 | **return** (m)
else if $\lambda_m \leq u < \lambda_\sigma$ **then**
 | **return** (σ)
else if $\lambda_\sigma \leq u < \lambda_\beta$ **then**
 | **return** (β)
else if $\lambda_\beta \leq u < \lambda_{b,\rho}$ **then**
 | **return** (b, ρ)
else
 | **return** ERROR

in which we only move that new point that just arrived. We call this **Sample 0P**²⁶. The second type consists of sampling 1 parameter of the IVP to explain the new data to map the change of price by a change of the economic climate. For example, the ATM is the point which arrived the most recently, then our PF will assume with one of its scenarios that this is due to a change in Vol of Vol (and therefore) all the point of the implied volatility should be adjusted accordingly. We denote by this sampling **Sample 1P**. The third type consists of assuming that the point change is the result of two economic factor happening at the same time. For example, imagine you are assigned the task to mark an in the money call for wheat and the economical climate is that there are political tension with Russia²⁷ (therefore Vol of Vol increases) and that at the same time we have information that there are possible droughts that are incoming (there is a change of skew). This leads our PF to assume with one of its scenarios that this is due to a change in Vol of Vol and Skew at the same time. We

²⁶Note that we could have called this one Sample 1P with the “P” meaning “point” but this could be also interpreted as “parameter” which we use for the second type of sampling. Therefore in order to limit confusion we preferred calling it Sample 0P for 0 parameters

²⁷Russia is one of the first exporters of wheat

denote this sampling **Sample 2P**.

Sample0P:

$$x_k^{(i)} = \arg \min_{\tilde{\sigma}_t(\tau, d)} \sum_{\tau} \sum_d [C(\sigma_t(\tau, d)) - C(\tilde{\sigma}_t(\tau, d))]^2$$

subject to:

$$\forall d \in C^k, C_1^{\mathcal{B}, d, \tau} < C_2^{\mathcal{B}, d, \tau}$$

$$\forall \tau \in C^{\tau}, C_1^{\mathcal{C}, d, \tau} < C_1^{\mathcal{C}, d, \tau}$$

$$\hat{\sigma}(p, q) = \dot{\sigma}(p, q)$$

$$1_{|a_t^i - \tilde{a}_t^i| > 0} + 1_{|b_t^i - \tilde{b}_t^i| > 0} + 1_{|\rho_t^i - \tilde{\rho}_t^i| > 0} + 1_{|m_t^i - \tilde{m}_t^i| > 0} +$$

$$1_{|\sigma_t^i - \tilde{\sigma}_t^i| > 0} + 1_{|\beta_t^i - \tilde{\beta}_t^i| > 0} = 0$$

Sample1P:

$$x_k^{(i)} = \arg \min_{\tilde{\sigma}_t(\tau, d)} \sum_{\tau} \sum_d [C(\sigma_t(\tau, d)) - C(\tilde{\sigma}_{\mathcal{H}(\lambda), t}(\tau, d))]^2$$

subject to:

$$\forall d \in C^k, C_1^{\mathcal{B}, d, \tau} < C_2^{\mathcal{B}, d, \tau}$$

$$\forall \tau \in C^{\tau}, C_1^{\mathcal{C}, d, \tau} < C_1^{\mathcal{C}, d, \tau}$$

$$\hat{\sigma}(p, q) = \dot{\sigma}(p, q)$$

$$1_{|a_t^i - \tilde{a}_t^i| > 0} + 1_{|b_t^i - \tilde{b}_t^i| > 0} + 1_{|\rho_t^i - \tilde{\rho}_t^i| > 0} + 1_{|m_t^i - \tilde{m}_t^i| > 0} + 1_{|\sigma_t^i - \tilde{\sigma}_t^i| > 0} + 1_{|\beta_t^i - \tilde{\beta}_t^i| > 0} = 1$$

$$C(K, \sigma_0(K, \tau)) < C(K - \Delta, \sigma_0(K - \Delta, \tau)),$$

$$C(Ke^{-r\Delta}, \sigma_0(Ke^{-r\Delta}, \tau)) < C(K, \tau + \Delta, \sigma_0(K, \tau + \Delta))$$

Sample2P:

$$x_k^{(i)} = \arg \min_{\tilde{\sigma}_t(\tau, d)} \sum_{\tau} \sum_d [C(\sigma_t(\tau, d)) - C(\tilde{\sigma}_{\mathcal{H}(\lambda), t}(\tau, d))]^2$$

subject to:

$$\forall d \in C^k, C_1^{\mathcal{B}, d, \tau} < C_2^{\mathcal{B}, d, \tau}$$

$$\forall \tau \in C^{\tau}, C_1^{\mathcal{C}, d, \tau} < C_1^{\mathcal{C}, d, \tau}$$

$$\hat{\sigma}(p, q) = \dot{\sigma}(p, q)$$

$$1_{|a_t^i - \tilde{a}_t^i| > 0} + 1_{|m_t^i - \tilde{m}_t^i| > 0} + 1_{|\sigma_t^i - \tilde{\sigma}_t^i| > 0} + 1_{|\beta_t^i - \tilde{\beta}_t^i| > 0} = 0$$

$$1_{|b_t^i - \tilde{b}_t^i| > 0} + 1_{|\rho_t^i - \tilde{\rho}_t^i| > 0} = 2$$

$$C(K, \sigma_0(K, \tau)) < C(K - \Delta, \sigma_0(K - \Delta, \tau)),$$

$$C(Ke^{-r\Delta}, \sigma_0(Ke^{-r\Delta}, \tau)) < C(K, \tau + \Delta, \sigma_0(K, \tau + \Delta))$$

Remark Note that we could employ the same methodology with three or more of the parameters but usually higher order Greeks beyond the second-order are considered negligible on the market and therefore it is not worth adding complexity as the ratio of the benefits over the latter does not invite such extension. However, for the sake of rigour, one may decide to apply higher order sampling as an exercise. We could very well imagine scenarios in which the parameters of different tenors react together at the same extent but the de-arbitraging methodology would partially take care of this specific additional type of sampling. We also thought that the methodology was

complex enough the way it is currently proposed.

4.6 Statistical Arbitrage for Vanilla Options

4.6.1 Point-wise Approach

One natural way to describe an implied volatility $\sigma_t(k, \tau)$ of price $C_t(Ke^{-r\Delta}, T)$ at time t of strike k and tenor τ where k and τ are observable pillar points is to do it as a function of the adjacent 4 points as seen by Figure 3.4 or by Table 4.2.

Remark For the sake of making the notation a bit more intuitive, we have used the notation from Table 4.2 with where equation (4.17) provides the relevant equivalences.

		$P_t^{i-1,j}$		
	$P_t^{i,j-1}$	$P_t^{i,j}$	$P_t^{i,j+1}$	
		$P_t^{i+1,j}$		

$$P_t^{i-1,j} := C_t(Ke^{-r2\Delta}, T - \Delta) \quad (4.17a)$$

$$P_t^{i,j-1} := C_t(K - \Delta)e^{-r\Delta}, T) \quad (4.17b)$$

$$P_t^{i,j} := C_t(Ke^{-r\Delta}, T) \quad (4.17c)$$

$$P_t^{i,j+1} := C_t(K + \Delta)e^{-r\Delta}, T) \quad (4.17d)$$

$$P_t^{i+1,j} := C_t(K, T + \Delta) \quad (4.17e)$$

Table 4.2: Implied Volatility Grid and Formula

We aim at studying the Bayesian probability problem of equation (4.18).

$$p\left(\mathbf{P}^{i,j} - l(\mathbf{P}^{i,j}) \middle| l(\mathbf{P}^{i,j}), l(\mathbf{P}^{i-1,j}), l(\mathbf{P}^{i,j-1}), l(\mathbf{P}^{i,j+1}), l(\mathbf{P}^{i+1,j})\right) \quad (4.18)$$

where, in the discrete space, $\mathbf{P} = P_{\tau \in [1,t]}$ and l represents the lag inducing function such that $l(P_{t+1}) = (P_t)$.

4.6.2 Simplified IVP & Closed Form Calibration

We employ the simplification suggested in the previous section to show how the closed-form calibration is made easier by choosing the parameters a , b and ρ . More specifically, we show how the calibration is done on the mid-IV parameters in equation (4.19) and how it is done on the liquidity parameters with equation (4.21). Let us however prior to that formalize the definition of the Simplified IVP (SIVP).

Definition (Simplified IVP): Let $\chi = \{\chi^M, \chi^L\}$ with $\chi^M = \{a_\tau, b_\tau, \rho_\tau, m_\tau\}$ and $\chi^L = \{\alpha_\tau, \psi_\tau\}$ a given parameter set for the function $f: \mathbb{R}^{+,*} \times \mathbb{R}^+ \times [-1, +1] \times \mathbb{R} \times$

$\mathbb{R}^+ \times \mathbb{R}^+ \longrightarrow \mathbb{R}^{+,*}$, the Simplified IVP (SIVP) is given by

$$\sigma_{IVP,o,\tau}^2(k) = a_\tau + b_\tau \left[\rho_\tau(k - m_\tau) + \sqrt{(k - m_\tau)^2} \right], \quad (4.19a)$$

$$\sigma_{IVP,\pm,\tau}^2(k) = a_\tau + b_\tau \left[\rho_\tau(z_{\pm,\tau} - m_\tau) + \sqrt{(z_{\pm,\tau} - m_\tau)^2} \right] \pm \alpha_\tau, \quad (4.19b)$$

$$z_{\pm,\tau} = k[1 \pm \psi_\tau]. \quad (4.19c)$$

Proposition 4.6.1 (Simplified IVP Mid's Closed-form Calibration) *The simplified IVP model's midprice presented by equation (4.19a) can be calibrated in closed form using equation 4.20.*

$$\hat{m}_\tau = \underbrace{\arg \min_k}_{k} \sigma_{o,\tau}^2(k), \quad (4.20a)$$

$$\hat{a}_\tau = \sigma_{o,\tau}^2(\hat{m}_\tau), \quad (4.20b)$$

$$\hat{b}_\tau = \frac{\sigma_{o,\tau}^2(\hat{m}_\tau + k) + \sigma_{o,\tau}^2(\hat{m}_\tau - k) - 2\hat{a}_\tau}{2|k|}, \quad (4.20c)$$

$$\hat{\rho}_\tau = \frac{\sigma_{o,\tau}^2(\hat{m}_\tau + k) - \sigma_{o,\tau}^2(\hat{m}_\tau - k)}{2\hat{b}_\tau k}. \quad (4.20d)$$

Proof Set $\hat{m}_\tau = \underbrace{\arg \min_k}_{k} \sigma_{o,\tau}^2(k)$, we know that $\hat{b}_\tau \geq 0$, therefore $\hat{a}_\tau = \sigma_{o,\tau}^2(\hat{m}_\tau)$. Now choose 2 symmetrical points around the vertical axis passing through the minimum point, use $\sigma_{o,\tau}^2(k) = a_\tau + b_\tau \left(\rho_\tau(k - \hat{m}_\tau) + |k - \hat{m}_\tau| \right)$, we have an equation with two unknowns and two observation that allow us to solve the system. ■

Proposition 4.6.2 (Simplified IVP Bid-Ask spread closed form calibration) *The simplified IVP model's Bid-Ask presented by equation (4.19b) can be calibrated in closed form using equations 4.21.*

$$\hat{\alpha}_\tau = \sigma_{+,\tau}^2(\hat{m}_\tau) - \hat{a}_\tau + \hat{m}_\tau \hat{b}_\tau \hat{\rho}_\tau + |\hat{m}_\tau| \hat{b}_\tau, \quad (4.21a)$$

$$\hat{\psi}_\tau = \frac{\sigma_{+,\tau}^2(\hat{m}_\tau) + \sigma_{-,\tau}^2(\hat{m}_\tau) - 2\hat{a}_\tau}{2|\hat{m}_\tau| \hat{b}_\tau}. \quad (4.21b)$$

Proof Using the results from proposition 4.6.1, we estimate \hat{a}_τ , \hat{b}_τ , $\hat{\rho}_\tau$ and \hat{m}_τ . Now we select a couple of reliable observed points and plug into the IVP equation to obtain $\sigma_{IVP,\pm,\tau}^2(k) = \hat{a}_\tau + \hat{b}_\tau \times \left[\hat{\rho}_\tau(k[1 \pm \psi_\tau] - \hat{m}_\tau) + \sqrt{(k[1 \pm \psi_\tau] - \hat{m}_\tau)^2} \right] \pm \alpha_\tau$. We now have a system of two equations with two unknowns that can be solved in a straightforward fashion. ■

4.6.2.1 Formal Bayesian Set-up

We aim at studying the following probability distribution function (4.22).

$$p\left(\mathbf{\Omega}^\tau - l(\mathbf{\Omega}^\tau) \middle| l(\mathbf{\Omega}^{L(\tau)}), l(\mathbf{\Omega}^{l(\tau)})\right) \quad (4.22)$$

where, in the discrete space, $\mathbf{\Omega}^\tau = \{m_{\tau,t}, a_{\tau,t}, b_{\tau,t}, \rho_{\tau,t}\}$ with $\tau \in \text{ON}, 1W, 2W, 1M, 3M, 6M, 1Y, 2Y^{28}$, $t \in [1, T]$ and l represents the lag inducing function such that $l(m_{\tau,t+1}) = (m_{\tau,t})$, $m_{l(\text{ON}),t+1} = m_{1W,t} \dots m_{l(1Y),t+1} = (m_{2Y,t})$ and vice versa for the $m_{L(1W),t+1} = (m_{\text{ON},t}) \dots m_{L(2Y),t+1} = (m_{1Y,t})$.

Remark One of the trade-offs in reducing the number of points of interest (three in this example) is that with a higher number of parameters we can obtain several combinations of parameters which may fit the three points (Figure 4.14). One may show that a combination of a, b, σ from equation (4.3) may fit these three points in different ways. This is the reason why we need to limit our strikes of interest (on

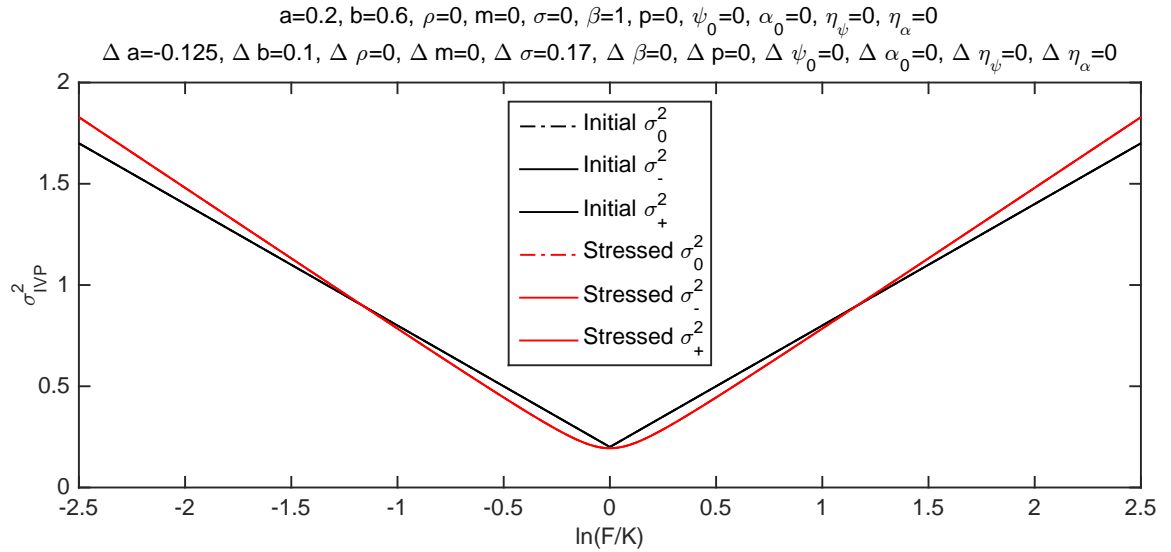


Figure 4.14: Sets of Fixed Points Modeled by Different Sets of IVP Parameters

which we ultimately trade) to three available highly liquid points: the ATM and a couple more at $\pm x$ strikes around the ATM.

We have seen in chapter 2 examples of mean reverting stochastic processes and we have seen in this chapter potentially mean reverting parameters. However, most risk systems assume normality or log-normality, therefore symmetry in the forecasted distribution, improper for these complicated risk factors. We propose in the next chapter a methodology to address this issue using clustering methodologies.

²⁸Arbitrary, but usually tenor pillars in most asset classes.

Chapter 5

Clustering for Distribution & Regime Change Forecasting

We have seen in chapter 2 an example of mean reverting stochastic process. However, most risk systems assume normality or log-normality in the distribution of the underlying variable. Irrespective of whether we are dealing with short term interest rates or complex implied volatility risk factors, both express some form of mean-reversion. In this section we show how to solve this issue via clustering, as a simple enhancement (rather than an opposition) of classic stochastic models from financial mathematics. Taking a general approach, we discuss how these clustering techniques can help us model complex risk factors such as for example the ones we introduced in Section 4 in a simple way. We also introduce in that effect the concept of Responsible VaR [7] to illustrate our findings.

5.1 Introduction

5.1.1 Objectives

5.1.1.1 Leading vs Lagging Risk System

The current risk models available to practitioners are at best Responsive and therefore lagging with respect to regime changes which means that one (or few depending on the quantile level) risk breaches is needed for the mathematical model to be able to adjust to the changing market condition. The first objective for this part is to lay down the mathematical specification for a risk system that would be leading as opposed to lagging.

5.1.1.2 SDE vs Gaussian Processes

Domain knowledge can lead the wise Machine Learner to replace a complex parameterized problem (with its design issues and limitations) with a Gaussian process that is statistically more optimal and with higher potential for robustness. Such examples exist in the literature [119] but none makes the connection with the world of stochastic differential equations (SDE) as it pertains to quantitative finance and therefore to the proxying examples we gave in the previous chapter.

5.1.2 Mathematical formulation of Anticipative VaR

In this section we formulate the concept of Anticipative VaR going from the intuitive SDE approach and trying to extend what is considered the classic approach towards what is more realistic. We take this approach to introduce intuitively the clustering mirror approach in Section 5.3.

Definition (Responsive VaR): A VaR model that will be able to adapt, *a posteriori*, to increased volatility conditions will be referred to as Responsive VaR.

Definition (Anticipative VaR): A VaR model that will be able to adapt, *a priori*, to increased volatility conditions will be referred to as Anticipative VaR.

Remark We have chosen to introduce the concept of clustering in the context of Risk as opposed to trading as we felt that this simple distribution forecasting element provided more innovation in the context of Risk than of Trading.

5.1.3 Agenda

In Section 5.2, we introduce the SDE approach. More specifically, using the generalized diffusion model methodology, split in primary and secondary parameters, we formalize a forecast distribution for risk purposes. We also summarize some of the benefits associated with respect to the current used models. We also provide the calibration methodology. In Section 5.3 we introduce the corresponding clustering methodology and go through the benefits and calibration of the latter as well. In Section 5.5 we go through an example while attempting to break few of the margining misconceptions in the process. We introduce in this context the concept of Anticipative VaR.

Remark The concept of VaR and Expected shortfall will be interchangeably used in this thesis (though the formal definition is different) as the use of either of these risk measures can be interchangeably used in the context of risk *anticipativity* and risk *responsibility*. Also note that this section along with Section 5.6.4 in which we lay down the foundations of Responsible VaR are sections that are interesting to read concurrently.

5.2 The SDE Approach

5.2.1 Generalized Diffusion for Risk Factors Model

In order to expose some of the benefits of the SDE approach we first need to introduce the Generalized Diffusion for Risk Factors model.

Definition (Generalized Diffusion for Risk Factors): We consider a risk factor process $(X_t)_{t \geq 0}$ with natural filtration $(\mathcal{F}_t)_{t \geq 0}$, and we define the forward risk factor process $(F_t)_{t \geq 0}$ by $F_t := \mathbb{E}(X_t | \mathcal{F}_0)$ generated by equation (5.1).

$$dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t, \quad (5.1)$$

where θ is the rolling speed of mean reversion, μ is the long term rolling mean, α is the positivity flag enforcer, $\beta \in [-1, +1]$ is boundary flag enforcer and $\{\bigcup dW_i\}_{i=t-\tau}^t$, the set of historical deviation of your assumed model's distribution (e.g. all the historical absolute bumps¹ in the context of a normal diffusion) and τ the tenor.

Remark In practice it may sometimes be worth tweaking the parameters to make them more flexible for back-testing efficiency purposes. Equation (5.2) represents this modified “backtesting friendly” version. With this in mind we introduce the alternative “Backtesting Friendly Generalized Diffusion for Risk Factors” which includes a couple of additional parameters, κ_θ and κ_W aiming at allowing a more conservative approach to risk by widening the distribution generated by equation (5.1).

Definition (Backtesting Friendly Generalized Diffusion for Risk Factors): We consider a risk factor process $(X_t)_{t \geq 0}$ with natural filtration $(\mathcal{F}_t)_{t \geq 0}$, and we define the forward risk factor process $(F_t)_{t \geq 0}$ by $F_t := \mathbb{E}(X_t | \mathcal{F}_0)$ generated by equation (5.2).

$$dX_t = \frac{\theta_t}{S_{l,t\theta}}(\mu_t - X_t)dt + \kappa_W \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t, \quad (5.2)$$

¹This is a jargon used sometimes in the industry to mean historical deviations from a model.

where θ , the rolling speed of mean reversion, μ is the long term rolling mean, α in the positivity flag enforcer, $\beta \in [-1, +1]$ boundary flag enforcer and $\{\bigcup dW_i\}_{i=t-\tau}^t$, the set of historical deviation of your assumed model's distribution (e.g. all the historical absolute bumps in the context of a normal diffusion), with τ is the rolling window length (e.g. three years), meaning that all of the primary parameters will now be function of this rolling window, κ_θ is the speed of mean reversion dampener, κ_W is the variance enhancer and μ_t function of a constant drift μ .

5.2.2 Calibration

As we will see in this section the calibration of the parameters of our bumping model from equation (5.2) comes in 4 steps. First we need to choose our assumptions and start with the model of (5.1) instead². This is where we decide whether the risk factor ought to be mean reverting or not, enforce positivity or not, enforce being in $[-1, +1]$ or not. Second we need to carefully choose the order of the sequential estimation of the primary parameters. Third, we need to fine tune the backtest thanks to the secondary parameters. And last we need to make the adjustment of the λ parameter in the Responsible VaR equation (5.11f) to fit the manager's risk appetite³.

5.2.2.1 Choice of Model Assumption

We recommend the following flags to be enforced according to the risk factor. For any, assumed, non mean reverting random process which has to stay positive (e.g. spot), in equation (5.1), set $\theta = 0$, $\alpha = 1$, $\beta = 0$ and we obtain the classic Log-Normal model. For any volatility related mean reverting bumps where we enforce positivity, set $\theta \neq 0$, $\alpha = 1$, $\beta = 0$. For interest rate assume, mean reverting bumps where we do not enforce positivity (like in the case of the OU [120] diffusion), set $\theta \neq 0$, $\alpha = 0$, $\beta = 0$. For assigning a diffusion on correlation itself or the ρ parameter in the SVI, gSVI, IVP [100, 12, 13], that is mean reverting bumps bounded in $[-1, 1]$, set $\theta \neq 0$, $\alpha = 0$, $\beta = 1$. For assigning a diffusion on the minimum of the moneyness axis in an implied Vol, we recommend a mean reverting process which can go negative and positive, set $\theta = 0$, $\alpha = 0$, $\beta = 0$ and we obtain the model with an underlier following the normal assumption. These choices for model assumption have been summarized in Table 5.1.

²We assume that $\kappa_\theta = \kappa_W = 1$.

³We see this step more in details in Section 5.6.4.

Distribution	θ	α	β	Risk Factor
Normal	$= 0$	$= 0$	$= 0$	Non Mean reverting Spread
Log-Normal	$= 0$	$= 1$	$= 0$	Equity
CIR-like	> 0	$= 1$	$= 0$	Positive Rates
OU Process	> 0	$= 0$	$= 0$	Positive & Negative Rates
$[-1, 1]$ Bounded OU Process	> 0	$= 0$	$= 1$	ρ Parameter in IVP

Table 5.1: Generalized SDE to known SDEs

5.2.2.2 Calibration of the Primary Parameters

If mean reversion is not enforced, the calibration of the model is trivial. If this is not clear we invite the reader to go back to equation (5.1) and think more about what each parameter does. Once we are in the context of mean reversion, we need to calibrate in sequence $\hat{\mu}$, $\hat{\theta}$ and finally the errors from the model $\{\bigcup d\hat{W}_i\}_{i=t-\tau}^t$. The calibration of μ is straight forward, done by equation (5.3).

$$\hat{\mu}_{t,\tau} = \mathbb{E}[X_{t,\tau}|\mathcal{F}_{t,\tau}] = \frac{1}{N} \sum_{i=t-\tau}^t X_i, \quad (5.3)$$

where $N = \text{card}\{\tau, \tau+1, \dots, t-1, t\}$ ⁴. The parameter θ happens to be tricky. Indeed if one rearranges equation (5.1), we obtain $\theta_{t,\tau} = \frac{dX_t - \sigma X_t^\alpha (1-X_t^2)^\beta dW_t}{(\mu_{t,\tau} - X_t)dt}$. However if one was to take all the available samples for θ and compute $\mathbb{E}[\theta_{t,\tau}|\mathcal{F}_{t,\tau}]$ like in equations (5.3), the estimation would quickly be dominated by instances where $\mu_{t,\tau}$ is very close to X_t and where θ “explodes” as a consequence. This naive approach creates a random bias in the estimation of $\theta_{t,\tau}$, which needs to be avoided. A solution to this issue is to incorporate a variance reduction technique. The idea comes from noticing the explosion effect described previously and in [10]. In the proposed methodology we deliberately choose to neglect zones in which the explosion is highly likely, for instance as it is described by figure 2.8. The relevant zones are $B_+ = |\frac{\max(X_i, i \in [t-\tau, t])}{2}|$, and $B_- = |\frac{\inf(X_i, i \in [t-\tau, t])}{2}|$ ⁵. These zones can be visualized in figure 2.8. In this figure, Z_θ represents situation in which sampling θ is very likely to be of high quality. When

⁴which is essentially the number of sample used to estimate μ .

⁵We note that the estimation of θ is noised when $Z_\sigma = B_+ > |X_i, i \in [t-\tau, t]| > B_-$. The reverse is true when $Z_\theta = |X_i, i \in [t-\tau, t]| > B_+ \cup |X_i, i \in [t-\tau, t]| < B_-$, sampling θ is a good idea. We will therefore sample θ in Z_θ .

we apply this idea of variance reduction technique, we obtain

$$\hat{\theta}_{t,\tau} = \frac{1}{n} \sum_{i \in Z_\theta} \frac{\Delta X_i - \sigma X_i^\alpha (1 - X_i^2)^\beta \Delta W_i}{(\hat{\mu}_{t,\tau} - X_i)} 1_{X_i \in Z_\theta} \quad (5.4)$$

as for a proper estimation of θ , where n is the cardinality of the set of all instances in which we sampled in the Z_θ zone. In practice doing this average over the 40th and 60th percentiles provides enough data and filters out enough outliers to make a better quality estimator than without the variance reduction technique. Once μ and θ have been estimated, we need to estimate the deviation from the model, this is done by isolating the shocks (i.e, increments of the Brownian motion) in equation (5.1). The following equation provides this estimation for the errors term.

$$\Delta \hat{W}_i = \frac{\Delta X_i - \hat{\theta}_{t,\tau} (\hat{\mu}_{t,\tau} - X_i)}{X_i^\alpha (1 - X_i^2)^\beta}. \quad (5.5)$$

5.2.2.3 Secondary Parameters in the Anticipative VaR

Equation (5.2) has three secondary parameters. The calibration of these secondary parameters are, as we will see, very much qualitative in approach and geared towards practitioners rather than pure probabilist. The first secondary parameter, τ , represents the “*rolling window of interest*”. This window can be chosen so as to either satisfy the regulatory constraints on model selection or/and based on how well your backtest performs with respect to your risk appetite. The second secondary parameter, from equations (5.2) and (5.1) is κ_θ . The latter can be, intuitively understood to be the “*elastic aging factor*”. The larger κ_θ , the more the speed of mean reversion estimated by the assumed model would get weaker and the more we converge towards a drift-less model (classic models). This is particularly useful in situation in which the risk manager believes that the long term mean reversion factor should be weaker than what is estimated by the data. The third and last secondary parameter, κ_W , can be thought of the “*returns beef-upper*”. It was created so as to create more conservative (but more penalizing at the same time) risk engines to match different risk appetite especially considering that this is a new proposed model.

5.2.3 Benefits

There are multiple benefits from choosing such generalized stress scenario formula. First it is versatile as it models all the known risk models on top of new ones. Second it is deployable and robust: once the estimation has been performed the same code works for every risk factor. Third it is leading as opposed to lagging as it allows for

anticipation in the regime change as opposed to waiting passively for responding to a regime change. Fourth it is more realistic from mean reverting risk factors (e.g. implied volatility, short term interest rates). This is because when these risk factors are high, applying relative shifts (the classic model) overestimates the moves on the upside but underestimates the moves on the downside. Finally it decreases arbitrage scenarios: since the diffusion of equation (5.1) is more realistic with respect to market observable phenomenon which are in turn function of arbitrage conditions that can occur at the portfolio level (implied volatility mean reverts).

5.2.4 Drawbacks

The choice of the boundaries of some these assumption is quite arbitrary. For instance if we had chosen the CIR like diffusion (in which $\theta > 0$, $\alpha = 1$ and $\rho = 0$) in 2014, our risk engine would not have been able to accommodate the regime change of 2015 in which we started entering the bizarre world of negative interest rates. There are also lots of practitioners who have difficulties understanding the concept of assumed distribution. More specifically the idea conveyed by the α and β parameters and the resulting diffusions induced by equation (5.1) is beyond many practitioners comfort zone as many feel that the mean reversion component introduces a bias in the estimation of the risk distribution. There are also some challenges in the calibration process even when considering the variance reduction techniques described in the Section 5.2.2.

5.3 The Machine Learning Approach

As we have seen from the previous Section 5.2 the selection of the parameters, their calibration and the practitioners culture makes conducting Risk Management using the SDE approach challenging on several levels. We introduce in this section a solution which we will show improves the status quo on many fronts.

5.3.1 Mathematical Specifications

We introduce first the concept of a simple band-wise clustering methodology to mimic some of the SDE properties.

Lemma 5.3.1 (Statistic of Order To Gaussian Mixture) *Let $R = \{x_1, \dots, x_n\}$ be a set of empirical random variables taken from equation 5.1 with cumulative distribution function $F(x)$ and density $f(x)$. Let $O = \{x_{(1)}, \dots, x_{(n)}\}$ be the ordered set of*

R such that $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ and $O_p^i = \{x_{(\lceil (n(i-1)+1)/p \rceil)}, \dots, x_{(\lfloor n(i)/p \rfloor)}\}$. Then the empirical distribution function for an SDE of the form of equation (5.1) can be approximated by a union of band-wise Bernoulli process given by

$$\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n} \sum_{j=1}^p \sum_{i=\eta}^{\zeta} \mathbf{1}_{x_i \in O_p^j} \quad (5.6)$$

with $\eta = \lceil n((i-1)+1)/p \rceil$ and $\zeta = \lfloor n(i)/p \rfloor$.

Remark In the case where $p = 3$, using a Gaussian Mixture such that $\hat{F}_n(x_i|\mathcal{F}_t) = \mathcal{N}(-3, 1)\mathbf{1}_{x_i \in O_3^1} + \mathcal{N}(0, 1)\mathbf{1}_{x_i \in O_3^2} + \mathcal{N}(3, 1)\mathbf{1}_{x_i \in O_3^3}$, we obtain the approximate stratification shown in figure 2.8. The stratification, in our case, is such that the cardinality in each O_p^j is approximately the same, as opposed to being the result of a geometrical separation function of $x_{(1)}$ and $x_{(n)}$. Figure 5.1 illustrates this remark.

Lemma 5.3.2 (SDE to p-Gaussian) *The distribution given by equation 5.6 converges towards a p-Gaussian Mixture.*

Proof The indicator function $\mathbf{1}_{x_i \in O_p^j}$ is a Bernoulli random variable with parameter p , and because the sum of Bernoulli random variable is Binomial distributed, $\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n} \sum_{j=1}^p \sum_{i=\eta}^{\zeta} \mathbf{1}_{x_i \in O_p^j}$. We can also see that in equation (5.1) that the $\lim_{n \rightarrow \infty, p \rightarrow \infty} (\mu_{t,\tau} - X_t) = \lambda_{t,\tau}$ and therefore $dX_t - \lambda_{t,\tau} = \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t$ becomes a locale martingale. Using Glivenko-Cantelli theorem [121, 122], we get $\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$. Therefore, the distribution from equation (5.1) can be approximated by $\cup_{i=1}^p \mathcal{N}(\lambda_i, \sigma_i)$. ■

Remark by looking at the difference between Figure 5.1 and Figure 5.2, we see how increasing the value of p can lead to a smoothing probability distribution function.

Theorem 5.3.3 *Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{Q})$ be our probability space, with $(\mathcal{F}_t)_{(t \geq 0)}$ where \mathbb{Q} is the risk neutral probability measure. The probability distribution $f(x|\mathcal{F}_t)$ induced by the SDE $dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t$ converges almost surely towards a p-Gaussian mixture as n and p converge towards infinity.*

Proof We split the proof in 2 steps. Indeed, lemmas 5.3.1 and 5.3.2 have been introduced in order to make the proof for this theorem more readable. In lemma 5.3.1 we have shown how ordering the returns can allow us to define zones where the returns are assumed to be sampled approximately through the same distribution. And in lemma 5.3.2 we have shown that in discrete time the SDE behaves in well defined limited zones like a martingale. ■

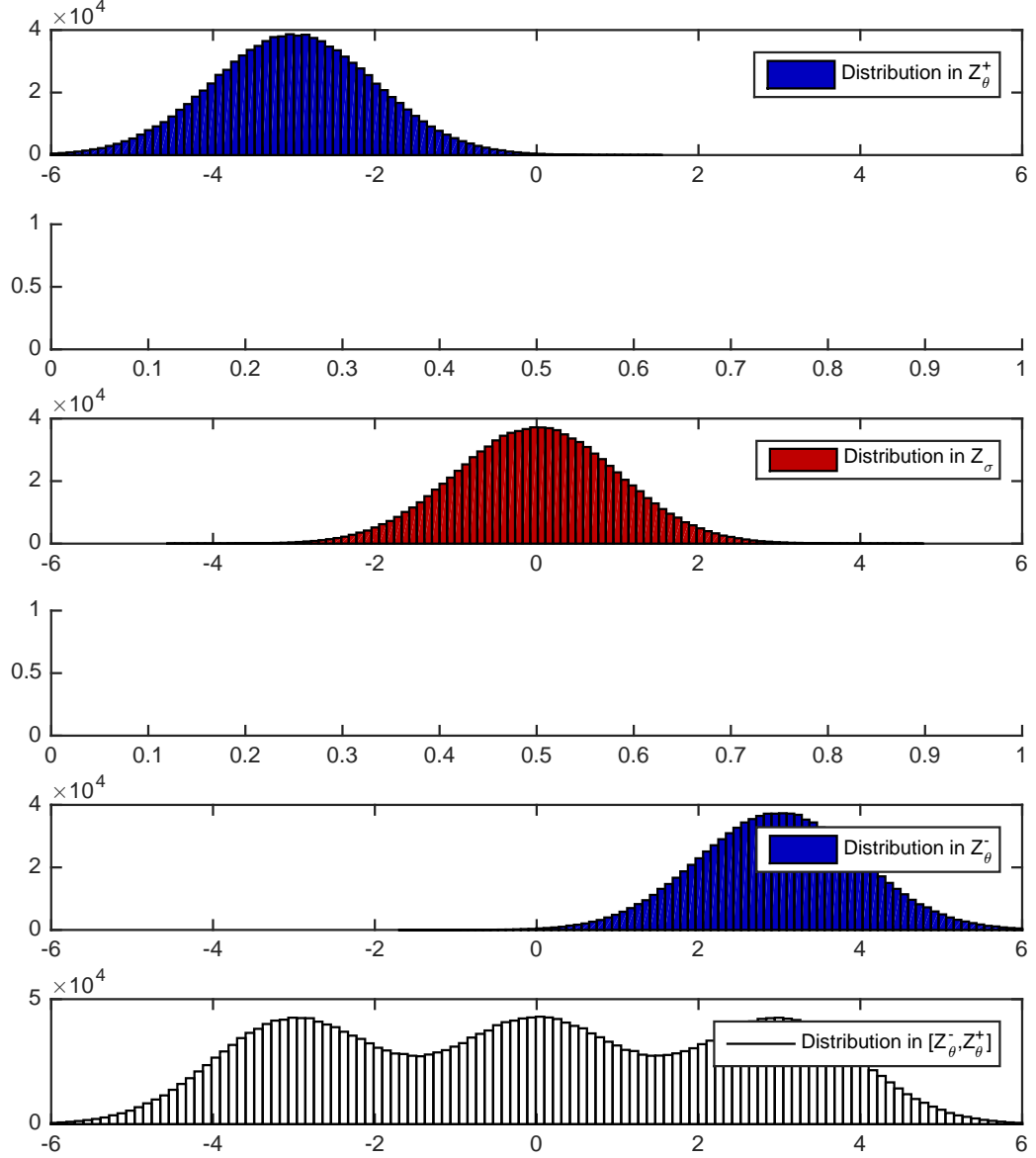


Figure 5.1: Gaussian distribution mimicking approximately Figure 2.8. The top figure, represents a zone in which our forecast by the mean reverting SDE is significantly above its historical mean (therefore the forecasted distribution is biased on the downside). The bottom figure represents the zone in which our mean reverting SDE is significantly below its historical (therefore the forecasted distribution is biased on the upside). In red we have the forecasted distribution which is centered around zero because we are at the historical long term mean

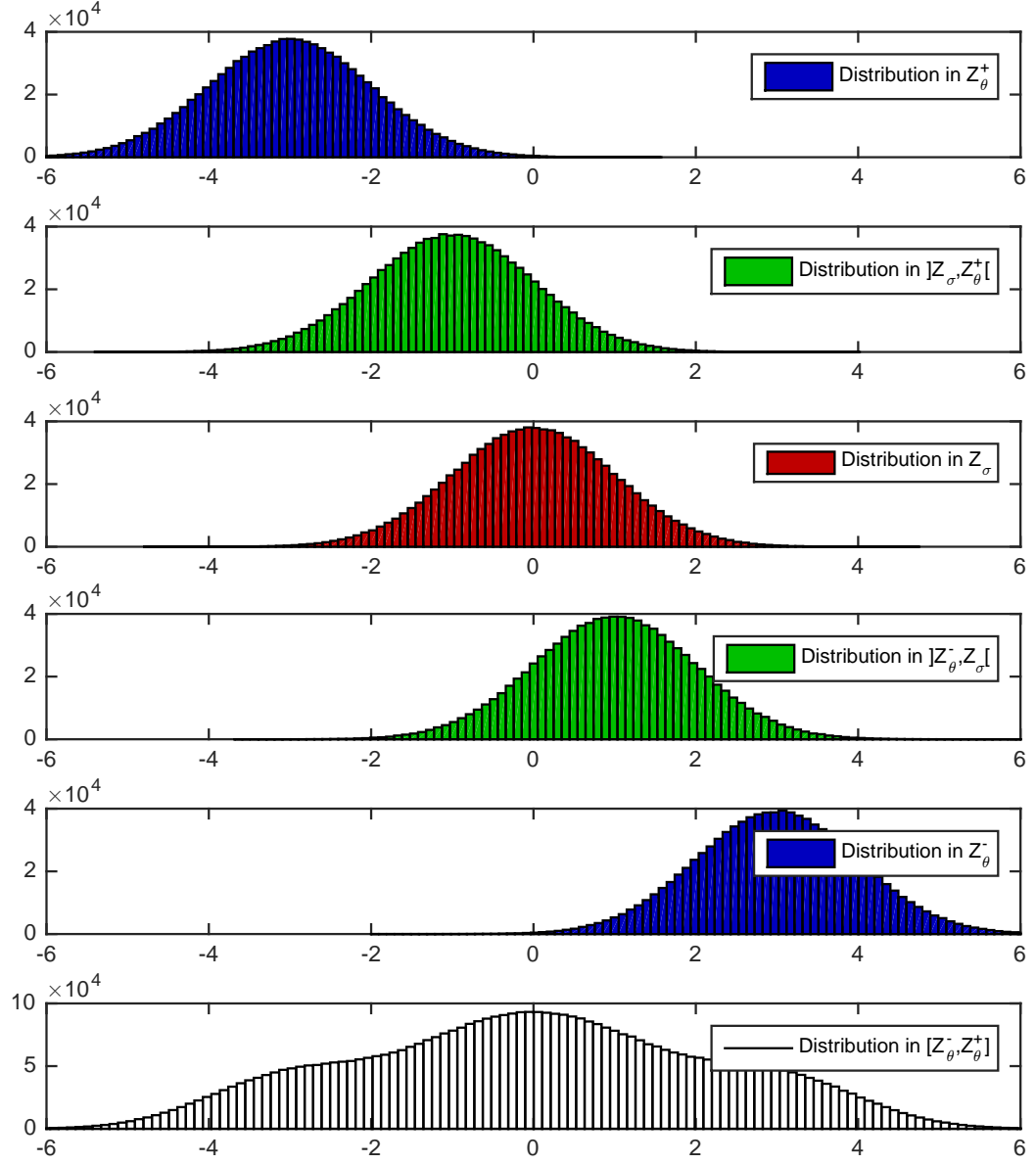


Figure 5.2: Exact same situation as in figure 5.1 but with two additional zones (in green)

5.3.2 Calibration

Note that we have included few simulations in section 2.4.9.

5.3.2.1 The Bayesian set-up

Let $B_t = \{B_{n,t}^+, B_{n-1,t}^+, \dots, B_{1,t}^+, B_{1,t}^-, \dots, B_{n-1,t}^-, B_{n,t}^-\}$, such that $B_{n,t}^+ > B_{n-1,t}^+ > \dots > B_{1,t}^+ > 0 > B_{1,t}^- > \dots > B_{n-1,t}^- > B_{n,t}^-$. We study X_t of equation (5.1) in the context of this Bayesian construction. We have seen that depending on the spread, the resulting approximated distribution of the samples differ [7]. The calibration algorithm consists of creating as many zones as possible whilst converging to the results from Theorem (SDE to Band-wise Gaussian Mixture) of page 140 introduced originally in [7].

5.3.2.2 A straightforward learning algorithm

The idea behind the calibration of the Band-Wise Gaussian Mixture is similar, though not exactly the same as to the variance reduction technique of Figure 2.8 explained in the same section. Namely, depending on the selected zone, the resulting approximated distribution of the samples differ. The calibration algorithm consists of creating as many zones as possible trying to converge to the results from the theorem 5.3.3. Algorithm (4) is a rough pseudo-code for the calibration process.

Remark Note that in algorithm (4), we use a QuickSort, which can be substituted by other sorting algorithm. We invite the motivated readers to investigate on their own this idiosyncratic issue. Also note that this algorithm has neither been optimized nor checked for data quality (e.g. the combination of n and p should be such that each band has enough data (e.g. minimum 30) for the statistical estimators to be significant).

5.3.3 Benefits

One important point to note about the additional benefits of the Machine Learning approach over the SDE approach⁶ is to take a look at the example associated to the way LCH was handling interest rates in Europe and USA up to 2014. Indeed, it was assumed that interest rates, in Europe and USA⁷ could never become negative. Risk managers, at LCH, would assume a $\beta = 0$ and an $\alpha = 1$ (in equation 5.1) in the SDE

⁶Beyond the obvious benefits associated to achieving the same results though through a simpler channel and also bypassing convoluted SDE calibration issues in the process.

⁷Note that interest rates had been negative in Japan for many years prior to that.

Algorithm 4: Band-Wise Gaussian Mixture

Input: array $X_{1:n}$ and number of bands p

Output: $\Omega^{(1:p)}$, $[B_{(1:p)}^+, B_{(1:p)}^-]$ are returned.

Sorting state:

$X_{(1:n)} \leftarrow \text{QuickSort}(X_{1:n})$

$[B_{(1:p)}^+, B_{(1:p)}^-] \leftarrow \text{FindPercentileBands}(X_{(1:n)}, p)$

$\Omega^{(1:\lceil n/p \rceil)} \leftarrow []$

Allocation state:

for $j = 1$ **to** p **do**

for $i = 1$ **to** n **do**
 if $B_{(1:p)}^- \leq X_{(i)} < B_{(1:p)}^+$ **then**
 Amend($\Omega^{(j)}$, $X_{(i)}$)

Checking Approximation state:

$\hat{\mu}_{1:p} \leftarrow \text{mean}(\Omega^{(1:p)})$

$\hat{\sigma}_{1:p} \leftarrow \text{stdev}(\Omega^{(1:p)})$

Print($\cup_{i=1}^p \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i)$)

Return state:

$\Omega^{(1:p)}$, $[B_{(1:p)}^+, B_{(1:p)}^-]$

approach: the latter two parameters enforcing positivity for the simulated scenarios of our risk factor. This very reasonable assumption (at least up to that point) crashed the whole risk engine in the most important clearing house in the world. The ML approach would have however been able to continue its dynamical learning scenario without any problem⁸.

5.4 Application to Pairs Trading

Remark Note that some of this material appeared in [5] but the thesis structure was such that part is instead presented in chapter 2 and more specifically Section 2.4 on portfolio optimization for cointelated pairs which the reader should get familiar with.

Now we present the trading signal that translates to investment strategy in machine learning approach. We set, from equation (2.5) $\Delta_t = X_t - Y_t$, assume for now that $r > 0$ and $B_t = \{B_{n,t}^+, B_{n-1,t}^+, \dots, B_{1,t}^+, B_{1,t}^-, \dots, B_{n-1,t}^-, B_{n,t}^-\}$, such that $B_{n,t}^+ > B_{n-1,t}^+ >$

⁸This point of view was not shared with all the members of the jury during the PhD defense [94].

$\dots > B_{1,t}^+ > 0 > B_{1,t}^- > \dots > B_{n-1,t}^- > B_{n,t}^-$. We have seen that depending on the spread, the resulting approximated distribution of the samples differ [43]. The calibration algorithm consists in creating as many zones as possible and testing for strategy optimality within these bands. For instance, in a situation in which we take a *direct* approach (see Remark 5.4) consisting of 4 strategies:

- Strategy S^{++} in which we are long both X and Y at time t in between bands $[a, b]$ and with P&L $V_{[a,b],t}^{++}$.
- Strategy S^{+-} in which we are long X and short Y at time t in between bands $[a, b]$ and with P&L $V_{[a,b],t}^{+-}$.
- Strategy S^{-+} in which we are short X and long Y at time t in between bands $[a, b]$ and with P&L $V_{[a,b],t}^{-+}$.
- Strategy S^{--} in which we are short both X and Y at time t in between bands $[a, b]$ and with P&L $V_{[a,b],t}^{--}$.

Where the P&Ls are defined as follows:

$$\begin{aligned} V_{[a,b],T}^{++} &= \sum_{t=0}^T [w_{[a,b],t}^{++} \Delta X_t + (1 - w_{[a,b],t}^{++}) \Delta Y_t] 1_{a < \Delta_t \leq b}, \\ V_{[a,b],T}^{+-} &= \sum_{t=0}^T [w_{[a,b],t}^{+-} \Delta X_t - (1 - w_{[a,b],t}^{+-}) \Delta Y_t] 1_{a < \Delta_t \leq b}, \\ V_{[a,b],T}^{-+} &= \sum_{t=0}^T [-w_{[a,b],t}^{-+} \Delta X_t + (1 - w_{[a,b],t}^{-+}) \Delta Y_t] 1_{a < \Delta_t \leq b}, \\ V_{[a,b],T}^{--} &= \sum_{t=0}^T [-w_{[a,b],t}^{--} \Delta X_t - (1 - w_{[a,b],t}^{--}) \Delta Y_t] 1_{a < \Delta_t \leq b}. \end{aligned}$$

Remark We call this approach *direct*, since ideally the number of strategies should consist of a more granular weight distribution. However for the sake of this example we wanted to keep the explanation more intuitive.

We define the maximum P&L achieved by each of these strategies by $V_{[a,b],T}^{\mp\mp,*}$, as given by equation (5.8) and define $S_{[a,b],T}^{**}$ of P&L $V_{[a,b],T}^{**}$ in equation (5.9), the optimal strategy using Gaussian Learning in band $[a, b]$.

$$V_{[a,b],T}^{\mp\mp,*} = \arg \max_{w_{[a,b],t}^{\mp\mp} \in [0,1]} V_{[a,b],T}^{\mp\mp}, \quad w_{[a,b],t}^{\mp\mp} \in [0,1], \quad (5.8)$$

$$V_{[a,b],T}^{**} = \max(V_{[a,b],T}^{++,*}, V_{[a,b],T}^{+-,*}, V_{[a,b],T}^{-+,*}, V_{[a,b],T}^{--,*}). \quad (5.9)$$

Algorithm 5: GMforCointelatedPairs(P, h)

Input: array $P_{1:n}$ and number of bands h

Output: $\Omega^{(1:h)}, [B_{(1:h)}^+, B_{(1:h)}^-]$ are returned

▷ Sorting state

$P_{(1:h)} \leftarrow \text{QuickSort}(P_{1:n}), [B_{(1:\frac{h}{2})}^+, B_{(1:\frac{h}{2})}^-] \leftarrow \text{FindPercentileBands}(P_{(1:n)}, h)$

$B_{(1:h)} \leftarrow [B_{(1:\frac{h}{2})}^+, B_{(1:\frac{h}{2})}^-], \Omega^{(1:\lceil n/h \rceil)} \leftarrow []$

▷ Allocation state **for** $j = 1$ *to* h **do**

for $i = 1$ *to* n **do**
if $P_{(i)} \in B^i$ **then**
Amend($\Omega^{(j)}, P_{(i)}$)

▷ Optimize the 4 types of P&L for each band

for $i = 1$ *to* h **do**

$V_{B_i,T}^{++,*} \leftarrow \arg \max_{w_{B_i,t}^{++}, t \in [0,T]} V_{B_i,T}^{++}, V_{B_i,T}^{+-,*} \leftarrow \arg \max_{w_{B_i,t}^{+-}, t \in [0,T]} V_{B_i,T}^{+-}$
 $V_{B_i,T}^{-+,*} \leftarrow \arg \max_{w_{B_i,t}^{-+}, t \in [0,T]} V_{B_i,T}^{-+}, V_{B_i,T}^{--,*} \leftarrow \arg \max_{w_{B_i,t}^{--}, t \in [0,T]} V_{B_i,T}^{--}$

▷ Rank and return best strategy for each band

for $i = 1$ *to* h **do**

$V_{B_i,T}^{**} \leftarrow \max(V_{B_i,T}^{++,*}, V_{B_i,T}^{+-,*}, V_{B_i,T}^{-+,*}, V_{B_i,T}^{--,*})$ $S_T^* \leftarrow$
 $(S_{B_i,T}^{++,*}, S_{B_i,T}^{+-,*}, S_{B_i,T}^{-+,*}, S_{B_i,T}^{--,*})$
 $S_{B_i,T}^{**} \leftarrow \text{returnCorrespondingStrat}(V_{B_i,T}^{**}, S_T^*),$

▷ Forecasting & Return buy/sell signals

$\text{signal}^S, \text{signal}^{S_l} \leftarrow \text{forecast}(S_{B_i,T}^{**}, S_t, S_{l,t})$

We further provide algorithm (5) as the pseudo-code for the calibration process. Note that in both algorithm (4) and (5), we have used a QuickSort which can be substituted by other sorting algorithm. We invite the motivated readers to investigate on their own this idiosyncratic issue. Also note that, in order to keep the algorithm easy to read and only focused on the optimization process, the pseudocode does not include a section on data quality or on the optimality of the band sizes⁹. Also note the use of self explanatory functions such as *returnCorrespondingStrat*(x, y) which given the set of strategies and the P&L that maximized that strategy returns as its name indicates the corresponding strategy. The function *forecast*(x, y, z) takes as input the set of trained strategies and the current level of S_t and $S_{l,t}$ and returns a prediction of where the signals for the latter two should be. Finally the use of the arg max function in lines 13-16 can be replaced by a simple for loop but in the interest of not making the pseudocode too crowded we have kept it this way.

⁹E.g., the combination of n and p should be such that each band has enough data (e.g. minimum 30) for the statistical estimators to be significant.

Remark We have seen [43] that a reasonable risk manager or trader can assume the generalized SDE (5.1) with $\beta = 0$ and an $\alpha = 1$, in order to enforce positivity for the simulated scenarios of our risk factor. This very reasonable assumption would have crashed the whole risk engine. The approach we advocate would have, however, been able to continue its dynamical learning scenario without any problem.

Finally note that some simulations have been included in section 2.4.9.

5.5 Application To Risk Management

5.5.1 Margining: Classic & Anticipative Methodologies

Before going further, it would be useful to gain acquaintance with some of the practitioners technical jargon.

Definition (Full Revaluation) : Let N denote the total number of risk factors relevant to a portfolio P and R_i each of the relevant risk factors (e.g. exchange rate, interest rate, skew, vol of vol, ATM vol etc.) of this portfolio. Then define f , the function that takes all the historical scenarios¹⁰ as input and would revalue the portfolio's¹¹ risk distribution. The reported risk statistics such as the VaR or the expected shortfall are then function of these stressed scenarios.

5.5.2 Drift in Risk & Misconceptions

The best way to describe the margining methodology is to see how it works with the currently models. This will help later when it comes to introducing the concept of Anticipative VaR. Table 5.2 is a good tool when it comes to understanding the classic methodology under the log-normality¹² and normality¹³ assumptions. The resulting returns at the portfolio level would be given by $P\&L_{t,\tau} = f\left(S_t(1 + \Delta S_\tau), \Sigma_{E,K,t} + \Delta \Sigma_{E,K,\tau}, \dots\right)$. Note that in the function f , we have incorporated the symbol¹⁴ “...” to signify that the same methodology is used for all the relevant vol points for all the relevant tenors.

Remark The stressed scenarios may be cleansed prior the revaluation using de-arbitraging methodologies [12] but, Table 5.2 was designed to be a simple example to introduce next the concept of Anticipative VaR.

¹⁰ $\cup_{i=1}^N \cup_{\tau=1}^T R_{i,t} + \Delta_{t-\tau}^{R_{i,\tau}}$.

¹¹Note that $\tau \in [0, T]$ with T being the length of the available relevant data (e.g. 10 years).

¹²“We use proportional bumps”.

¹³“We use absolute bumps”.

¹⁴A convenient notation.

Date τ	$dS_t = S_t d\hat{W}_t^a$	$d\Sigma_{E,K,\tau} = d\hat{W}_t^b$	\cdots	P&L $_{t,\tau}$
$t - 1$	$\frac{\Delta S_{t-1}}{S_{t-1}}$	$\Delta \Sigma_{E,K,t-1}$	\cdots	-1.7%
$t - 2$	$\frac{\Delta S_{t-2}}{S_{t-2}}$	$\Delta \Sigma_{E,K,t-2}$	\cdots	+ 0.7%
\vdots	\vdots	\vdots	\vdots	\vdots
$t - 750$	$\frac{\Delta S_{t-750}}{S_{t-750}}$	$\Delta \Sigma_{E,K,t-750}$	\cdots	-1.4%

Table 5.2: Example of full revaluation table under classic Risk engines

5.5.3 Margining in the context of Anticipative VaR

Table 5.3 is the translation of Table 5.2 but in an Anticipative format: under the assumption of mean reversion using equation (5.2). Usually the part which is most

Date τ	$dS_t = S_t d\hat{W}_t^a$	$d\Sigma_{E,K,\tau} = \hat{\theta}(\hat{\mu} - \Sigma_{E,K,t})dt + \Sigma_{E,K,\tau} d\hat{W}_t^b$	\cdots	P&L $_{t,\tau}$
t-1	$\frac{\Delta S_{t-1}}{S_{t-1}}$	$\Delta \Sigma_{E,K,t-1}$	\cdots	-1.9%
t-2	$\frac{\Delta S_{t-2}}{S_{t-2}}$	$\Delta \Sigma_{E,K,t-2}$	\cdots	+0.8%
\vdots	\vdots	\vdots	\vdots	\vdots
$t - 750$	$\frac{\Delta S_{t-750}}{S_{t-750}}$	$\Delta \Sigma_{E,K,t-750}$	\cdots	-1.6%

Table 5.3: Example of full revaluation table under anticipative Risk engines

misunderstood in this methodology is how $\{\bigcup dW_i\}_{i=t-\tau}^t$ behaves in either of the models. The key in understanding this part is to realize that $\{\bigcup dW_i\}_{i=t-\tau}^t$ represent the “deviations/error” from an assumed model. For instance you could mix a model in which you try to capture the risk of a call and try to define the corresponding risk factor. You assume that the underlier follows a log normal distribution (you enforce $\theta = 0$ but $\alpha = 1$), you assume that the ATM vol is mean reverting and positive (you enforce $\theta \neq 0$ but set $\alpha = 1$), you assume interest rates are mean reverting but could be negative (you enforce $\theta \neq 0$ but $\alpha = 0$). The correlation at the infinitesimal level of these errors from these models $\{\bigcup dW_i\}_{i=t-\tau}^t$ will be preserved if the time stamp is preserved. The resulting returns at the portfolio level would be given by $P\&L_{t,\tau} = f\left(S_t(1 + \Delta S_\tau), \Sigma_{E,K,t-1} + \theta(\hat{\mu} - \Sigma_{E,K,t}) + \Delta S_\tau, \cdots\right)$. We invite the reader to take a look at the column labeled P&L $_{t,\tau}$ in both Table 5.2 and 5.3. The numbers in Table 5.2 have been chosen randomly but the ones from Table 5.3 are chosen in terms of what one would expect in terms of approximate difference had the vol been different from its historical mean. So the P&L is shifted because the risk is now

asymmetric. Figure 5.3 describes how the forecasted risk distribution behaves as a result of where the risk factor is positioned with respect to its long term mean.

5.5.4 Results & Rolling Conditional Distribution

We can observe how the historical deviations of equation (5.1) behaves as a function of where the stochastic process stands with respect to its long term mean. Indeed, in a situation where one assigns $\beta = 0$, $\alpha = 1$ for the ATM implied volatility for the 2 years expiry we obtain Figure 5.3. What this figure attempts at exposing is the

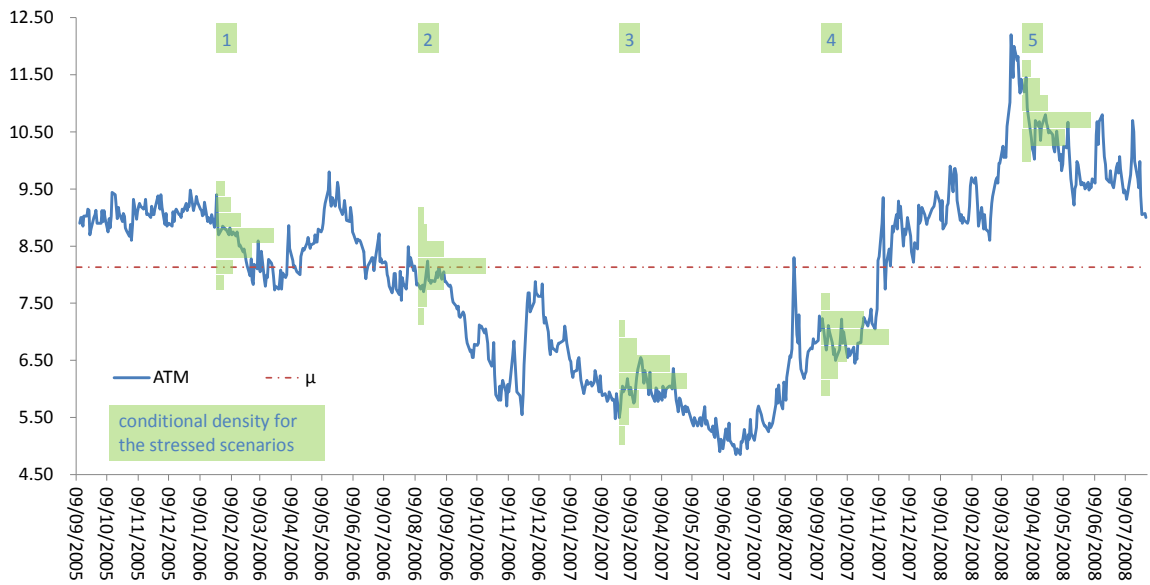


Figure 5.3: Example of forecasted distributions, in green under the Anticipative VaR hypothesis. In zones 1 and 5, the ATM is about the historical mean and therefore the forecasted distribution is biased on the downside. In zones 3 and 4, we have the exact opposite case and finally in zone 2 the distribution is symmetrical because the ATM is equal to the long term mean μ .

asymmetric/skewed behavior for the distribution of the generated stressed scenarios as a function of where the risk factor stands with respect to its long term mean¹⁵. Indeed, we can see in zones 3 and 4, that the risk factor is biased towards the upside whereas in zones 1 and 5 in which the risk factor is above its historical rolling mean, the simulated distribution is biased on the downside. Note here that in a situation in which you are below your historical mean, you can still have simulations that take you below your current value (the reverse is true for the symmetric situation in which you are above your historical mean). A second interesting point to note is that the

¹⁵We have assumed for display purposes that the mean will not be rolling in this graph but fixed.

more the risk factor is significantly above its historical mean, the more skewed is the resulting distribution (for example the distribution of zone 5 is more skewed towards the downside compared to the one of zone 2 because the deterministic side of the stochastic differential equation pressures the simulations more on the downside, even though it still allows moves on the upside). From these few zones, the reader can now understand the terminology chosen (“Anticipative”) for this new risk concept. This can be opposed to the concept of Responsive VaR in which one always takes a symmetric approach and re-scales the risk factors returns once a big market move has already occurred (so the VaR is Responsive as opposed to Anticipative).

5.6 Reconciling Responsive & Stable VaR

We reconcile in this section the concept of Responsive VaR and Stable VaR which are, as we have seen in the introduction two discordant concepts. They remain, however, equally desirable by practitioners.

5.6.1 Generating Risk Scenarios

The area of Risk scenarios generation is a complex one on its own. Choosing the underlying assumptions around the diffusion process has a great deal of influence on the Risk profile we obtain. More specifically, we see some of these challenges in details in Section 5. In this section we assume that we have generated these scenarios and that the task is now to make sure the risk profile is both responsive and stable.

5.6.2 Responsive VaR

Once the underlying assumptions of the stress testing scenario diffusions (for example: log-normal, normal or mean-reversion) has been decided the stressed scenarios need to be adjusted in order to address the market change of volatility, so as to get the label of “responsiveness”.

Definition (Responsive VaR): A VaR model that will be able to adapt, *a posteriori*, to increased volatility conditions will be referred to as Responsive VaR.

There are few ways to conceptually address this concept and they all rely on a scaling of long term volatility compared to recent volatility.

5.6.2.1 The Exponential Weighting Approach

An Exponential Weighting Approach (EWMA) is often used as a solution since the latter does not have an arbitrary cut off point but rather older data's contribution in the calculation of σ_c (the current volatility) must decrease exponentially. In this situation the unnormalized weights are calculated using $w_1 = 1$, $w_t = \lambda w_{t-1}$ and the normalized weights are given by $\tilde{w}_t = w_t / \sum_{i=1}^N w_i$. Though regulations change on a regular basis [44], at the time this thesis was written, the instructions were that $\sum_{i=1}^N w_i \geq 728$ [44] so λ is chosen in order to respect this regulatory constraint. The latter essentially means that the calculation of a measure of risk ought to be at least two years of data [44].

5.6.2.2 The Local to Historical Volatility Approach

We define σ_h from equation (5.10a) as being the standard deviation over our entire history of relevant data $[1, \dots, N]$, and σ_c from equation (5.10b) the standard deviation over more recent history, with $N > p > 1$, the responsive VaR formula is given by equation (5.10c).

$$\sigma_h = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}, \quad (5.10a)$$

$$\sigma_c = \sqrt{\frac{\sum_{i=p}^N (x_i - \bar{x})^2}{N - p - 1}}, \quad (5.10b)$$

$$\text{RVaR}_\alpha(X) = \frac{\sigma_c}{\sigma_h} \text{VaR}_\alpha(X), \quad (5.10c)$$

where α the quantile level and $\text{RVaR}_\alpha(X) = \frac{\sigma_c}{\sigma_h} \inf\{x \in \mathbb{R} : P(X + x < 0) \leq 1 - \alpha\} = \frac{\sigma_c}{\sigma_h} \inf\{x \in \mathbb{R} : 1 - F_X(-x) \geq \alpha\}$.

Remark Note that p in equation (5.10b) is a free parameter, which is usually the result of a constrained optimization problem in which the financial institution calculating its VaR tries to minimize its capital requirement¹⁶ with the constraints being set by the regulators in order to make the relevant statistics significant¹⁷.

5.6.3 Responsive vs. Stable VaR

Technical documents written by practitioners exposed the conflicting properties of Responsive and Stable VaR [27, 123].

¹⁶And chooses the best p to minimize that VaR.

¹⁷The Basel committee usually likes to see 2 years of data, therefore $N - p$ needs to be at least two years.

Definition (Stable VaR): A VaR model that remains robust will be referred to as Stable VaR.

Youngman's [27] is an example of how the duality between Responsive and Stable VaR are understood and used by practitioners. Figure 5.4 plots three graphs for the 99% VaR of triple B rated corporate bond (so a linear product) using 3 different lookback periods¹⁸. The VaR model in this situation is quite simple. We assume that

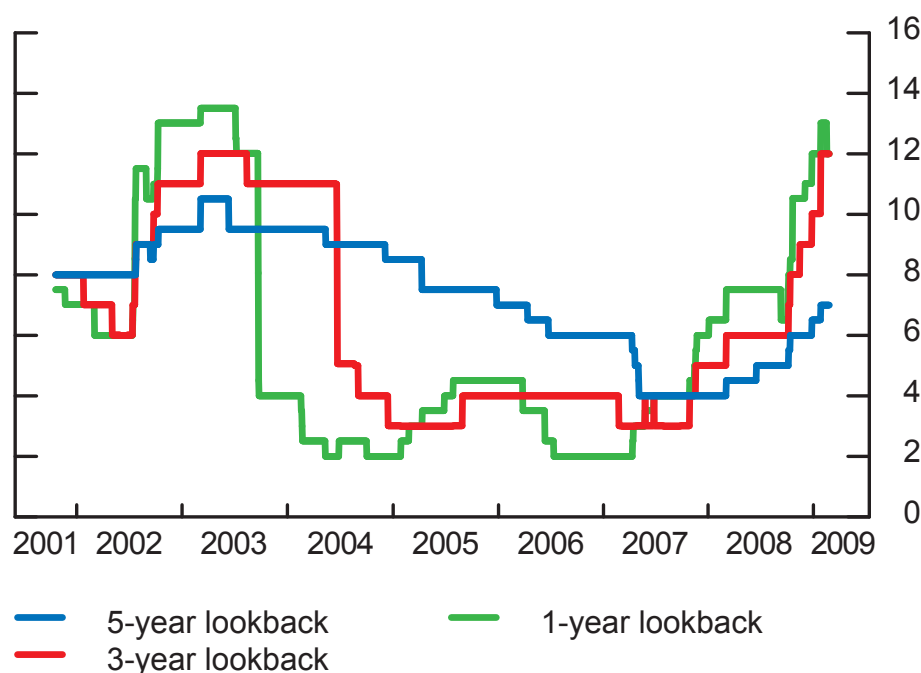


Figure 5.4: Youngman's [27] example of One-Day 99% VaR BBB Corporate Bonds.

the underliers follows a rolling log-normal distribution in which the rolling windows are, in this example 1 (in green), 3 (in red) and 5 (in blue) years. What we can see is that the green plot, which is the one with the shortest rolling window, happens to be the most responsive to market events whereas the blue plot (the biggest rolling window) happen to be the most stable. The rational is that, for VaR stability, the green plots fluctuates too much and can create liquidity congestion in the case where the market would get used to a low VaR market environment. For VaR responsiveness, the blue line is too conservative in low volatility environment and not reactive enough in situations of increased volatility.

¹⁸For example the p parameter in equation 5.10b.

5.6.4 Responsible VaR

Reconciling the conflicting properties of a Responsive VaR and Stable VaR is the second problem we solve by introducing to the concept of Responsible VaR.

Definition (Responsible): A VaR model that is **Responsive** on the upside and **Stable** on the downside will be called Responsible. Let ν_t represent the VaR at the α level at time t . Responsible VaR is jointly defined by the stochastic processes $\tilde{\nu}_t^+$ and $\tilde{\nu}_t^-$ summarized by

$$\alpha = \int_{-\infty}^{\nu_t^+} p_t(x) dx, \quad (5.11a)$$

$$\tilde{\nu}_0^+ = \nu_0^+, \quad (5.11b)$$

$$\tilde{\nu}_t^+ = \max \left(\nu_t^+, \lambda \tilde{\nu}_{t-1}^+ + (1 - \lambda) \nu_t^+ \right), \quad (5.11c)$$

$$1 - \alpha = \int_{\nu_t^-}^{+\infty} p_t(x) dx, \quad (5.11d)$$

$$\tilde{\nu}_0^- = \nu_0^-, \quad (5.11e)$$

$$\tilde{\nu}_t^- = \min \left(\nu_t^-, \lambda \tilde{\nu}_{t-1}^- + (1 - \lambda) \nu_t^- \right). \quad (5.11f)$$

where λ is a scalar that controls the stability of our risk measure on the downside.

Note the lower the λ , the more stable is the Responsible VaR. Figures 5.5, 5.6, 5.7 and 5.8 illustrate how λ impacts the risk measure. This methodology can in fact be used independently of whether we are in a Responsive or an Anticipative VaR context.

Remark Adaptive Smoothing Methods for non stationary data [124], though optimal mathematically are sometimes less useful in a challenging regulatory environment. The concept of Responsible VaR can be summarized by a system of two controlled Snell Envelopes of the VaR level at their respective quantile level. In a practical point of view one must record the VaR of a specific portfolio in time and record the instantaneous Anticipative or Responsive VaR and adjust it based on equation (5.11f). That particular last point may be deterrent in direct use if the IT constraints are not flexible or too slow.

This concludes the first part¹⁹ of the thesis in which we illustrated how data-driven models can be gently²⁰ “apposed” to classic FM models. In the next Part of the thesis²¹ we take the challenging opposite view and try to illustrate how the new models could be a revolution in QF.

¹⁹Part II.

²⁰where we presented Machine Learning as a granular enhancement tool to QF.

²¹Part III.

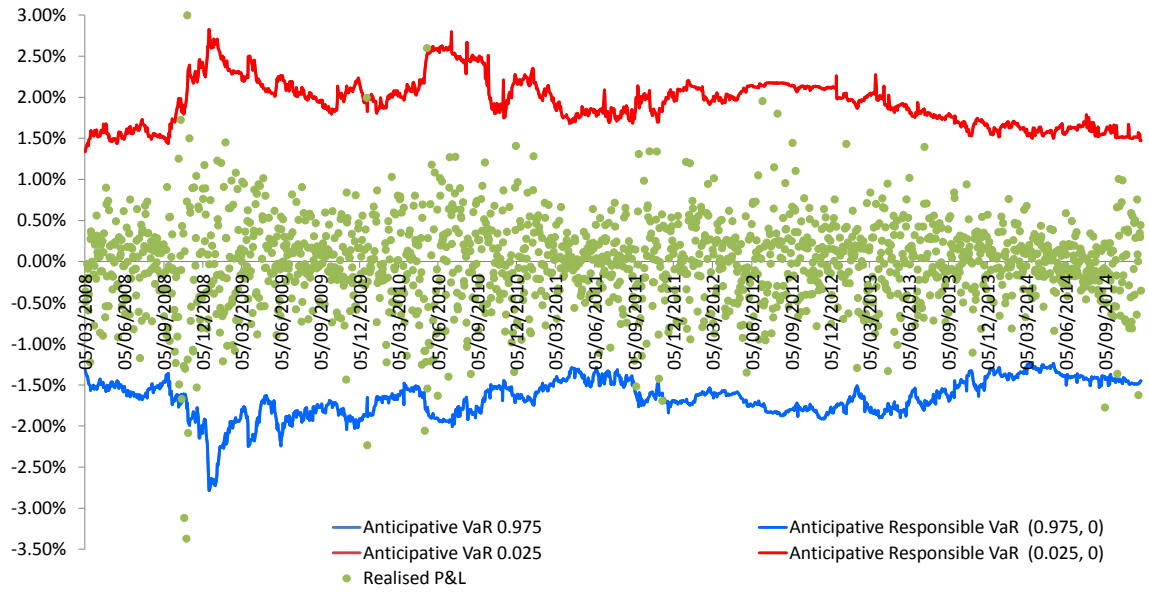


Figure 5.5: Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.000$

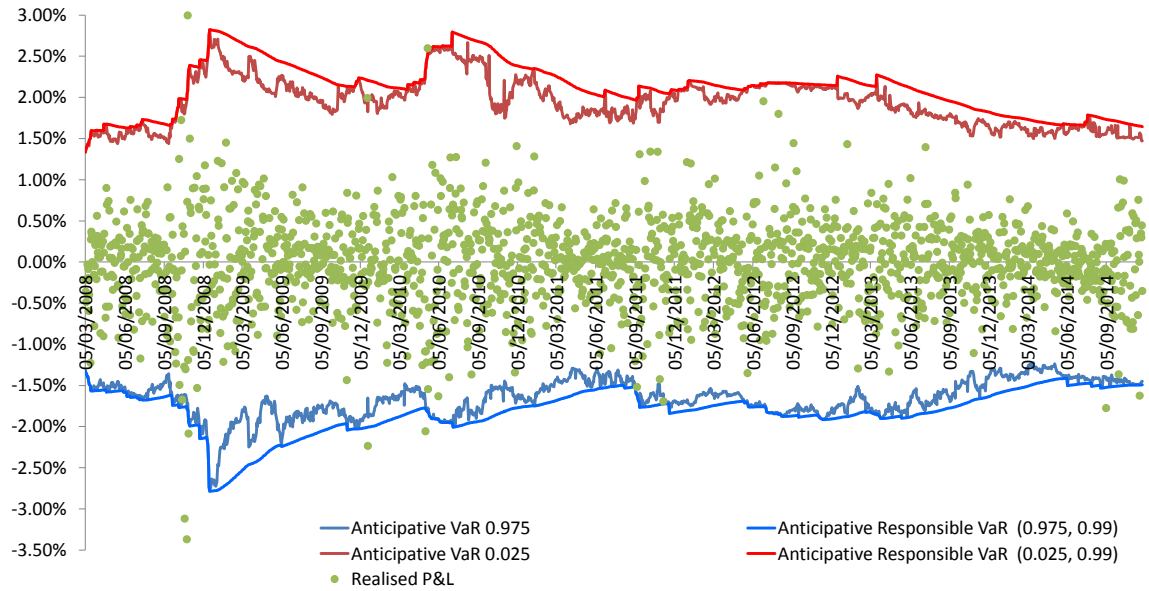


Figure 5.6: Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.990$

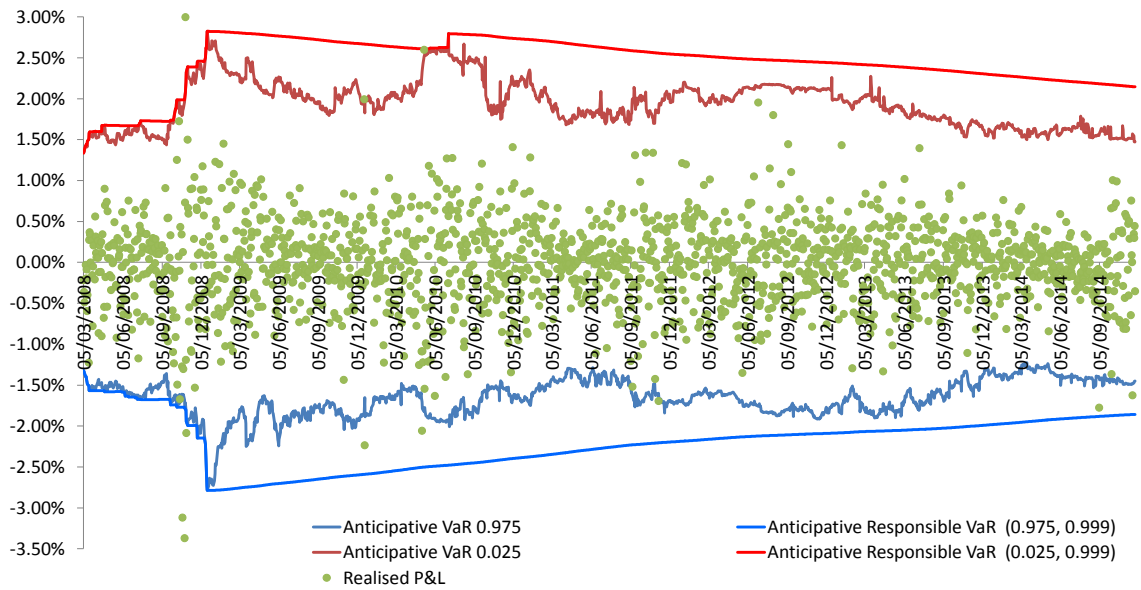


Figure 5.7: Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 0.999$

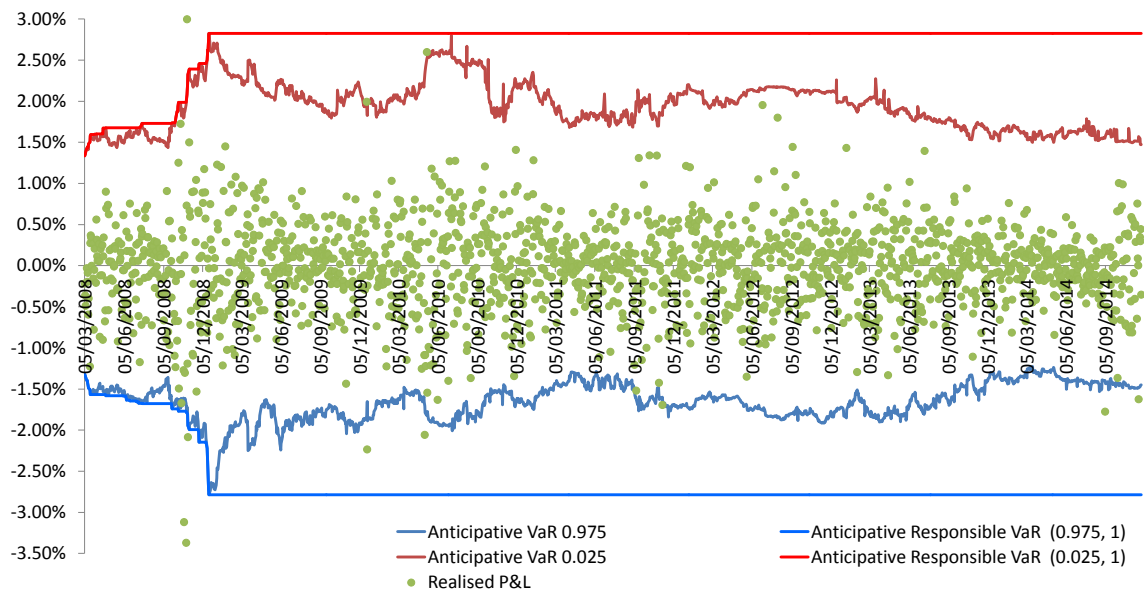


Figure 5.8: Anticipative Responsible VaR USD/EUR straddle backtest: $\lambda = 1.000$

Part III

From a Top-Down to a Bottom-Up Approach in Financial Modelling

Chapter 6

Agent-Based Intelligent System, from Shallow to Deep Learning

Following Bouchaud's¹ call for a revolutionary change in economics [16], taking a bottom-up approach via Agent-Based Models, instead of the traditional top-down approach², this section will focus on an example in which the traditional Financial Mathematics approach can be revolutionized by tools from Machine Learning and therefore provide an argument opposing these two fields rather than apposing them. Served more as an introductory chapter to the second part of the thesis, our aim here is to showcase methods that are associated to the other STEM fields that we believe have their space in twenty first century QF as well as translate them in QF jargon. More specifically, first we make the parallel with the scientific method used in Conway's Game of Life [52, 53] in Section 6.1. Then we delve into some formalization of Electronic Trading in Section 6.3. In Section 6.4 we go through a very brief literature review of Neural Networks in order to introduce the rational of certain types of architecture and their learning potential. In Section 6.5 we introduce the core DNA for our financial strategies and expose how this structure can model many of the well known financial strategies. Finally, in Section 6.6, in the context of Electronic Trading, we reflect on the relationship between architecture and meaning in order to explain the incentive for Deep Learning (DL). We use in that occasion the same methods used in adversarial algorithms in order to expose how Deep Learning can naturally result from simpler strategies in Shallow Learning.

¹Quant of the year 2017 and 2018.

²as best symbolized by the Brownian Motion assumption in Financial Mathematics

6.1 The Financial Game of Life

In chapters 6, 7 and 8 of this thesis we will take a methodological approach similar to the Game of Life, a well known 4 rules cellular automaton published by Conway [52] in the mid 70's³. More specifically, we inspire ourselves from the scientific method

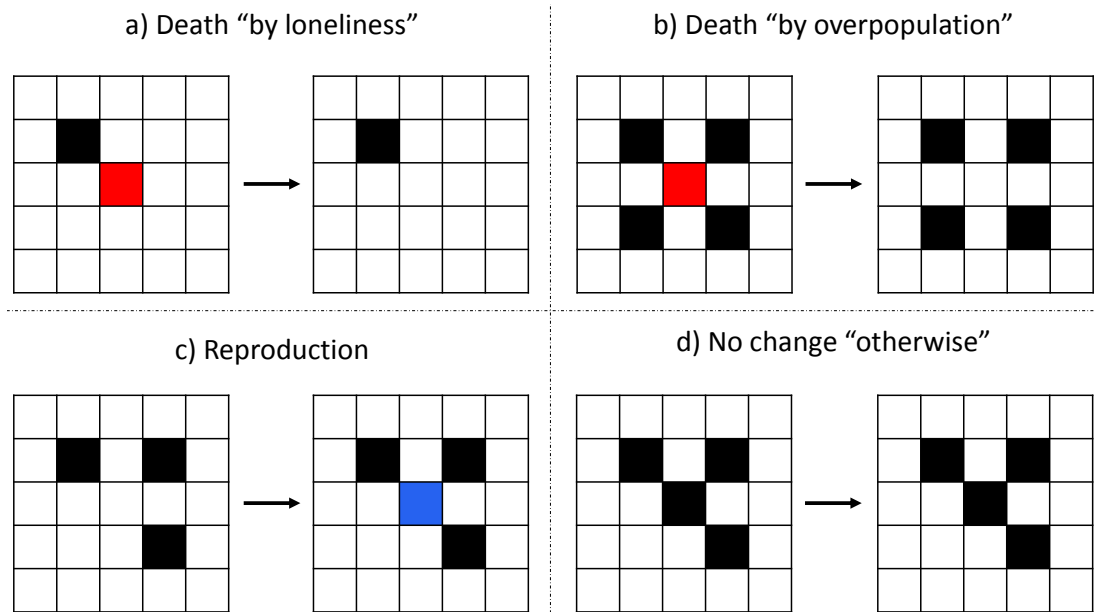


Figure 6.1: Conway's Game of Life rules illustrated

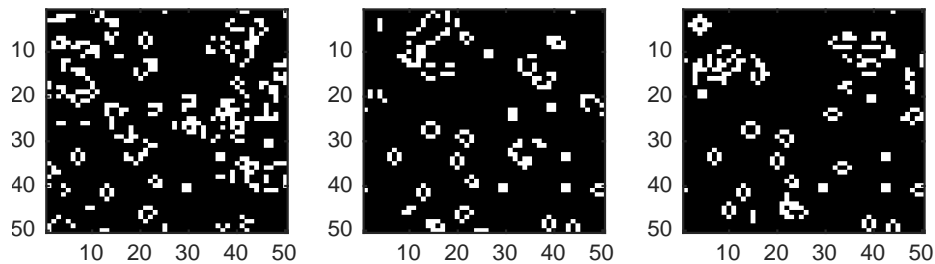


Figure 6.2: Three snapshots of a simulation of Conway's Game of Life

but use it instead to the world of High Frequency Trading (HFT) while adjusting some of the idiosyncratic parts of the exercise. As a reminder, Conway's Game of Life assumes that complexity in an ecosystem⁴ arises from simple rules. To illustrate

³Figure 6.1 provides the rules of the Automaton and Figure 6.2 represents 3 snapshot of one random simulation.

⁴In this thesis Ecosystem and Market are interchangeable since the former is taken to be an intuitive image of the latter.

this Simplicity-to-Complexity path, in the Game of Life, the four simple rules of Figure 6.1 can lead to many different families of complex automata. As a reminder, starting with random seeds and after few iterations, the simulation leads to stable, oscillating and moving forms. For stable forms⁵, the concept of financial stability may be raised through a similar methodology. The reader may guess that the concept of financial cycles or HF oscillations (Figure 1.1) may be induced through a similar methodology: for example in the Game of Life we also have oscillating forms⁶. Finally the moving forms⁷ may have different sizes and speeds⁸. The parallel to the world of quantitative financial strategies would be the following. First interacting agents lead to market price fluctuations. More specifically their interaction determines the stability or instability of the market⁹. Second, the market will not necessarily follow the rules of a zero-sum game¹⁰ (with, however, random seeds). Third agents (e.g. strategies) will follow a simple rule for their births and deaths.

6.2 Generative Adversarial Networks

Introduced in 2014 [54], Generative Adversarial Networks (GANs) are a relatively recent, but promising field in Machine Learning. Usually classified as among the algorithms used in unsupervised learning, they usually involve a system of two neural networks competing in a zero-sum game settings. More specifically these two networks, already trained with the same objective (e.g., being good at a strategy game for instance), but trained with slightly different data compete in a zero-sum game and this in turn becomes new data, sometimes called latent data, which in turn is used to improve the two networks. This process can continue as long as needed since the lack of data is no longer a problem. The £400 million buyout of DeepMind by Google popularized the model in 2014 [125].

⁵For example the “Block”, the “Beehive”, the “Loaf”, the “Boat”.

⁶For example the “Blinker (2 period iteration)”, the “Toad” (2 period iteration), the “Beacon” (2 period iteration), the “Pulsar” (3 period iteration), the “Pentadecathlon” (15 period iteration).

⁷for example the “Glider” and the “Lightweight spaceship” (LWSS)

⁸This latter family is arguably conceptually not really providing a useful comparison to our problem.

⁹Depending on what the market is made of in terms of the strategies involved as well as the evolving order-book.

¹⁰Meaning that its evolution is determined by its initial state, requiring no further input.

6.3 Electronic Trading

6.3.1 Description

An order book, traditionally consists of a list of orders that a trading venue such as an exchanges uses to record the market participants' interests in a particular financial product [126]. Typically a rule based algorithm records these interests taking into account, the price and the volume proposed (on either side of the Bid-Ask) as well as the time in which that interest was recorded (in situations in which interest at the same price is recorded by few different market participants, a referee decides which would win the trade: usually FIFO).

6.3.2 Variable Definition

Definition (Order-Book): We label by a_1^t and b_1^t the best ask and bid total volumes at time t . By extension a_i^t, b_i^t with $i \in \{1, 2, 3, 4\}$ would correspond to total volume at the relevant depths' of the order book with the special case where $i = 4$ which then would represent the total volume at the 4th depth level in addition to all the other market depths superior in price (in the case of the Asked price and vice versa for the bid price). We denote by m_t the midprice of the best bid/ask at time t . The price increment, at the different levels, l is usually 1bps¹¹. Equation (6.1) formalises the price as a function of the level. Figure 6.3 represents our version of the order book.

$$p_l^t = m_t[1 + (-1 \times 1_{l \in b_i^t} + 1 \times 1_{l \in a_i^t}) \times 0.001\%]^i \quad (6.1)$$

Remark We will assume that the Leading Indicators for the price process can only be taken from the order book which is a reasonable assumption in the higher frequencies. Some usually accepted leading indicators are the price of the underlier itself and the accumulated volume at different market depths of the order book (4 on the bid side and 4 on the ask side for a total of 9 leading indicators with the price process: see Figure 6.3 for visual representation).

6.4 Neural Net Architecture & Learning Potential

6.4.1 A Brief Qualitative History

In the spirit of explaining the complex through simple incremental steps, like in Econophysics [127], this particular subsection is dedicated to how complexifying sim-

¹¹bps stands for basis points or in terms of percentage: 0.01%.

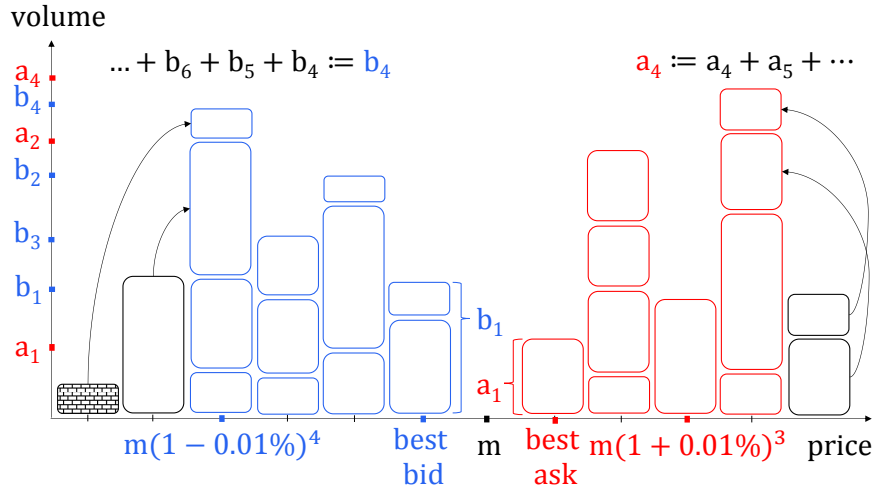


Figure 6.3: Order-book visual representation

ple Artificial Neural Network's (ANN) architecture with more hidden layers¹² or with feedback connection can lead to useful enhanced learning potential such as the one offered by Deep Learning (DL). More specifically, taking this approach allows us to slowly move towards Deep Learning and unweave¹³ the black box associated to the latter challenging¹⁴ mathematical concept. With this in mind, two well known, but important milestones in Machine Learning are worth reminding of. Especially for the beginners, these two milestones can shed light on why the core building blocks of our HFTE model is a certain way and also prepare intuitively the reader for sections 6.5 and 6.6. First, Warren McCulloch and Walter Pitts [128] introduced their threshold logic model in 1943 which is agreed to have guided the research in NN for more or less a decade. Second, Rosenblatt [129] formally introduced the perceptron concept in 1962 though some early stage work had started in the 1950s. The idea of the perceptron was one in which the inputs x_1 and x_2 as depicted in Figure 6.4 could act as separators¹⁵ and therefore a direct equivalence could be made to the Multi-Linear Regression (MLR) which we will provide more details in Section 6.5.3. One observed limitation of the perceptron, as described by Rosenblatt in 1969, was that a simple yet critical well known functions such as the XOR function could not be modeled [130]. This resulted in a loss of interest in the field until it was shown that a

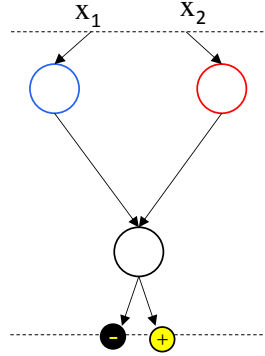
¹²“Deep Learning” is arguably just a fancy word for a Perceptron with many hidden layers.

¹³Note that the unweaving analysis goes forward instead of backward.

¹⁴The current status quo is that Deep Learning works for many applications but that we do not necessarily know why in the details.

¹⁵The exact research was one in which the methodology acted as a 1, 0 through a logistic activation function $f(x) = \frac{1}{1+e^{-x}}$ as opposed to a linear one. However that small distinction is not significant enough in this context to delve too much into it.

Network Structure for LR



LR function representation

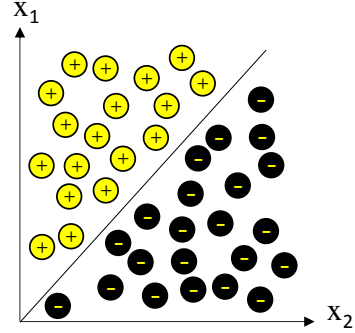
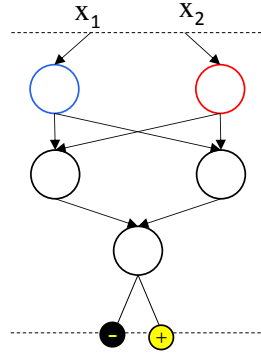


Figure 6.4: Simple Neural Network Modelling a Linear Regression

Feedforward Artificial Neural Network (ANN) with two or more layers could in fact model these functions (see Figure 6.5 for the illustration). Added, to this we have the

Network Structure for XOR



XOR function representation

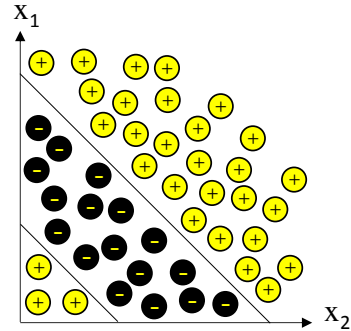


Figure 6.5: Feedforward ANN with 1 Hidden Layer & the XOR Function

well known over-fitting [131] problems when it comes to supervised learning which, to some extent would like to keep adding hidden layers when the learning potential has been absorbed. This problem of learning potential to over-fitting has been there since inception, though regular progress is being made in that domain without real breakthrough¹⁶. A real breakthrough happened, however, in 1956 with what is now known as Kolmogorov's superposition theorem [132] which we formalize next.

¹⁶We refer here the reader to area of ML known as Regularization.

6.4.2 Kolmogorov-Arnold's Superposition Theorem

Born in what is speculated as a heated supervisor/supervisee relation¹⁷, the Kolmogorov-Arnold's superposition theorem [132, 133] is perhaps the most remarkable result in formalized mathematical machine learning of the 20th century. It states that every multivariate continuous function can be represented as a superposition of continuous functions of two variables. First designed to address Hilbert's thirteenth problem that he presented in Paris in a mathematics conference in 1900, the theorem ended up being a generalization¹⁸ of what was considered one of the top 23 most important problems defined by Hilbert¹⁹. The theorem below formalizes the results.

Theorem 6.4.1 *Let $f : \mathcal{I}^n := [0, 1]^n \rightarrow \mathcal{R}$ be an arbitrary multivariate continuous function. Then the function f has the representation:*

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right) \quad (6.2)$$

with continuous one-dimensional inner and outer functions Φ_q and $\phi_{q,p}$. The functions Φ_q and $\phi_{q,p}$ are defined on the real line. The inner functions $\phi_{q,p}$ are independent of the function f .

Proof The full proof and potential minor improvements of the latter can be found in the mathematics literature [134].

Remark Note the sentence “*every multivariate continuous function can be represented as a superposition of continuous functions of two variables*” is the way Kolmogorov's paper was translated into English. It may be a tad confusing as two functions are involved in equation (6.2) each involving n and $2n$ variables (so not necessarily two variables). One can think about each of the possible pairs of variables as being combined into simpler intermediate functions which is then recombined either directly²⁰ or indirectly²¹. The inner functions are usually not represented with two variables in the NN literature (because trivial) which may make this sentence a tad confusing today. In order to understand the rational of why Kolmogorov expressed

¹⁷Kolmogorov and Arnold published separately their results [132, 133].

¹⁸The initial problem was to solve 7th degree equation using algebraic continuous functions of two parameters.

¹⁹One of the most influential mathematicians of the 20th century.

²⁰Through the inner function for example linearly or non linearly (e.g: Sigmoid function).

²¹Through function composition via the outer function.

the theorem this way, we need to understand the context in the past. Indeed the theorem was initially designed to solve 7th degree equation using algebraic continuous functions of *two* parameters, and though it can be seen, today, as a generalisation of this problem, the initial objective was focused on *two* parameters only, hence the wording.

Besides his own student [134] which he shares the main results of the theorem with, the theorem prompted several contributions. They can be categorized as focused on the inner or outer functions. Notably, Lorentz relaxed the constraint on the outer functions Φ_q and noticed that they could be the same [135, 136]. Sprecher proved that the inner functions $\phi_{q,p}$ can be replaced by $\lambda_p \phi_{q,p}$ with some rules around the scalar λ_p [137, 138]. Another notable technicality around the interpretation came with Hecht-Nielsen [139, 140, 137] who translated the theorem into a feed-forward neural networks with an input layer, one hidden layer and an output layer.

6.5 Intelligent Agents & Financial Strategies

6.5.1 The High Frequency Financial Funnel

The pillars associated to the construction of the High Frequency Trading Ecosystem (HFTE) model has in its inspirational roots the idea that strategies in the market interact, or to choose an alternative jargon “Mutually Excite” [141], and it is their interaction that creates the fluctuations in the prices (the same way interaction create complexity in the Game of Life [52]). It also assumes that strategies can invade others and therefore the study of the financial market partially comes to studying an n -species predator prey stochastic model. Another pillar is that the construction of each of these strategies must have the same DNA²²: a financial funnel (e.g. Figure 6.6). The idea of a common DNA is its ability to model many classic strategies and more. For instance our proposed DNA should be able to model Trend Following (TF) strategies, Moving Average Convergence Divergence (MACD), Multi-Linear Regression (MLR) or XOR like strategies like it can be seen by figures 6.7, 6.8, 6.9, and 6.12 respectively. Because of its ability to model all these strategies with a relatively shallow neural network (as opposed to using a Deep Neural Network²³) is the main drivers which have led us to propose the Funnel, introduced by Martin Nowak [30], as the simplest possible network to model (therefore which minimizes over-fitting) the key functions

²²Called HFFF: a Neural Network Architecture we explain next.

²³We will discuss more in details the Bias-Variance Dilemma in Section 6.6.

for our application. The area of evolutionary graph theory is quite rich. Many graphs provide interesting properties but Nowak's suggestion seems adequate as we illustrate next but let us first introduce the network formally.

Definition (High Frequency Financial Funnel): We can formalize the learning process from all of our strategies using the HFFF of Figure 6.6 by providing a set \mathcal{H} , as described by equation (6.3) of weights corresponding to all the possible weights of this particular figure.

$$\mathcal{H} \triangleq \left\{ \begin{array}{cc} \bigcup_{j \in [1,9]} w_{\bar{s},j}^i & \bigcup_{j \in [1,9]} w_{s,j}^i \\ \bigcup_{j \in [1,9], i \in [1,3]} w_{\bar{s},i,j}^{h_1} & \bigcup_{j \in [1,9], i \in [1,3]} w_{s,i,j}^{h_1} \\ \bigcup_{j \in [1,3]} w_{\bar{s},j}^{h_2} & \bigcup_{j \in [1,3]} w_{s,j}^{h_2} \\ w_{\bar{s},j \in [1,9]}^o & w_{s,j \in [1,9]}^o \end{array} \right\} \quad (6.3)$$

with w^i , w^h and w^o , respectively the weights associated to the input, hidden and output layers. More formally let the HFFF [4] be a NN architecture of 9 inputs, 3 hidden layers and 1 output layer. Each node connects to the next layer and to itself. Each connection to itself will be labelled by w_s and the others by $w_{\bar{s}}$. We will admit that $w_{\bar{s}} \sim \mathcal{U}[-1, 1]$ and that $w_s \sim \mathcal{U}[0, 1]$ and therefore the more generally we have equation (6.4).

$$w_x \sim \mathcal{U}[-1_{x=\bar{s}}, 1] \quad (6.4)$$

Remark Note that in the context of this paper we have chosen to work with Martin Nowak's [30] funnel, which modification is described in Figure 6.6. This architecture offers the advantage of linking some interesting bridges between the worlds of information theory, evolutionary dynamics and biology. Indeed in information theory it also resembles the classic structure of a Neural Network and can therefore easily accommodate the mapping of classic and less classic financial strategies. In evolutionary dynamics, Moran like Processes can easily be formalized through similar means. In biology the topological structure is a potent amplifier of selection [30].

Note also that the HFFF from Figure 6.6 can easily be trained using a classic error back propagation algorithm like the one described in algorithm (6)²⁴. We see next how the HFFF models classic financial strategies starting with the Exponentially Weighted Moving Average (EWMA).

²⁴Where the activation function would be linear so as to make sure the MLR strategy can be exactly replicated.

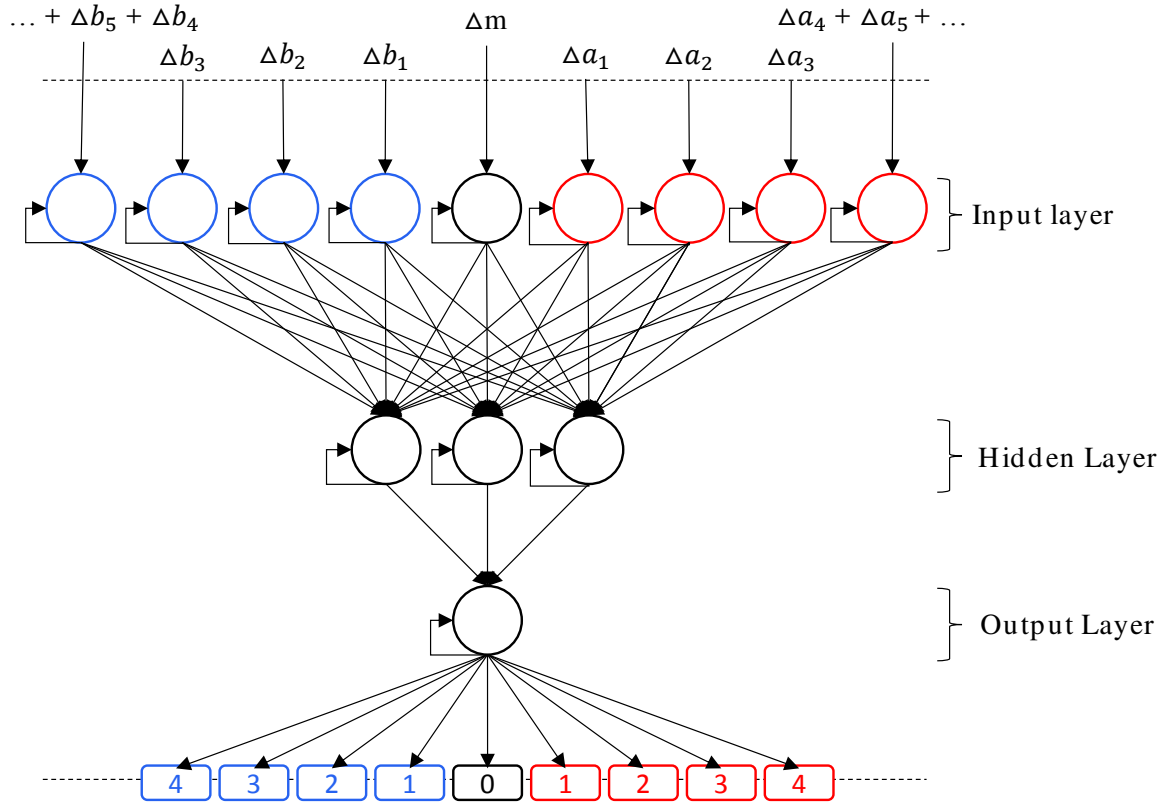


Figure 6.6: The High Frequency Financial Funnel

6.5.2 The EWMA Architectures

6.5.2.1 Trend Following

A very common trading strategy is known as trend following (TF). The idea of TF is that if the price has been going a certain way (e.g., up or down) in the recent past, then it is more likely to follow the same trend in the immediate future.

Definition (TF): The mathematical formulation of a TF can be diverse but in the context of this thesis we use an exponentially weighted moving average (EWMA):

$$\hat{x}_t = (1 - \lambda)x_t + \lambda\hat{x}_{t-1}, \quad \lambda \in [0, 1] \quad (6.5)$$

Remark In this equation, $\lambda \in [0, 1]$ represents the smoothness parameter. The lower the λ , the more the next move will be conditional to the immediately adjacent previous move. Conversely, the higher the λ , the more the future move will be function to the long term trend. The idea being that through a simple filtering process, the noise is extracted from the signal which then returns a clean time series \hat{x}_t traders like to seldom use directly or sometimes by using it with couple of other similar equations

Algorithm 6: Backpropagation

Input: topological structure \mathcal{H} with unoptimized weights

Output: topological structure \mathcal{H} with optimized weights

for d in data **do**

Forwards Pass

 ▷ Starting from the input layer, do a forward pass through the network, computing the activities of the neurons at each layer.

Backwards Pass

 ▷ Compute the derivatives of the error function with respect to the output layer activities

for layer in layers **do**

 ▷ Compute the derivatives of the error function with respect to the inputs of the upper layer neurons. Compute the derivatives of the error function with respect to the weights between the outer layer and the layer below. Compute the derivatives of the error function with respect to the activities of the layer below.

▷ Updates the weights

▷ Return Net with updated architecture

with a different value of λ and therefore defining a signal as a difference of these various filtered time series.

Proposition 6.5.1 *The HFFF can model trend following strategies.*

Proof We set to 0 all the unnecessary connections: $\cup_{j \in [1,4]} w_{\bar{s},j}^i = 0$, $\cup_{j \in [1,4]} w_{s,j}^i = 0$, $\cup_{j \in [6,9]} w_{\bar{s},j}^i = 0$, $\cup_{j \in [6,9]} w_{s,j}^i = 0$, $\cup_{j \in [1,4], i \in [1,3]} w_{\bar{s},i,j}^{h_1} = 0$, $\cup_{j \in [1,4], i \in [1,3]} w_{s,i,j}^{h_1}$, $\cup_{j \in [6,9], i \in [1,3]} w_{\bar{s},i,j}^{h_1} = 0$, $\cup_{j \in [6,9], i \in [1,3]} w_{s,i,j}^{h_1} w_{\bar{s},3}^h = 0$, $w_{s,1}^h = 0$ and $w_{s,3}^h = 0$. ■

The proof is illustrated in Figure 6.7 (the weights equal to 0 have not been represented²⁵). We will address the problem of rigorously formalizing mathematically what constitutes a trend following in a subsequent paper. However for now, in order to keep the discussion intuitive, we consider a trend following strategy to have an architectural DNA which would look like the one in Figure 6.7.

6.5.2.2 Mean Reversion

One of the current hurdles in our research is our classification issue and the Moving Average Convergence/Divergence (MACD) strategy is a good example as to why. Indeed the MACD strategy which is technically associated to the EWMA family tries

²⁵Note that there is different ways to achieve the same numerical results though with a different NN format.

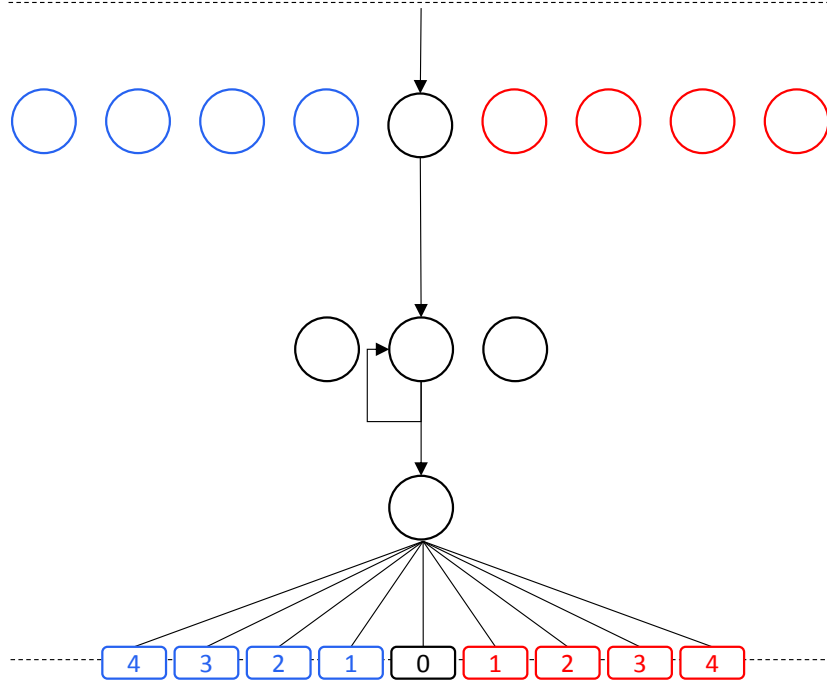


Figure 6.7: The EWMA Strategy in HFFF format

to capture a behavioural finance meaning which can potentially be classified as an antithetic TF strategy (which are in the EWMA family).

Proposition 6.5.2 *The HFFF can model the MACD strategy.*

Proof The MACD was designed to reveal changes in the direction and duration of a trend. It essentially models difference between a “fast” EWMA $(S_t^{N_f})$ and another “slower” EWMA $(S_t^{N_s})$. For instance the popular MACD(12,26), $M_t^{12,26}$ is given by:

$$M_t^{N_f, N_s} = S_t^{N_f} - S_t^{N_s}, \quad (6.6)$$

$$S_t^\alpha = \begin{cases} S_1, & t = 1 \\ \alpha \cdot S_t + (1 - \alpha) \cdot S_{t-1}^\alpha, & t > 1, \end{cases} \quad (6.7)$$

$$\alpha = 2/(N_\alpha + 1), \quad (6.8)$$

$$N_\alpha = \{N_f, N_s\}. \quad (6.9)$$

The MACD(12,26) (e.g, $\{N_f, N_s\} = \{12, 26\}$) in particular has gained recent attraction with practitioners. For instance it has gained a great deal of momentum for algorithmic traders. Figure 6.8 represents a generic MACD. If one is looking specifically for a MACD(12,26), then the weights of the hidden layers must be such that $\alpha_{12} = 2/13$ and $\alpha_{26} = 2/27$ and the ones of the output layers must be a simple subtraction to abide by the above definition.

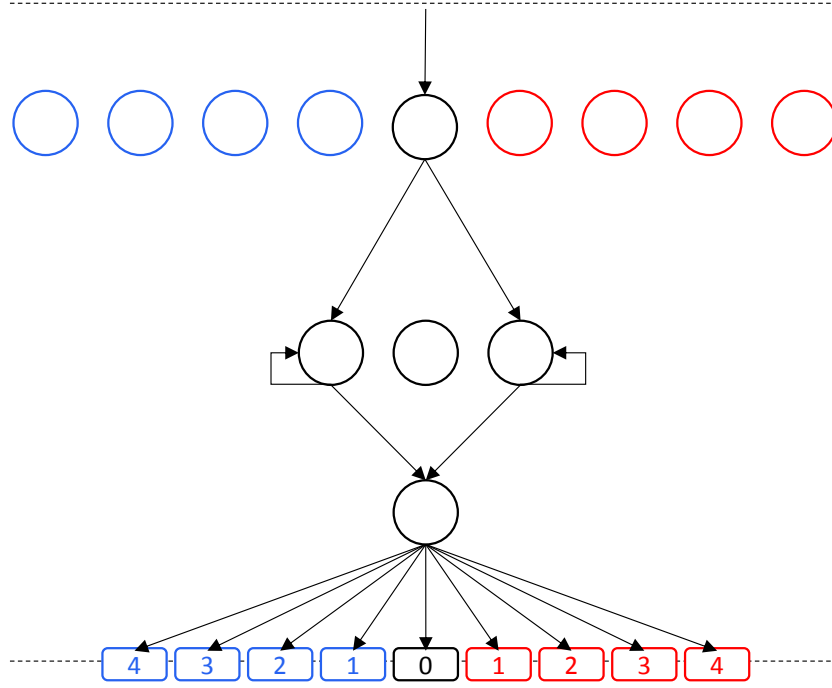


Figure 6.8: MACD Strategies (or a difference of EWMA's) in HFFF format

6.5.3 Multi Linear Regression NN Format

The Multi Linear Regression (MLR) is another well known, relatively simple strategy traders have been using in the industry.

Definition (MLR): Given $\{y_i, x_{i-1,1}, \dots, x_{i-1,9}\}_{i=1}^n$ where n is the sample size, \mathbf{x}_i the explanatory variable, y_i the dependent variable and ε_i the error term, then our MLR is given by:

$$\begin{aligned} y_i &= \beta_1 x_{i-1,1} + \dots + \beta_9 x_{i-1,9} + \varepsilon_i \\ &= \mathbf{x}_{i-1}^T \beta + \varepsilon_i, \quad i = 1, \dots, n. \end{aligned} \tag{6.10}$$

where T denotes the transpose, so that $\mathbf{x}_{i-1}^T \beta$ is the inner product between vectors x_i and β . The best unbiased estimator of β is given by $\hat{\beta} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T y$ and sometimes also referred to β^{OLS} .

Proposition 6.5.3 *The HFFF can model multi linear regression like strategies.*

Proof Set $\cup_{j \in [1,4]} w_{\bar{s},j}^i = 0$, $\cup_{j \in [1,4]} w_{s,j}^i = 0$, $\cup_{j \in [6,9]} w_{\bar{s},j}^i = 0$, $\cup_{j \in [6,9]} w_{s,j}^i = 0$, $w_{\bar{s},1}^h = 0$, $w_{\bar{s},3}^h = 0$, $w_{s,1}^h = 0$, $w_{s,3}^h = 0$. ■

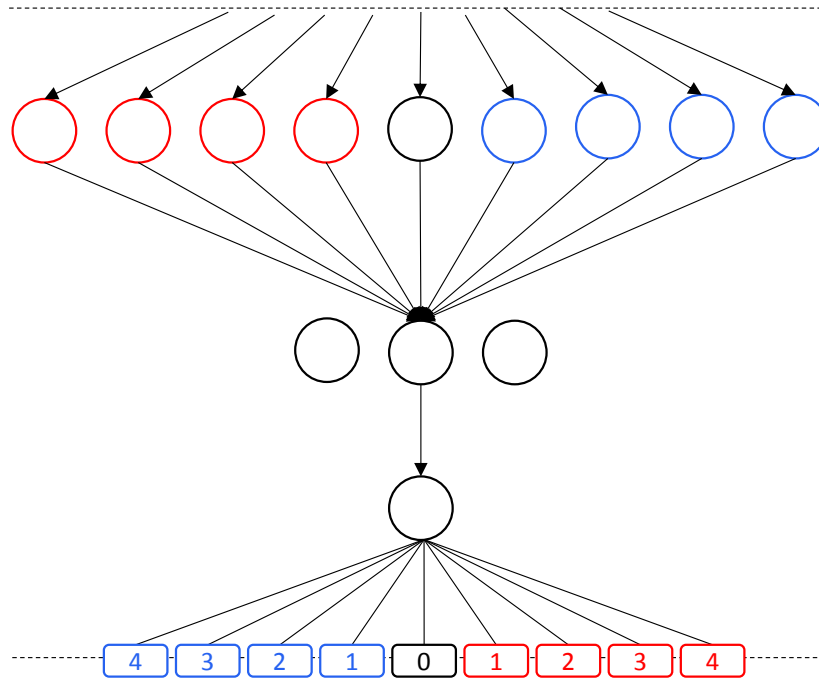


Figure 6.9: The MLR strategy in HFFF format

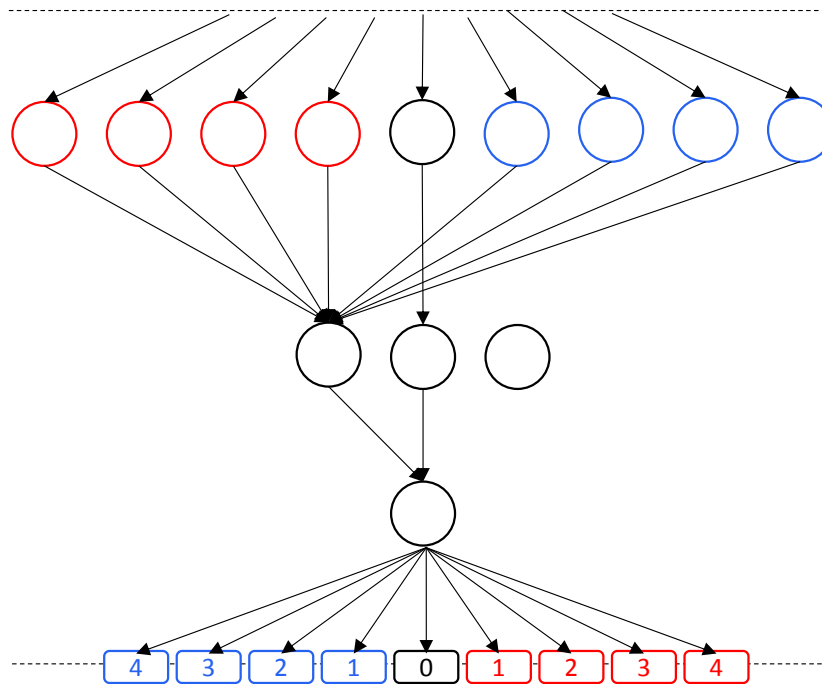


Figure 6.10: Another MLR strategy in HFFF format

Figure 6.9 represents a possible form of the MLR. As discussed before, different NN formats may lead to more or less the same strategy (figure 6.10 can translate

mathematically to what figure 6.9 translates into). We also note that, logistic or weighted MLR can be modeled through the same HFFF of Figure 6.9 by simply changing respectively the activation function (from linear to logistic) and the weights.

6.5.4 Regularized NN Format & Lasso Regression

The bias-variance dilemma (BVD) is a technical term representing the optimization by constraints problem which aims at simultaneously minimizing the error from erroneous assumptions (bias) in our learning algorithm or commonly called “under-fitting” and the error from the out of sample analysis (variance) or commonly called “over-fitting”. One of the properties of DL is its dual ability to learn the most complicated functions but also makes it prone to over-fitting. It is therefore recommended that one applies conscious efforts in studying carefully the associated benefits to complexity ratio in the context of the BVD. Regularization is usually the term employed for the methodology that aims at finding the optimal model according to the BVD. The mathematical formalization suggests that we calibrate a function f which takes as input a potential infinite number of explanatory variable x_1, x_2, \dots, x_n so as to minimize the distance to a target y under some cost measure V subject to a penalization, or regularization term²⁶ $R(f)$. Equation (6.11) refers to this generic Regularization.

$$\min_f \sum_{i=1}^n V(f(x_i), y_i) + \lambda R(f) \quad (6.11)$$

Within the family of Regularized methodologies the Lasso²⁷ methodology is the most common one and usually associated with the MLR we have seen in the previous paragraph. They have been gaining momentum in the past few years as they represent the simplest ML technique which has the reputation to work in systematic trading provided the strategy and the input variables are sound.

Definition (Lasso Regression): Given $\{y_i, x_{i-1,1}, \dots, x_{i-1,9}\}_{i=1}^n$ where n is the sample size, \mathbf{x}_i the explanatory variable, y_i the dependent variable and ε_i the error term, then our Lasso Regression is formalized by equation (6.12).

$$y_i = \beta_1 x_{i-1,1} + \dots + \beta_9 x_{i-1,9} + \varepsilon_i \quad \text{subject to} \quad \sum_{j=1}^9 |\beta_j| \leq \lambda. \quad (6.12)$$

where λ is an input parameter that determines the amount of regularisation desired.

²⁶or regularizer.

²⁷Short for Least Absolute Shrinkage and Selection Operator.

Proposition 6.5.4 *The HFFF can model Lasso regression like strategies.*

Proof Simply set $w_{s,2}^h = 0$, make sure the regularization is done exclusively on one of the remaining hidden layer and finally make sure the remaining hidden layer calibrates its weight the same way as the β^{OLS} from the MLR. Figure 6.11 gives an illustration.

■

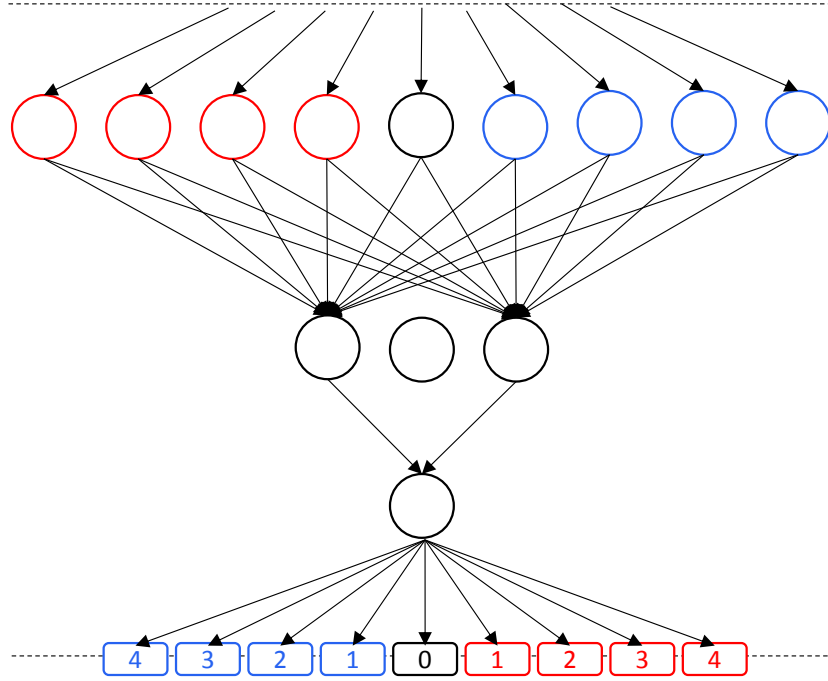


Figure 6.11: A Lasso regression in HFFF format

Remark Figure 6.11 and Figure 6.13 look the same but the weights and activation functions are different.

6.5.5 XOR NN Format

We recall here the truth table associated by the XOR function in Table 6.1. Let us look at the following known HF relationship. This will justify why the HFFF must be able to model the XOR function.

Definition (Open Interest): We define the Open Interest (OI) as being the difference between the total volume on the bid side minus the total volume on the asked side.

Remark The concept of OI, in equation (6.13a), and orderbook imbalance [126], in equation (6.13b), are conceptually the same though the latter is normalised in $[-1, 1]$ and the former is not. For mally we have

$$I^d = O_b^d - O_a^d, \quad (6.13a)$$

$$\tilde{I}^d = \frac{I^d}{O_b^d + O_a^d}, \quad (6.13b)$$

where O_b^d and O_a^d represent the total OI on the bid and ask side of the orderbook in the d 's depth.

It is known that when the price and the OI are rising then the market is bullish, when the Price is rising but the Open Interest is Falling then the market is bearish, when the Price is falling but the Open Interest is rising then the market is bearish, and finally when the Price is falling and the Open Interest is falling then the market is bullish. These 4 market situations can be summarized by Table 6.1.

I_1	I_2	O	Price (I_1)	Open Interest (I_2)	Signal (O)
1	1	0	Rising	Rising	Buy
1	0	1	Rising	Falling	Sell
0	1	1	Falling	Rising	Sell
0	0	0	Falling	Falling	Buy

Table 6.1: XOR Relationship Between Open Interest, Price & Signal

Proposition 6.5.5 *The HFFF can model XOR like strategies.*

Proof Simply set $\cup_{j \in [1,4]} w_{s,j}^i = 0$, $\cup_{j \in [1,4]} w_{s,j}^i = 0$, $\cup_{j \in [6,9]} w_{s,j}^i = 0$, $\cup_{j \in [6,9]} w_{s,j}^i = 0$, $w_{s,1}^h = 0$, $w_{s,3}^h = 0$, $w_{s,1}^h = 0$, $w_{s,3}^h = 0$. ■

Remark We make two observations. First, the preceding proof is visually illustrated by Figure 6.12 (the weights equal to 0 have not been represented). Second, the XOR HFFF can be designed in various ways. We will address the problem of rigorously formalizing mathematically what constitutes an XOR in a subsequent paper. However for now, in order to keep the discussion intuitive, we will consider an XOR strategy to have an architectural DNA which would look like the one from Figure 6.12. Figure 6.13 represents an equivalent alternative example of an XOR strategy.

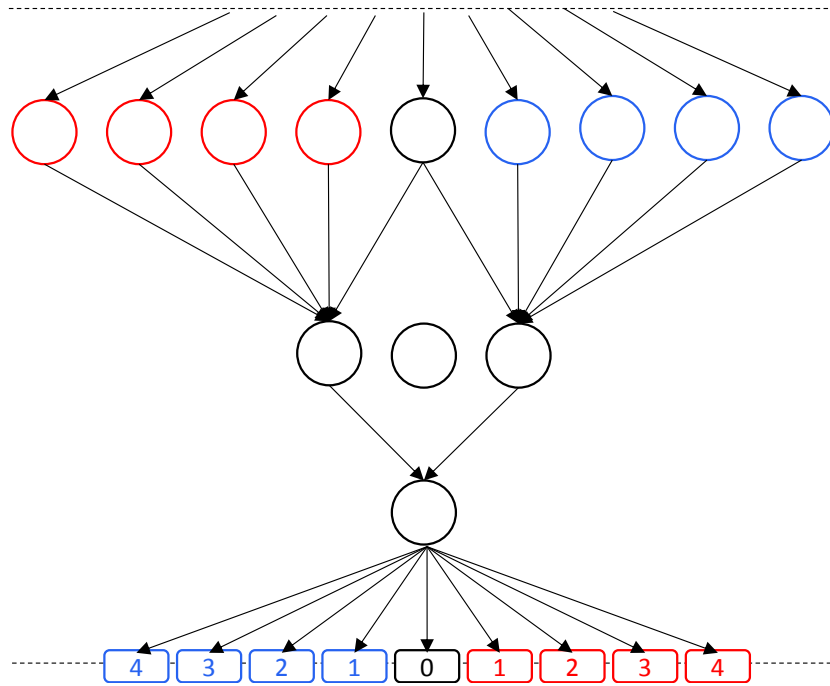


Figure 6.12: The XOR strategy in HFFF format

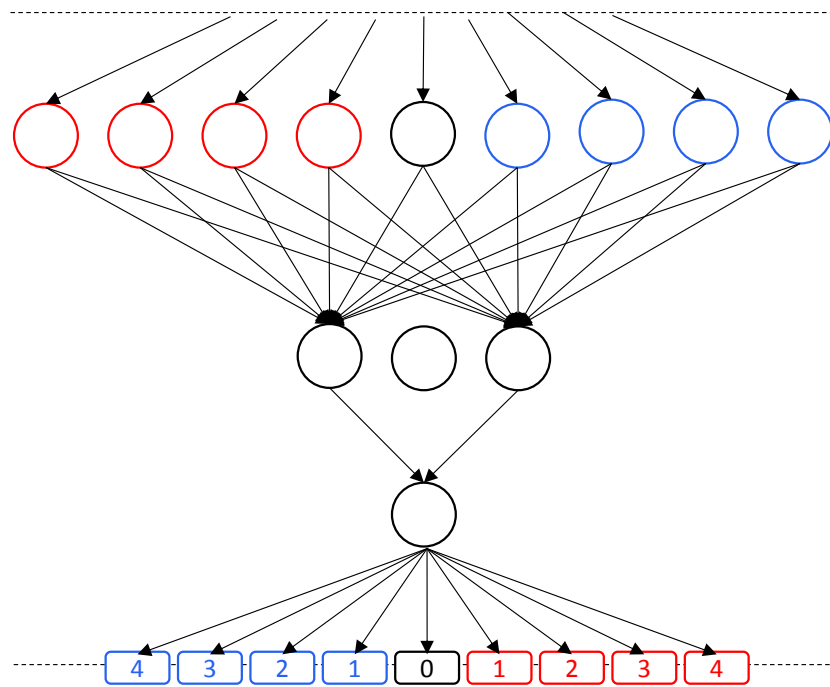


Figure 6.13: Another XOR strategy in HFFF format

6.6 Network HFFF & Deep Learning Thoughts

6.6.1 Deep Learning Brief Recent History

Deep learning (DL) has been gaining a great deal of momentum as a subbranch of Machine Learning for very good reasons. For instance, Deep-Mind was created in 2011 and subsequently bought by Google in 2014 for \$400M [125]. DL became famous for building an AI algorithm, AlphaGo, which outperformed the best Go master in the world. Though, an event in which a Machine Learning algorithm would beat a master was not an original feature²⁸, the complexity of the game and the number of possible moves made the Deep-Mind director speculate that the algorithm worked intuitively rather than using a pure logic based approach like it was done with chess and Deep Blue. This extraordinary feature was achieved through an initial deep neural network architecture in which an initial data set is used for training purposes and the algorithm, once the data is exhausted, would play an older version of itself and get incrementally better this way. The ultimate goal behind this self learning A.I. is to create a general purpose Algorithm.

6.6.2 Methodology

The scientific methodology behind the construction of the game of Go is one we wish to apply to our HFFFs and create a dynamical ecosystem of quantitative financial strategies. For instance increasingly advanced strategies compete with each other and we eventually get an interesting portfolio of strategies as well as their co-evolution. However, the HFFF itself potentially suffers from similar kind of limitations that prevented the XOR function to be learnt without 1 hidden layer (see Figure 6.5 and 6.4 as well as paragraph 6.4). A legitimate question can be asked on whether a single hidden layer is enough. The answer to this question is in fact negative as Convolutional Neural Network (CNN) have shown more potential compared to shallow learning [143]²⁹ architectures. Some other studies reveal universal features of price formation [144] but lack a study on simpler intermediate benchmarks. For instance in [144] a logistic regression is used as a benchmark. It would have been interesting to see some more complex intermediate³⁰ benchmarks. We have arbitrarily taken as hypothesis the HFFF to be good enough to model few critical strategies in

²⁸The computer “Deep Blue” beat Kasparov in May 1997 [142].

²⁹I am however personally skeptical on the results of these published studies but I do accept the potential of CNN in trading.

³⁰Starting with a shallow NN and increasing in complexity in order to understand whether the universal features learnt are because the NN is deep or is it because it has a hidden layer.

the domain of QF and above all proceeding this way is important in unweaving the black box associated DL.

6.6.3 Our Hypothesis

Ensuing the above we make the following hypothesis: increased NN architectural complexity leads to the domination of simpler strategies architecturally. For example, the TF strategy “should dominate” a random swarm of strategies. In turn the MLR strategy “should dominate” the TF. The idea here is that the MLR strategy “should capitalize” on areas of the orderbook the TF strategy does not have access to because the DNA is limiting (information on the OI in absent in the TF but present in the MLR). Similarly the XOR strategy “should dominate” the MLR by splitting the OI surface in additional zones that the MLR cannot understand (lacking the necessary hidden layer). Figure 6.14 illustrates these hypotheses. In some way you



Figure 6.14: Illustration for intuitive strategy invasion: “denser NN” lead to increased potential for invasion.

could extrapolate this “invasion” to “increased network complexity” tendencies to a system that could potentially converge towards a DL infrastructure. It is however usually accepted in the industry that the likelihood of overfitting increases as one adds hidden layers. However, we have also seen with Shallow Learning that adding hidden layers can also allow us to do regularization which removes to some extent the latter argument against. Natural questions arise here. Are the hypothesis between “invasion” to “increased network complexity” verified empirically? How does these different architectures impact each others performance as well as the market? We study this latter point in the next Chapter.

Chapter 7

Evolutionary Dynamics & Strategy Ecosystem

In the context of the bottom-up approach for algorithmic trading for which we discussed the strategy architecture in chapter 6, here we formalize the interaction rules at the ecosystem level. More specifically, we go from the premise that a good theory can be simulated but that a simulation can also help bring intuition on what the theory might be, and these two research tools can go back and forth until the theory is ironed out [145]. In doing so, we will first do a very brief review of relevant inference and dynamical models in Section 7.1, Game Theory in Section 7.2 and Theoretical Biology in Section 7.3. Indeed, we first let the random states of the latter HFFF strategies interact through the primitive strategies swarm via the market order-book. We then show, in Section 7.4, that although not necessarily optimal, the simulation provides a great deal of intuition that help us formalize additional simulations through a simple MCMC. More specifically we perform few simulations in Section 7.5 in order to illustrate how the computer tournament presented here would work in order to expose some complexity hurdles. The latter will lead to Section 7.6 in which we introduce the concept of *Path of Interaction*¹ which we present as the better option for studying the market using the bottom-up approach.

7.1 Review of Markov Chain Monte Carlo Models

In this section we go over a brief overview of classic Inference and Dynamical models focusing on Markov Chain Monte Carlo (MCMC). MCMC algorithms [146] sample from a probability distribution based on a Markov chain that has a desired equilibrium distribution, the quality of the sample improving at each additional iteration.

¹Defined formally later in this Chapter.

7.1.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is a MCMC method that aims at obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult [147] because of high dimensions. At each iteration x_t , the proposal next point x' is sampled through a proposed distribution $g(x'|x_t)$. We then calculate:

- $a_1 = \frac{P(x')}{P(x_t)}$: the probability ratio between the proposed and the previous sample,
- $a_2 = \frac{g(x_t|x')}{g(x'|x_t)}$: the ratio of the proposal density in both directions²,

and set $a = \max(a_1 a_2, 1)$. We then accept $x_{t+1} = x'$ if $r \sim U[0, 1] \geq a$ which essentially means that if $a = 1$, accept is always true otherwise you accept with a probability $a_1 a_2$. The algorithm works best if the proposal distribution is similar to the real distribution (so feature engineering can be helpful here). Note that the seed is slowly forgotten as the number of iterations increases.

7.1.2 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo³ [148], is an MCMC method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. It serves to address the limitations of the Metropolis-Hastings algorithm by adding few more parameters. The idea of the methodology is to reduce the correlation between successive samples using a Hamiltonian evolution process⁴.

7.1.3 Gibbs Sampling

Perhaps one of the simplest MCMC algorithms, the Gibbs Sampling (GS), was first introduced in Geman & Geman [149] in the context of an application to image processing. Later it was discussed in the context of missing data problems [150]. The benefit of the Gibbs algorithm for Bayesian analysis was demonstrated in Tanner and Wong [150]. To define the Gibbs sampling algorithm, let the set of full conditional distributions⁵ be: $\pi(\psi_1|\psi_2, \dots, \psi_p), \dots, \pi(\psi_d|\psi_1, \psi_2, \dots, \psi_{d-1}, \psi_{d+1}, \dots, \psi_p), \dots, \pi(\psi_p|\psi_1, \dots, \psi_{p-1})$. One cycle of the GS, described in algorithm (7), is completed by sampling $\{\psi_k\}_{k=1}^p$ from the mentioned distributions, in sequence and refreshing the conditioning variables. When d is set to 2 we obtain the two block Gibbs sampler

²Equal to 1 is the proposal density is symmetric.

³Sometimes also referred to as hybrid Monte Carlo though more in the past.

⁴Targeting states with a higher acceptance rate

⁵We define $\pi(\cdot)$ as the probability distribution and by ψ_i the variable i .

Algorithm 7: Gibbs Sampling

Input: Specify an initial value $\boldsymbol{\psi}^{(0)} = (\psi_1^{(0)}, \dots, \psi_p^{(0)})$

Output: $\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \dots, \boldsymbol{\psi}^{(M)}\}$

▷ Sample:

for $j = 1, 2, \dots, M$ **do**

 Generate $\psi_1^{(j+1)}$ from $\pi(\psi_1 | \psi_2^{(j)}, \psi_3^{(j)}, \dots, \psi_p^{(j)})$

 Generate $\psi_2^{(j+1)}$ from $\pi(\psi_2 | \psi_1^{(j+1)}, \psi_3^{(j)}, \dots, \psi_p^{(j)})$

 ⋮

 Generate $\psi_d^{(j+1)}$ from $\pi(\psi_d | \psi_1, \psi_2, \dots, \psi_{d-1}, \psi_{d+1}, \dots, \psi_p)$.

 ⋮

 Generate $\psi_p^{(j+1)}$ from $\pi(\psi_p | \psi_1^{(j+1)}, \dots, \psi_{p-1}^{(j+1)})$

▷ Return the values:

$\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \dots, \boldsymbol{\psi}^{(M)}\}$

described by Tanner & Wong [150]. If we take general conditions, the chain generated by the GS converges to the target density as the number of iterations goes towards infinity. The main drawback with this method however is its relative computational heavy aspect because of the burn-in period. The model is described in pseudo-code in algorithm (7).

7.1.4 Ordered Over-relaxation

Over-relaxation is usually a term associated with a Gibbs Sampler but in the context of this subsection we discuss Ordered Over-relaxation. The methodology aims at addressing the slowness associated in performing a random walk with inappropriately selected step sizes. The latter problem was addressed by incorporating a momentum parameter which consist of sampling n random variables (20 is considered a good [151] number for n), sorting them from biggest to smallest, looking where x_t ranks: say at p 's position, amongst the n variables, and then picking $n - p$ for the subsequent sample x_{t+1} . This form of optimal “momentum” parameter design is a central pillar of research in MCMC [152].

7.1.5 Slice Sampling

Slice sampling is one of the remarkably simple methodologies [152] of MCMC which can be considered as a mix of Gibbs sampling, Metropolis-Hastings and rejection sampling methods. It assumes that the target density $P^*(x)$ can be evaluated at

any point x but is more robust compared with the Metropolis-Hastings especially when it comes to step size. Like rejection sampling it draws samples from the volume under the curve. The idea of the algorithm is that it switches vertical and horizontal uniform sampling by starting horizontally, then vertically performing “slices” based on the current vertical position. MacKay made quality contributions especially when it comes to its visual representation [151].

7.1.6 Multiple-try Metropolis

One way to address the curse of dimensionality is the Multiple-try Metropolis which can be thought of as an enhancement of the Metropolis-Hastings algorithm. The former allows multiple trials at each point instead of one by the latter. By increasing both the step size and the acceptance rate, the algorithm helps the convergence rate of the sampling trajectory [153]. The curse of dimensionality is another central area of research in MCMCs [152].

7.1.7 Reversible-Jump

Another variant of the Metropolis-Hastings, and perhaps most promising methodology when it comes to our application is the Reversible-jump MCMC (RJ-MCMC) developed by Green [154]. One key factor of the RJ-MCMC is that it is designed to address changes of dimensionality issues. We face a dual type issues around change of dimensionality. The first being the frequency of each strategy in an ecosystem and the second element being the HFFF which branching structure and size changes as a function of the strategy⁶. More formally, let us define $n_m \in N_m = \{1, 2, \dots, I\}$, as our model indicator and $M = \bigcup_{n_m=1}^I \mathbb{R}^{d_m}$ the parameter space whose number of dimensions d_m is function of model n_m (with our model indicators not needing to be finite). The stationary distribution is the joint posterior distribution of (M, N_m) that takes the values (m, n_m) . The proposal m' can be constructed with a mapping $g_{1mm'}$ of m and u , where u is drawn from a random component U with density q on $\mathbb{R}^{d_{mm'}}$. The move to state (m', n'_m) can thus be formulated as $(m', n'_m) = (g_{1mm'}(m, u), n'_m)$. Function $g_{mm'} := (m, u) \mapsto (m', u')$, with $(m', u') = (g_{1mm'}(m, u), g_{2mm'}(m, u))$ must be one to one, differentiable, and have a non-zero support: $\text{supp}(g_{mm'}) \neq \emptyset$, in order to enforce the existence of the inverse function $g_{mm'}^{-1} = g_{m'm}$ (itself differentiable as well). Consequently (m, u) and (m', u') must have the same dimension, which is enforced if the dimension criterion $d_m + d_{mm'} = d_{m'} + d_{m'm}$ is verified ($d_{mm'}$ is

⁶See in section III and Figures 6, 9, 10 and 11.

the dimension of u). This criterion is commonly referred to as dimension matching. Note that if $\mathbb{R}^{d_m} \subset \mathbb{R}^{d_{m'}}$ then the dimensional matching condition can be reduced to $d_m + d_{mm'} = d_{m'}$, with $(m, u) = g_{m'm}(m)$. The acceptance probability is given by $a(m, m') = \min \left(1, \frac{p_{m'm} p_{m'} f_{m'}(m')}{p_{mm'} g_{mm'}(m, u) p_m f_m(m)} \left| \det \left(\frac{\partial g_{mm'}(m, u)}{\partial (m, u)} \right) \right| \right)$, where $p_m f_m$, the posterior probability is given by $c^{-1} p(y|m, n_m) p(m|n_m) p(n_m)$ with c being the normalizing constant. Many problems in data analysis require the unsupervised partitioning. Roberts, Holmes and Denison [155] re-considered the issue of data partitioning from an information-theoretic viewpoint and shown that minimization of partition entropy may be used to evaluate the most probable set of data generators which can be employed using a RJ-MCMC.

7.2 Game Theory Review

In this Section we present relevant concepts from the world of Game Theory.

7.2.1 Prisoner's Dilemma

The prisoner's dilemma (PD) is a well known standard example of a game. The way it is usually explained is in the context of a situation involving 2 prisoners who have organized illegal actions for which they have been caught by a third party (the police) who however needs confessions from either of the prisoners in order to abide by the complex legal proceedings. The prosecutor wants to close the case and send someone to prison (at least one of the two suspects), so he offers a deal involving a confession against a more lenient judgment. Both captives are offered this deal independently and away from each other. If the criminals both cooperate (C), nobody goes to prison but they each get a heavy fine. If one denounces⁷ (D) the other, then he will be free without any fine, but the one being denounced has to go to prison and get a fine. If they each denounce each other they go to prison without a fine. Broadly speaking that little story can be formalized into a 2 by 2 matrix⁸ with CC, CD, DC and DD with respective payoffs (2,2), (0,3), (3,0) and (1,1). Although the prisoners should clearly cooperate here, given that they do not know what the other is going to do, by expectation (with equal probability for a C and a D) any of the two users should denounce the other given that the expectation of the payoff for denouncing is 2 as opposed to a 1 for a cooperation. This is the reason why this game theory concept is referred to as a "dilemma".

⁷Sometimes also referred in the literature as "Deceits".

⁸Figure 7.1a.

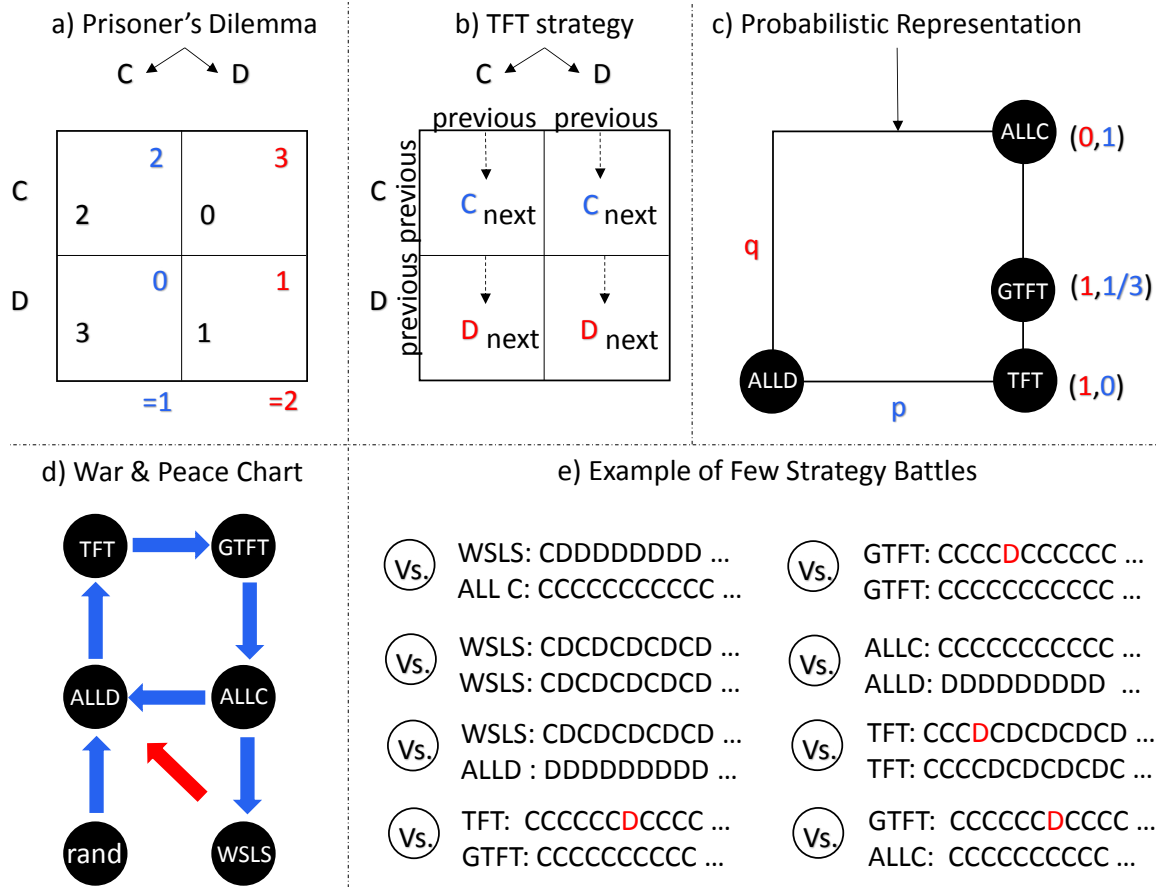


Figure 7.1: Some classic Game theory Representations [28, 29, 30, 31]

7.2.2 Axelrod's Computer Tournament

However this optimal strategy, in a “single iteration” presented in Section 7.2.1, changes when the game becomes iterative. This concept was formalized by Robert Axelrod [28, 29]. Indeed, he designed a computer tournament aiming at understanding what makes a strategy optimal in the context of an ecosystem in an iterative format. In that occasion he invited Mathematicians, Computer Scientists, Economists and Political Scientists to code a strategy they believed could win such tournament with the constraints of PD rules in which it is not known when the tournament will stop⁹. Many strategies were thrown into this ecosystem in this form of computer tournament. The range of strategies went from being very simplistic like “Always Deceit” (AD)¹⁰ to many other more complicated strategies which generic representation can

⁹E.g., it is by expectation best to deceit if one plays the PD only once. By iteration he should always deceit on the last move, but knowing this, the opponent should also deceit. Using this logic each player should deceit on the next to the last move and the same logic kicks in and very quickly one is led to the conclusion that he/she should deceit from the very first move.

¹⁰or its mirror: the AC “Always Cooperate” (AC) strategy.

be looked at in Figure 7.1b. Surprisingly the Tit For Tat (TFT) strategy came at the top of this tournament. The TFT is considered, in the literature, to be a nice strategy, meaning that it is never the first to deceit (its first move is by design to be a C), but it is also a strategy that is able to retaliate in situation in which it was previously deceived. Finally, it is a strategy that is able to forgive: meaning that if it sees that the adversary algorithm has decided to cooperate after a deceit, then he switches back to a C.

7.2.3 Evolutionary Dynamics

Martin Nowak [30] enhanced some of Axelrod's work by introducing new strategies and further developing the concepts of invasion/dominance¹¹ within a competitive strategic ecosystem. For instance, we can see from Figure 7.1d that some strategies invade others but these latter strategies can be, in turn, invaded by other ones which in turn can be invaded by the very first strategy mentioned and induce cycles¹². Indeed an ecosystem composed of a set of unbiased random strategies (that would randomly cooperate (C) or deceit (D)) would invite the invasion of an ALLD (always defect) kind. In turn the frequency of ALLD would take the ecosystem which would invite the TFT strategy which would benefit from the mutual cooperation within the same proximity. This process continues in a similar fashion. Figure 7.1d) exposes how some of these strategies may interact with each other. The following additional information may help in refreshing what some of these acronyms mean: The main takeaway from

Acronym	Strategy	Description
TFT	Tit for Tat	Developed in the previous Section
GTFT	Generous Tit for Tat	Less grudge prone than TFT
WSLS	Win-Stay, Lose-Shift	Outperforms TFT [30, 31]
ALLC	Always Cooperates	Self explanatory
ALLD	Always Deceits	Self explanatory
rand	Random Strategy	Outputs a C or a D with equal probability

Table 7.1: Evolutionary Dynamics Related Strategies

this parallel was to expose how the rise and fall of strategies can easily be engineered

¹¹by extension when applied to finance some strategies may dominate and invade others.

¹²one may extrapolate that economical cycles may be influenced by similar kind of processes.

through simple systematic rules based on an ecosystem and how complexity can be induced from simple rules. Figure 7.1 summarizes some of the main messages from Axelrod's [28, 29] and Nowak's [30, 31] work.

7.2.4 Minority Game

7.2.4.1 A Short Introduction

Extending some of the ideas we discussed in an economical context is not new. For instance the Minority Game (MG) [156], considered to be a simple congestion game, involves players who need to choose between two options (+1, -1). Those who have selected the option chosen by the minority “win”. Figure 7.2 illustrates the game. The

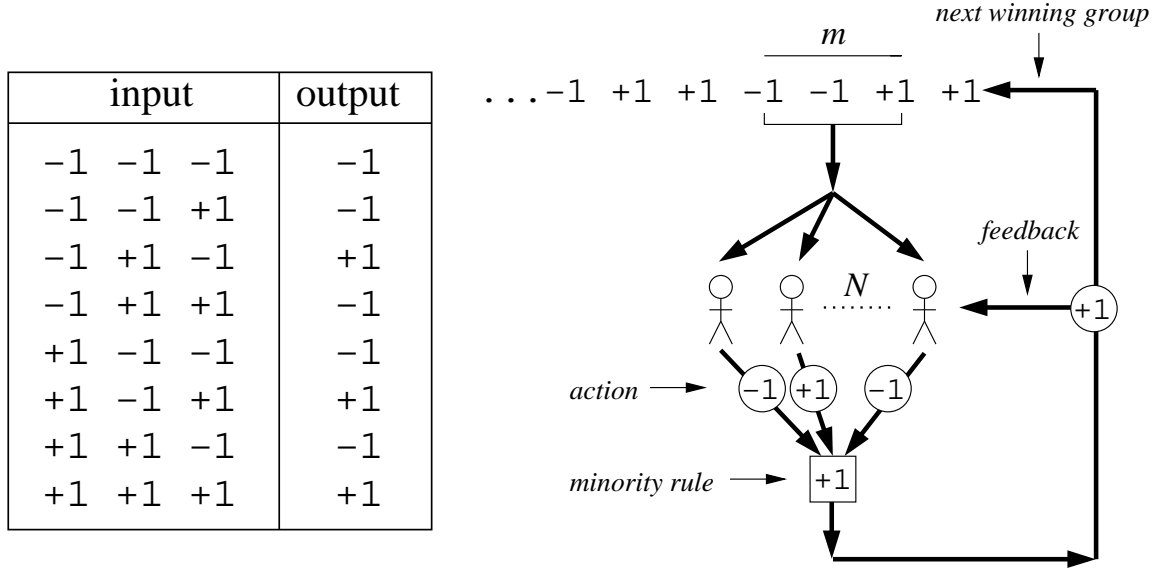


Figure 7.2: Minority Game (right) with $m = 3$ & possible strategy Table (left) [32]

game was latter formally studied in its mathematical details in [157] with a statistical mechanics perspective. More formally, the game consists of N (odd) agents. At each time step, t the N agents take an action $a_i(t) \in \{-1, +1\}$ with $i = 1, \dots, N$. The total “Market” action is calculated according to equation (7.1).

$$A(t) = \sum_{i=1}^N a_i(t). \quad (7.1)$$

Each of the N agents is given a payoff given by equation (7.2).

$$p_i(t) = -a_i(t)g[A(t)], \quad (7.2)$$

where $g(x) = \text{sgn}(x)$ though this latter function can find other forms. The winning outcome $W(t+1) = \text{sgn}[A(t)]$. At this stage of the analysis, we assume arbitrarily that the agents all have the same memory and they can only remember the m last winning outcomes, $W(t+1), W(t), \dots, W(t+1-m)$ and that their actions can only be sampled from a set of strategies, all using a weighted average of these outcomes. The number of available strategies is by construction $2^{2^m} = s$ and their prediction is given by its $\mu(t) \in \{1, \dots, 2^m\}$ components (number whose representation $\{-1, 1\}$ is the last winning group). Let $\vec{I}(t)$ be the vector with zero components except its $\mu(t)$ component. The payoff of each strategy α of agent i is given by equation (7.3).

$$p_i^\alpha(t+1) = p_i^\alpha(t) - \vec{r}_i^\alpha \cdot \vec{I}(t) g[A(t)]. \quad (7.3)$$

These are considered intermediate calculations as only the best performing strategy gets to decide what to do next. We call $\beta_i(t) \in \{1, \dots, \}$ the best strategy of agent i then the action in given by equation (7.4).

$$a_i(t) = \vec{r}_i^{\beta_i(t)} \cdot \vec{I}(t). \quad (7.4)$$

Though the optimal strategy can be obviously changing, and therefore ultimately modify the dynamic of the market some interesting seasonality were observed [32]. For instance in Figure 7.3 we can clearly see patterns appearing $A(t)$. In Figure 7.4

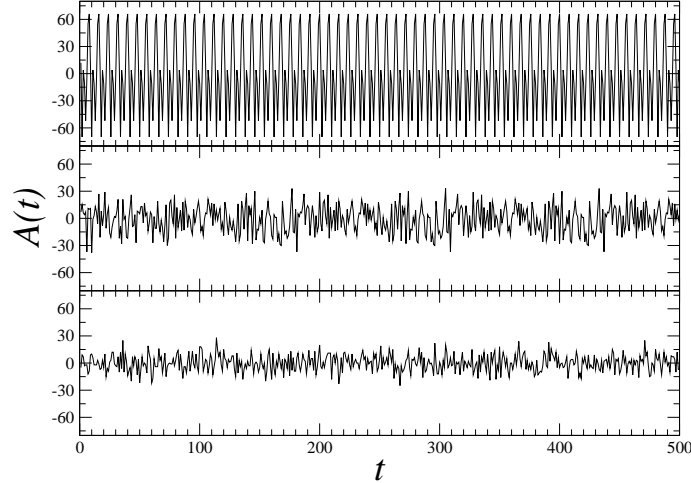


Figure 7.3: $A(t)$ with $m = 15, 7, 2$ with visible seasonality [32].

we can observe that σ^2/N is only a function of $\alpha = 2m/N$, which considering the complexity of the interactions between the set of agents can be quite remarkable. Generally speaking we observe that simple economical games can yield very complex systems but that “physical laws” (e.g. Figure 7.4) can emerge out of this complexity.

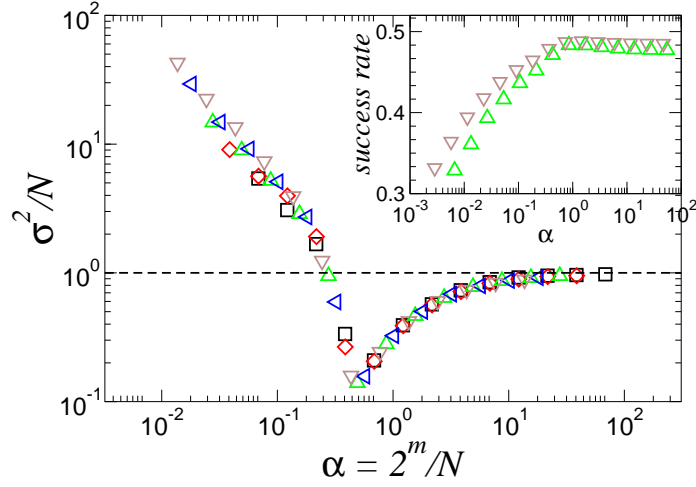


Figure 7.4: Relation between σ and m $N = 101, 201, 301, 501, 701$ [32].

This latter point has been elegantly raised in the finance’s scientific literature [16] but not perhaps enough incorporated. More specifically though the problem is clearly understood [16, 158], the current approach to the scientific method is still top-down.

7.2.4.2 Criticism

Though rigorous in the scientific method¹³ and incorporating interesting features such as the possibility of non rational agents [157], the rules of the game seem to have been chosen to reveal something interesting about the physics of the game rather than attempts at genuinely reproducing a realistic economical model¹⁴. This point of view is, however, personal and disputed¹⁵ [70]. Another weak argument for the Minority Game is that it assumes perfect synchronisation. The latter¹⁶ *cannot be achieved; even worse, all the trade orders arrive sequentially, hence simultaneity is a theoretician’s phantasm at best* [159]. Also in a pure trading point of view it is really the combined notional of the market participants that decides on the direction of the market. To clarify, let us imagine an ecosystem of algorithm trading on the market: the majority of these algorithm are TFs following an upward trend and there is within these algorithms one (the minority) which has a contrarian view of the market. This algorithm would do very poorly in this ecosystem but would prevail in

¹³Rules at the microscopic level reveal laws at the macroscopic level [157].

¹⁴The minority game remain an economical model since it is a game theoretical model but perhaps not realistic when applied to the Financial markets.

¹⁵Jean Philippe Bouchaud believes that this point of view is “harsh”. More specifically he feels that some of the findings associated to the Majority Game [159, 160, 161] would help nuance our opinion.

¹⁶Like it is the case the Minority Game.

the MG. Choosing the more realistic assumption, and perhaps renaming the exercise the majority game instead, would yield uninteresting mathematical results but would be slightly more realistic in its economical assumptions¹⁷. Also, a very important feature of real economic models for algorithmic trading is the volume traded¹⁸ but this latter point is entirely neglected in the MG. Generally speaking, there are also additional points orthogonal to the volume argument. For instance, economic theory has been dominated in the past two centuries with the idea that in acting selfishly, agent contribute in a way that is most advantageous to society [162]. This latter point contrasts with the MG, where the minority wins but the extend at which the minority wins is lower than the total lost by the majority. This means that iteratively the game is not “win-win” or even “zero-sum” but “lose-lose” which does not reflect the increased utility the economy has had on our lives as real agents.

7.3 Theoretical Biology Review

It is widely accepted, in the context of ecosystem modelling, that complexity should always arise from simplicity [163, 53]. The stability of an ecosystem, once complex is however more prone to a debate. More specifically, it was discussed in the 1960s [164] that complexity in an ecosystem insures its stability or to keep the same jargon communities not being sufficiently complex to damp out oscillations have a higher likelihood of vanishing [165, 166]. The diversity-stability debate has in fact been ongoing since the 1950s [167] with no consensus being ever reached. It was initially believed that nature was infinitely complex and therefore more diverse ecosystem should insure more stability [167, 168, 169]. This assertion was however ultimately challenged through rigorous mathematical specification [163, 170, 171] in the 1970s and 1980s by using Lotka-Volterra’s Predator/Prey model initially published in the 1920’s [172, 173] with similar “non-intuitive” results. More recently the work has been extended to small ecosystems of three-species food chain [174]. The intuitive three species example we have chosen to discuss is the one containing Sharks (denoted by z parameter), Tunas (denoted by y parameter) and Small Fishes (denoted by x parameter), the idea being that tunas eat small fishes and tunas are eaten by sharks. Without loss of generality sharks are assumed to die of natural causes and their decomposing bodies join an infinite supply of food for the small fishes. This idea can be formalized by a set of differential equations summarized in equation (7.5).

¹⁷Note that this latter point has been disputed with the rational that the market favours the sellers when most want to buy and vice versa [157].

¹⁸Though it does increase ultimately its complexity.

Definition (Lotka-Volterra 3-Species Predator Prey Model): Let a be the natural growth rate of species $x(t)$ (with $\mathbb{R} \xrightarrow{x} \mathbb{R}$) in the absence of predator, d the one of $y(t)$ (with $\mathbb{R} \xrightarrow{y} \mathbb{R}$) in the absence of $z(t)$ (with $\mathbb{R} \xrightarrow{z} \mathbb{R}$). We also have b representing the negative predation effect of $y(t)$ on a and e the one of $z(t)$ on $y(t)$. We also have g which mirrors the efficiency of reproduction of $z(t)$ in the presence of prey $y(t)$. The relationship between $x(t)$, $y(t)$ and $z(t)$ is given by equation (7.5).

$$\begin{cases} \frac{dx(t)}{dt} &= ax(t) - bx(t)y(t), \\ \frac{dy(t)}{dt} &= -cy(t) + dx(t)y(t) - ey(t)z(t), \\ \frac{dz(t)}{dt} &= -fz(t) + gy(t)z(t). \end{cases} \quad (7.5)$$

Remark Note that we assume that $x(t)$ never dies of natural causes (if it's too old then it can't run fast enough to outrun its $y(t)$ predators) but this is not the case for $z(t)$ since it is an alpha predator and therefore needs some natural death rate which is controlled by f .

This relatively simple system of three equations has been studied extensively [167] for stability, for example via Lyapunov coefficient [175] and the eigenplane of the Jacobian matrix [174]. There are different traditional ways to represent stability or instability for these kind of equations, for example Figure 7.5 represents a 2D stable representation and Figure 7.8 represents a particular 3D unstable representation. For the latter case, we can notice that the oscillations between the three species increases to the point, here not shown, where the amplitudes are so big that z goes extinct and at which point x and y start oscillating, with however a constant amplitude. We refer the motivated reader back to the original paper [167] for similar examples with interesting (idiosyncratic) properties. One interesting point to notice is that in cases of “relative best stability”, in which $a = b = c = d = e = f = g = 1\%$ from Figure 7.5, we have oscillations which are stable through time with the highest peak from the ultimate prey (x) coming first and the lowest peak of the ultimate predator (z) coming last. This suggests that sophisticated working trading strategies need enough prey like strategies in the same ecosystem to get them to be profitable. One other interesting observation is that the total ecosystem population as depicted in the thick black line from the same figure suggests that it itself oscillates which may not necessarily be intuitive. Indeed one could have speculated that the loss of a species directly benefits the other and that therefore the total population should stay constant. This interesting observation suggests that the oscillations of a financial market *may likewise be subject to similar dynamics*: a financial ecosystem may go through periods in which it thrives followed by period in which it declines. The

economy itself is somewhat a noisy version of Figure 7.8. The stunning similarities

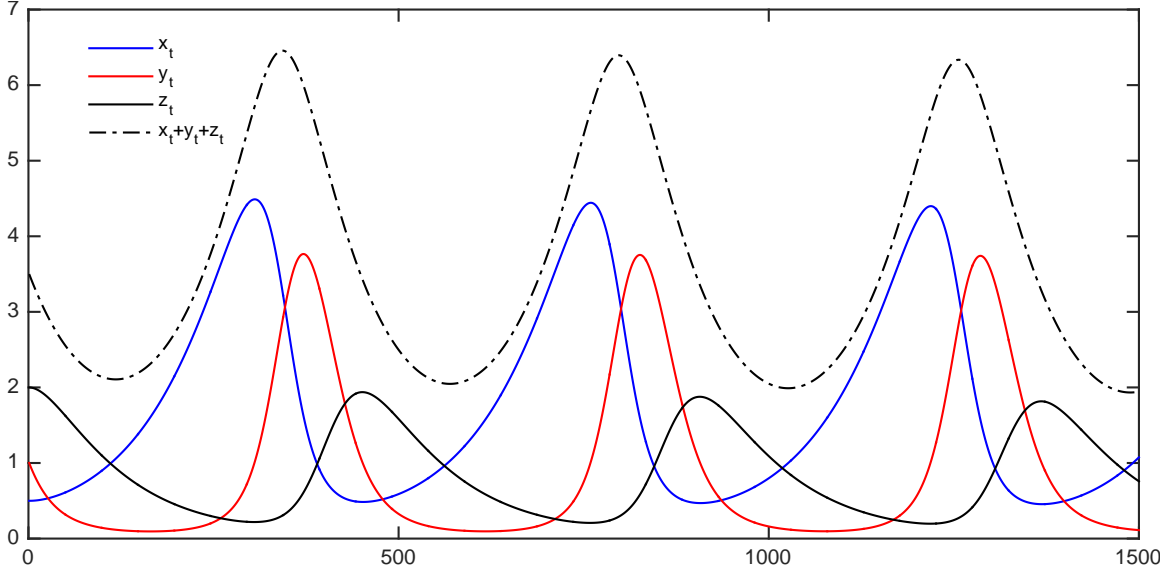


Figure 7.5: Stable 3-species Lotka-Volterra Simulation

of the competitive resource driven biological ecosystem along with some compelling similarities in some of its cyclical behavior makes the Lotka-Volterra n -species food chain equation an interesting candidate when it comes to studying the stability of the financial market especially the electronic trading markets because of its systematic rule based approach and non zero sum game like roots. However, fees and transaction costs are obviously not taken into consideration in the biological system described.

7.4 Formalizing the Evolutionary Process

With the aim of providing intuition with respect to the sort of interaction that may occur between strategies, we need to formalize the Evolutionary Process (EP), but first, we go through few definitions.

Definition (Evolutionary Process): In the context of our study, the set of rules that control the continuous change of an ecosystem and more specifically its agents (e.g. strategies), will be arbitrarily called Evolutionary Process.

Definition (Iteration Types): We define two types of iterations. The first type of iteration will be called **Micro**, corresponding to an infinitesimal increment in our environment namely, an increment in which a strategy S analyses and in turn changes the order book by placing an order itself. The second type of iteration will be called

Macro, corresponding to a generational increment in our environment namely, *a certain equal number* of Micro increment per strategy leading to a calculation of profit and loss (P&L) and a survival process¹⁹ based on this P&L.

Definition (Strategy): A strategy will consist of an HFFF \mathcal{H} , a rolling P&L \mathcal{P} and a common orderbook \mathcal{O} as shown by equation (7.6).

$$\mathcal{S} \triangleq \{\mathcal{P}, \mathcal{H}, \mathcal{O}\}. \quad (7.6)$$

Definition (Alive/Survived Strategy): A strategy is defined as alive if it does currently take action on the EP.

Definition (Dead Strategy): A strategy is defined as dead if it no longer takes actions on the EP.

Definition (Born Strategy): A strategy is born if it is about to take action on the EP for the first time.

Definition (Strategy Classification): We will label N_k the number of total alive strategies, N_k^e the number of trend following like strategies, N_k^m the number of **multi**-linear regression like strategies, N_k^r the number of **xor** like strategies and N_k^o the number of *other unclassified* strategies²⁰. The relationship between these entities can be summarized by equation (7.7).

$$N_k = N_k^e + N_k^m + N_k^r + N_k^o. \quad (7.7)$$

Remark One may ask why we have not chosen the first letters of each of the strategies (“t” for trend following, “m” for multi-linear regression and “x” for XOR strategy). The reason why this has been named this way is because as we will see in Section 7.3, we wanted to hypothesize a parallel between the biological Predator-Prey model. Namely we propose that N_k^e behaves in mathematical biology like the number of preys in a Lotka-Volterra (LV) three species equations [174], that N_k^m behaves in mathematical biology like the number of **mixed** (both prey and predator) in the same system of equation and that N_k^r behaves in mathematical biology like the number of super predators as the third species of that system of equations. The different possible permutations, constraints on the first letters being different for each type of strategy and the association to the LV three species equation, made the choice of e , m and r at first glance the most optimal in this qualitative optimization by constraint problem.

¹⁹Explained next.

²⁰This label will be the same in Section 7.3.

7.4.1 Survival & Birth Processes

The survival, death & birth processes are a set of processes which impact the number of live strategies N_k at any time k the following way. Let $\mathcal{S}_{N_k} = S_{(1)}, S_{(2)}, \dots, S_{(n)}, S_{(n+p)}, \dots, S_{(N_k)}$ be the strategies ranked with respect to their P&L from highest to lowest, we will admit the following definitions:

Definition (Survivor Set): The Survivor set²¹ is the set of strategies with a positive P&L. Namely if $\mathcal{S}_a = S_{(1)}, S_{(2)}, \dots, S_{(s)}$ with $S_{(s)} \geq 0$ and $S_{(s+1)} < 0$. We will subdivide this set by distinguishing the secondary survivors set which cardinality $a_2 = \lfloor \frac{s}{2} \rfloor$, survive without reproducing and the primary survivors set which cardinality $a_1 = s - a_2$, which survive and have one offspring with a “slightly different DNA” in form of a conditional resampling of their NN format.

Definition (Birth Process): We denote by the Birth process, the first half of survived strategies. Namely, if $a_1 = b = \lfloor \frac{s}{2} \rfloor$ the strategies $\mathcal{S}_1 \dots \mathcal{S}_{a_1}$ will both survive and reproduce and create a set of equal size but with a slightly different HFFF and with cardinality $b = a_1$.

Definition (Death Process): We denote by the Death process, the set of strategies with a negative P&L. Namely if $\mathcal{S}_d = S_{(s+1)}, S_{(s+2)}, \dots, S_{(N_k)}$ will disappear from the market at Macro iteration $k + 1$.

Remark We can easily see that $s = a_1 + a_2$, $a_1 \geq a_2$, $a_1 = b$. Figure 7.6 illustrates these few definitions.

7.4.2 Inheritance with Mutations

The intuition about the mutation process is that each birth is function of a successful strategy (the positive P&L of parents $\mathcal{S}_1 \dots \mathcal{S}_{a_1}$) and should therefore resemble a great deal to that single parent²² which produced it but be at the same time a bit different to allow the ecosystem to evolve. We have seen in Section 6.5.1 that the DNA of our strategies is essentially their HFFF \mathcal{H} (which is itself a combination of weights). We therefore concentrate on performing the re-sampling on the weights of the offspring. Recall that the pdf of the beta distribution is given by

$$\mathcal{B}(x, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} \quad (7.8)$$

²¹Or alternatively alive process.

²²So no crossover in this model.

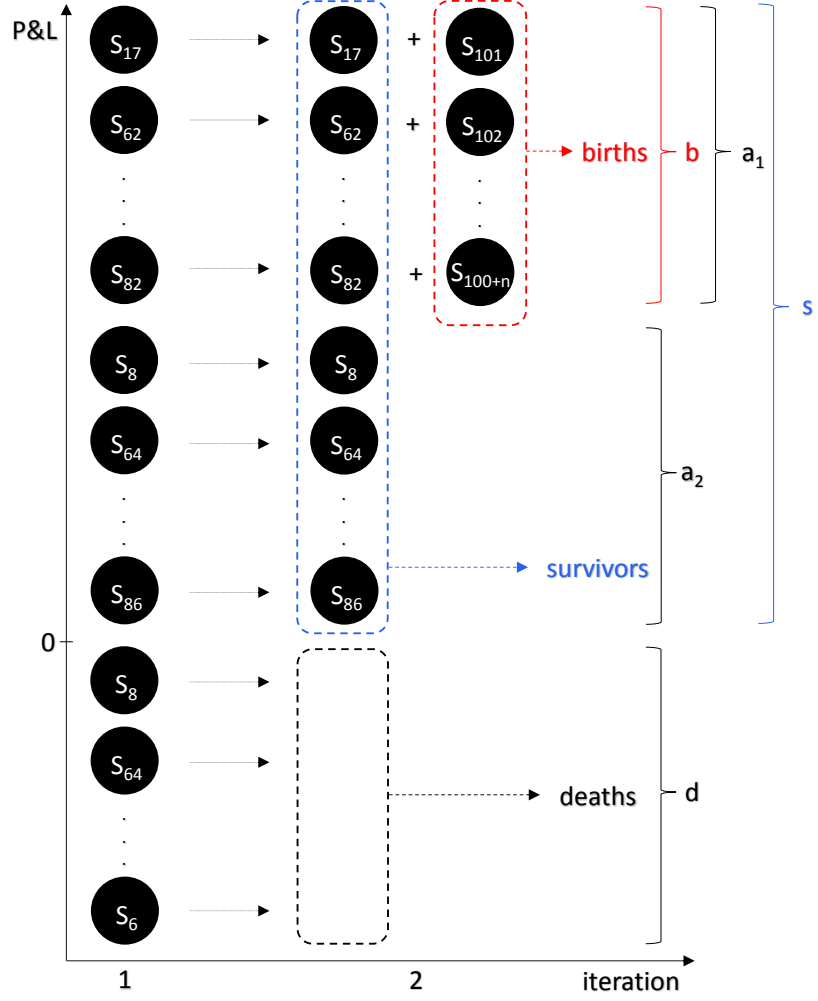


Figure 7.6: Illustration for the Death and Birth processes in our EP.

with $\Gamma(n) = (n-1)!$, $0 \leq x \leq 1$, and $\alpha, \beta > 0$, the shape parameters. This distribution is interesting because it is defined in a closed interval $[0, 1]$ and can therefore be rescaled easily through a change of variable to $[-1, 1]$, an interval which is a basic way of formalizing a normalized importance of each node in the NN format decision making of Figure 6.6. It also offers a broad range of interesting shapes allowing the possibility to code a conditional resampling model and therefore make clever proximity changes around the symbolic levels: -1 , 0 and 1 . This way we can prevent too large deviations and rather select small incremental changes and intuitively follow the principles of selection. We can see that the $\mathcal{B}(x, 1, 7)$ or $\mathcal{B}(1 - x, 1, 7)$ both concentrate a great deal of the mass of the distribution towards 0 and 1 respectively. Likewise $\mathcal{B}(x, 3, 7)$ and $\mathcal{B}(x, 5, 7)$ provide a more unbiased modification of the deviations since more symmetrical.

Definition (Mutation Sampling): Each HFFF is composed of a collection of branches formalized by equation (6.3). During a birth each of these branches is sampled according to equation (7.9).

$$\mathcal{D}(\tilde{x}) = \mathcal{B}\left(\frac{\tilde{x}+1}{2}, \alpha(\tilde{x}), \beta\right) 1_{\tilde{x} \leq -\frac{1}{5}} + \mathcal{B}\left(1 - \frac{\tilde{x}+1}{2}, \alpha(\tilde{x}), \beta\right) 1_{\tilde{x} > \frac{1}{5}} + f(\tilde{x}) 1_{|\tilde{x}| \leq \frac{1}{5}}, \quad (7.9)$$

$$\text{where } \alpha(\tilde{x}) = \begin{cases} 1, & \text{if } 1 > |\tilde{x}| \geq \frac{3}{4} \\ 3, & \text{if } \frac{3}{4} > |\tilde{x}| \geq \frac{1}{2} \\ 5, & \text{if } \frac{1}{2} > |\tilde{x}| \geq \frac{1}{5} \end{cases}, \quad F(k) = \begin{cases} \frac{1}{10}, & \text{if } k \leq -\frac{1}{5} \\ \frac{8}{10}, & \text{if } |k| < \frac{1}{5} \\ 1, & \text{if } k \geq \frac{1}{5} \end{cases}, \quad \tilde{x} \in]-1, 1[\text{ and } \beta = 7.$$

Remark The function $\alpha(\tilde{x})$ models the interval of condition and is arbitrarily chosen, though constructed by noticing that the mode of the Beta distribution is given by $\frac{\alpha-1}{\alpha+\beta-2}$ and also so as to make the intervals loosely equal.

7.5 Evolutionary Dynamics Simulation

7.5.1 Observations

Following Cedric Villani's [145] comment on the relationship between theory and simulation, more specifically around how simulations can give us good intuition about the theory, we lay forward the results which helped us formalize the hypothesis that we investigate latter on in the chapter more specifically when it comes to the kind of interactions that may take place. These interactions will be formally addressed through the concept of Path of Interaction that we will introduce in Subsection 7.6. However in the meantime, in order to discuss the matter at the intuitive level only we denote by "HFFF 1" *Trend Following* (TF), "HFFF 2" *Multi-Linear Regression* (MLR) and "HFFF 3" *XOR*.

Remark Figure 7.7 has been included for illustration purpose only and not as a supporting material. The strategies were conveniently classified in order to expose the idea that there ought to be a direct correlation between NN density and likelihood of being a predator. There is therefore a strong bias in the results (that we are the first to stress here). These results cannot possibly be taken as a strong supporting evidence but rather as an illustration of what we are trying to test. They could almost be seen as "philosophy"²³ [70]. We describe next those illustrations.

²³This is the wording used by Jean Philippe Bouchaud during the viva.

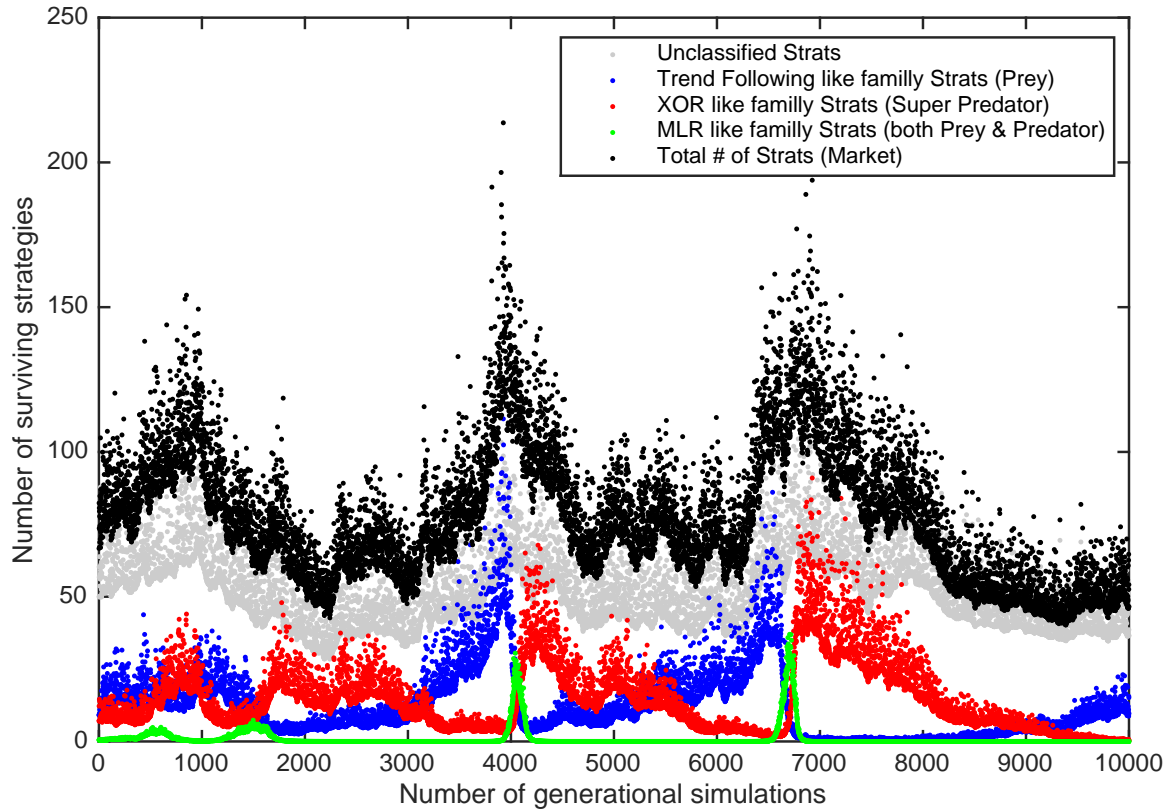


Figure 7.7: Hypothesis of how an HFTE Simulation should perform.

We may hypothesize the following. First the market is bullish in the first part then becomes bearish in the next part, while the market is bearish the TF type strategies, first increases in frequencies then diminishes in the middle of the zones, at which point the MLR type strategies increases but ultimately decreases when the XOR strategies frequency increases.

Remark A great deal of simulation were performed to verify the hypothesis, Figure 7.7 represents one of them. However none was conclusive enough for us to present them in a way that would be transparent. More specifically some classification issues²⁴ and lack of consistency in the simulations results discouraged us to publish some of these findings. However, in the interest of keeping the intuition alive, let us assume, temporarily that the simulations were of better quality²⁵.

²⁴We have mentioned in chapter 6.3.

²⁵We will eventually abandon this path and take a more granular approach in Section 7.6.

7.5.2 Philosophical Interpretation

At the early stages of research, we hypothesized the following incomplete interpretation: there ought to be a direct correlation between NN density and invasion. Indeed, TF strategies are what people commonly call self-fulfilling prophecies meaning that they only work as long as everyone making up the competitive environment follow the same trend. The biological mirror, as described from Section 7.3, would be an ultimate prey which given an environment without any predator would never die and actually grow exponentially. The XOR strategy is hypothesized as being a super predator strategy (similar to the z parameter in Section 7.3) and feeds on the MLR strategies. MLR is hypothesized as being both a predator and a prey strategy. It is supposed to feed onto the TF strategies but is used as prey by the XOR strategies. The way the MLR should dominate the TF strategy is due to the fact that it looks at additional leading information on the orderbook (the volumes at the different depth of the order book) so it should be leading in the trend whereas the TF should be lagging on the trend. XOR strategies could in this situation only survive if enough prey (MLRs) are present in the ecosystem otherwise they die. The way the XOR strategy is assumed to dominate the MLR strategy is due to its ability to “hide its cards” better and is able to better decipher spurious positions at higher depths of the orderbook. The XOR strategy is hypothesises as unable to invade the TF strategies on its own since the sophistication of its bait (the systematic strategy built to bait the MLR) is too complex to “trick” the TF. An analogy could be made with a professional poker player playing with a beginner whose moves are almost random.

Remark It has been speculated that the need for a bigger brain in the human species is partly due to the need for human to elaborate deceitful strategies with their rivals and cooperative strategies with their allies. It is therefore not entirely ridiculous to associate increased neural network branching (to be roughly understood as increase in cranial size) with increased strategy complexity. However, increased intelligence does not necessarily equate to survival. A way to illustrate this is to observe the shark population, which is considered like an apex predator in the sea but with a relatively small brain that has not evolved for millennia. By analogy, we could speculate that strategies with increased complexity may win in the short run but may not necessarily prevail in the long run. We will tackle these questions more rigorously in Section 7.6 but wanted in the meantime to stress that the problem may be more complex and less intuitive than what is may look like initially.

7.5.3 Regulation

7.5.3.1 Optimal Control Theory

The Hamilton-Jacobi-Bellman (HJB) partial differential equation [176] was developed in 1954 and is widely considered as a central theme of optimal control theory. Its solution is the value function giving the minimum cost for a given dynamical system and its associated cost function. Solved locally, the HJB is a necessary condition, but when, over the entire state space, it is referred to as necessary and sufficient for an optimum. Its method can be generalized to stochastic systems. Its discrete version is referred to as the Bellman equation and its continuous version, the Hamilton-Jacobi equation. Formally we consider the problem in deterministic optimal control over the time period $[0, T]$ the equation (7.10):

$$V(x(0), 0) = \min_u \left\{ \int_0^T C[x(t), u(t)] dt + D[x(T)] \right\}, \quad (7.10)$$

where $C[., .]$ is the scalar cost rate function, $D[., .]$ is the utility at the final state, $x(t)$ the system state vector with $x(0)$ usually given, and finally $u(t)$ where $0 \leq t \leq T$ is called the control vector we aim at finding. The system of equations is also subject to $\dot{x}(t) = F[x(t), u(t)]$ where $F[., .]$ is a deterministic vector describing the evolution of the state vector over time.

7.5.3.2 Partial Differential equation Specification

The HJB partial differential equation is given by:

$$\dot{V}(x, t) + \min_u \{ \nabla V(x, t) \cdot F(x, u) + C(x, u) \} = 0, \quad (7.11)$$

subject to the terminal condition $V(x, T) = D(x)$. The function $V(x, t)$, commonly known as the Bellman value function (our unknown scalar), represents the cost incurred from starting in x at time t and controlling the system optimally until T . The function $V(x(t), t)$ is the optimal cost-to-go function, then by Bellman's principle of optimality from time t to $t + dt$, we have $V(x(t), t) = \min_u \{ V(x(t + dt), t + dt) + \int_t^{t+dt} C(x(s), u(s)) ds \}$. The Taylor expansion of the first term is $V(x(t + dt), t + dt) = V(x(t), t) + \dot{V}(x(t), t) dt + \nabla V(x(t), t) \cdot \dot{x}(t) dt + o(dt)$ where $(o)(dt)$ denotes the higher order terms of the Taylor expansion. Canceling $V(x(t), t)$ on both sides and dividing by dt , and taking the limit as dt approaches zero, we obtain the HJB equation. Its resolutions is done backwards in time which can be extended to its stochastic version. In this latter case we have $\min_u \mathbb{E} \left\{ \int_0^T C(t, X_t, u_t) dt + D(X_T) \right\}$, with this time

$(X_t)_{t \in [0, T]}$ being stochastic²⁶ and $(u_t)_{t \in [0, T]}$ the control process. By first using Bellman and then expanding $V(X_t, t)$ with Ito's rule, one finds the stochastic HJB equation $\min_u \{\mathcal{A}V(x, t) + C(t, x, u)\} = 0$ where \mathcal{A} represents the stochastic differentiation operator, and subject to the terminal condition $V(x, T) = D(x)$ ²⁷.

7.5.3.3 Algorithm Trading Systemic Risk

Given that this thesis proposes that the fluctuations of the markets are linked to the frequency of the strategies composing the ecosystem of the market, we propose a model which would take advantage of this assumptions to propose to build the first few steps of a theory that would help high level regulations. The exercise would consist of monitoring these strategies interactions and flagging the market when necessary. This may sound a bit grand or overly ambitious but for the sake of opening up a discussion or at least exposing the benefits of future research let us develop a bit the argument. Suppose now that we label strategies of Figure 6.7, 6.9 and 6.13 by respectively x , y and z and that we use equation (7.5). If we can somehow correctly classify and guess what the frequency of x , y and z are in the ecosystem, then we can study whether or not the ecosystem is stable [174]. Now going back to the actual mathematical study of the stability of the financial market. Answering if a financial market composed of 3 strategies is stable would come to studying the Jacobian matrix J from equation (7.12).

$$J(x, y, z) = \begin{bmatrix} a - by & -xb & 0 \\ yd & -c + dx - ez & -ye \\ 0 & -zg & -f + gy \end{bmatrix} \quad (7.12)$$

More specifically, by examining the eigenvalues of $J(x, y, z)$ we can indirectly gain information about the future equilibrium of our financial system, an important element at the regulatory level²⁸. More specifically if all eigenvalues of $J(x, y, z)$ have negative real parts then our system is asymptotically stable. Figure 7.8 gives an illustration of a situation in which one of the eigenvalues is negative. Many questions could be raised here: how can the regulators gain information on the parameters composing the system of equations (7.5)? Also, the market has surely more than 3 types of strategies, how many exactly? Are these strategies easily classifiable in terms of prey, predator and super predator or can you find more subtle instances? It is very likely that trading desks especially in the high-frequency domain refuse to provide their sets of strategies for the regulators to study the Jacobian matrix in order to take the

²⁶and needing optimization.

²⁷The randomness has disappeared.

²⁸We assume for the sake of this example that we only have 3 strategies.

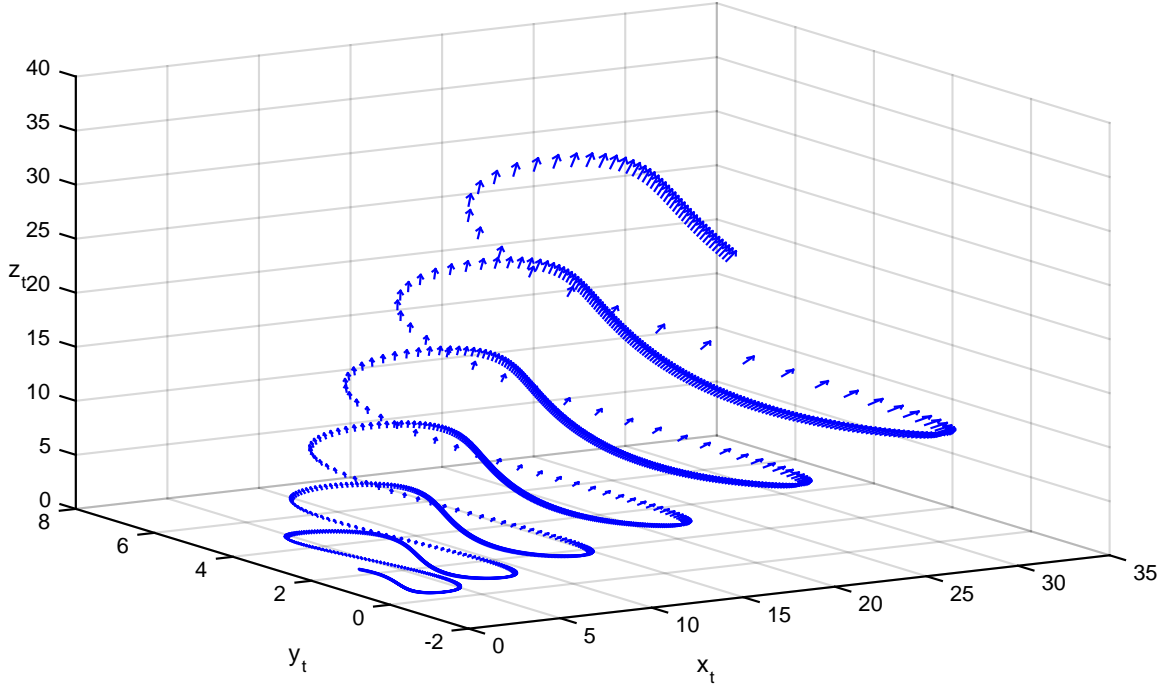


Figure 7.8: Unstable 3-Species Lotka-Volterra Simulation

relevant actions²⁹. However having the strategies on their own would not do much. It is really the relationship of these strategies that need to be better understood. We take this opportunity to recall the hypothesis we introduced in our last paper [4]:

Hypothesis: Diversity in financial strategies in the market lead to its instability.

7.6 Path of Interaction

Our first few simulations, despite not fulfilling the burden of proof, opened our eyes up to issues associated to optimality, need for more scientific rigour and perhaps an alternative way to fulfill this burden of proof. The concept of Path of Interaction that we introduce next is an attempt at addressing this alternative methodology.

7.6.1 HFTE Game

7.6.1.1 Definitions

One way to control our simulation issues, is to perhaps take a step back in complexity in order to gain momentum in constructing a theory with more rigor. With this in

²⁹Instruct the trading desks to increase or decrease their notional so as to enforce a manual intervention for the sake of the market's stability.

mind we have chosen to inspire ourselves from the scientific method used by Axelrod [28, 29] extended by Nowak's [30, 31], and to introduce a mathematical object, similar in spirit to the PD matrix used as a battle ground (Figure 7.1) by the name of Path of Interaction. In order to do this rigorously. Let us first provide a few definitions.

Definition (Dynamic Mini Order-Book): We denote by Dynamic Mini Order-Book o , the sequence of length l of static snapshots of the order-book $^{a_2, a_1}M_{b_1, b_2}$ of asked, a_i , and bid, b_i , volumes where i corresponds to the depth of the order book and M its mid price.

Remark In the context of our study we take $l = 4$.

Definition (Ranking Rule): A Ranking Rule are the set of directives that decides the Birth, Death and Survival processes of any Strategy Ecosystem.

Definition (Environment): We call an Environment e of size i a set of evolving strategies, $S = s^a, s^b, \dots, s^i$ of HFFF spanning the one from Figure 6.6 with potential to interact with each other one after the other via an order-book, $^{a_2, a_1}M_{b_1, b_2}$. This will contain a “seed” consisting of different volumes at different depth that would need to be clarified for each game.

Remark Note that the Ranking Rules we assume going forward are the one described by Figure 7.6. The environment can then evolve according to a set of Ranking Rules.

Definition (HFTE Game): We call an HFTE Game the sequence of Environments composed of 2 strategies, $S = s^a, s^b, \dots, s^i$ of HFFF spanning the one from Figure 6.6 with a dynamic mini order-book and P&L.

Definition (Full Order-Book (FOB)): An OB is called full if and only if it has a volume of 1 on all the depth of the OB.

Definition (Path of Interaction Table): We call a Path of Interaction Table an HFTE Game decomposed in its most infinitesimal steps.

The top row of the table points to the strategies involved. The row below (2nd row from the top) is the stage of the HFTE Game. The 3rd row corresponds to the trading signal. The game starts in a state in which none of the two strategies has a position (Signal = “N/A”) on the order book. Because each strategy needs some form of information on the order book, we assume that there is a random seed on the

order book. There are four possibilities of random seeds corresponding to whether the price has been going up or down and whether the order book has increased its OI or decreased it. These four situations are symbolized by the following set of symbols: \Uparrow , \Downarrow , \Updownarrow and \Downarrow . We have chosen the case of \Uparrow to illustrate our examples arbitrarily. The 4th row corresponds to the order book state. The latter can be either full or not. We will see that this latter point matters but for now let us illustrate this point with an example.

Example In Table 7.2, we start with $P^{1,1,1,1,1,1}$ meaning that at the current price P , we have one order to sell at the first 6 depths of the order book. The 5th row corresponds to the current price (last completed order) or the midprice if no order was completed in the current iteration. The 6th row corresponds to the OI. If the buy side of the order book has one of its orders matched then the OI decreases by 1 (-1 if the opposite occurs). The 7th row corresponds to the price change. If no order is matched, then the price is approximated by the mid price. The last row corresponds to the profit and loss. In order to illustrate the Path of Interaction we propose to go through the details of a TF strategy interacting with another TF. Algorithm (8) represents our simplified TF strategy and Table 7.2 represent the Path of Interaction of two strategies following the systematic rules of algorithm (8). In this table,

Algorithm 8: Simplified TF Strategy

Input: $s, \Delta O, \Delta P$

Output: o

▷ Our simple TF Strategy copies last update's trend while disregarding OI

if $\Delta P > 0$ **then**

| order \leftarrow 1

else if $\Delta P < 0$ **then**

| order \leftarrow -1

else

| order \leftarrow 0

return order

for simplicity³⁰, we represent only one side of the order book: $P^{1,1,1,1,1,1}$ (for display purposes seeing that the price only takes one direction in the simulations). Since both strategies follow the trend, and the order book is full, the price keeps increasing, their respective P&L keeps increasing and the OI imbalance keeps decreasing³¹. Table 7.2

³⁰The price dynamics goes in only one direction in this case.

³¹Note that during the viva of this thesis it was pointed out that the quality of the simulation would be greatly enhanced by a market making strategy [94] which would incorporate market orders to insure a minimum bid-ask spread. This is certain a valid point and would need investigation.

Strategy	seed $\uparrow\uparrow$	TF1	TF2	TF1	TF2	TF1	TF2
Iteration	0		1		2		3
Signal	N/A	+1	+1	+1	+1	+1	+1
OB	$P^{1,1,1,1,1,1}$	$_0P^{1,1,1,1,1}$	$_{0,0}P^{1,1,1,1}$	$_{0,0,0}P^{1,1,1}$	$_{0,0,0,0}P^{1,1}$	$_{0,0,0,0,0}P^1$	$_{0,0,0,0,0,0}P$
Mid	100	101	+102	103	104	105	106
ΔOI	+1	-1	-2	-3	-4	-5	-6
$\Delta Price$	+1	+1	+1	+1	+1	+1	+1
P&L	[0, 0]		[1, 0]		[2, 1]		[3, 2]

Table 7.2: Path of Interaction for 2 TF Strategies with $\uparrow\uparrow$ Seeds and Full OB

can therefore be seen as a way to illustrate³² that the TF strategies interacting with each other is “self fulfilling”, a terminology we introduce next.

We introduced, in Section 7.2.3, the concept of Invasion Flow Chart. We translate, next the same method, for quantitative financial strategies. We go first through few formal definitions.

Definition (Invasion): A strategy, s , is invasive with respect to an environment, e when the P&L of s increases in the environment e and if the frequency of s in e increases as a result. An invasion is always done in the context of an order-book containing a seed (that needs to be clarified) and a selection process (e.g. Figure 7.6) that may change from game to game.

Remark Note that the P&L alone does not make a strategy invasive but rather the latter in conjunction with a genetic algorithm which would discriminate against strategies with negative P&L and promote through reproduction the ones with a positive P&L. With time the frequency of successful strategies would increase to the point in which the ecosystem would evolve. One of these algorithms was presented in Figure 7.6. Also note that if a strategy A invades B in an ecosystem, e containing

However, the method presented in this chapter of the thesis is mainly a translation of Axelrod’s computer tournament [29, 28, 30]. Indeed, as a friendly reminder, Axelrod himself did not propose all the strategies from inception but rather constructed the platform on which, fellow scientist would incorporate theirs.

³²The formal proof may require more rigour.

a seed s_1 , B could still invade A in the context of another seed s_2 . The length of the interaction is also critical. For instance A invades B in an ecosystem, e with s_1 only in the first x iterations but the invasion could be reversed if the iteration increases (and the order-book therefore changes significantly).

Example For instance, if we assume that the more complex a network is, the more likely it is to invade, we would expect to see an invasion flowchart like the one in Figure 7.9. Indeed, if we assume that TF strategies bring some sort of innovation from a random swarm and if we assume that the MLR sees more information than the TF (and so on ...) then Figure 7.9 represents a flow chart that exhibits the idea that complexity and invasion are linked. This chart also assumes that beyond XOR strategies, the complexity would be such that it would equate to a random strategy or would alternatively take a complex path which would lead to a “farmer” like strategy. We will illustrate later on in this chapter that the hypothesis illustrated by Figure 7.9 is not necessarily verified.

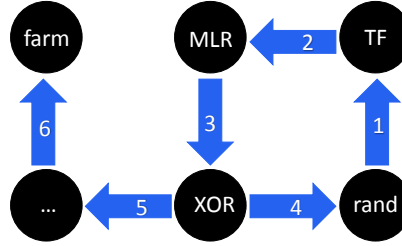


Figure 7.9: Illustration for a hypothetical Strategy Invasion Map [4, 3]

Definition (Self-Fulfilling): We call a strategy, s Self-Fulfilling when its frequency in the environment increases (through a genetic algorithm for example) as a result of the presence of other instances of the same strategy.

7.6.2 Strategy Tournament

7.6.2.1 Foreword

Before we discuss our Strategy Tournament, in order to avoid the classification issues mentioned earlier in the thesis, we take their simplest forms. First, let us introduce the simplified MLR strategy formalized in algorithm (9). The idea of this simplified version is that Price and OB imbalance both contribute in defining the trading signal. The last simplified strategy is a simplified XOR strategy formalized in algorithm (10). A Path of Interaction tournament was implemented in the context of 15 possible

Algorithm 9: Simplified MLR Strategies

Input: $s, \Delta O, \Delta P$ **Output:** o \triangleright Simplified MLR Strategy follows the trend until basic OB imbalance**if** $\Delta O + 2 \times \Delta P > 0$ **then**| order \leftarrow 1**else if** $\Delta O + 2 \times \Delta P < 0$ **then**| order \leftarrow -1**else**| order \leftarrow 0 \triangleright Return order

Algorithm 10: Simplified XOR Strategies

Input: $s, \Delta O, \Delta P$ **Output:** o \triangleright Defining simplified XOR Strategy**if** $(\Delta O > 0) \ \& \ (\Delta P > 0)$ **then**| order \leftarrow 1**else if** $(\Delta O > 0) \ \& \ (\Delta P < 0)$ **then**| order \leftarrow -1**else if** $(\Delta O < 0) \ \& \ (\Delta P > 0)$ **then**| order \leftarrow -1**else if** $(\Delta O < 0) \ \& \ (\Delta P < 0)$ **then**| order \leftarrow 1**else**| order \leftarrow 0 \triangleright Return order

games on 7 different timescales: $[0, 2, 3, 5, 11, 23, 47]$. The choice of these timescales may be a little odd at first glance but the idea was to increase the timescale on average by a factor of two while at the same time picking prime numbers. The idea of the latter is related to an intuition that we had over potential cycles occurring in these games. And we thought that if the timescales were to be chosen on common factors, it would create a potential additional layer of complexity which would have prevented us from seeing the bigger picture³³.

Remark In order to use some conventions around strategy sequences for HFTE games we have chosen the following notation $\xrightarrow{s_1} s_2$ and $s_2 \xrightarrow{s_1} s_3$ to mean, for the first case, that strategy s_1 changes first the OB, then s_2 (and the sequence continues until the end of the timescale) and, for the second case s_3 impacts the OB after s_2 (before, again going back to s_1). For example, $TF \hookrightarrow TF$ means that the environment

³³But this again was an intuitive approach.

is composed of two TF strategies and $MLR \xrightarrow{TF} XOR$ refers to an HFTE game composed of a TF, MLR and XOR strategy which OB impact sequence is one which mimics the intuitive order laid down by the \hookrightarrow symbol (TF, first, MLR, second and XOR, third). These symbols are expended into their full form in Tables 7.3 and 7.4 but we thought it would be useful to have a text friendlier version for the analysis.

7.6.2.2 Results

Table 7.3 reports the results of these games for two strategies interacting and Table 7.4 represents the same for 3 strategies. We make several interesting observations.

Proposition 7.6.1 *The TF strategy is self-fulfilling on a OB that is full.*

Proof We have illustrated this point with Table 7.2. Though only on 4 iterations, the proof can be expanded on longer timescales using recursion. ■

Remark The intuition we had [4] around the TF acting like a prey increasing exponentially in frequency in the absence of predator is confirmed. The first connections to the Lotka-Volterra 3-species predator/prey model is established. It is worthy to note however that there is a benefit in starting first as the TF1 does better at the end in this HFTE game. Also note that the Limit Order Book (LOB) is filled for this example. The rational for filling the LOB is to have some snapshot of an order-book that would not be biased (no order book imbalance³⁴ is allowed at the game's inception). The alternative, an empty order book, would have been less interesting³⁵.

Proposition 7.6.2 *A strategy A can invade a strategy B but the latter can invade the same strategy B if the seed or and the sequence in which these strategies are started changes.*

Example The MLR strategy invades the TF strategy on the longer times scales (in column s_2 of Table 7.3 we can see that for timesclares greater than 3 but lower than 47, the MLR strategy achieves a higher P&L) but when the MLR starts the HFTE game (column s_4 of Table 7.3), then the TF strategy invades the MLR strategy. The same remark can be made when the XOR strategy take the MLR spot in the same HFTE set up (column s_3 and s_7 of Table 7.3).

³⁴Please see [126] in order to get acquainted with this specific terminology

³⁵Though not uninteresting.

Scenario	$\begin{array}{c} \text{TF} \\ \curvearrowright \\ \text{TF} \end{array}$	$\begin{array}{c} \text{TF} \\ \curvearrowright \\ \text{MLR} \end{array}$	$\begin{array}{c} \text{TF} \\ \curvearrowright \\ \text{XOR} \end{array}$	$\begin{array}{c} \text{MLR} \\ \curvearrowright \\ \text{TF} \end{array}$	$\begin{array}{c} \text{MLR} \\ \curvearrowright \\ \text{MLR} \end{array}$	$\begin{array}{c} \text{MLR} \\ \curvearrowright \\ \text{XOR} \end{array}$	$\begin{array}{c} \text{XOR} \\ \curvearrowright \\ \text{TF} \end{array}$	$\begin{array}{c} \text{XOR} \\ \curvearrowright \\ \text{MLR} \end{array}$	$\begin{array}{c} \text{XOR} \\ \curvearrowright \\ \text{XOR} \end{array}$
Key	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
Round									
P&L ₀	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$
ΔP_0	0	0	0	0	0	0	0	0	0
P&L ₂	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$
ΔP_2	0	0	0	0	0	0	0	0	0
P&L ₃	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$
ΔP_3	3	3	-1	3	3	-1	3	3	-1
P&L ₅	$\begin{bmatrix} 9 \\ 7 \end{bmatrix}$	$\begin{bmatrix} -9 \\ 8 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 9 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 9 \end{bmatrix}$	$\begin{bmatrix} 7 \\ -2 \end{bmatrix}$
ΔP_5	6	-3	-3	5	-5	-4	5	-4	3
P&L ₁₁	$\begin{bmatrix} 45 \\ 40 \end{bmatrix}$	$\begin{bmatrix} -39 \\ 21 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -50 \\ 78 \end{bmatrix}$	$\begin{bmatrix} 15 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 14 \\ 7 \end{bmatrix}$	$\begin{bmatrix} -9 \\ 26 \end{bmatrix}$	$\begin{bmatrix} -16 \\ 21 \end{bmatrix}$	$\begin{bmatrix} 7 \\ -2 \end{bmatrix}$
ΔP_{11}	15	11	-3	13	15	-6	11	11	3
P&L ₂₃	$\begin{bmatrix} 198 \\ 187 \end{bmatrix}$	$\begin{bmatrix} -216 \\ 48 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -326 \\ 387 \end{bmatrix}$	$\begin{bmatrix} 39 \\ 38 \end{bmatrix}$	$\begin{bmatrix} 74 \\ 11 \end{bmatrix}$	$\begin{bmatrix} -87 \\ 122 \end{bmatrix}$	$\begin{bmatrix} -21 \\ 54 \end{bmatrix}$	$\begin{bmatrix} 7 \\ -2 \end{bmatrix}$
ΔP_{23}	33	26	-3	28	36	-10	23	27	3
P&L ₄₇	$\begin{bmatrix} 828 \\ 805 \end{bmatrix}$	$\begin{bmatrix} -703 \\ 354 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -1580 \\ 1707 \end{bmatrix}$	$\begin{bmatrix} 96 \\ 99 \end{bmatrix}$	$\begin{bmatrix} 290 \\ 19 \end{bmatrix}$	$\begin{bmatrix} -459 \\ 530 \end{bmatrix}$	$\begin{bmatrix} -27 \\ 54 \end{bmatrix}$	$\begin{bmatrix} 7 \\ -2 \end{bmatrix}$
ΔP_{47}	69	-75	-3	58	78	-18	47	27	3

Table 7.3: P&L in Path of Interaction for 2 Strategies with $\uparrow\uparrow$ Seeds and Full OB

Remark Note that what we call “Round” in tables 7.3 and 7.4 corresponds to the iteration of the game. For instance $P\&L_{23}$ corresponds to the cumulative profit each strategy has made, competing in the environment up to the 23rd iteration.

Proposition 7.6.3 *The Dominance relation is not transitive.*

Example This comes to exposing that if a strategy A dominates a Strategy B and Strategy B dominates Strategy C, this does not mean that Strategy A will dominate Strategy C. A counterexample for this point is given by s_2 , s_6 and s_3 of Table 7.3.

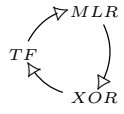
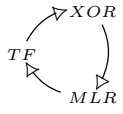
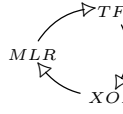
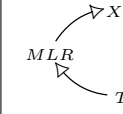
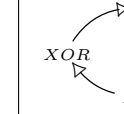
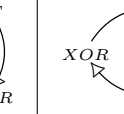
Scenario						
Key	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}
Round						
P&L ₀	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$
ΔP_0	0	0	0	0	0	0
P&L ₂	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$
ΔP_2	0	0	0	0	0	0
P&L ₃	$\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$
ΔP_3	3	-2	3	-1	-1	3
P&L ₅	$\begin{bmatrix} 4 \\ 0 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -2 \\ -8 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 8 \\ -5 \\ 7 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 6 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -11 \\ -3 \\ 20 \end{bmatrix}$	$\begin{bmatrix} -3 \\ 7 \\ -4 \end{bmatrix}$
ΔP_5	4	3	-4	-5	-6	-3
P&L ₁₁	$\begin{bmatrix} -26 \\ 60 \\ -21 \end{bmatrix}$	$\begin{bmatrix} -108 \\ 61 \\ -36 \end{bmatrix}$	$\begin{bmatrix} 17 \\ -7 \\ -34 \end{bmatrix}$	$\begin{bmatrix} 14 \\ -18 \\ -37 \end{bmatrix}$	$\begin{bmatrix} 10 \\ -26 \\ 25 \end{bmatrix}$	$\begin{bmatrix} -48 \\ 57 \\ -47 \end{bmatrix}$
ΔP_{11}	17	-14	7	6	18	16
P&L ₂₃	$\begin{bmatrix} -65 \\ -13 \\ 119 \end{bmatrix}$	$\begin{bmatrix} -164 \\ 131 \\ -39 \end{bmatrix}$	$\begin{bmatrix} 57 \\ -35 \\ -198 \end{bmatrix}$	$\begin{bmatrix} 54 \\ -62 \\ -201 \end{bmatrix}$	$\begin{bmatrix} 128 \\ -93 \\ -74 \end{bmatrix}$	$\begin{bmatrix} 137 \\ 145 \\ -96 \end{bmatrix}$
ΔP_{23}	31	-32	15	14	45	-46
P&L ₄₇	$\begin{bmatrix} -3127 \\ 2500 \\ -231 \end{bmatrix}$	$\begin{bmatrix} -720 \\ 250 \\ -289 \end{bmatrix}$	$\begin{bmatrix} 233 \\ -187 \\ -910 \end{bmatrix}$	$\begin{bmatrix} 230 \\ -246 \\ -913 \end{bmatrix}$	$\begin{bmatrix} 187 \\ -230 \\ -588 \end{bmatrix}$	$\begin{bmatrix} 553 \\ 621 \\ 13 \end{bmatrix}$
ΔP_{47}	-54	-54	31	30	97	-104

Table 7.4: P&L in Path of Interaction for 3 Strategies with $\uparrow\uparrow$ Seeds and Full OB

Proposition 7.6.4 *Having a more complex strategy does not mean it will be invasive.*

Example We can observe in column s_7 of Table 7.3, that the TF strategy invades the XOR strategy over the first 47 iterations even-though the XOR strategy involves a hidden layer, on the contrary to the TF strategy that consist of only 1 input.

Proposition 7.6.5 *All strategies can make money even if the market goes down.*

Example See s_6 example in Table 7.3. ■

Proposition 7.6.6 *Starting first is not always an advantage.*

Example See s_5 in Table 7.3 for the example (even with twin strategies). ■

7.6.2.3 Few Interesting Hypotheses

Finally we wanted to end this chapter by suggesting few hypotheses based on some of our observations.

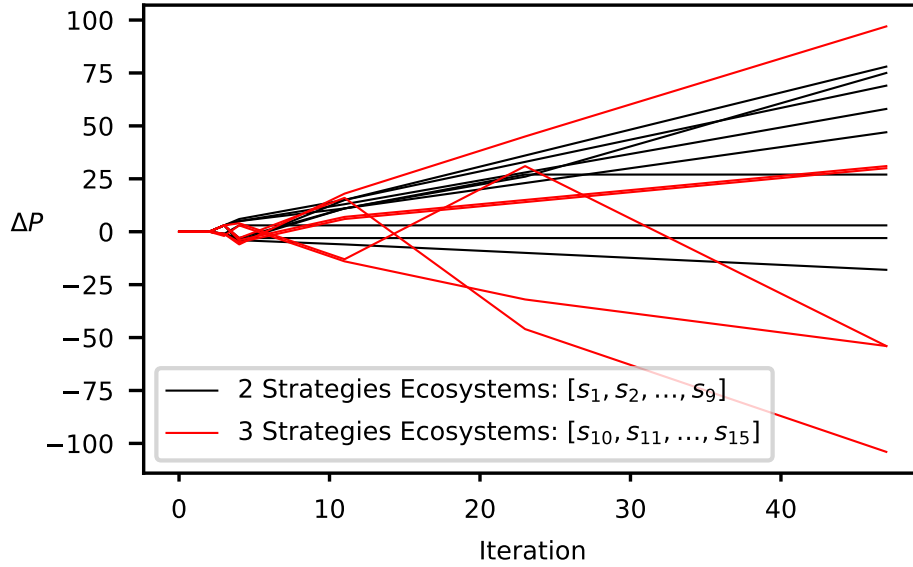


Figure 7.10: Instability increases with an additional strategy.

Hypothesis: In the finite horizon we have looked at, we noticed that all strategies in our ecosystems can increase their P&L at the same time but all cannot decrease their P&L at the same time. This observation, again on a finite horizon may not be true as $t \rightarrow \infty$.

We noticed this interesting fact with our relatively small sample of HFTE games but have not been able to find a counter example yet nor been able to rigorously prove it. The proof might be easier than it seems, using perhaps the pigeonhole principle but we have not been able to formalize the proof or a sketch.

Hypothesis: Similar physical laws drive *morality*³⁶ and HFTE games.

We explain next this seemingly odd and unexpected terminology. In the spirit of using simple rules at the agent level as a triggering point to complex interactions in the ecosystem that can turn into laws, we thought that this second hypothesis would also be inspiring. This latter proposed hypothesis might not be immediately obvious but there seems to be interesting connection between the TF strategy in an HFTE game and the TFT strategy in Axelrod's [28, 28] computer tournament described in Section 7.2.3. As a reminder the TFT strategy cooperates first and continues doing so until it is deceived, upon which it deceits on the next move. However, the TFT has the ability to forgive. This means that, if the opponent agent decides to cooperate again then the TFT, starts cooperating on the next encounter. The TFT is therefore considered a nice strategy [48] but adaptable at the same time [30]. So how does that relate to the TF in finance? Both are successful strategies yet are very simple. They both replicate the last agent's move: so they are both "cooperative"³⁷ but are adaptable³⁸.

Remark This comparison may not seem quite apropos at first or at least it may not be intuitive. This may be related to the negative bias we have against the moral aspects of Finance. These are due to many elements but one contributing factor is due to some of the misconducts in HFT which are more related to unfair advantages in technology or immoral actions [177, 178] taken in an unfair game based on asymmetric information about the market. In our research technological advantages are not taken into consideration.

Hypothesis: Diversity in financial strategies in the market lead to its instability.

Remark Finally we proposed a previously introduced hypothesis we wanted to raise before concluding this chapter: we noticed, in Figure 7.10 that the 3 strategies ecosystems exhibited more fluctuations than the 2 strategies ecosystems which tend to support the hypothesis than more diversity in an ecosystem of strategies induces more instability³⁹ to the market. It does however suggests it empirically but this example does not constitute obviously a proof. Also, a bigger ecosystem allows for more time

³⁶We refer here to some of the work associate the formal mathematical definition of morality [48], more specifically in the context of cellular automaton and the iterative prisoners dilemma [29, 28, 30].

³⁷Cooperative in evolutionary dynamics seem to translate into "trending" in QF.

³⁸The TF can change his position on the market if the trend changes.

³⁹This assertion could be challenged with a simulation which additional strategies (such as potentially a market making strategy) would be designed in order to create this stability [94].

for a random walk to depart from its expected mean so to some extent the fact that the 3 strategies ecosystem has a higher variance is explained by that increased timescale but the increased fluctuation seems to go beyond what is expected by the addition of an additional step. It is also important to point that the stability and diversity debate has had an interesting breakthrough recently [179]. More specifically, the stability of the equilibria reached by ecosystems formed by a large number of species with strong and heterogeneous interactions (therefore more realistic ecological niches) the system displays multiple equilibria which are all marginally stable. Though this studies is applied to biological niches, it is not difficult to imagine that a similar result could be found in an algorithmic trading ecosystem. This would therefore contradict the hypothesis we have put forward.

In this chapter we have built the humble start of a schemes involving the Bottom-Up approach to algorithmic trading. We first attempted to reach that objective by tackling the problem using a simple genetic algorithm methodology. Though intuitive an interesting, we abandoned this approach because of a series of problem associated, but not limited to, classification, lack of visibility and lack of optimality. We however, took this opportunity to shown possible connections to other STEM fields and how they could be brought in the world of QF through the regulatory door. To study the problem with more visibility, rigour, and in order to gain momentum, we took a step back in the scientific approach and formalized the HFTE game as well as the Path of Interaction concepts. We have also given 15 different kinds of HFTE games split on 7 different timescales and also presented few interesting observations about the interesting complexity in the relationship of these strategies, even when simplified. This study was done with the premise that we knew what strategies were involved in the ecosystem and in which sequence they act upon this ecosystem. Though simplistic, in the choice of the available strategies⁴⁰, the current model give a good overview of how the method could be enhanced by simply adding more strategies including market makers. However, market participants are quite secretive in reality when it comes to their financial strategies. The only observable data on the market is essentially the price dynamics and the order book. We explore in the next chapter how inference can be constructed in the Bottom-Up approach when the price dynamic alone is available.

⁴⁰The presence of market makers would make the results more interesting [94].

Chapter 8

Stability of Financial Systems and Multi-Target Tracking

In this section we take a look at another example in which ML can revolutionize classic Mathematical Finance as it lays down the foundations for controlling systemic risk in a challenging electronic trading environment where speed and secrecy are of utmost importance. More specifically, in Section 8.1, we offer a literature review of Multi-Target Tracking methodologies starting with the linear, and then moving to non linear methods. In Section 8.2 we expand the study by connecting some of the concepts in the previous chapter by constructing a particle filtering methodology applied to the tracking of ecosystem of financial strategies.

8.1 Classic Methods in Multi-Target Tracking

Multi-Target Tracking (MTT), which deals with state space estimation of moving targets, has applications in different fields [180, 181, 182], the most intuitive ones being perhaps military related such as for radar and sonar design. Let us first review some of the classic models associated to a single target.

8.1.1 Linear Methods

8.1.1.1 Kalman Filter

The Kalman Filter (KF) is a mathematical tool which provides the best estimation (in a MSE sense) of some dynamical process, (x_k) , perturbed by noise and influenced by a controlled process (u_k) . The estimation of (x_k) , is based upon observations, (y_k) , which are function of these dynamics. A review can be found in [183] and the

dynamics of (x_k) are given by:

$$x_k = F_k x_{k-1} + B_k u_k + w_k, \quad (8.1)$$

where F_k is the state transition model which is applied to the previous state x_{k-1} ; B_k is the control-input model which is applied to the vector u_k ; w_k is the noise process, which is assumed to be drawn from a zero mean multivariate normal distribution with covariance matrix Q_k , so $w_k \sim N(0, Q_k)$. At time k an observation of y_k is made according to equation (8.2).

$$y_k = H_k x_k + v_k, \quad (8.2)$$

where H_k is the observation model which maps the true state space into the observed space. The noise, v_k , induced by the observation is assumed to be zero mean Gaussian white noise with $v_k \sim N(0, R_k)$. We also assume that the noise vectors $(\{w_1, \dots, w_k\},$

Algorithm 11: Kalman Filter

Input: array of weights w_1^N

Output: array of weights w_1^M resampled

$\hat{x}_{k|k-1} \leftarrow F_k \hat{x}_{k-1|k-1} + B_{k-1} u_{k-1}$
 $P_{k|k-1} \leftarrow F_k P_{k-1|k-1} F_k^T + Q_{k-1}$

▷ Predicted state

Innovation (or residual)

$\tilde{y}_k \leftarrow y_k - H_k \hat{x}_{k|k-1}$

Covariance

$S_k \leftarrow H_k P_{k|k-1} H_k^T + R_k$

Optimal Kalman gain

$K_k \leftarrow P_{k|k-1} H_k^T S_k^{-1}$

Updated state estimate

$\hat{x}_{k|k} \leftarrow \hat{x}_{k|k-1} + K_k \tilde{y}_k$

Updated estimate covariance

$P_{k|k} \leftarrow (I - K_k H_k) P_{k|k-1}$

▷ Update state

▷ Return state

Return w_1^M

$\{v_1 \dots v_k\}$ at each step are mutually independent¹. The KF being a recursive estimator, we only need the estimated state from the previous time step and the current measurement to compute the estimate for the current state. \hat{x}_k will represent the estimation of our state x_k at time up to k . The state of our filter is represented by two variables: $\hat{x}_{k|k}$, the estimate of the state at time k given observations up to

¹That is, $cov(v_k, w_k) = 0$ for all k .

and including time k ; $P_{k|k}$, the error covariance matrix (a measure of the estimated accuracy of the state estimate). The KF has two distinct phases: predict and update. The predict phase uses the state estimate from the previous timestep to produce an estimate of the state at the current timestep. In the update phase, measurement information at the current timestep is used to refine this prediction to arrive at a new, more accurate state estimate, again for the current timestep. The formula for the updated estimate covariance above is only valid for the optimal Kalman gain. Usage of other gain values require a more complex formula. The KF methodology has been summarized by algorithm (11).

Remark Please see original papers [184, 185] for the formal derivation of the model.

Although the KF presents obviously lots of benefits in tracking, its linear constraints makes it not a the ideal choice for non linear applications.

8.1.1.2 Extended Kalman Filter

The EKF is essentially an approximation of the KF for “non-severely-non-linear” models. In the EKF the state transition and observation models need not be linear functions of the state but may instead be differentiable functions. The dynamics and measurements of the model is presented in equation (8.3).

$$\begin{cases} x_k &= f(x_{k-1}, u_k) + w_k, \\ y_k &= h(x_k) + v_k. \end{cases} \quad (8.3)$$

where $f(\cdot)$ and $h(\cdot)$ are derivable functions and represent the dynamics of x and its indirect observation respectively. The algorithm is very similar to the one described in algorithm (11) but with couple of modifications highlighted in algorithm (12)².

Remark The derivation of the model from algorithm (12) is very similar to the one of algorithm (11) with couple of exceptions, first F_k and H_k approximations at the first order of F_k and H_k , we obtain a truncation error which can be bounded and satisfies the inequality known as Cauchy’s estimate: $|R_n(x)| \leq M_n \frac{r^{n+1}}{(n+1)!}$, here $(a-r, a+r)$ is the interval where the variable x is assumed to take its values and M_n is a positive real constant such that $|f^{(n+1)}(x)| \leq M_n$ for all $x \in (a-r, a+r)$. The value of M_n increases as the curvature or non-linearity of $f(\cdot)$ is more pronounced. When $|R_n(x)|$ increases it is possible to improve our approximation at the cost of complexity by increasing by

²Note that here $F_k = \frac{\partial f}{\partial x} \bigg|_{\hat{x}_{k-1|k-1}, u_k}$ and $H_k = \frac{\partial h}{\partial x} \bigg|_{\hat{x}_{k|k-1}}$.

one degree our Taylor approximation, i.e: $F_k = \frac{\partial f}{\partial x} \Big|_{f(\hat{x}_{k-1|k-1}, u_k)} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \Big|_{f(\hat{x}_{k-1|k-1}, u_k)^2}$
and $H_k = \frac{\partial h}{\partial x} \Big|_{f(\hat{x}_{k|k-1})} + \frac{1}{2} \frac{\partial^2 h}{\partial x^2} \Big|_{f(\hat{x}_{k|k-1})^2}$. ■

Algorithm 12: Extended Kalman Filter

Input: array of weights w_1^N

Output: array of weights w_1^M resampled

▷ Predicted state

$$\begin{aligned}\hat{x}_{k|k-1} &\leftarrow f(\hat{x}_{k-1|k-1}, u_k) \\ P_{k|k-1} &\leftarrow F_k P_{k-1|k-1} F_k^T + Q_{k-1}\end{aligned}$$

▷ Update state:

$$\begin{aligned}&\text{Innovation (or residual)} \\ \tilde{y}_k &\leftarrow y_k - h(\hat{x}_{k|k-1}) \\ &\text{Covariance} \\ S_k &\leftarrow H_k P_{k|k-1} H_k^T + R_k \\ &\text{Optimal Kalman gain} \\ K_k &\leftarrow P_{k|k-1} H_k^T S_k^{-1} \\ &\text{Updated state estimate} \\ \hat{x}_{k|k} &\leftarrow \hat{x}_{k|k-1} + K_k \tilde{y}_k \\ &\text{Updated estimate covariance} \\ P_{k|k} &\leftarrow (I - K_k H_k) P_{k|k-1}\end{aligned}$$

Remark Though the EKF tries to address some of the limitations of the KF by relaxing some of the linearity constraints it still needs to assume that the underlying function dynamics are both known and differentiable.

8.1.2 Non-Linear Methods

8.1.2.1 Importance Sampling

Importance sampling (IS) was first introduced in [186] and was further discussed in several books including in [187]. The objective of importance sampling is to sample the distribution in the region of importance in order to achieve computational efficiency via lowering the variance of an estimator. The idea of importance sampling is to choose a proposal distribution $q(x)$ in place of the true, harder to sample probability distribution $p(x)$. The main constraint is related to the support of $q(x)$ which is assumed to cover that of $p(x)$.

$$\int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx, \quad (8.4a)$$

$$\hat{f} = \frac{1}{N_p} \sum_{i=1}^{N_p} W(x^{(i)})f(x^{(i)}). \quad (8.4b)$$

In equation (8.4a) we write the integration problem in discrete time with equation (8.4b) the numerical approximation where N_p , usually describes the number of independent samples drawn from $q(x)$ to obtain a weighted sum to approximate \hat{f} ,

$$W(x^{(i)}) = \frac{p(x^{(i)})}{q(x^{(i)})}, \quad (8.5a)$$

$$W(x^{(i)}) \propto p(x^{(i)}) q(x^{(i)}), \quad (8.5b)$$

and where $W(x^{(i)})$ in equation (8.5a) is the Radon-Nikodym derivative of $p(x)$ with respect to $q(x)$ also known in the engineering literature as the importance weights. Equation (8.5b) suggests that if the normalizing factor for $p(x)$ is not known, the importance weights can only be evaluated up to a normalizing constant. To ensure that $\sum_{i=1}^{N_p} W(x^{(i)}) = 1$, we normalize the importance weights to obtain equation (8.6).

$$\hat{f} = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} W(x^{(i)}) f(x^{(i)})}{\frac{1}{N_p} \sum_{i=1}^{N_p} W(x^{(i)})} = \frac{1}{N_p} \sum_{i=1}^{N_p} \tilde{W}(x^{(i)}) f(x^{(i)}), \quad (8.6)$$

where $\tilde{W}(x^{(i)}) = \frac{W(x^{(i)})}{\sum_{i=1}^{N_p} W(x^{(i)})}$ are called the normalized importance weights. The variance of importance sampler estimate [188] in equation (8.6) is given by

$$\begin{aligned} Var_q[\hat{f}] &= \frac{1}{N_p} Var_q[f(x)W(x)] \\ &= \frac{1}{N_p} Var_q[f(x)p(x)/q(x)] \\ &= \frac{1}{N_p} \int \left[\frac{f(x)p(x)}{q(x)} - \mathbb{E}_p[f(x)] \right]^2 q(x) dx \\ &= \frac{1}{N_p} \int \left[\left(\frac{f(x)p(x)}{q(x)} \right)^2 - 2p(x)f(x)\mathbb{E}_p[f(x)] \right] dx + \frac{(\mathbb{E}_p[f(x)])^2}{N_p} \\ &= \frac{1}{N_p} \int \left[\left(\frac{f(x)p(x)}{q(x)} \right)^2 \right] dx - \frac{(\mathbb{E}_p[f(x)])^2}{N_p} \end{aligned}$$

The variance can be reduced when an appropriate $q(x)$ is chosen to either match the shape of $p(x)$ so as to approximate the true variance; or to match the shape of $|f(x)|p(x)$ so as to further reduce the true variance of \hat{f} . To see this we know that

$$\frac{\partial Var_q[\hat{f}]}{\partial q(x)} = -\frac{1}{N_p} \int \left[\frac{(f(x)p(x))^2}{q(x)^2} \right] dx = -\frac{1}{N_p} \int \left[\frac{(f(x)p(x))^2}{q(x)q(x)} \right] dx$$

with $q(x)$ having the constraint of being a probability measure that is $\int_{-\infty}^{+\infty} q(x) dx = 1$, we find that $q(x)$ must match the shape of $p(x)$ or of $|f(x)|p(x)$.

8.1.3 Resampling Methods

Resampling methods are usually used to avoid the problem of weight degeneracy in our algorithm. Avoiding situations where our trained probability measure tends towards the Dirac distribution must be avoided because it does not give much information on all the possibilities of our state. There exists many different resampling methods, Rejection Sampling, Sampling-Importance Resampling, Multinomial Resampling, Residual Resampling, Stratified Sampling, and the performance of our algorithm can be affected by the choice of our resampling method. The stratified resampling proposed by Kitagawa [189] is optimal in terms of variance. Figure 8.1 gives an illustration of the Stratified Sampling and the corresponding algorithm is described in algorithm (13). We see at the top of the Figure 8.1 the discrepancy

Algorithm 13: Resample

Input: array of weights w_1^M

Output: array of weights w_1^M resampled

▷ Sample:

$$u^0 \sim \mathcal{U}[0, 1/M]$$

▷ Resample:

for $m = 1$ **to** N **do**

$$\left[\begin{array}{l} i^{(m)} \leftarrow \left\lfloor (w_n^{(m)} - u^{(m-1)}m) \right\rfloor + 1 \\ u^{(m)} \leftarrow u^{(m)} + \frac{i^{(m)}}{M} - w_n^{(m)} \end{array} \right]$$

between the estimated pdf at time t with the true pdf and the corresponding CDF of our estimated PDF. Random numbers from $[0, 1]$ are drawn, depending on the importance of these particles and are moved to more useful places as a result.

8.1.3.1 Sequential Monte Carlo Methods

Sequential Monte Carlo methods (SMC), also known as Particle Filters (PF) are statistical model estimation techniques based on simulation. They are the sequential (or “on-line”) analogue of Markov Chain Monte Carlo (MCMC) methods and similar to importance sampling methods. If they are elegantly designed they can be much faster than MCMC. Because of their non linear quality they are often an alternative to the Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF). They however have the advantage of being able to approach the Bayesian optimal estimate with sufficient samples. They are technically more accurate than the EKF or UKF. The aims of the PF is to estimate the sequence of hidden parameters, x_k

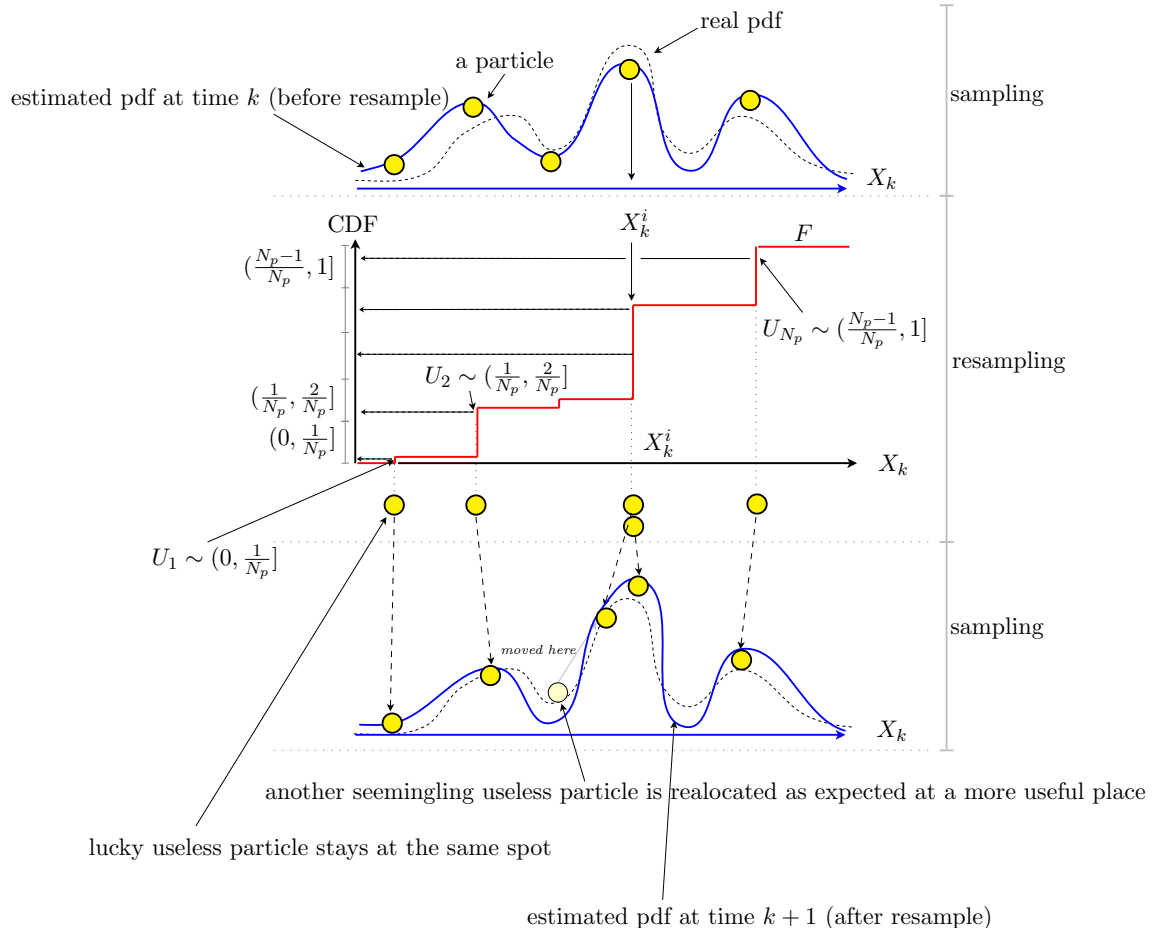


Figure 8.1: Stratified Sampling illustration

for $k = 1, 2, 3, \dots$, based on the observations y_k . The estimates of x_k are done via the posterior distribution $p(x_k|y_1, y_2, \dots, y_k)$. PF do not care about the full posterior $p(x_1, x_2, \dots, x_k|y_1, y_2, \dots, y_k)$ like it is the case for the MCMC or importance sampling (IS) approach. Let us assume x_k and the observations y_k can be modeled in the following way: $x_k|x_{k-1} \sim p_{x_k|x_{k-1}}(x|x_{k-1})$ and with given initial distribution $p(x_1)$, $y_k|x_k \sim p_{y|x}(y|x_k)$. Equations (8.7a) and (8.7b) gives an example of such system.

$$x_k = f(x_{k-1}) + w_k, \quad (8.7a)$$

$$y_k = h(x_k) + v_k. \quad (8.7b)$$

It is also assumed that $cov(w_k, v_k) = 0$ or w_k and v_k mutually independent and i.i.d. with known probability density functions. $f(\cdot)$ and $h(\cdot)$ are also assumed known functions. Equations (8.7a) and (8.7b) are our state space equations. $f(\cdot)$ and $h(\cdot)$ are linear functions, with w_k and v_k both Gaussian, the KF is the best tool to find the exact sought distribution. If $f(\cdot)$ and $h(\cdot)$ are non linear then the Kalman filter (KF)

is an approximation. PF are also approximations, but convergence can be improved with additional particles. PF methods generate a set of samples that approximate the filtering distribution $p(x_k|y_1, \dots, y_k)$. The expectations under the probability measure is given by:

$$\int f(x_k)p(x_k|y_1, \dots, y_k)dx_k \approx \frac{1}{N_P} \sum_{L=1}^{N_P} f(x_k^{(L)}) \quad (8.8)$$

where N_P denotes the number of samples. Sampling Importance Resampling (SIR) is the most commonly used PF algorithm, which approximates the probability measure $p(x_k|y_1, \dots, y_k)$ via a weighted set of N_P particles $(w_k^{(L)}, x_k^{(L)}) : L = \{1, \dots, N_P\}$. The importance weights $w_k^{(L)}$ are approximations to the relative posterior probability measure of the particles such that $\sum_{L=1}^P w_k^{(L)} = 1$. SIR is essentially a recursive version of importance sampling. Like in IS, the expectation of a function $f(\cdot)$ can be approximated by:

$$\int f(x_k)p(x_k|y_1, \dots, y_k)dx_k \approx \sum_{L=1}^{N_P} w^{(L)} f(x_k^{(L)}) \quad (8.9)$$

The algorithm performance is dependent on the choice of the proposal distribution, $\pi(x_k|x_{1:k-1}, y_{1:k})$, where the optimal proposal distribution is $\pi(x_k|x_{0:k-1}, y_{0:k})$:

$$\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1}, y_k). \quad (8.10)$$

Because it is easier to draw samples and update the weight calculations the transition prior, $\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1})$, is often used as importance function. The technique of using transition prior as importance function is commonly known as Bootstrap Filter and Condensation Algorithm. Figure 8.1 illustrates algorithm (14). Note that on line 5 of the latter, $\hat{w}_k^{(L)}$, simplifies to $w_{k-1}^{(L)}p(y_k|x_k^{(L)})$, when $\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k}) = p(x_k^{(L)}|x_{k-1}^{(L)})$.

8.1.4 Scenario Tracking Algorithm

8.1.4.1 Introduction

Recently, SMC methods [190, 191, 192], especially when it comes to the data association issue, have been developed. Particle Filters (PF) [193, 194], have recently become a popular framework for MTT, because able to perform well even when the data models are nonlinear and non-Gaussian, as opposed to linear methods used by the classical methods like the KF/EKF [195]. Given the observations and the previous target state, SMC can employ sequential importance sampling recursively

Algorithm 14: Sequential Monte Carlo

Input: array of weights w_p^N , $\pi(x_k|x_{1:k-1}^{(L)}, y_{1:k})$

Output: array of weights w_p^N resampled

▷ Sample

for $L = 1$ *to* N_P **do**

$x_k^{(L)} \sim \pi(x_k|x_{1:k-1}^{(L)}, y_{1:k})$

for $L = 1$ *to* N_P **do**

$\hat{w}_k^{(L)} \leftarrow w_{k-1}^{(L)} \frac{p(y_k|x_k^{(L)})p(x_k^{(L)}|x_{k-1}^{(L)})}{\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k})}$

for $L = 1$ *to* N_P **do**

$w_k^{(L)} \leftarrow \frac{\hat{w}_k^{(L)}}{\sum_{J=1}^P \hat{w}_k^{(J)}}$

$\hat{N}_{eff} \leftarrow \frac{1}{\sum_{L=1}^P (w_k^{(L)})^2}$

▷ Resample

draw N_P particles from the current particle set with probabilities

proportional to their weights. Replace the current particle set with this new one.

if $\hat{N}_{eff} < N_{thr}$ **then**

for $L = 1$ *to* N_P **do**

$w_k^{(L)} \leftarrow 1/N_P$.

and update the posterior distribution of our target state. The Probability Hypothesis Density (PHD) filter [196, 197, 198], which combines the Finite Set Statistics (FISST), an extension of Bayesian analysis to incorporate comparisons between different dimensional state-spaces, and the SMC methods, was also proposed for joint target detection and estimation [199]. The M-best feasible solutions is also a new useful finding in SMC [199, 200, 201, 202, 203]. Articles [204, 205] were proposed to cope with both the multitarget detection and tracking scenario but according to [206] they are not robust if the environment becomes noisier and more hostile, such as having a higher clutter density and a low probability of target detection. To cope with these problems a hybrid approach and its extensions were implemented [206]. The aim of these methods is to stochastically estimate the number of targets and therefore the multitarget state. The soft-gating approach described in [207] is an attempt to address the complex measurement-to-target association problem. To solve this issue of detection in the presence of spurious objects a new SMC algorithm is presented in [208]. That method provided a solution to deal with both time-varying number of targets, and measurement-to-target association issues. Currently, tracking

for multiple targets faces a couple of major challenges that are yet to be answered efficiently. We explore these issues in the next two subsections.

8.1.4.2 Time-Varying Number of Targets Problem

The first of these two main challenges is the modelling of the time-varying number of targets in an environment high in clutter density and low in detection probability (hostile environment). To some extent the PHD filter [209, 204, 205], based on the FISST, has proved ability in dealing with this problem with unfortunately a significant degradation of its performance when the environment is hostile [206].

8.1.4.3 Measurement-to-Target Association Problem

The second main challenge is the measurement-to-target association problem. Because there is an ambiguity between whether the observation consists of measurements originating from a true target or a clutter point, it becomes essential to identify which one is which. The typical and popular approach to solve this issue is the Joint Probabilistic Data Association (JPDA) [180, 210]. Its major drawback though is that its tracks tend to coalesce when targets are closely spaced [211] or intertwined. This problem has been, however, partially studied. Indeed the sensitivity of the track coalescence may be reduced if we use a hypothesis pruning strategy [212, 213]. Unfortunately the track swap problems still remain. Also performance of the EKF [195] is known to be limited by the linearity of the data model on the contrary to SMC based tracking algorithms developed by [214, 215, 216, 217]. This issue of data association can also be sampled via Gibbs sampling [217].

Remark Note that other problems subsist. Some of these problems are related to the ones we mentioned above and some others are different. The literature for MTT is quite rich. For the sake of keeping the literature review to a reasonable length we refer to other noticeable contributions [218, 219, 220, 216, 221, 222, 149, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 180, 210, 180, 234, 235, 236, 191, 237, 238, 239, 240, 241].

8.2 HFTE SMC Tracking Methodology

Now that we have done a comprehensive review of tracking methodologies we would like to apply our findings to the HFTE formulated problem from Subsection 7.6.2.

8.2.1 Direct Approach

The direct approach consists of tracking not only the number of alive strategies and number of births but on top of that the HFFF of each of the live strategies, namely, if each particle is associated to our state space θ then we can summarize our state space by:

$$\theta \triangleq \left\{ N_t^s, N_t^b, N_t^d, \cup_{i=1}^{N_t^a} S_i, \cup_{i=1}^{N_t^a} \mathcal{H}_i, \cup_{i=1}^{N_t^a} \mathcal{P}_i, \mathcal{O} \right\} \quad (8.11)$$

with N_t^s , the number of survived strategies, N_t^b , the number of born strategies, N_t^d , the number of dead strategies and N_t^a , the number of alive strategies³. As we can see from equation (8.11), not only do we need to keep track of the alive strategies through time but also of their type which may be overly ambitious at this stage. We would rather apply a simpler approach and make instead more transparent progress.

8.2.2 Simplified Simulation

We present here an application of the results from Section 7.6 to our tracking methodology introduced in this Chapter. We assume the state space is limited to a set of 15 scenarios spanned by up to 3 different types of strategy⁴ acting on the OB in different sequences. To manage complexity we assume that there is no birth or death processes involved in our scenarios. Algorithm (15) describes our simplified study in pseudo code. Note that the traditional resampling algorithm as developed by Doucet [242] has been substituted by the term $W_{t-1}^s + \lambda_r$ in the line $w_t^s \rightarrow \lambda_e \times \frac{L_s}{W} + (1 - \lambda_e - \lambda_r) \times W_{t-1}^s + \lambda_r \times 1/15$. We also added a small noise function to the market observed prices to make observations more realistic. The results from the series of simulations are presented in Figure 8.2. We observe that every scenario had already clearly emerged by iteration 23 (second row from the bottom on all 15 scenarios). By iteration 47, the density is very clear, so much so that the only reason it is not a Dirac function is due to the resampling methodology introduced in that effect. Though simplistic, this specific method lays down the foundations on how this specific problem should be tackled in a methodology point of view. Adding layers of complexity (e.g., increase in the number of strategies, OB dynamics, births and death) becomes more of a mathematical technicality rather than a conceptual one.

Remark This concludes the second part⁵ of the thesis in which we illustrated how data-driven models can be fundamentally “opposed” to classic financial mathematics.

³ $N_t^a = N_t^s + N_t^b$ or $N_t^a = N_{t-1}^a + N_t^b - N_t^d$.

⁴Exact formalization has been given by Algorithms (8), (9) and (10).

⁵Part III.

Algorithm 15: Particle Filter on Simplified HFTE Strategies State Space

Input: $\Delta P, I, w_t, \lambda_e, \lambda_r$ **Output:** w_t $w_{t-1} \leftarrow w_t, W \leftarrow 0$ \triangleright Sample**for** $0 \leq s \leq 15$ **do** $\quad L_s \leftarrow \exp(-\Delta P - \Delta P_{(H_I^s)})$
 $\quad W = W + L_s$ \triangleright Resample**for** $0 \leq s \leq 15$ **do** $\quad w_t^s \leftarrow \lambda_e \times \frac{L_s}{W} + (1 - \lambda_e - \lambda_r) \times W_{t-1}^s + \lambda_r \times 1/15$ \triangleright returnreturn w_t

Algorithm 16: Scenarios Hash Table H_I^s

Input: I **Output:** array position**if** $I == 0$ **then** \quad return 0**else if** $I == 2$ **then** \quad return 1**else if** $I == 3$ **then** \quad return 2**else if** $I == 5$ **then** \quad return 3**else if** $I == 11$ **then** \quad return 4**else if** $I == 23$ **then** \quad return 5**else if** $I == 47$ **then** \quad return 6**else** \quad return ‘issue with iteration recognition’

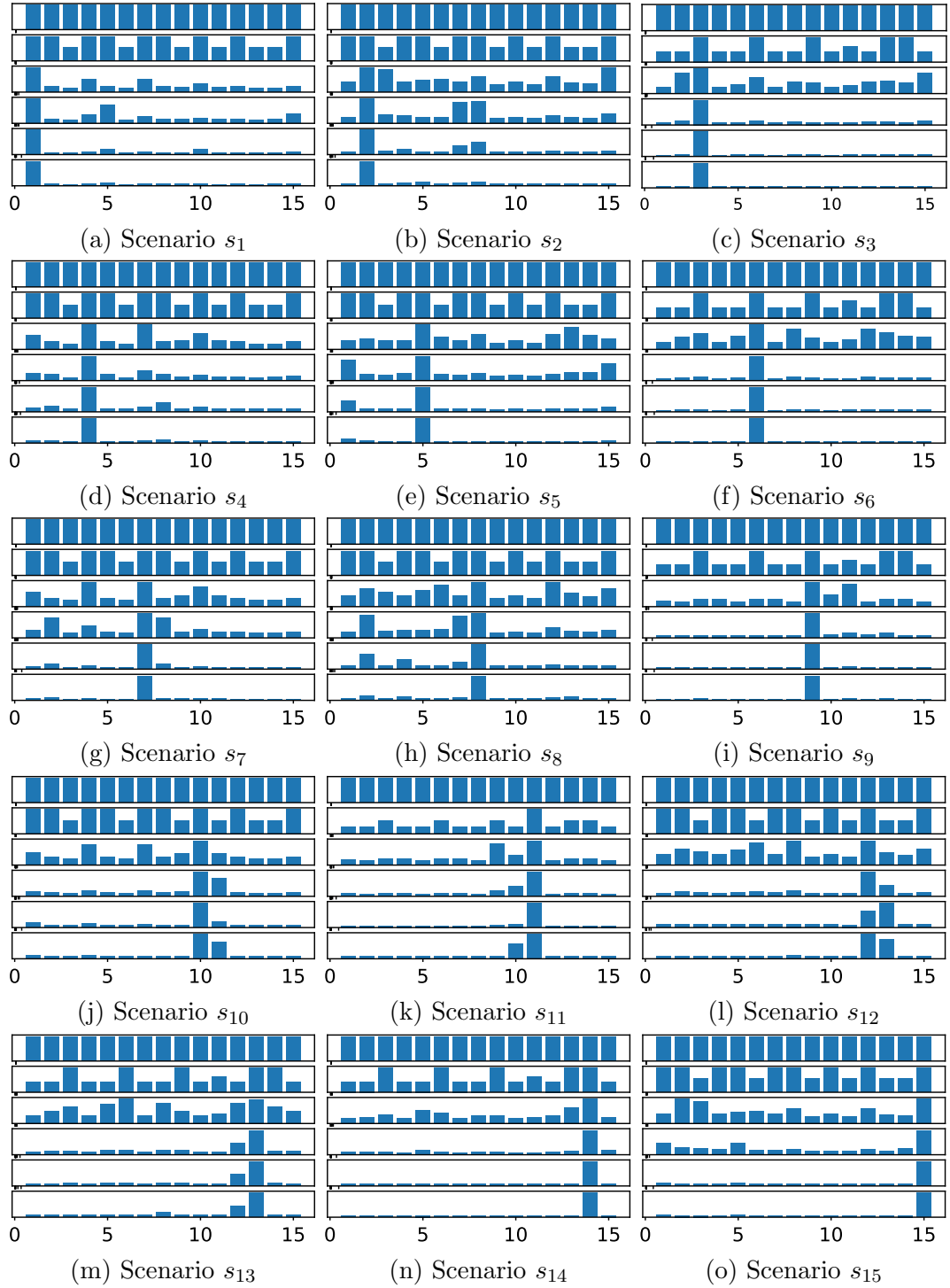


Figure 8.2: Particle Filter on market scenarios on milestones $[2, 3, 5, 11, 23, 47]$. Each of the 6 rows of the 15 ecosystems corresponds to the probability of ecosystem S_i at milestone m

Part IV

Conclusion, Appendices & Bibliography

Chapter 9

Conclusion

9.1 Covariance and Data Reassuming Models

9.1.1 Summary

The first part of the thesis was dedicated to illustrating how the triggering effect of the multiple crises has brought us into an interesting era of Quantitative Finance. More specifically, we showed that we are moving out a of a world in which models assumed data to one in which data is slowly but surely re-assuming the models. More specifically we introduced the Cointelation model in order to expose how by taking a more descriptive approach [11] of the markets instead of the mathematically convenient Black Scholes Log-Normal diffusion model [37, 46], we may measure a correlation of -1 , where in fact the real long term correlation may be $+1$. We introduce the concept of inferred correlation [10] which we explained, can be understood as the more realistic estimation of the real risk associated to holding two Cointelated pairs. We also showed how clustering can help us in the context of Cointelated pairs [5]. In the context of portfolio optimization we saw a hybrid methodology between classic Financial Mathematics and what people consider pure Machine Learning can help the resolution of the problem while illustrating and example in which FM's and ML's whole can outperform their individual sums. More specifically, we try to solve our nonlinear partial differential equation of the end of Section 2.4.6 with a deep learning algorithm¹. In the spirit of exposing additional examples of classic models (which assume data), we introduced an enhancement [12, 13] of the SVI, a classic Financial Mathematics model. This enhancement, the IVP model [13, 6], was designed to adjust the SVI model based on data driven observations on the wings of the implied

¹Using the DGM [23].

volatility surface as well as the incorporation of a liquidity overlay. Finally we proposed simple enhancements of these traditional Financial Mathematics models using clustering and illustrated our findings by introducing the concept of Responsible VaR [7] model, a portmanteau neologism that aims at reconciling discordant risk measures.

9.1.2 Future Research

9.1.2.1 Cointelated Portfolio Performance measure

We have seen that there are few performance functions that can be used (SPT, Sharpe Ratio, Utility Function) when calibrating cointelated pairs. We have decided to choose two different performance measures in our methodology. More specifically, even though the idea of a performance measure in the form of SR seems straightforward for half of the problem (the Markowitz part of the trading signal), the use of the latter in the other half of the trading signal (the OU part) seems inadequate. This made us question whether we could come up with a more robust methodology.

9.1.2.2 The n -Cointelated case

There are few extensions or improvements that can be performed on the optimization process for cointelated pairs research. We can first ask ourselves the question of the n -Cointelated case for which the trio has been specified by equation (9.1).

Definition (Three-asset Cointelation Model): Let $(\Omega, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$, be our probability space with $(\mathcal{F}_t)_{(t \geq 0)}$ where \mathbb{P} the historical probability measure under which the discounted price, rS , is not necessarily a martingale. The three-asset Cointelation model is given by

$$\begin{aligned} dS_t^a &= \sigma S_t^a dW_t^a, \\ dS_t^b &= \theta(S_t^a - S_t^b)dt + \sigma S_t^a dW_t^b, \\ dS_t^c &= \theta(S_t^a - S_t^c)dt + \sigma S_t^c dW_t^c. \end{aligned} \tag{9.1}$$

One natural question is how to define this trio. For instance, would equation (9.1) be more in-line with the pair from equation (2.5) or would

$$\begin{aligned} dS_t^a &= \sigma S_t^a dW_t, \\ dS_t^b &= \theta(S_t^a - S_t^b)dt + \sigma S_t^a dW_t^b, \\ dS_t^c &= \theta(S_t^b - S_t^c)dt + \sigma S_t^c dW_t^c, \end{aligned} \tag{9.2}$$

be better (or equivalent)? This is subject of future research.

9.1.2.3 Application to Cryptocurrencies

Following the above point, we propose to study the model using cryptocurrencies as data and relate this exercise with the n -Cointelation case. We believe that Cryptocurrencies could be an interesting case study. The reasons why Cryptocurrencies came to be is a disputed debate and we will instead refer to an external literature review [243] and focus instead on the business opportunities that have arose from this new market. Though small relative to other asset classes, the market cap is growing much faster [243] (already at a \$300 Billion Market) than the others. Cryptocurrencies offer a source of orthogonal alpha at the low frequency domain and, many asset managers want to have, as a result an overall² exposure to them. As a result, there has been an emergence of cryptocurrency indexes in the recent past with construction methodology closely related to Risk Parity and SPT [243]. Given the spectacular volatility of the cryptocurrency market, even though the point of the index is to reduce the overall volatility, the position remains fundamentally long. The resulting volatile fees (both management and performance) can be such that building a self-sufficient business in which a steady income is expected to finance the basic operations can become very challenging. However, using a beta neutral approach could be the solution and using the cointelation model here could be an interesting application. The Altcoins themselves are perhaps cointelated with respect to each other and are overall cointelated to Bitcoin. This is the hypothesis that we would like to study in the future. However we need to wait until enough data is gathered to conduct this study.

9.1.2.4 Band-wise Gaussian mixture

We have used the Band-wise Gaussian mixture methodology in chapter 5. Is this methodology optimal or can we find better? More specifically can we come up with a methodology that would be more continuous in terms of its bands definition. Ideed in Section 5, we have defined our fixed band so as to make the sample set of equal size in each band defined by $O_p^i = \{x_{(\lceil (n(i-1)+1)/p \rceil)}, \dots, x_{(\lfloor n(i)/p \rfloor)}\}$. It seems it would not require much effort to rather make these bands dynamic in such a way that the current point of investigation x_c falls exactly in the middle of O_p^i . There are also some questions around the boundary of definition that needs to be answered especially if that latter new proposal is implemented.

²Without really looking at the fundamentals.

9.1.2.5 Normalizing Options Contracts

In Section 3.2 we looked at the issues associated to Normalizing Options Contracts into standardized pillars. In doing so we used interpolation and extrapolation techniques. More specifically, to fill the sparse data we used some of the material associated to the IVS parametrization. Though these methodologies provide great approximations, they however lack robustness when it comes to insuring arbitrage at the portfolio level in between pillars (though the pillars themselves remain mutually non arbitrageable). We have seen that there exist asymptotic results. More specifically we discussed necessary but not sufficient theoretical extrapolation and interpolation techniques that have created confusions in the literature. The initial idea has still great potential which deserves continuing research in.

9.1.2.6 Particle Filter for Implied Volatility MTT

We have raised several limitation to our current model. First we have limited our model to a span of stress tests in i different parameters³. This means that very complex co-movements of three or more parameters are not taken into account. Though, these do not matter for vanilla options, they may matter for more complex exotics. More scenarios could be included.

9.1.2.7 Additional Liquidity issues for Implied Volatility

The IVP model was introduced in [13] with the aim of modelling the Bid Ask spread. It generalized the gSVI. In fact the gSVI is a special case of IVP (one in which the parameters of the liquidity component are set to superpose the mid price). Indeed the IVP model makes the necessary adjustment on the wings to incorporate bid ask spread on top of taking into account the position size. The downside transform defined by equation (4.1a) can certainly be more tailored to asset classes and perhaps fine tuned. Also, the parameters $\beta_{o,\tau}, \beta+, \tau$ and $\beta-, \tau$ model the idea that the further away you are from the ATM, the bigger the Bid Ask spread (curvature adjustment). However this change of variable could yield a Bid Ask spread of 0 ATM. This issue was addressed by adding an ATM bid ask factor function $\min(\frac{\alpha_\tau}{2}, \alpha_\tau)$, where α_τ is a guess ATM Bid Ask half spread but adjusted if its value is such that it will be higher than the lowest point of the implied volatility. However, this methodology is not the most elegant and would certainly benefit from more clever a priori analysis.

³Where $i = \{(0), (a), (b), (\rho), (m), (\sigma), (\beta), (b, \rho)\}$.

9.1.2.8 De-arbitraging the FX case

We exposed some of the asset class idiosyncratic hurdles of the de-arbitraging methodology in Section 3.5.2. More specifically we shown how the traditional de-arbing methodology fails with the FX pillars. The latter pillars are misaligned (like a trapezoid instead of the rectangle: see Figure 3.5). This problem can be addressed with a more convoluted optimization by constraint methodology. More specifically, the extrapolated pillars would still need to be mutually arbitrage free and their interpolation would still need to go through the observed points. This optimization by constraint would however suffer from convergence issues because more degrees of freedom. We will study this interesting problem in a future paper.

9.1.2.9 Harmonizing Stochastic & Local Volatility

We have seen that both the Heston and SVI models are popular in the industry and converge asymptotically to each other [59]: see equation (9.3).

$$dS_t = \sqrt{v_t} S_t dW_t^1, \quad S_0 \in \mathbb{R}_+^* \quad (9.3a)$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma v_t^{\frac{1}{2}} dW_t^2, \quad v_0 \in \mathbb{R}_+^*, \quad (9.3b)$$

$$d\langle W^1, W^2 \rangle_t = \rho dt, \quad (9.3c)$$

$$v(k, t) \rightarrow a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}]. \quad (9.3d)$$

The SVI was made obsolete at Bank of America for its inability to accurately model the sub-linearity of the wings. At the same time the Heston model is known for its inability to model the smile. This convergence is mathematically convenient for the SVI but does not solve an economic and data driven fact (see Figure 4.7). Only an implied correlation surface could reconcile the current models to the data. The proposed stochastic volatility model that we propose as being able to converge towards the IVP model is given by equation (9.4).

$$dS_t = \sqrt{v_t} S_t dW_t^1, \quad S_0 \in \mathbb{R}_+^*, \quad (9.4a)$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma \sqrt{v_t} dW_t^2, \quad v_0 \in \mathbb{R}_+^*, \quad (9.4b)$$

$$d\langle W^1, W^2 \rangle_t = \rho(t, S_t)dt, \quad (9.4c)$$

$$\rho(t, S_t) = \rho_+(t) + [\rho_-(t) - \rho_+(t)] [1 - \exp(-\beta(t)|S_t - K|)], \quad (9.4d)$$

$$v(k, t) \rightarrow a + b[\rho(z - m) + \sqrt{(z - m)^2 + \sigma^2}], \quad (9.4e)$$

where $z = h(k)$ reflect a change of variable to control the sub-linearity of the wings. The main differences as compared to the classic Heston equation model is that the parameter ρ becomes stochastic and given by $\rho(t, S_t)$.

9.2 Bottom-Up Approach to Trading

9.2.1 Summary

We started this thesis by pointing to a puzzling observation from the newly born high frequency commodities market which, because of its extreme youth and therefore immaturity makes it a great case study for a high-frequency market at inception and therefore for our purpose. More specifically as we have seen with Figure 1.1, unusual patterned oscillations occurred in the commodities market. These oscillations cannot be explained by the TD assumption in Quantitative Finance. We have proposed in this thesis to study these oscillations with the BU approach instead staying in line with the recommendations of a modelling revolution [16] to occur post subprime crisis. The latter theory was developed in 3 Sections. We first expressed, in chapter 6, classic Financial Strategies in HFFF format and shown the incentive for going from a simple perceptron, to shallow and finally deep learning. We also established connections to fields that are traditionally associated to mathematical biology, in chapter 7, namely predator-prey models and evolutionary dynamics. This was done in order to expand the mathematical tools that we believe have value in 21st century Quantitative Finance. These helped us express the bottom-up approach at the infinitesimal level. More specifically we developed the concept of Path of Interaction in an HFTE Game and proposed 3 hypothesis as a mean to inspire future researchers. Finally, in chapter 8 we looked at how the financial market composition could be tracked through time with MTT.

9.2.2 Current & Future Research

Our first few simulations opened issues associated to optimality and need for more scientific rigor. We have classified these points in half a dozen issues listed below.

9.2.2.1 Classification Simplification

The direct initial simulation approach [4] was too challenging and the results perhaps too convoluted to filter out the essence of the research. For this reason we proposed to study the problem using fixed HFFFs, of Figures 6.7, 6.9 and 6.12. Though this simplifies the problem, it also means that there is human intervention in the strategy pool chosen. This latter intervention, though convenient is not ideal. Less human interventions should take place going forward.

9.2.2.2 The State Space can be improved

Following the previous point, choosing three types of strategies limits our state space which makes our tracking methodology easier but not as realistic as we wish. Additional strategies must be incorporated and more HFTE games must be included in our database of scenarios. This could be the work of many years. It could be partially addressed by creating an online database in which interested scientists could deposit their findings in object oriented format for simulation purposes. Note that this kind of collaborations are already in place in cellular automata [52].

9.2.2.3 Order-book Dynamics

Many of the markets are driven by different rules for the OB. We need to incorporate these different rules in our HFTE games as the latter rules obviously impact the outcome of the games.

9.2.2.4 Increased HFFF complexity does not equate to Invasion

It has been speculated that the need for a bigger brain in humans is partly due to the need for humans to elaborate deceitful strategies with their rivals and cooperative strategies with their allies. It is therefore not entirely ridiculous to associate increased neural network branching (to be roughly understood as increase in cranial size) with increased strategy complexity. However, increased intelligence does not necessarily equate to survival as we can see in the shark population, considered like an apex predator in the sea (but with a relatively small brain), has not evolved for millennial. We are very much at the early stages of defining NN complexity and dominance. A clear picture did not necessarily emerge from the first simulations though an interesting comparison can be made with Axelrod's computer tournament [29]. Indeed, Axelrod showed that it was not necessarily the most complicated strategies that prevailed at the end⁴. The TF strategy shares some aspects of the TFT strategy in the sense that they are both simple and adaptive. However, taking the argument in reverse ("complexity pays off" instead of "simple adaptable strategies are best"), can we think about a farming strategy? By this we mean: can we come up with a strategy that would understand the state of the ecosystem and would take actions based on that ecosystem, deliberately avoiding acquiring alpha on the ecosystem if it

⁴Please see TFT strategy in Section 7.2.1: it is a strategy that is simple in the way it adapts to the other strategies.

felt that it would be beneficial for the long term health of the financial ecosystem? These are questions that we may figure out sooner than expected.

9.2.2.5 Birth & Death Processes

We need to incorporate a Birth and Death Process to our MTT to study more realistic scenarios. To this end, we need to incorporate the OB in the likelihood function instead of using it only in the price dynamics. This will undoubtedly make the programming exercise more challenging but will, at the same time, bring more value to the research in the long run.

9.2.2.6 Complex Food Webs

We have seen in Section 7.6.1.1 that we have taken $l = 4$ in our Path of Interaction sequence. Would the Path of Interaction results change if we increase the sequence's length? In the context of the Path of Interaction study, is there a more rigorous way to connect some of the Lotka-Volterra predator prey models to these interactions? It seems intuitively more likely that the strategy ecosystem should rather be a complex food web. Can we enhance the idea of the simple Lotka-Volterra predator prey model to more complex food webs? More specifically what are the strategies that would create a stable and unstable food web? The concept of Path of Interaction is meant to be a bridge connecting the gap between strategy formalization to evolutionary dynamics but this bridge is not entirely specified yet.

9.2.2.7 Diversity & Stability

One other legitimate question that we can ask ourselves is whether the HFFF is a complex enough network to model most financial strategies? And if not all, does it encompass enough strategies to convey something interesting and meaningful when you make the strategies interact with each other. In this context our first chapter and paper [4] ended with a hypothesis, currently an open problem, that is interesting to mention in the context of future research:

Hypothesis: Diversity & the Financial Markets Diversity in financial strategies in the market leads to its instability.

Note that this hypothesis can be perhaps indirectly studied, or at least intuitively, using some of the finding in mathematical biology (MB). More specifically a hypothesis “diversity in ecosystem and the induced stability”⁵ exists in MB. However, we are still

⁵Though no consensus is reached there either.

potentially a long way before being able to answer this question though interesting take approaches have been taken at the human scale⁶ [244] as well as at the order size level [245].

9.2.2.8 Profitable & Unprofitable Ecosystem Asymmetry

The second hypothesis that we introduced is as follows:

Hypothesis: Profitable & Unprofitable Ecosystem Asymmetry All strategies in an ecosystem can make money at the same time but all cannot lose money at the same time.

As we mentioned earlier we noticed this interesting fact with our relatively small sample of HFTE games but have not been able to find a counter example yet nor have been able to rigorously prove it. It would be relatively easy to incorporate more simulations involving more strategies to see if we can find a counter example. Alternatively, if the hypothesis can be proven then we recommend searching proof around the pigeonhole.

9.2.2.9 Morality & HFTE Games

The last hypothesis we introduced is as follows:

Hypothesis: Morality & HFTE games Similar physical laws drive morality and HFTE games.

We explained that the TF strategy in an HFTE game and the TFT strategy in Axelrod's [28, 28] computer tournament described in Section 7.2.3 seems to have interesting similarities: they are both, simple, successful, cooperative but adaptive. It would be worth spending more time thinking about a relationship that is quite puzzling as the connections between the two applications Morality and Finance is very odd but the similarities are definitively there and they both emerge through the means of interacting agents.

9.2.2.10 Regularization & Regulation

In the spirit of some of the work we introduced with the Cointelation model in chapter 2.2, more specifically on proper market conduct, one important element of our current research is the one associated to socially responsible and consumer finance. With this

⁶To be understood in opposition to faster high frequency robot scale.

in mind we developed the concept of *unfortunate cost of pattern recognition: the genetic disorder of the financial industry (UTOPE-ia)* which we recently formally introduced [9]. This concept revolves around the development of the UTOPE [9] concept, which is a proposed methodology to build regulation taking as reference point the cost of pattern recognition instead of the benefits of the latter. More specifically we propose a regulation methodology that takes as a reference, not the problems with the market itself, but rather the issues associated with a human, analyzing these latter market's issues. Though, we take a more qualitative approach, we expose how the arrival of ML, and more specifically the understanding of the costs of pattern recognition, may force us to adopt a 180 degree change in the way we approach financial regulation. In fact, understanding over-fitting (e.g. Figure 9.1) can shed light to our own human limitations and more specifically how the latter can be arguably considered the virus of the financial industry.

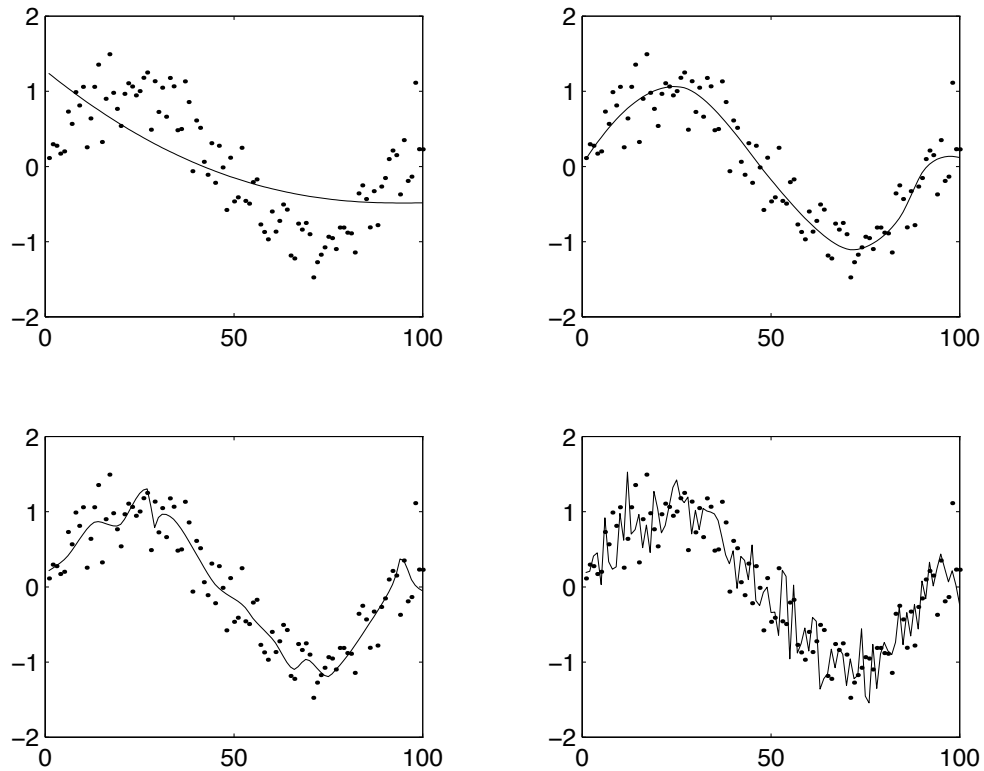


Figure 9.1: Under-fitting, just right and over-fitting examples [33].

9.3 Closing Words

“The more efficient you are at doing the wrong thing, the wronger you become. It is much better to do the right thing wronger than the wrong thing righter. If you do the right thing wrong and correct it, you get better.”

– Russell L. Ackoff

We thought that this quote from Russell Ackoff would be an inspiring closing statement as it relates to some of the content of this thesis. Indeed, we have chosen to take two approaches in the scientific method.

In Part II, we have taken a more conservative approach and tried to build on the last 100 years of Mathematical Finance. We did so by incorporating some of the new Machine Learning methods for data science applications. We believe that these methods have immediate engineering applications. Indeed, these man made systems are a function and a reflection of an imperfect sequence of events which have been handed to us by the previous generations and for which a practical solution is needed. The proposed solutions need to take into consideration constraints that are as much budgeting as they are social related.

In Part III, we have chosen instead to question the most fundamental assumption of Quantitative Finance: the market is stochastic. In doing so, we choose to take the bottom-up approach in the scientific method: simple deterministic rules at the microscopic⁷ level create the illusion of stochasticity at the macroscopic⁸ level. Though the approach we have taken pushes us to take a significant step back in our study of the market, we believe that it is in order to gain momentum for the future and in the best interest of our societies to do so. This will help turn Quantitative Finance into a field that follows a more traditional scientific method and will naturally solve many other problems.

To that extend we know that *all models are wrong, but some are useful*⁹ and in that spirit we have arguably done the right thing wronger in the second part of the thesis¹⁰ but the wrong thing righter in the first part¹¹.

⁷Systematic strategies interacting.

⁸The stock market.

⁹A quote that is generally attributed to George Box.

¹⁰Labelled Part III in the table of contents.

¹¹Labelled Part II in the table of contents.

Appendices

DGM NN Architecture

We can see the bird's-eye view of the DGM [23, 24] method in Figure 2 and its details in Figure 3. The details of this architecture can be found in [23, 24].

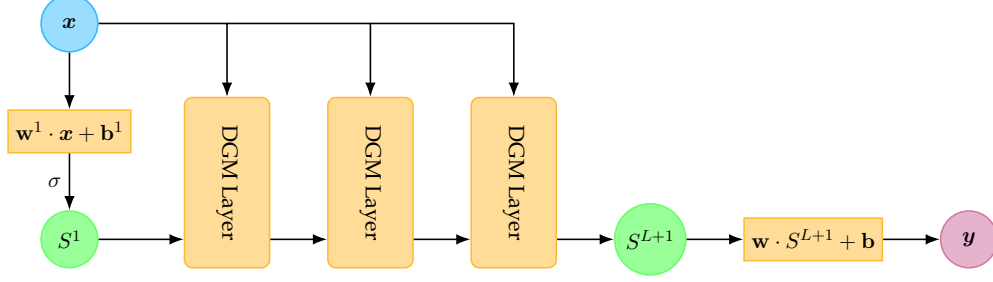


Figure 2: Bird's-eye view of our DGM [23]

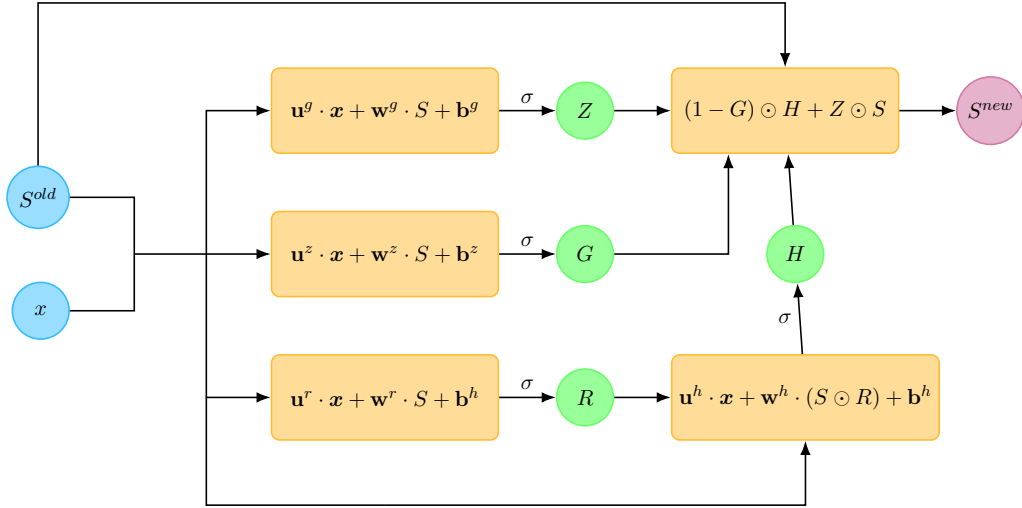


Figure 3: Detailed view of our DGM [23]

Proof of Lemma 2.4.1

$$E[r_p(t)] = w_S(t)E[r(S_t)] + w_{S_l}(t)E[r(Y_t)], \quad (5)$$

where $r(S_t) = \ln\left(\frac{S_t}{S_{t-\Delta t}}\right)$ and $r(S_{l,t}) = \ln\left(\frac{S_{l,t}}{S_{l,t-\Delta t}}\right)$ are daily log returns of assets S and S_l . Since S_t is a geometric Brownian motion, we have

$$E[r(S_t)] = (\mu - \frac{\sigma^2}{2})\Delta t - \ln(S_{t-\Delta t}) \quad (6)$$

where $S_{t-\Delta t}$ is a known constant at time $t - \Delta t$. The expectation of log return of asset S_l is

$$E[r(S_{l,t})] = E[\ln(S_{l,t})] - \ln(S_{l,t-\Delta t}), \quad (7)$$

where $S_{l,t-\Delta t}$ is a known constant at time $t - \Delta t$. We use Taylor expansion to approximate expected value and variance of logarithm of $S_{l,t}$ and covariance of logarithm $S_{l,t}$ and logarithm S_t (see [246]):

$$E[\ln(S_{l,t})] \approx \ln(E[S_{l,t}]) - \frac{\sigma^2[S_{l,t}]}{2E[S_{l,t}]^2}, \quad (8)$$

$$\sigma^2[\ln(S_{l,t})] \approx \frac{\sigma^2[S_{l,t}]}{E[S_{l,t}]^2}, \quad (9)$$

$$\sigma[\ln(S_{l,t}) \ln(S_t)] \approx \ln\left(1 + \frac{\sigma[S_t S_{l,t}]}{E[S_t]E[S_{l,t}]}\right). \quad (10)$$

Now, we need to derive $E[S_{l,t}]$. We have a set of stochastic differential equations from (2.5)

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t dW_t, \\ dS_{l,t} &= \theta[S_t - S_{l,t}]dt + \sigma_l S_{l,t} dZ_t, \\ d\langle W, Z \rangle_t &= \rho dt. \end{aligned} \quad (11)$$

We have

$$S_{l,t} = S_{l,0} + \theta \int_0^t (S_s - S_{l,s})ds + \sigma_l \int_0^t S_{l,s} dZ_s,$$

Taking expectation on both sides we have

$$E[S_{l,t}] = S_{l,0} + \theta \int_0^t E[S_s - S_{l,s}]ds.$$

Differentiating on both sides

$$\frac{dE[S_{l,t}]}{dt} = \theta S_0 e^{\mu t} - \theta E[S_{l,t}].$$

Denoting $E[S_{l,t}] = y(t)$ we have the ordinary differential equation (ODE):

$$y' = -\theta y + \theta S_0 e^{\mu t} \quad (12)$$

The solution is given by

$$y(t) = E[S_{l,t}] = a e^{\mu t} + (S_{l,0} - a) e^{-\theta t}, \quad (13)$$

where $a = \frac{\theta S_0}{\mu + \theta}$. In order to derive $E[S_{l,t}^2]$ we need to first compute $E[S_t S_{l,t}]$. Applying integration by parts (IBP) to (2.5) we have

$$\begin{aligned} d(S_t S_{l,t}) &= S_t dS_{l,t} + S_{l,t} dS_t + dS_t dS_{l,t} \\ &= \theta S_t^2 dt - \theta S_t S_{l,t} dt + \sigma_l S_t S_{l,t} dW_t + \mu S_t S_{l,t} dt + \sigma S_t S_{l,t} dZ_t + \sigma \sigma_l \rho S_t S_{l,t} dt. \end{aligned}$$

Thus we have:

$$S_t S_{l,t} = S_0 S_{l,0} + \theta \int_0^t S_s^2 ds + \sigma_l \int_0^t S_s S_{l,s} dW_s (\mu - \theta + \sigma \sigma_l \rho) \int_0^t S_s S_{l,s} ds + \sigma \int_0^t S_s S_{l,s} dZ_s.$$

Taking expectation and differentiating on both sides

$$\frac{dE[S_t S_{l,t}]}{dt} = \theta E[S_t^2] + (\mu - \theta + \sigma \sigma_l \rho) E[S_t S_{l,t}].$$

Denoting $E[S_t S_{l,t}] = x(t)$ we have ODE

$$x' = \theta E[S_t^2] + (\mu + \sigma \sigma_l \rho - \theta) x. \quad (14)$$

Since S_t is GBM, we have

$$E[S_t^2] = E[S_0^2 e^{(2\mu - \sigma^2)t + 2\sigma W_t}] = S_0^2 e^{(2\mu + \sigma^2)t}.$$

Thus (14) becomes

$$x' = \theta S_0^2 e^{(2\mu + \sigma^2)t} + (\mu - \theta + \sigma \sigma_l \rho) x.$$

Using variation of parameters method we get the solution

$$x(t) = E[S_t S_{l,t}] = b e^{(\mu - \theta + \sigma \sigma_l \rho)t} + (S_0 S_{l,0} - b) e^{(\mu - \theta + \sigma \sigma_l \rho)t},$$

where $b = \frac{\theta S_0^2}{\mu + \sigma^2 + \theta - \sigma \sigma_l \rho}$. Now we are ready to compute $E[S_{l,t}^2]$. By Itô's lemma the dynamics of $S_{l,t}^2$ is $dS_{l,t}^2 = 2S_{l,t} dS_{l,t} + (dS_{l,t})^2 = (\sigma_l^2 - 2\theta) S_{l,t}^2 dt + 2\theta S_t S_{l,t} dt + 2\sigma_l S_{l,t}^2 dZ_t$.

Integrating on both sides

$$S_{l,t}^2 = S_{l,0}^2 + (\sigma_l^2 - 2\theta) \int_0^t S_{l,s}^2 ds + 2\theta \int_0^t S_s S_{l,s} ds + 2\sigma_l \int_0^t S_{l,s}^2 dZ_s.$$

Taking expectation on both sides and differentiating

$$\frac{dE[S_{l,t}^2]}{dt} = (\sigma_l^2 - 2\theta) E[S_{l,t}^2] + 2\theta E[S_t S_{l,t}]$$

Defining $E[S_{l,t}^2] = z(t)$ and plugging in the value for $E[S_t S_{l,t}]$ from equation (15) we have ODE

$$z' = (\sigma_l^2 - \theta) z + 2\theta b e^{(\mu + \sigma^2)t} + 2\theta (S_0 S_{l,0} - b) e^{(\mu - \theta + \sigma \sigma_l \rho)t}.$$

Using variation of parameters we obtain the following solution

$$z(t) = E[S_{l,t}^2] = c e^{(\mu + \sigma^2)t} + d e^{(\mu - \theta + \sigma \sigma_l \rho)t} + (S_{l,0}^2 - c - d) e^{(\sigma_l^2 - 2\theta)t},$$

with $c = \frac{2\theta b}{2\mu + \sigma^2 - \sigma_l^2 + 2\theta}$ and $d = \frac{2\theta(S_0 S_{l,0} - b)}{\mu - \sigma_l^2 + \theta + \sigma \sigma_l \rho}$. Now we are ready to approximate $E[\ln(S_{l,t})]$. From (8) we have

$$\begin{aligned} E[\ln(S_{l,t})] &\approx \ln[E[S_{l,t}]] - \frac{E[S_{l,t}^2]}{2E[S_{l,t}]^2} + \frac{1}{2} \\ &= \ln(ae^{\mu t} - (S_{l,0} - a)e^{-\theta t}) + \frac{1}{2} - \frac{ce^{(2\mu + \sigma^2)t} + de^{(\mu - \theta + \sigma \sigma_l \rho)t}}{2(ae^{\mu t} - (S_{l,0} - a)e^{-\theta t})^2} \\ &\quad + \frac{(S_{l,0}^2 - c - d)e^{(\sigma_l^2 - 2\theta)t}}{2(ae^{\mu t} - (S_{l,0} - a)e^{-\theta t})^2} \end{aligned}$$

From (9) we have

$$Var[\ln(S_{l,t})] \approx \frac{E[S_{l,t}^2]}{E[S_{l,t}]^2} - 1 = \frac{ce^{(2\mu + \sigma^2)t} + de^{(\mu + \sigma \sigma_l \rho - \theta)t}}{(ae^{\mu t} - (S_{l,0} - a)e^{-\theta t})^2} + \frac{(S_{l,0}^2 - c - d)e^{(\sigma_l^2 - 2\theta)t}}{(ae^{\mu t} - (S_{l,0} - a)e^{-\theta t})^2} - 1.$$

And from (10) we obtain the covariance of between two log-asset prices

$$\begin{aligned} Cov[\ln(S_t) \ln(S_{l,t})] &\approx \ln\left(\frac{E[S_t S_{l,t}]}{E[S_t]E[S_{l,t}]}\right) \\ &\approx \ln\left(\frac{be^{(\mu + \sigma^2)t} + (S_0 S_{l,0} - b)e^{(\sigma \sigma_l \rho - \theta)t}}{aS_0 e^{\mu t} + (S_{l,0} S_0 - a S_0)e^{-\theta t}}\right). \end{aligned}$$

With $r_S = \ln\left(\frac{S_T}{S_0}\right)$ and $r_{S_l} = \ln\left(\frac{S_{l,T}}{S_{l,0}}\right)$ being the log returns of asset S and S_l over $[0, T]$, the expected return of portfolio over $[0, T]$ is

$$\begin{aligned} E[r_p(T)] &= h_1 E[r_S] + h_2 E[r_{S_l}] + h_3 E[r_B] \\ &= h_1 \left[\left(\mu - \frac{\sigma^2}{2} \right) T \right] + h_2 \left[\ln(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T}) - \right. \\ &\quad \left. \frac{ce^{(2\mu + \sigma^2)T} + de^{(\mu - \theta + \sigma \sigma_l \rho)T}}{2(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} + \frac{(S_{l,0}^2 - c - d)e^{(\sigma_l^2 - 2\theta)T}}{2(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} - \ln(S_{l,0}) + \frac{1}{2} \right] + h_3 rT. \end{aligned}$$

Now the variance of the portfolio return over $[0, T]$ is given by:

$$Var(r_p(T)) = h_1^2 \sigma_S^2 + h_2^2 \sigma_{S_l}^2 + 2h_1 h_2 \sigma_{SS_l}$$

$$\begin{aligned} \text{Expanding, we get } Var(r_p(T)) &= h_2^2 \sigma^2 T + \frac{de^{(\mu + \sigma \sigma_l \rho - \theta)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} + \frac{(S_{l,0}^2 - c - d)e^{(\sigma_l^2 - 2\theta)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} - \\ &1] + h_1^2 \left[\frac{ce^{(2\mu + \sigma^2)T}}{(ae^{\mu T} - (S_{l,0} - a)e^{-\theta T})^2} + 2h_1 h_2 \left(\ln\left(\frac{be^{(\mu + \sigma^2)T} + (S_0 S_{l,0} - b)e^{(\sigma \sigma_l \rho - \theta)T}}{aS_0 e^{\mu T} + (S_{l,0} S_0 - a S_0)e^{-\theta T}} \right) \right) \right]. \end{aligned}$$

Analytical solution for HJB equation Derivation

We are reducing a three dimension HJB equation below into a two dimensional Cauchy problem by using separation anzats. For ease of notation let $\tilde{\sigma} = \sigma^2 - 2\sigma\sigma_l\rho + \sigma_l^2$ and rewrite (2.51):

$$G_t + \sup_{h_1} \left\{ \frac{1}{2} (h_1^2 \tilde{\sigma} v^2 G_{vv} + \tilde{\sigma} z^2 G_{zz} + 2h_1 \tilde{\sigma} v z G_{vz}) + (h_1 [\mu - \theta(z-1)] + r) v G_v + (\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(z-1)) z G_z \right\} = 0. \quad (15)$$

The first order condition for the maximization is

$$h_1^* \tilde{\sigma} v G_{vv} + \tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v = 0 \quad (16)$$

Note that the constants 1/2 and 2 are dismissed because absorbed by the other parameters. Now assuming $G_{vv} < 0$ the first order condition is sufficient, yielding

$$h_1^* = - \frac{\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v}{\tilde{\sigma} v G_{vv}}. \quad (17)$$

Plugging (17) back into (15) yields:

$$\begin{aligned} G_t &+ \frac{1}{2} \left\{ \frac{(\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v)^2}{\tilde{\sigma}^2 v^2 G_{vv}^2} \tilde{\sigma} v^2 G_{vv} + \tilde{\sigma} z^2 G_{zz} \right. \\ &- 2 \frac{\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v}{\tilde{\sigma} v G_{vv}} \tilde{\sigma} v z G_{vz} \left. \right\} \\ &- \frac{\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v}{\tilde{\sigma} v G_{vv}} [\mu - \theta(z-1)] v G_v \\ &+ r v G_v + [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(z-1)] z G_z = 0. \end{aligned} \quad (18)$$

Multiplying both sides of equation by $\tilde{\sigma} G_{vv}$ we get:

$$\begin{aligned} &\tilde{\sigma} G_t G_{vv} + \frac{1}{2} (\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v)^2 - \\ &(\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v) \tilde{\sigma} z G_{vz} + \frac{1}{2} \tilde{\sigma} z^2 G_{zz} G_{vv} - \\ &(\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v) [\mu - \theta(z-1)] G_v + \\ &\tilde{\sigma} r v G_v G_{vv} + \tilde{\sigma} [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(z-1)] z G_z G_{vv} = 0. \end{aligned}$$

Expanding we have:

$$\begin{aligned} &\tilde{\sigma} G_t G_{vv} + \frac{1}{2} \tilde{\sigma}^2 z^2 G_{vz}^2 + \frac{1}{2} [\mu - \theta(z-1)]^2 G_v^2 + \\ &\tilde{\sigma} z [\mu - \theta(z-1)] G_v G_{vz} - \tilde{\sigma}^2 z^2 G_{vz}^2 - \\ &\tilde{\sigma} z [\mu - \theta(z-1)] G_v G_{vz} + \frac{1}{2} \tilde{\sigma} z^2 G_{zz} G_{vv} - \\ &\tilde{\sigma} z [\mu - \theta(z-1)] G_v G_{vz} - [\mu - \theta(z-1)]^2 G_v^2 + \\ &\tilde{\sigma} r v G_v G_{vv} + \tilde{\sigma} [\mu + \sigma_l^2 - \sigma\sigma_l\rho - \theta(z-1)] z G_z G_{vv} = 0 \end{aligned}$$

Simplifying we get:

$$\begin{aligned}
& \tilde{\sigma} G_t G_{vv} - \frac{1}{2} \tilde{\sigma}^2 z^2 G_{vz}^2 - \frac{1}{2} [\mu - \theta(z-1)]^2 G_v^2 + \\
& \frac{1}{2} \tilde{\sigma} z^2 G_{zz} G_{vv} - \tilde{\sigma} z [\mu - \theta(z-1)] G_v G_{vz} + \\
& \tilde{\sigma} r v G_v G_{vv} + \tilde{\sigma} [\mu + \sigma_l^2 - \sigma \sigma_l \rho - \theta(z-1)] z G_z G_{vz} = 0.
\end{aligned} \tag{19}$$

At this stage we were able to turn our four variable PDE into three but we can get rid of one more. To obtain the closed form solution, we consider the following separation ansatz:

$$G(t, v, z) = f(t, z) v^\gamma, \tag{20}$$

with the terminal condition

$$f(T, z) = 1 \quad \forall z. \tag{21}$$

We have the following derivatives of (20):

$$G_t = f_t v^\gamma, G_v = f v^{\gamma-1} \gamma, G_z = f_z v^\gamma, G_{vv} = f v^{\gamma-2} \gamma(\gamma-1), G_{vz} = f_z v^{\gamma-1} \gamma, G_{zz} = v^\gamma f_{zz}$$

Plugging them back into (20) and divide by $v^{2(\gamma-1)} \gamma$:

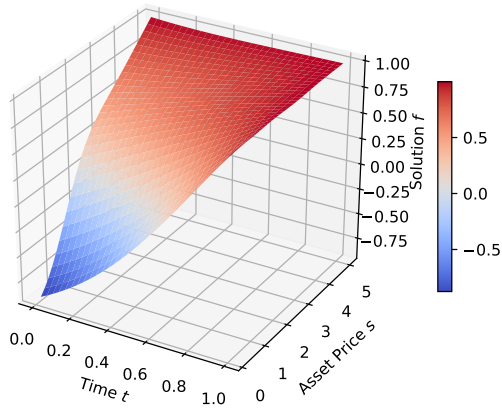
$$\begin{aligned}
& \tilde{\sigma} (\gamma-1) f f_t - \frac{1}{2} \tilde{\sigma}^2 \gamma z^2 f_z^2 - \frac{1}{2} \gamma [\mu - \theta(z-1)]^2 f + \frac{1}{2} \tilde{\sigma} (\gamma-1) z^2 f f_{zz} - \\
& \tilde{\sigma} \gamma [\mu - \theta(z-1)] z f f_z + \tilde{\sigma} \gamma (\gamma-1) r f^2 + \tilde{\sigma} (\gamma-1) [\mu + \sigma_l^2 - \sigma \sigma_l \rho - \theta(z-1)] f f_z = 0.
\end{aligned}$$

We now have our PDE with only two variables instead of four. The issue at this stage is that this PDE does not have a closed for solution. This is a non standard PDE, which is not high dimensional but is non linear which makes using finite difference methods or any standard numerical methods difficult. For this reason we propose to use the DGM to solve equation (22). Then optimal strategy is

$$\begin{aligned}
h_1^* &= - \frac{\tilde{\sigma} z G_{vz} + [\mu - \theta(z-1)] G_v}{\tilde{\sigma} v G_{vv}} \\
&= - \frac{\tilde{\sigma} z (f_z v^{\gamma-1} \gamma) - [\mu - \theta(z-1)] (f v^{\gamma-1} \gamma)}{\tilde{\sigma} v (f v^{\gamma-2} \gamma(\gamma-1))} \\
&= - \frac{\tilde{\sigma} z f_z - [\mu - \theta(z-1)] f}{\tilde{\sigma} f (\gamma-1)} \\
&= - \frac{z f_z}{(\gamma-1) f} - \frac{[\mu - \theta(z-1)]}{\tilde{\sigma} (\gamma-1)}.
\end{aligned} \tag{22}$$

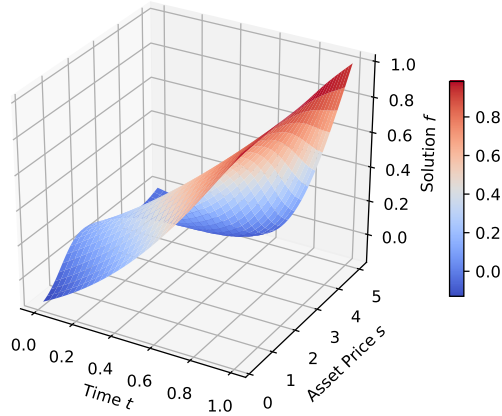
Note: to calculate the optimal control we do not divide by second order derivative, so the propagation of errors is avoided.

$$\sigma = \eta = 0.1, \kappa = 0.1, \rho = 0, \mu = 0.2, \gamma = 0.5$$



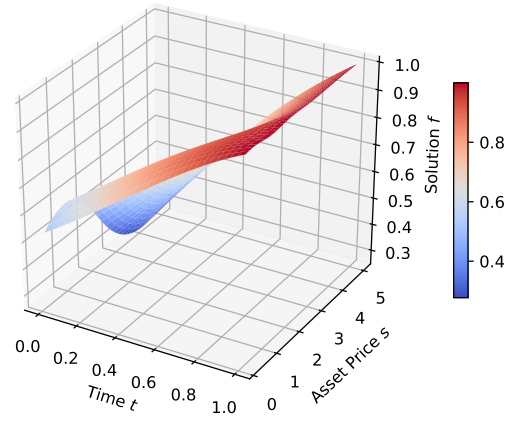
(a) Simulation: low σ with low θ

$$\sigma = \eta = 0.1, \kappa = 0.3, \rho = 0, \mu = 0.2, \gamma = 0.5$$



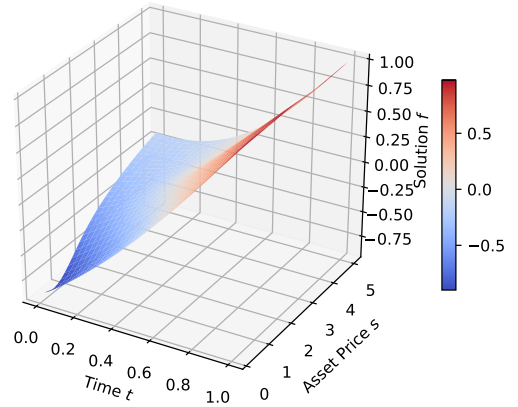
(c) Simulation: low σ with high θ

$$\sigma = \eta = 0.3, \kappa = 0.1, \rho = 0, \mu = 0.2, \gamma = 0.5$$



(b) Simulation: high σ with low θ

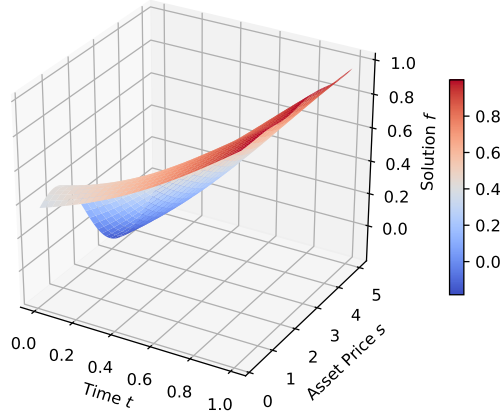
$$\sigma = \eta = 0.3, \kappa = 0.3, \rho = 0, \mu = 0.2, \gamma = 0.5$$



(d) Simulation: high σ with high θ

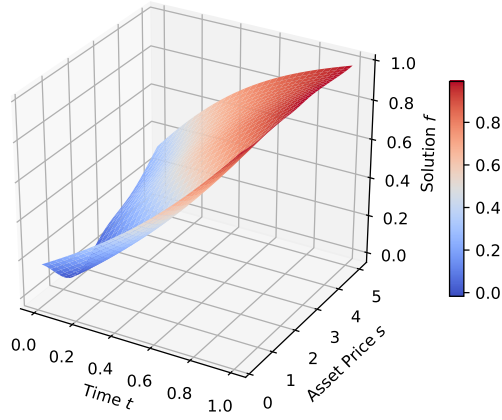
Figure 4: Four scenarios of σ and θ from high to low

$$\sigma = \eta = 0.2, \kappa = 0.2, \rho = -0.5, \mu = 0.0, \gamma = 0.5$$



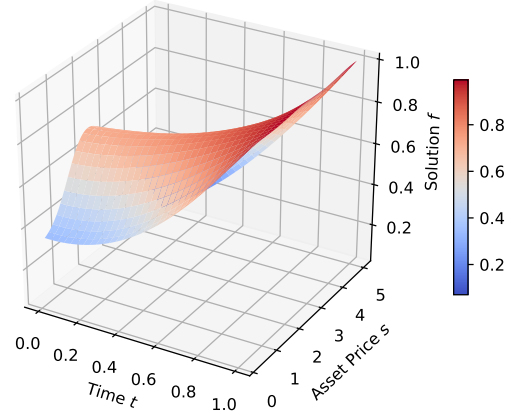
(a) Simulation: low μ with low ρ

$$\sigma = \eta = 0.2, \kappa = 0.2, \rho = -0.5, \mu = 0.4, \gamma = 0.5$$



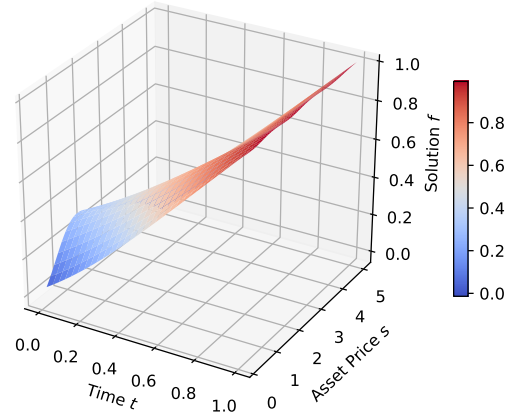
(c) Simulation: high μ with low ρ

$$\sigma = \eta = 0.2, \kappa = 0.2, \rho = 0.5, \mu = 0.0, \gamma = 0.5$$



(b) Simulation: low μ with high ρ

$$\sigma = \eta = 0.2, \kappa = 0.2, \rho = 0.5, \mu = 0.4, \gamma = 0.5$$



(d) Simulation: high μ with high ρ

Figure 5: Four scenarios of ρ and μ from high to low

Convergence of Heston to natural SVI

Heston implied variance

Forde et al. proved¹² that the implied volatility surface [247] behaves asymptotically like equation (23).

$$\sigma_\infty^2(x) = \begin{cases} 2 \left(2V^*(x) - x + \sqrt{V^*(x)^2 - kV^*(x)} \right), & \text{for } x \in \mathbb{R} \setminus \left[-\frac{1}{\hat{\theta}}, \frac{1}{\hat{\theta}} \right], \\ 2 \left(2V^*(x) - x - \sqrt{V^*(x)^2 - kV^*(x)} \right), & \text{for } x \in \left(-\frac{1}{\hat{\theta}}, \frac{1}{\hat{\theta}} \right). \end{cases} \quad (23)$$

with $\hat{\theta} := \kappa\theta/(\kappa - \rho\sigma)$, and the function $V^* : \mathbb{R} \mapsto \mathbb{R}_+$ is defined by

$$V^*(x) := p^*(x)x - V(p^*(x)), \quad \text{for all } x \in \mathbb{R}, \quad (24)$$

where

$$V(p) := \frac{\kappa\theta}{\sigma^2} (\kappa - \rho\sigma p - d(p)), \quad \text{for all } p \in (p_-, p_+), \quad (25a)$$

$$d(p) := \sqrt{(\theta - \rho\sigma p)^2 + \sigma^2 p(1 - p^2)}, \quad \text{for all } p \in (p_-, p_+), \quad (25b)$$

$$p^*(x) := \frac{\sigma - 2\kappa\rho + (\kappa\theta\rho + x\sigma)\sigma_l(x^2\sigma^2 + 2x\kappa\theta\rho\sigma + \kappa^2\theta^2)^{-\frac{1}{2}}}{2\sigma\hat{\rho}^2}, \quad \text{for all } x \in \mathbb{R}, \quad (25c)$$

$$\sigma_l := \sqrt{4k^2 + \sigma^2 - 4\kappa\rho\sigma}, \quad (25d)$$

$$p_\pm := \left(-2\kappa\rho + \sigma \pm \sqrt{\sigma^2 + 4\kappa^2 - 4\kappa\rho\sigma} \right) / (2\sigma\hat{\rho}^2), \quad (25e)$$

$$\hat{\rho} := \sqrt{1 - \rho^2} \quad (25f)$$

Remark Note that here the implied variance corresponds to European options with maturity T and maturity-dependent strike $K = S_0 \exp(xT)$.

Heston to natural SVI

The material associated with this part of the proof has been taken from [100].

Lemma .0.1 (Heston to natural SVI) *The natural SVI parametrization for implied variance is given by equation¹³ (26)*

$$\sigma_{SVI}^2(x) = \frac{w_1}{2} \left(1 + w_2\rho x + \sqrt{(w_2x + \rho)^2 + (1 - \rho^2)} \right), \quad \forall x \in \mathbb{R}, \quad (26)$$

¹²We refer here to Forde's original paper [247] for this long and convoluted proof. We will assume in that paper that the proof is correct.

¹³Note that here the Δ parameter in equation (4.5), which appears in natural SVI formula in [100] is set to be equal 0 in equation (26).

where x corresponds to a time-scaled log-moneyness. With the choice of SVI parameters in terms of the Heston parameters are the following:

$$w_1 := \frac{4\kappa\theta}{\sigma^2(1-\rho^2)} \left(\sqrt{(2\kappa - \rho\sigma)^2 + \sigma^2(1-\rho^2)} - (2\kappa - \rho\sigma) \right), \quad \text{and} \quad w_2 := \frac{\sigma}{\kappa\theta}. \quad (27)$$

and under Assumption $k - \rho\sigma > 0$ (27), $\sigma_{SVI}^2(x) = \sigma_\infty^2(x)$ for all $x \in \mathbb{R}$.

Proof Denote $\Delta(x) := \sqrt{\sigma^2 x^2 + 2\kappa\theta\rho\sigma x + \kappa^2\theta^2}$, where $\sigma_l = \sqrt{4k^2 + \sigma^2 - 4\kappa\rho\sigma}$ and $\hat{\rho}$ is defined by (25f). By plugging (27) into (26) the SVI implied variance formula is of the following form

$$\sigma_{SVI}^2(x) = \frac{2}{\sigma^2\hat{\rho}^2} (\sigma_l - (2\kappa - \rho\sigma)) (\kappa\theta + \rho\sigma x + \Delta(x)), \quad \text{for all } x \in \mathbb{R}. \quad (28)$$

In order to simplify the expression for σ_∞^2 in (23), first, the expression for $V^*(x)$ in 23 is rewritten in the following form:

$$V^*(x) = \frac{A(x)\Delta(x) + B(x)\sigma_l}{2\sigma^2\hat{\rho}^2\Delta(x)} \quad (29)$$

with

$$A(x) := x\sigma^2 - 2x\kappa\rho\sigma - 2\kappa^2\theta + \kappa\theta\rho\sigma, \quad \text{and} \quad B(x) := 2x\sigma\kappa\theta\rho + x^2\sigma^2 + \kappa^2\theta^2\rho^2 + \kappa^2\theta^2\hat{\rho}^2. \quad (30)$$

Since $B(x) = \Delta^2(x) \implies V^*(x) = (A(x) + \Delta(x)\sigma_l)/(2\sigma^2\hat{\rho}^2)$. And

$$2V^*(x) - x = \frac{A(x) + \Delta(x)\sigma_l - x\sigma^2\hat{\rho}^2}{\sigma^2\hat{\rho}^2} = \frac{\Delta(x)\sigma_l - (2\kappa - \rho\sigma)(\kappa\theta + x\rho\sigma)}{\sigma^2\hat{\rho}^2}, \quad (31)$$

where the factorization $A(x) - x\sigma^2\hat{\rho}^2 = -(2\kappa - \rho\sigma)(\kappa\theta + x\rho\sigma)$ is used.

Now in (23) denote $\Phi(x) := V^*(x)^2 - xV^*(x)$. We have

$$\Phi(x) = \left(\frac{\Delta(x)\sigma_l}{2\sigma^2\hat{\rho}^2} \right)^2 + \alpha(x)\Delta(x) + \beta(x), \quad (32)$$

where

$$\alpha(x) := -\frac{\sigma_l(2\kappa - \rho\sigma)(\kappa\theta + x\rho\sigma)}{2\sigma^4\hat{\rho}^4}, \quad \text{and} \quad \beta(x) := \frac{1}{4\sigma^4\hat{\rho}^4} ((2\kappa - \rho\sigma)^2(\kappa\theta + x\rho\sigma)^2 - x^2\sigma^4\hat{\rho}^4).$$

Using the following factorization:

$$\Delta^2(x) = (\kappa\theta + x\rho\sigma)^2 + x^2\sigma^2\hat{\rho}^2, \quad \text{and} \quad \sigma_l^2 = (2\kappa - \rho\sigma)^2 + \sigma^2\hat{\rho}^2 \quad (33)$$

we can write $\beta(x) = (4\sigma^4\hat{\rho}^4)^{-1} ((2\kappa - \rho\sigma)^2\Delta^2(x) - x^2\sigma^2\hat{\rho}^2\sigma_l^2)$ and hence

$$\begin{aligned}\Phi(x) &= \frac{1}{4\sigma^4\hat{\rho}^4} \left([(2\kappa - \rho\sigma)^2 + \sigma^2\hat{\rho}^2] \Delta^2(x) + \alpha(x)\Delta(x) + (\sigma_l^2 - \sigma^2\hat{\rho}^2)(\Delta^2(x) - x^2\sigma^2\hat{\rho}^2) - x^2\sigma^4\hat{\rho}^4 \right) \\ &= \frac{1}{4\sigma^4\hat{\rho}^4} \left((2\kappa - \rho\sigma)^2\Delta^2(x) + \alpha(x)\Delta(x) + \sigma_l^2(\kappa\theta + x\rho\sigma)^2 \right) \\ &= \frac{1}{4\sigma^4\hat{\rho}^4} (\sigma_l(\kappa\theta + x\rho\sigma) - (2\kappa - \rho\sigma)\Delta(x))^2,\end{aligned}\tag{34}$$

where $a(x) := 4\sigma^4\hat{\rho}^4\alpha(x)$. By taking the square root of $\Phi(x)$ we have

$$\begin{aligned}\sigma_l(\kappa\theta + x\rho\sigma) - (2\kappa - \rho\sigma)\Delta(x) &= (\kappa\theta + x\rho\sigma)\sqrt{(2\kappa - \rho\sigma)^2 + \sigma^2\hat{\rho}^2} - (2\kappa - \rho\sigma)\sqrt{(\kappa\theta + x\rho\sigma)^2 + x^2\sigma^2\hat{\rho}^2} \\ &= \sqrt{\gamma(x) + \sigma^2\hat{\rho}^2(\kappa\theta + x\rho\sigma)^2} - \sqrt{\gamma(x) + x^2\sigma^2\hat{\rho}^2(2\kappa - \rho\sigma)^2},\end{aligned}$$

where $\gamma(x) := (2\kappa - \rho\sigma)^2(\kappa\theta + x\rho\sigma)^2$. Now, because $\gamma(x) \geq 0$ for all $x \in \mathbb{R}$, then the sign of this expression is given by the difference $\psi(x) := \sigma^2\hat{\rho}^2(\kappa\theta + x\rho\sigma)^2 - x^2\sigma^2\hat{\rho}^2(2\kappa - \rho\sigma)^2$. Note further that we have $\psi(x) = \kappa\sigma^2\hat{\rho}^2(2x + \theta)(2x\rho\sigma + \kappa\theta - 2\kappa x)$, that this polynomial has exactly two real roots $-\theta/2$ and $\hat{\theta}/2$, and that its second-order coefficients reads $-4\kappa\sigma^2\hat{\rho}^2(\kappa - \rho\sigma) < 0$ under assumption $\kappa - \rho\sigma > 0$. So plugging (31) and (34) into (23), we obtain (28) and the proposition follows.

Remark In their original paper [247], Forde, Jacquier and Mijatovic show that

$$\hat{\sigma}_t^2(x) = \hat{\sigma}_\infty^2(x) + t^{-1} \frac{8\sigma_\infty^4(x)}{4x^2 - \hat{\sigma}_\infty(x)} \ln \left(\frac{A(x)}{A_{BS}(x, \hat{\sigma}_\infty(x), 0)} \right) + o(t^{-1}) \tag{35}$$

with $A(x)$ a complicated function which detailed specification can be found in the original paper [247]. Using Forde, Jacquier and Mijatovic [247] notation $\hat{\sigma}_t^2(x) = \hat{\sigma}_\infty^2(x) + \varepsilon(t, x)$, where $\varepsilon(t, x)$ represents the error term $t^{-1} \frac{8\sigma_\infty^4(x)}{4x^2 - \hat{\sigma}_\infty(x)} \times \ln \left(\frac{A(x)}{A_{BS}(x, \hat{\sigma}_\infty(x), 0)} \right) + o(t^{-1})$ then it is tempting to assume:

1. $\varepsilon(t, x)$ behaves like the downside transform family introduced by equation (4.1a).
2. the 2 in $\hat{\sigma}_t^2(x)$ can be replaced by $4p$ with $0 < p < \frac{1}{2}$ and assume the sub-linearity would magically come out through a tedious derivation.

However none of these 2 tempting guess would work because the moments are now all finite and there is no finite interval of study needed and the SVI and IVP equations are slightly different. You have to derive the problem from scratch.

Proof for right and left wing sub-linearity

The following two theorems are taken from [55].

We consider risk-neutral returns X with cumulative distribution function F , with $\bar{F} = 1 - F$ and f for the probability density function of X . The class of regularly varying functions at $+\infty$ of index α is denoted by \mathbb{R}_α .

Theorem .0.2 (Theorem (Right-Tail-Wing Formula)) Assume $\alpha > 0$ and

$$\exists \epsilon > 0 : \quad \mathbb{E} [e^{(1+\epsilon)X}] < \infty. \quad (36)$$

Then $(i) \rightarrow (ii) \rightarrow (iii) \rightarrow (iv)$,

where

$$-\ln f(k) \in \mathbb{R}_\alpha \quad (37a)$$

$$-\ln \bar{F}(k) \in \mathbb{R}_\alpha \quad (37b)$$

$$-\ln c(k) \in \mathbb{R}_\alpha \quad (37c)$$

and¹⁴

$$V(k)^2/k \sim \psi [-1 - \ln \bar{F}(k)/k]. \quad (38)$$

If (37b) holds then $-\ln c(k) \sim -k - \ln \bar{F}$ and

$$V(k)^2/k \sim \psi [-1 - \ln \bar{F}(k)/k], \quad (39)$$

if (37a) holds, then $-\ln f \sim -\ln \bar{F}$ and

$$V^2(k)/k \sim \psi [-1 - \ln f(k)/k]. \quad (40)$$

Finally, if either $-\ln f(k)/k$ or $-\ln \bar{F}(k)/k$ or $-\ln c(k)/k$ goes to infinity as $k \rightarrow \infty$ then $V^2(k)$ behaves sublinearly. More precisely,

$$V^2(k)/k \sim \frac{1}{-2 \ln f(k)/k} \quad \text{or} \quad \frac{1}{-2 \ln \bar{F}(k)/k} \quad (41)$$

Theorem .0.3 (Theorem (Left-Tail-Wing Formula)) Assume $\alpha > 0$ and

$$\exists \epsilon > 0 : \quad \mathbb{E} [e^{-\epsilon X}] < \infty. \quad (42)$$

Then $(i) \rightarrow (ii) \rightarrow (iii) \rightarrow (iv)$,

where

$$-\ln f(-k) \in \mathbb{R}_\alpha \quad (43a)$$

$$-\ln F(-k) \in \mathbb{R}_\alpha \quad (43b)$$

$$-\ln p(-k) \in \mathbb{R}_\alpha \quad (43c)$$

¹⁴ $g(k) \sim h(k)$ means $g(k)/h(k) \rightarrow 1$ as $k \rightarrow \infty$

and

$$V(-k)^2/k \sim \psi[-1 - \ln p(-k)/k]. \quad (44)$$

If (43b) holds then $-\ln p(k) \sim k - \ln F(-k)$ and

$$V(-k)^2/k \sim \psi[-\ln F(-k)/k], \quad (45)$$

if (43a) holds, then the $-\log f(-k) \sim -\ln F(-k)$ and

$$V(k)^2/k \sim \psi[-\ln f(-k)/k]. \quad (46)$$

Finally, if either $-\ln f(-k)/k$ or $-\ln F(-k)/k$ or $-\ln p(-k)/k$ goes to infinity as $k \rightarrow \infty$ then $V^2(-k)$ behaves sublinearly. More precisely,

$$V^2(-k)/k \sim \frac{1}{-2 \ln f(-k)/k} \quad \text{or} \quad \frac{1}{-2 \ln F(-k)/k} \quad \text{or} \quad \frac{1}{-2 \ln p(-k)/k}. \quad (47)$$

Acronyms (A-E)	
AC	Asset Class or Assumed Correlation
AFVS	Arbitrage Free Volatility Surface
ALLC	Always Cooperates
ALLD	Always Deceits
ANN	Artificial Neural Network
AOTSPST	Average Of The Same Parameters Surrounding Tenors
Aug	Augmenting
a.s	Almost Surely
BD	Big Data
BPS	Basis Point
BS	Black-Scholes
BU	Bottom-Up
BVD	Bias-Variance Dilemma
CDF	Cumulative Distribution Function
C&IC	Cointelation & Inferred Correlation
DAPD	Data Analysis and Patterns in Data
DGM	Deep Galerkin Method
DMUAA	Decision Making under Uncertainty and Asset Allocation
EKF	Extended Kalman Filter
EM	Expectation Maximisation
EMRPMA	Efficient Markets, Risk Premia and Market Anomalies
EP	Evolutionary Process

Acronyms (E-M)	
ES	Expected Shortfall or Economic sector
ET	Electronic Trading
FIFO	First In First Out
FISST	Finite Set Statistics
gSVI	generalised Stochastic Vol Inspired
GL	Geographical Location
GSF	Gaussian Sum Filter
GTFT	Generous Tit For Tat
HF	High Frequency
HFFF	High Frequency Financial Funnel
HFT	High Frequency Trading
HFTE	High Frequency Trading Ecosystem
HMM	Hidden Markov Model
I_i	Input “i”
IC	Inferred Correlation
i.i.d.	Identically and Independently Distributed
IPPF	Independent Partition Particle Filter
IS	Importance Sampling
IVP	Implied Volatility Parametrization
JMLS	Jump Markov Linear Systems
JPDA	Joint Probabilistic Data Association
JMS	Jump Markov Systems
KF	Kalman Filter
Lasso	Least Absolute Shrinkage & Selection Operator
LSTM	Long Short-Term Memory
LPSOTC	Liquidity Profile And Size Of The Company
MACD	Moving Average Convergence Divergence
MB	Mathematical Biology
MC-JPDAF	Monte Carlo Joint Probabilistic Data Association Filter
MCMC	Markov Chain Monte Carlo
MDITP	May Do If Time Permits
MG	Minority Game
MHT	Multiple Hypotheses Tracker
MLR	Multi-Linear Regression
MMSE	Minimum Mean Square Error
MSR	Multi-Sigmoid Regression
MSMC	Multi Sequential Monte Carlo
MSE	Mean Square Error
MTT	Multi Target Tracking
MVC	Mean Variance Criterion

Acronyms (N-Z)	
NICtV	Needs Internet Connection to View
NN	Neural Network
NNSF	Nearest Neighbour Standard Filter
O	Output
OB	Order-Book
OI	Open Interest
OMC	Order May Change
OMI	Oxford-Man Institute of Quantitative Finance
OU	Ornstein-Uhlenbeck
P&L	Profit and Loss
PDF	Probability Distribution Function
PDT	Product Diffusion Types
PF	Particle Filter
PHD	Probability Hypothesis Density
PT	Product Type
rand	Random
Rev	Revolutionizing
RIN	Review If Necessary
SABR	Stochastic Alpha Beta Rho
SFS	Stability of Financial Systems
SIR	Sampling Importance Resampling
SIVP	Simplified IVP
SMC	Sequential Monte Carlo
SMFM	Single Molecule Fluorescence Microscopy
SR	Stephen Roberts
SRCF	Socially Responsible and Consumer Finance
SSPF	Sequential Sampling Particle Filter
STEM	Science Technology Engineering Mathematics
STRAT	Strategy
SVI	Stochastic Volatility Inspired
TBD	track-before-detect
TD	Top-Down
TF	Trend Following
TFT	Tit For Tat
TSD	Tentative Submission Date
UKF	Unscented Kalman Filter
UTOPE	Unfortunate cosT Of Pattern rEcognition
VaR	Value at Risk
w.r.t.	with respect to
WSLS	Win-Stay, Lose-Shift

Bibliography

- [1] Babak Mahdavi-Damghani. A Bottom-up Approach to the Financial Markets Agent-Based Quantitative Algorithmic Strategies: Ecosystem, Dynamics & Detection. *Europe Risk Summit Conference*, pages 1–37, 2019.
- [2] Babak Mahdavi-Damghani. Guidelines for Building a Realistic Algorithmic Trading Market Simulator for Backtesting while incorporating Market Impact: Agent-Based Strategies in Neural Network Format, Ecosystem Dynamics & Detection. *Oxford-Man Institute of Quantitative*, pages 1–18, 2019.
- [3] Babak Mahdavi-Damghani and Stephen Roberts. Deciphering Price Formation in the High Frequency Domain: Systems & Evolutionary Dynamics As Keys for Construction of the High Frequency Trading Ecosystem. *Oxford-Man Institute of Quantitative Finance*, 2017.
- [4] Babak Mahdavi-Damghani. Introducing the HFTE Model: a Multi-Species Predator Prey Ecosystem for High Frequency Quantitative Financial Strategies. *Wilmott Magazine*, 89:52–69, 2017.
- [5] Babak Mahdavi-Damghani, Konul Mustafayeva, Stephen Roberts, and Cristin Buescu. Portfolio Optimization in the Context of Cointelated Pairs: Stochastic Differential Equation vs. Machine Learning Approach. *Mathematics & Finance: Research in Options Conference*, 2018.
- [6] Babak Mahdavi-Damghani and Stephen Roberts. Machine Learning Techniques for Deciphering Implied Volatility Surface Data in a Hostile Environment: Scenario Based Particle Filter, Risk Factor Decomposition & Arbitrage Constraint Sampling. *Oxford-Man Institute of Quantitative Finance*, 2018.
- [7] Babak Mahdavi-Damghani and Stephen Roberts. A Proposed Risk Modeling Shift from the Approach of Stochastic Differential Equation towards Machine Learning Clustering: Illustration with the Concepts of Anticipative & Responsible VaR. *Oxford-Man Institute of Quantitative Finance*, 2017.

- [8] Babak Mahdavi-Damghani, Konul Mustafayeva, Cristin Buescu, and Stephen Roberts. Convergence of Heston to SVI Proposed Extensions: Rational and Conjecture for the Convergence of Extended Heston to the Implied Volatility Surface Parametrization. *Oxford-Man Institute of Quantitative Finance*, 2017.
- [9] Babak Mahdavi-Damghani. UTOPE-ia. *Wilmott Magazine*, 60:28–37, 2012.
- [10] Babak Mahdavi-Damghani. The Non-Misleading Value of Inferred Correlation: An Introduction to the Cointelation Model. *Wilmott Magazine*, 67:50—61, 2013.
- [11] Babak Mahdavi-Damghani, D. Welch, C. O’Malley, and S. Knights. The Misleading Value of Measured Correlation. *Wilmott Magazine*, 62:64—73, 2012.
- [12] Babak Mahdavi-Damghani and Andrew Kos. De-arbitraging with a weak smile: Application to skew risk. *Wilmott Magazine*, pages 40–49, 2013.
- [13] Babak Mahdavi-Damghani. Introducing the Implied Volatility surface Parametrisation (IVP): Application to the FX market. *Wilmott Magazine*, 77:68–81, 2015.
- [14] Babak Mahdavi-Damghani. Machine learning methods for financial forecasting: application to the S&P500. 2006.
- [15] Timothy C. Johnson. Finance and Mathematics: Where is the Ethical Malaise? *The Mathematical Intelligencer*, 37(4):8–11, Dec 2015.
- [16] Jean-Philippe Bouchaud. Economics needs a scientific revolution. *Nature*, 455:1181, 2008.
- [17] J. Doyne Farmer and Duncan Foley. The economy needs agent-based modelling. *Nature*, 460:685–686, 2009.
- [18] Mark Buchanan. Meltdown modelling: Could agent-based computer models prevent another financial crisis? *Nature*, 460:680–682, 2009.
- [19] Kenneth Cukier and Viktor Mayer-Schoenberger. The Rise of Big Data: How It’s Changing the Way We Think About the World. *Foreign Affairs*, 92(3):28–40, 2013.
- [20] Omar Al-Jarrah, Paul Yoo, Sami Muhaidat, George Karagiannidis, and Kamal Taha. Efficient machine learning for big data: A review. 2:87–93, 09 2015.

- [21] Nanex. Strange days june 8'th, 2011 - natgas algo. 2011.
- [22] Paul Wilmott. Frequently asked questions in Quantitative Finance. *Wiley*, pages 336–337, 2009.
- [23] Ali Al-Aradi, Adolfo Correia, Danilo Naiiff, and Gabriel Jardim. Solving Non-linear and High-Dimensional Partial Differential Equations via Deep Learning. 2018.
- [24] Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339 – 1364, 2018.
- [25] Axel Vogt. Counter example published on Wilmott's internet forum. *Wilmott*, 2001.
- [26] Bloomberg. 1 year expiry S&P 500 index options. 2016.
- [27] Peter Youngman. Procyclicality and Value at Risk. *Bank of Canada Financial System Review*, 2009. Internet Link.
- [28] Robert Axelrod. The Evolution of Cooperation. *Basic Books*, 1984.
- [29] Robert Axelrod. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration. *Princeton University Press*, 1997.
- [30] Martin Nowak. *Evolutionary dynamics: exploring the equations of life*. 2006.
- [31] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, 364:56–58, 1993.
- [32] Esteban Moro. The Minority Game: an introductory guide, 2004.
- [33] Stephen Roberts. Machine Learning & Pattern Recognition. 2008. Course notes.
- [34] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, and Yuanzhuo Wang. Significance and challenges of big data research. *Big Data Res.*, 2(2):59–64, June 2015.
- [35] Jae-Gil Lee and Kang Minseo. Geospatial big data: Challenges and opportunities. 2, 02 2015.

- [36] Shaokun Fan, Raymond Y.K. Lau, and J. Leon Zhao. Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Res.*, 2(1):28–32, March 2015.
- [37] Louis Bachelier. Theorie de la speculation. *Annales Scientifiques de l'École Normale Supérieure*, 3 (17):21–86, 1900.
- [38] Hamed Amini, Rama Cont, and Andrea Minca. Resilience to Contagion in Financial Networks. *Mathematical Finance*, 26:329–365, 2016.
- [39] Nassim Nicholas Taleb. *The Black Swan*. 2007.
- [40] Iacopo Mastromatteo, Michael Benzaquen, Zoltan Eisler, and Jean-Philippe Bouchaud. 02 2017.
- [41] Clara Leonard. Après la crise, l'enseignement de la finance repensé. Sep 2013.
- [42] Stephen Roberts and Martin Tegner. Sequential Sampling of Gaussian Latent Variable Models. *University of Oxford*, 2018.
- [43] Babak Mahdavi-Damghani and Stephen Roberts. A Proposed Risk Modeling Shift from the Approach of Stochastic Differential Equation towards Machine Learning Clustering: Illustration with the concepts of Anticipative & Responsible VaR. *Working paper*, <http://dx.doi.org/10.2139/ssrn.3039179>.
- [44] Basel Committee on Banking Supervision. Basel 3: A global regulatory framework for more resilient banks and banking systems. *Bank of International Settlements*, page 11, 2011.
- [45] Dan Immergluck. From the subprime to the exotic: Excessive mortgage market risk and foreclosures. *Journal of the American Planning Association*, 74(1):59–76, 2008.
- [46] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- [47] A Hall. A suggestion in the theory of Mercury. *Astronomical Journal*, 14:49–51, 1894.
- [48] Richard Dawkins. *The Selfish Gene*. 1976.
- [49] Bruno Dupire. Pricing with a smile. *Risk*, pages 17–20, 1994.

- [50] Bruno Dupire. Pricing and hedging with a smile. *Cambridge University Press*, 1997.
- [51] Jim Gatheral. The volatility surface. *Wiley Finance*, page 37, 2006.
- [52] Martin Gardner. Mathematical Games. *Scientific American*, 1970.
- [53] John Horton Conway. *On Numbers and Games*. 1976.
- [54] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. 2014.
- [55] Shalom Benaim and Peter Friz. Regular Variation and Smile Asymptotics. *Mathematical Finance*, 19:1–12, 2009.
- [56] Shalom Benaim, Peter Friz, and Roger Lee. *On Black-Scholes Implied Volatility at Extreme Strikes*, chapter 2, pages 19–45. John Wiley & Sons, Ltd, 2012.
- [57] Chris Rogers and Mike Tehranchi. The implied volatility surface does not move by parallel shifts. *University of Cambridge*, pages 7–10, 2009.
- [58] Roger Lee. The Moment Formula for Implied Volatility at Extreme Strikes. *Mathematical Finance*, 14:469–480, 2004.
- [59] Jim Gatheral and Antoine Jacquier. Convergence of Heston to SVI. *Quantitative Finance*, 11:1129–1132, 2011.
- [60] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.
- [61] Steven L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6:327–343, 1993.
- [62] K. Pearson. *Royal Society Proceedings*, 58:241, 1895.
- [63] K. Pearson. Notes on the History of Correlation. *Biometrika*, 13:25–45, 1920.
- [64] R. F. Engle and C.W.J. Granger. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica, Econometric Society*, 55:251–276, 1987.

- [65] Álvaro Cartea and Sebastian Jaimungal. Algorithmic Trading of Co-integrated Assets. *International Journal of Theoretical and Applied Finance*, 19, 07 2016.
- [66] Álvaro Cartea, Luhui Gan, and Sebastian Jaimungal. Trading Cointegrated Assets with Price Impact, 07 2018.
- [67] A.H. Tewfik, M. Kim, and M. Deriche. 19 Multiscale signal processing techniques: A review. In *Signal Processing and its Applications*, volume 10 of *Handbook of Statistics*, pages 819 – 881. 1993.
- [68] A. Benveniste, M. Basseville, A.S. Willsky, K.C. Chou, and R. Nikoukhah. Multiscale Statistical Signal Processing and Random Fields on Homogeneous Trees. *IFAC Proceedings Volumes*, 25(15):405 – 410, 1992. 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation 1991 , Budapest, Hungary, 8-12 July 1991.
- [69] Michèle Basseville and Albert Benveniste. Multiscale statistical signal processing. volume 4, pages 2065 – 2068 vol.4, 06 1989.
- [70] Jean-Philippe Bouchaud. Feedback during this thesis’ viva at the University of Oxford’s Mathematical Institute (witness Pr. Álvaro Cartea), September 2019.
- [71] Iain Murray. Lecture notes in Gaussian Processes and Kernels, 2016.
- [72] Ivan Zezula. On multivariate Gaussian copulas, 2007.
- [73] Dag Tjøstheim and Karl Ove Hufthammer. Inferring agent objectives at different scales of a complex adaptive system. *Journal of Econometrics*, 172:33–48, 2013.
- [74] J.C. Rodrigues. Measuring financial contagion. a copula approach. *Journal of Empirical Finance*, 14:401–423, 2007.
- [75] A.C. Inci, H.C. Li, and J. McCarthy. Financial contagion: a local correlation analysis. *Research in International Business and Finance*, 25:11–25, 2011.
- [76] R. Campbell, C.S. Forbes, K. Koedijk, and P. Kofman. Increasing correlations or just fat tails? *Journal of Empirical Finance*, 15:287–309, 2008.
- [77] Y. Hong, J. Tu, and G. Zhou. Asymmetries in stock returns: statistical tests and economic evaluation. *Review of Financial Studies*, 20:1547–1581, 2007.

- [78] FSA Handbook. 2008. Available on line.
- [79] H. Markowitz. Portfolio Selection. *The Journal of Finance*, 7:77–91, 1952.
- [80] S. Mudchanatongsuk, J. Primbs, and W. Wong. Optimal Pairs Trading: A Stochastic Control Approach. In *American Control Conference*, Seattle, Washington, USA, 6 2008.
- [81] J. Tobin. Liquidity preference as behavior towards risk. The review of economic studies. *The Review of Economic Studies*, 25:65—86, 1958.
- [82] W. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19:425–442, 1964.
- [83] J. Lintner. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47:13–37, 1965.
- [84] G.E. Uhlenbeck and L.S. Ornstein. On the theory of Brownian Motion. *Phys.Rev*, 36:23–41, 1930.
- [85] R. Bellman. *Dynamic programming*. Princeton, 1957.
- [86] Y. Kom Samo and A. Vervuurt. Stochastic Portfolio Theory: A Machine Learning Perspective. *Forthcoming in Quantitative Finance*.
- [87] R. Fernholz. *Stochastic Portfolio Theory*. New York, 2002.
- [88] E. Fama and F. Kenneth. The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives-Volume*, 18:25–46, 2004.
- [89] R. Fernholz, I. Karatzas, and C. Kardaras. Diversity and relative arbitrage in equity markets. *Finance and Stochastics*, 9:1–27, 2005.
- [90] A. Vervuurt and I. Karatzas. Diversity-weighted portfolios with negative parameter. *Annals of Finance*, 11:411–432, 2015.
- [91] M. Lo  ve. *Probability Theory*, volume 45. 1997.
- [92] E. Soeryana, N. Fadhlina, Sukono, E. Rusyaman, and S. Supian. Mean-variance portfolio optimization by using time series approaches based on logarithmic utility function. *IOP Conference Series: Materials Science and Engineering*, 166:012003, jan 2017.

- [93] R. Korn and H. Kraft. A Stochastic Control Approach to Portfolio Problems with Stochastic Interest Rates. *SIAM Journal on Control and Optimization*, 4:1250–1269, 2002.
- [94] Álvaro Cartea. Feedback during this thesis’ viva at the University of Oxford’s Mathematical Institute (witness Pr. Jean-Philippe Bouchaud), September 2019.
- [95] Harold Bierman. The valuation of stock options. *The Journal of Financial and Quantitative Analysis*, 2(3):327–334, 1967.
- [96] Álvaro Cartea and Sebastian Jaimungal. Irreversible Investments and Ambiguity Aversion. *International Journal of Theoretical and Applied Finance*, 20, 09 2017.
- [97] Ki Hosam, Byungwook Choi, Kook-Hyun Chang, and Miyoung Lee. Option pricing under extended normal distribution. *The Journal of Futures Market*, 25:845–871, 2005.
- [98] Mark B. Garman and Steven W. Kohlhagen. Foreign currency option values. *Journal of International Money and Finance*, 2(3):231–237, 1983.
- [99] Peter Tankov and Nizar Touzi. Calcul Stochastique en Finance. *École Polytechnique Paris, Département de Mathématiques Appliquées*, page 146, 2010.
- [100] Jim Gatheral and Antoine Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.
- [101] A.D. Fokker. Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. *Ann. Phys.*, pages 810–820, 1914.
- [102] Max Planck. Zur Theorie des Rotationsspektrums. *Annalen der Physik*, 358(11):241–256, 1917.
- [103] Nicole El Karoui. Couverture des risques dans les marchés financiers. *Course at the École Polytechnique in Paris*, page 92, 2003.
- [104] Álvaro Cartea, Sebastian Jaimungal, and Jamie Walton. Foreign Exchange Markets with Last Look. *Mathematics and Financial Economics*, 06 2018.
- [105] Hadrien De March and Pierre Henry-Labordere. Building arbitrage-free implied volatility: Sinkhorn’s algorithm and variants. 02 2019.

- [106] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: numerical solution of the inverse problem. *Journal of Computational Finance*, 1997.
- [107] S. Crepey. Calibration of the local volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, pages 91–127, 2003.
- [108] J. N. Dewynne, S. D. Howison, P. Wilmott, and Philipp J. Schonbucher. A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1758):2071–2092, 1999.
- [109] J. Cox. Notes on Option Pricing I: Constant Elasticity of Diffusions. *Unpublished draft, Stanford University*, 1975.
- [110] Patrick S. Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward. Managing smile risk. *Wilmott*, pages 84–108, 2002.
- [111] Alexandre Antonov, Michael Konikov, and Michael Spector. The free boundary sabr: Natural extension to negative rates. *Numerix*, 2015.
- [112] Chikio Hayashi. What is Data Science? Fundamental Concepts and a Heuristic Example. In *Data Science, Classification, and Related Methods*, pages 40–51, Tokyo, 1998.
- [113] Vasant Dhar. Data Science and Prediction.
- [114] Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.
- [115] Paul A. Samuelson. Mathematics of speculative price. *SIAM Review*, 15(1):1–42, 1973.
- [116] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–394, 1963.
- [117] R. R. Hocking. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1):1–49, 1976.
- [118] Lynda Flom and David L. Cassell. Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. 01 2007.

- [119] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. Gaussian Processes for Timeseries Modelling. *Philosophical Transactions of the Royal Society (Part A). Phil. Trans. R. Soc.*, 371:1–27, 2013.
- [120] G.E. Uhlenbeck and L.S. Ornstein. On the theory of Brownian Motion. *Phys.Rev.*, 36:823–841, 1930.
- [121] Howard Tucker. A Generalization of the Glivenko-Cantelli Theorem. *The Annals of Mathematical Statistics*, page 828–830, 1959.
- [122] A. W. Van der Vaart. Asymptotic Statistics. *Cambridge University Press*, page 265, 1998.
- [123] Meera Sharma. What is wrong with quantitative standard d.?
- [124] A Schlogl, Stephen Roberts, and Gert Pfurtscheller. A criterion for adaptive autoregressive models. 2:1581–1582, 02 2000.
- [125] Matt Burgess. Google’s deepmind wins historic go contest 4–1. *Wired*, 2016.
- [126] Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. Enhancing trading strategies with order book signals. *Applied Mathematical Finance*, 25:1–35, 02 2018.
- [127] Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. 2 edition, 2003.
- [128] Warren McCulloch and Walter Pitts. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [129] Frank Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. 1962.
- [130] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. 2nd edition with corrections, first edition 1969 edition, 1972.
- [131] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

- [132] Andrey Kolmogorov. On the representation of continuous functions of several variables as superpositions of functions of smaller number of variables. *Soviet Math. Dokl*, 4:2065–2068, 1956.
- [133] Andrey Arnold. On functions of three variables. *Proceedings of the USSR Academy of Sciences*, 114:679–681, 1957.
- [134] Jürgen Braun and Michael Griebel. On a Constructive Proof of Kolmogorov’s Superposition Theorem. *Constructive Approximation*, 30(3):653, May 2009.
- [135] G. Lorentz. *Approximation of functions*. Athena series: Selected topics in mathematics. 1966.
- [136] G. Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive approximation : advanced problems*. 1996. Includes bibliographical references (p. [621-640] and indexes.
- [137] David Sprecher. On the Structure of Continuous Functions of Several Variables. *Transactions of The American Mathematical Society*, 115:340–355, 03 1965.
- [138] David Sprecher. An improvement in the superposition theorem of Kolmogorov. *Journal of Mathematical Analysis and Applications*, 38:208 – 213, 1972.
- [139] Robert Hecht-Nielsen. Counterpropagation networks. *Appl. Opt.*, 26(23):4979–4984, Dec 1987.
- [140] Robert Hecht-Nielsen. Kolmogorov’s mapping neural network existence theorem. *Proceedings of the International Conference on Neural Networks III*, pages 11–14, 1987.
- [141] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Algorithmic Trading, Stochastic Control, and Mutually Exciting Processes. *SIAM Review*, 60:673–703, 01 2018.
- [142] Anderson Mark Robert. Twenty years on from deep blue vs kasparov: how a chess match started the big data revolution. *The Conversation*, 2017.
- [143] M. R. Vargas, B. S. L. P. de Lima, and A. G. Evsukoff. Deep learning for stock market prediction from financial news articles. In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 60–65, June 2017.

- [144] Justin Sirignano and Rama Cont. Universal features of price formation in financial markets: perspectives from Deep Learning. *Quantitative Finance*, 2018.
- [145] Cedric Villani. la simulation numérique est un enjeu majeur pour la société. *Atos France Interview*, 2012.
- [146] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.
- [147] David D. L. Minh and Do Le (Paul) Minh. Understanding the Hastings algorithm. *Communications in Statistics - Simulation and Computation*, 44(2):332–349, 2015.
- [148] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.
- [149] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [150] M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–549, 1987.
- [151] David MacKay. Information theory, inference, and learning algorithms. *Cambridge University Press*, 2003.
- [152] Radford M. Neal. Suppressing random walks in Markov Chain Monte Carlo using ordered overrelaxation. *University of Toronto, Department of Statistics*, 1995.
- [153] Jun S. Liu, Faming Liang, and Wing Hung Wong. *Journal of the American Statistical Association*, (449):121–134.
- [154] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.
- [155] S. J. Roberts, C. Holmes, and D. Denison. Minimum-entropy data partitioning using reversible jump Markov chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):909–914, Aug 2001.

- [156] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority Games: Interacting Agents In Financial Markets*. 2014.
- [157] A.C.C. Coolen. *The Mathematical Theory of Minority Games: Statistical Mechanics of Interacting Agents (Oxford Finance Series)*. New York, NY, USA, 2005.
- [158] Iacopo Mastromatteo, Bence Toth, and Jean-Philippe Bouchaud. Agent-based models for latent liquidity and concave price impact. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 89:042805, 04 2014.
- [159] Inter-pattern speculation: Beyond minority, majority and \$-games. *Journal of Economic Dynamics and Control*, 32(1).
- [160] J. Vitting Andersen and D. Sornette. The \$-game. *The European Physical Journal B - Condensed Matter and Complex Systems*, 31(1):141–145, Jan 2003.
- [161] I. Giardina and J.-P. Bouchaud. Bubbles, crashes and intermittency in agent based market models. *The European Physical Journal B - Condensed Matter and Complex Systems*, 31(3):421–437, Feb 2003.
- [162] Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. 1776.
- [163] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, pages 459–467, 1976.
- [164] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, 2011.
- [165] C. S. Elton. *The Ecology of Invasions by Animals and Plants*. Methuen, 1958.
- [166] G. E. Hutchinson. Homage to Santa Rosalia, or why are there so many kinds of animals? *Nature*, pages 145–159, 1959.
- [167] Kevin Shear McCann. The diversity-stability debate. *Nature*, pages 228–233, 2000.
- [168] R. H. MacArthur. Fluctuations of animal populations and a measure of community stability. *Ecology*, 36:533–536, 1955.
- [169] C.S. Elton. *Ecology of Invasions by Animals and Plants*. 1958.

- [170] P. Yodzis. The stability of real ecosystems. *Nature*, 289:674–676, 1981.
- [171] S.L. Pimm and J. H. Lawton. On feeding on more than one trophic level. *Nature*, 275:542–544, 1978.
- [172] Vito Volterra. Variazioni e fluttuazioni del numero d’individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, pages 31–113, 1926.
- [173] Alfred J. Lotka. Elements of Physical Biology. *Williams & Wilkins Co.*, 1925.
- [174] Erica Chauvet, Joseph E. Paultet, Joseph P. Previte, and Zac Walls. A Lotka-Volterra Three-species Food Chain. *Mathematics Magazine*, 75:243–255, 2002.
- [175] Alexandr Mikhailovich Liapunov. The General Problem of the Stability of Motion. *Doctoral dissertation, University of Kharkov*, 1892.
- [176] Richard Bellman. Dynamic programming and a new formalism in the calculus of variations. *Proceedings of the National Academy of Sciences of the United States of America*, 40(4):231–235, 1954.
- [177] Ricky Cooper, Michael Davis, and Ben Van Vliet. The Mysterious Ethics of High-Frequency Trading. *Business Ethics Quarterly*, 26(1):1–22, 2016.
- [178] Steven McNamara. The Law and Ethics of High-Frequency Trading. *SSRN Electronic Journal*, 17, 01 2015.
- [179] Giulio Biroli, Guy Bunin, and Chiara Cammarota. Marginally stable equilibria in critical ecosystems. *New Journal of Physics*, 20(8):1–12, aug 2018.
- [180] Y. Bar-Shalom and T. E. Fortmann. Tracking and Data Association. *Academic Press*, 1988.
- [181] X.R. Li, Y. Bar-Shalom, and T. Kirubarajan. Estimation, Tracking and Navigation: Theory, Algorithms and Software. *New York: John Wiley and Sons*, June, 2001.
- [182] L.D. Stone, C.A. Barlow, and T.L. Corwin. Bayesian Multiple Target Tracking.
- [183] S. Roweis and Z. Ghahramani. A unifying review of linear Gaussian models. *Neural Comput*, 11(2):305–345, 1999.
- [184] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35–45, 1960.

- [185] R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Transactions of the ASME - Journal of Basic Engineering*, 83:95–107, 1961.
- [186] A. Marshall. The use of multi-stage sampling schemes in Monte Carlo computations. in *Symposium on Monte Carlo Methods*, M. Meyer Ed. New York: Wiley, pages 123–140, 1956.
- [187] J.E. Handschin and D.Q. Mayne. Monte Carlo techniques to estimate conditional expectation in multi-state non-linear filtering. *Int. J. Contr.*, 9(5):547–559, 1969.
- [188] J.A. Bucklew. Large Deviation Techniques in Decision, Simulations, and Estimation. Wiley, 1990.
- [189] Genshiro Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Non-linear State Space Models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.
- [190] A. Doucet, N. de Freitas, and N. Gordon. Sequential Monte Carlo in Practice. New York: Springer-Verlag, 2001.
- [191] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(9):197–208, 2000.
- [192] J. Liu and R. Chen. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*.
- [193] N. Gordon, D. Salmond, and A. Smith. Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, 1993.
- [194] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.
- [195] S. Haykin. Adaptive Filter Theory. Englewood Cliffs, NJ:Prentice Hall, 4, 2000.
- [196] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.

- [197] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.
- [198] R. P. S. Mahler. Multitarget Bayes filtering via firstorder multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4), 2003.
- [199] William Ng, Jack Li, Simon Godsill, and Jaco Vermaak. A Hybrid Approach For Online Joint Detection And Tracking For Multiple Targets. *IEEEAC paper, Department of Engineering, University of Cambridge*, 1, 2004.
- [200] J. Larsen and I. Pedersen. Experiments with the auction algorithm for the shortest path problem. *Nordic Journal of Computing*, 6(4):403–421, 1998.
- [201] D.P. Bertsekas. Auction algorithms for network flow problems: A tutorial introduction. *Computational Optimization and Applications*, 1:7–66, 1992.
- [202] R.D. Leone and D. Pretolani. Auction algorithms for shortest hyperpath problems. *Society for Industrial and Applied Mathematics*, 11(1):149–159, 2001.
- [203] S. Blackman and R. Popoli. Design and Analysis of Modern Tracking Systems. *Norwood, MA: Artech House*, 1999.
- [204] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.
- [205] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.
- [206] W. Ng, J. Li, S. Godsill, and J. Vermaak. A hybrid approach for online joint detection and tracking for multiple targets. *University of Cambridge, Department of Engineering*, 2005.
- [207] William Ng, Jack Li, Simon Godsill, and Jaco Vermaak. Multiple target tracking using a new soft-gating approach and sequential Monte Carlo methods. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 4:1049–1052, 2005.
- [208] Jack Li, William Ng, Simon Godsill, and Jaco Vermaak. Online Multitarget Detection and Tracking Using Sequential Monte Carlo Methods. *Department of Engineering University of Cambridge*.

- [209] R.P.S. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4), 2003.
- [210] T. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8:173–184, 1983.
- [211] R.J. Fitzgerald. Track biases and coalescence with probabilistic data association. *IEEE Tr. Aerospace*, 21:822–825, 1985.
- [212] H.A.P. Blom and E.A. Bloem. Probabilistic data association avoiding track coalescence. *IEEE Tr. Automatic Control*, 45:247–259, 2000.
- [213] A.P. Henk, H.A.P. Blom, and E.A. Bloem. Joint IMM and coupled PDA to track closely spaced targets and to avoid track coalescence. *Proceedings of the Seventh International Conference on Information Fusion*, 1:130–137, 2004.
- [214] N. Ikoma and S.J. Godsill. Extended object tracking with unknown association, missing observations, and clutter using particle filters. *Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing*, September:485–488, 2003.
- [215] N.J. Gordon, D.J. Salmond, and D. Fisher. Bayesian target tracking after group pattern distortion. *O. E. Drummond, editor, Signal and Data Processing of Small Targets, SPIE 3163*, pages 238–248, 1997.
- [216] N.J. Gordon and A. Doucet. Sequential Monte Carlo for maneuvering target tracking in clutter. *Proceedings SPIE*, pages 493–500, 1999.
- [217] C. Hue, J-P. Le Cadre, and P. Perez. Sequential Monte Carlo methods for multiple target tracking and data fusion. *IEEE Tr. Signal Processing*, Feb(50):309–325, 2002.
- [218] J. Vermaak, S. Godsill, and P. Perez. Monte Carlo Filtering for Multi-Target Tracking and Data Association. *Signal Processing Engineering Department, University of Cambridge*, 2004.
- [219] M.G. Rutten, N.J. Gordon, and S. Maskell. Particle-based track-before-detect in rayleigh noise. *Proceedings of SPIE Conference on Signal Processing of Small Targets*, 2004.

- [220] A. Doucet, N. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001.
- [221] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. Markov Chain Monte Carlo in Practice. *Chapman and Hall*, 1996.
- [222] F. Dellaert, S.M. Seitz, C. Thorpe, and S. Thrun. EM, MCMC, and chain flipping for structure from motion with unknown correspondence. *Machine Learning*, 50(1-2):45–71, 2003.
- [223] D. Avitzour, W. Burgard, and D. Fox. Stochastic simulation Bayesian approach to multitarget tracking. *IEE Proceedings on Radar and Sonar Navigation*, 142(2):41–44, 1995.
- [224] N. J. Gordon. A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):353–358, 1997.
- [225] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1665–1670, 2001.
- [226] D. Schulz, W. Burgard, and D. Fox. People tracking with mobile robots using sample-based joint probabilistic data association filters. *in Proceedings of the IEEE International Conference on Robotics and Automation*, 22(2), 2003.
- [227] O. Frank, J. Nieto, J. Guivant, and S. Scheduling. Multiple target tracking using Sequential Monte Carlo methods and statistical data association. *Proceedings of the IEEE / RSJ International Conference on Intelligent Robots and Systems*, 2003.
- [228] R. Karlsson and F. Gustafsson. Monte Carlo data association for multiple target tracking. *in Proceedings of the IEE Seminar Target Tracking: Algorithms and Applications*, pages 13/1–13/5, 2001.
- [229] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

- [230] H. Gauvrit, J-P. Le Cadre, and C. Jauffret. A formulation of multitarget tracking as an incomplete data problem. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1242–1257, 1997.
- [231] R.L. Streit and T.E. Luginbuhl. Maximum likelihood method for probabilistic multi-hypothesis tracking. in *Signal and Data Processing of Small Targets, SPIE 2235*, O. E. Drummond, Ed., 1994.
- [232] L.Y. Pao. Multisensor multitarget mixture reduction algorithms for target tracking. *AIAA Journal of Guidance, Control and Dynamics*, 17, 1994.
- [233] D.J. Salmond. Mixture reduction algorithms for target tracking in clutter. *Signal and Data Processing of Small Targets, SPIE 1305*, O. E. Drummond, Ed., pages 434–445, 1990.
- [234] D. Read. An algorithm for tracking multiple targets. *IEEE Transactions on Automation and Control*, 24(6):84–90, 1979.
- [235] R.L. Popp, T. Kirubarajan, and K.R. Pattipati. Survey of assignment techniques for multitarget tracking. in *Multitarget/Multisensor Tracking: Applications and Advances III*, Y. Bar-Shalom and W. D. Blair, Eds. Artech House, 2000.
- [236] M. Isard and A. Blake. CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.
- [237] J.S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.
- [238] G. Kitagawa. Non-Gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, 82(400):1032–1063, 1987.
- [239] D.L. Alspach and H.W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximation. *IEEE Transactions on Automatic Control*, 17(4):439–448, 1972.
- [240] S.J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. *Proceedings of AeroSense: The 11th International Symposium on Aerospace / Defence Sensing, Simulation and Controls*, vol. Multi Sensor Fusion, Tracking and Resource Management II, 1997.

- [241] B.D.O. Anderson and J.B Moore. Optimal Filtering. *Englewood Cliffs: Prentice-Hall*, 1979.
- [242] A. Doucet. On Sequential Simulation-Based Methods for Bayesian Filtering. Technical report, 1998.
- [243] Babak Mahdavi-Damghani. Oxbridge capital partners yearly review. *Oxford Man Institute of Quantitative Finance*, 2018.
- [244] José Moran and Jean-Philippe Bouchaud. Will a Large Economy Be Stable? *arXiv e-prints*, page arXiv:1901.09629, Jan 2019.
- [245] Jean-Philippe Bouchaud. Chapter 7 - Agent-Based Models for Market Impact and Volatility. In Cars Hommes and Blake LeBaron, editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, pages 393 – 436. Elsevier, 2018.
- [246] H. Benaroya, S.M. Han, and M. Nagurka. *Probability Models in Engineering and Science*, volume 40. 2005.
- [247] Martin Forde, Antoine Jacquier, and A. Mijatovic. Asymptotic formulae for implied volatility in the Heston model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2124:3593–36205, 2010.

Index

- 3-assets Cointelation, 225
- AC, 182
- AD, 182
- Adaptive Smoothing Methods, 153
- Agent-Based Models, 157
- Agricultural, 123
- AI, 175
- Alfred Nobel, 28
- Alive Strategy, 190
- ALLC, 183
- ALLD, 183
- AlphaGo, 175
- Always Cooperate, 182
- Always Cooperates, 183
- Always Deceit, 182
- Always Deceits, 183
- Always Defect, 183
- ANN, 161
- Anomaly Detection, 81
- Anticipative VaR, 134, 153
- Approved Persons, 53
- Arbitrage, 102
- Arbitrage Free Implied Volatility, 86
- Artificial Neural Network, 161, 162
- ATM, 114, 227
- Axelrod's Computer Tournament, 182
- Bank of America, 35
- Basis Points, 160
- Bayesian Band-wise Learning, 133
- Beacon, 159
- Beehive, 159
- Bernoulli, 140
- Beta Distribution, 192
- Bias-Variance Dilemma, 171
- Bid-Ask, 113–115, 160
- Big Data, 32, 99
- Birth & Death Processes, 191
- Birth Process, 191
- Bisection Method, 86
- Black Box, 161
- Black-Scholes, 35, 101
- Black-Scholes-Merton, 81
- Black-Scholes, 123
- Blinker, 159
- Block, 159
- Boat, 159
- Bootstrap Filter, 217
- Bottom-Up, 26, 27, 157
- BP, 53
- Brent, 41
- Brent Algorithm, 86
- Brownian Motion, 26, 157
- BSM, 81
- Butterfly, 112
- BVD, 171
- Calendar Spread, 117
- CCP, 87
- Cellular Automaton, 3, 158
- Central Counter Party, 87

CEV, 101
 Clearing Methodology, 85
 Client Best Interest Rule, 56
 Closest Arbitrage Free Volatility, 117
 Clustering, 133
 CNN, 175
 Co-Movement, 46
 Cointegration, 44
 Cointelation, 35, 39, 47, 50, 54
 Commodities, 56, 122, 123
 Commodities Market, 101
 Condition on the Strike, 89
 Condition on the Tenor, 93
 Constant Elasticity of Variance, 101
 Contract Normalization, 33
 Convolutional Neural Network, 175
 Cooperation, 183
 Corporate Bond, 152
 Correlated Log-normal Pairs, 39
 Crossover, 191
 Crude Oil, 41
 Cycle, 183

 Data Science, 87
 De-Arbitraging, 117
 Dead Strategy, 190
 Death Process, 191, 271
 Deep Blue, 175
 Deep Learning, 36, 39, 157, 175
 Delta Space, 86
 Dependence, 41
 Developed Countries, 123
 Diffusion, 123
 Diversity-Stability Debate, 187
 DNA, 176
 Dominance, 183

 Donsker theorem, 29
 Downside Transform, 110
 Dynamic Mini Order-Book, 199

 Economic Factor, 122
 Economic Sector, 123
 Econophysics, 160
 Eigenvalue, 197
 EKF, 217, 219
 Elasticity of the Liquidity, 115
 Electronics, 123
 Emerging Markets, 123
 Energy, 123
 Envelop, 119
 Environment, 199
 Error Back Propagation, 165
 European Call, 83
 Eurostoxx, 107
 Evolution, 56
 Evolutionary Dynamics, 177, 183, 193
 Evolutionary Process, 189
 EWMA, 151, 166
 Extended Kalman Filter, 217
 Extrapolation, 98

 Falling Variance, 117
 FCA, 56
 Feature Engineering, 178
 Fetching, 117
 FIFO, 160
 Financial Funnel, 164
 Financial Mathematics, 26
 Financial Services and Markets Act, 53
 Financial Services Authority, 53
 Financial System, 197
 Finite Set Statistics, 218, 219
 FISST, 218, 219

Flash Crash, 4
FOB, 199
Front Office, 85
FRTB, 25, 87
FSA, 53
FSMA 2000, 53
Full Order-Book, 199
Full Revaluation, 147, 148

Game of Life, 30, 158
Game Theory, 177
GAN, 30
GANs, 159
Garman-Kohlhagen, 82, 84, 123
Generalized SVI, 107
Generative Adversarial Networks, 159
Generous Tit for Tat, 183
Genetic Algorithm, 177, 189
Geographical Location, 123
Gibbs Sampling, 178, 179
GK, 84
Goldman Sachs, 122
GS, 178
gSVI, 105, 107, 110, 136
GTFT, 183

Hadamard Multiplication, 73
Hamiltonian Monte Carlo, 178
Heston, 35
Heteroscedasticity, 41
HFFF, 165, 175, 231
HFTE, 161, 164
HFTE Game, 198, 199
High Frequency Financial Funnel, 165, 166
Hilbert, 163
Historical Volatility, 151
homoscedasticity, 40

i.i.d., 29, 39, 216
Implied Correlation, 44
Implied Volatility, 33, 108
Implied Volatility Parametrization, 111
Inferred Correlation, 35, 39
Inheritance, 191
Instantaneous Correlation, 47
Interpolation, 33, 88, 98, 124
Invasion, 176, 183, 201
Invasion Flow Chart, 201
Iteration Type, 189
Ito's Lemma, 83, 85, 89
IVP, 105, 111, 136
IVS, 33

Jacobian Matrix, 197
JPDA, 219

Kasparov, 175

Lagging Process, 40, 45, 47
Lagrange Multiplier, 120
Lasso, 171
Lasso Regression, 171
Leading Indicator, 160
Leading Process, 40, 45, 47
Levy Process, 121
LIBOR, 101
Limit Order Book, 204
Liquidity, 119
Liquidity Adjustment, 119
Loaf, 159
LOB, 204
Local Volatility, 151
Log-Normal, 82
Log-Normal Diffusion, 82
long short-term network, 73

Lotka-Volterra, 187, 190
 Lotka-Volterra 3-Species, 188
 LSTM, 73
 LV, 190

 MACD, 164, 168
 Macro Iteration, 190
 Majority Game, 187
 MAR, 47
 margining methodology, 147
 Markov Chain Monte Carlo, 177
 Markowitz, 56
 MCMC, 178
 Mean Reversion, 43
 Merrill Lynch, 35
 Metals, 123
 Metropolis-Hastings, 178
 Micro Iteration, 189
 Mills Ratio, 91
 Minority Game, 184
 Misleading Statement and Actions, 56
 MLR, 161, 164, 169, 176
 Moment Formula, 35
 Moran Process, 165
 MTT, 210
 Multi Linear Regression, 169
 Multi-Linear Regression, 161, 193
 Multiple-try Metropolis, 180
 Multiscale Autoregressive Model, 47
 Multitarget Tracking, 210
 Mutation Process, 191
 Mutation Sampling, 193

 N-Species Predator-Prey Stochastic Model, 164
 Natural SVI, 103
 Neural Net, 160
 Neural Network, 161
 Newton-Raphson, 86
 Normal, 83
 Normal Diffusion, 82
 Normalized Tenors, 89
 Normalizing, 117
 Number of Crosses, 50
 Number of Turning Points, 50

 OI, 176
 Oil, 53
 Open Interest, 172
 Optimization, 35
 Optimization by Constraint, 120
 Order Book, 160
 Order-Book, 160, 177
 Orderbook, 195
 Oscillation, 159
 OTC, 89
 OTM, 107
 Over-fitting, 162, 171
 Over-relaxation, 179

 p-Gaussian Mixture, 140
 P&L, 190
 Parametrization, 99
 Partial Differential Equations, 99
 Particle Filter, 217
 Path of Interaction, 198, 199
 PCA, 120
 PD, 181
 PDE, 39, 99
 Pearson's Correlation, 40, 49
 Pentadecathlon, 159
 Perceptron, 161
 PF, 217
 Pharmaceuticals, 123

PHD, 218, 219
 Physics, 26
 Pillars, 33
 Portfolio, 39, 175
 Position Size, 115
 Predator, 197
 Predator-Prey, 187, 190
 Predator/Prey models, 187
 Prey, 197
 Principal Component Analysis, 120
 Prisoner's Dilemma, 181
 Probability Hypothesis Density, 218, 219
 Probability Theory, 26
 Product Control, 85
 Proxying, 121
 Pulsar, 159

 Quantitative Finance, 26, 35

 Random Strategy, 183
 Ranking Rule, 199
 Raw SVI, 102, 105
 Regularization, 162, 171
 Regulation, 32, 196
 Resampling, 191, 192, 215
 Responsible, 153
 Responsible VaR, 135, 153
 Responsive, 36
 Responsive VaR, 134, 150, 153
 Reversible-jump, 180
 Risk Factor Decomposition, 122
 Risk Management, 85
 RJ-MCMC, 180
 Rolling Contract, 33
 Rolling Contracts, 87, 88

 SABR, 101

 Sampling Zones, 52
 Schonbucher, 100, 102
 Scientific Method, 30
 SDE, 50
 Secant method, 86
 Self-Fulfilling Strategy, 202
 Sequential Bootstrapping, 124
 Sequential Monte Carlo, 217
 Shallow Learning, 157, 175
 Sharpe Ratio, 59
 Simplified IVP, 130
 Sinkhorn's Algorithm, 97
 SIVP, 130
 Skew, 103
 Slice Sampling, 179
 SMC, 217
 Smoothing, 118, 119
 Snell Envelope, 153
 Sparse Data, 121
 Stability, 197
 Stable VaR, 152
 Standardized Strike, 87
 Standardized Tenors, 87
 Statistical Mechanics, 26
 STEM, 26
 Stochastic Alpha, Beta, Rho, 101
 Stochastic Volatility, 228
 Stochastic Volatility Inspired, 35, 102
 Strategy, 190
 Strategy Classification, 190
 Strategy Ecosystem, 177
 Stressed Scenarios, 147
 Sub-Prime Crisis, 87
 Super Predator, 197
 Superposition Theorem, 163
 Survival Process, 191

Survivor Set, 191
 SVI, 35, 102, 136
 SVI Jump-Wings, 107
 SVI-JW, 107
 Switching Flag, 121
 Systemic Risk, 197

 Target, 42
 Target association, 218
 Tenor Pillars, 33
 TF, 164, 166, 176, 195
 TFT, 183
 Theoretical Biology, 177, 187
 Threshold Logic Model, 161
 Tit For Tat, 183
 Tit for Tat, 183
 Toad, 159
 Top-Down, 26, 27, 157
 Transform, 106, 107
 Trend Following, 164, 166

 Under-fitting, 171
 Updating Volatility, 118
 UTOPE, 54

 VaR, 36
 Variance Reduction, 51
 Vol of Vol, 103
 Volatility Surface de-Arbitraging, 81

 Walmart, 42
 Wiener Process, 28
 Win-Stay Lose-Shift, 183
 Wing, 32, 35, 102, 113
 Wing Curvature, 115
 WSLS, 183

 XOR, 164, 176

 Zero-Sum Game, 159

