CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

Contents

Contributors xvi

Preface xix

INTERACTING WITH INVESTORS AND ASSET OWNERS 1

Part I Robo Advisors and Automated Recommendation 3

1 Introduction to Part I. *Robo-advising as a Technological Platform for Optimization and Recommendations*
Lisa L. Huang 5

2 New Frontiers of Robo-Advising: Consumption, Saving, Debt Management, and Taxes
Francesco D'Acunto and Alberto G. Rossi 9
2.1 Robo-advice and the balance-sheet view of the household 9
2.2 Robo-advising for consumption-saving choices 12
2.2.1 Open areas of inquiry in robo-advising for consumption-saving choices 16
2.3 Robo-advising and durable spending choices 18
2.3.1 Robo-advising for housing choices 18
2.3.2 Robo-advising for the purchase of vehicles 19
2.3.3 Open areas of inquiry in robo-advising for durable spending 20
2.4 Robo-advising and consumers' lending decisions 21
2.5 Areas of consumer finance with a scarce presence of robo-advising 22
2.5.1 Robo-advising and consumer credit management 23
2.5.2 Robo-advising and human capital investments 25
2.5.3 Robo-advising and tax management 26
2.6 E pluribus unum: Is the Holistic Robo-Advisor the future of robo-advising? 27
2.7 Conclusions 28

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

3 Robo-Advising: Less AI and More XAI? Augmenting Algorithms with Humans-in-the-Loop 33
Milo Bianchi and Marie Brière
3.1 Introduction 33
3.2 Why so popular? 34
3.2.1 Fintech revolution 36
3.2.2 Fundamental problems with investors 36
3.2.3 Fundamental problems with advisors 37
3.3 Promises 38
3.3.1 Accountable procedures and tailored recommendations 38
3.3.2 Make investors better off 41
3.3.3 Reach under-served investors 42
3.4 Open questions 43
3.4.1 Why not more AI/big data? 43
3.4.2 How far shall we go into personalization? 45
3.4.3 Can humans trust robots? 47
3.4.4 Do robots replace or complement human decisions? 51
3.5 The next generation of robo-advisors 51

4 Robo-advisory: From investing principles and algorithms to future developments 60
Adam Grealish and Petter N. Kolm
4.1 From investing principles to algorithms 60
4.1.1 Client assessment and onboarding 62
4.1.2 Implementation of the investment strategy 66
4.1.3 Ongoing management of the investment strategy 69
4.2 Automated tax management for retail investors 70
4.2.1 Tax-loss harvesting 71
4.2.2 Asset location 73
4.2.3 Tax lot management 74
4.3 Investor interaction 74
4.3.1 Investor education 74
4.3.2 Data collection and split testing 75
4.4 Expanding service offerings 75
4.4.1 Goals-based investing 75
4.4.2 Retirement planning 76
4.4.3 Responsible investing 77
4.4.4 Smart beta and factor investing 78
4.4.5 Risk parity 79
4.4.6 User-defined portfolios 79
4.4.7 Cash management 80
4.5 Conclusion 81

5 Recommender Systems for Corporate Bond Trading 86
Dominic Wright, Artur Henrykowski, Jacky Lee and Luca Capriotti
5.1 Introduction 86
5.2 Bond recommender systems 87
5.2.1 Content-based filtering 88
5.2.2 Collaborative filtering 90

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

5.3 Testing 93
5.3.1 Hyperparameter optimization 94
5.3.2 Testing results 94
5.4 Explaining recommendations 95
5.5 Conclusions 96

Part II How Learned Flows Form Prices 99

6 Introduction to Part II. *Price Impact: Information Revelation or Self-Fulfilling Prophecies?*
Jean-Philippe Bouchaud 101
6.1 Liquidity hide-and-seek 101
6.2 Information efficiency vs. statistical efficiency 102
6.3 Price "Discovery" vs. price "formation" 104

7 Order Flow and Price Formation
Fabrizio Lillo 107
7.1 Introduction 107
7.2 The limit order book 109
7.3 Modeling approaches 110
7.4 Order flow 115
7.5 Cross impact 117
7.6 Market impact of metaorders 119
7.7 Co-impact 124
7.8 Conclusion 127

8 Price Formation and Learning in Equilibrium under Asymmetric Information
Umut Çetin 130
8.1 Introduction 130
8.2 The Kyle model 131
8.2.1 A toy example 131
8.2.2 The Kyle model in continuous time 133
8.3 The static Kyle equilibrium 137
8.4 The static Kyle model with multiple insiders 141
8.5 Dynamic Kyle equilibrium 143
8.6 The Kyle model and default risk 144
8.7 Glosten–Milgrom model 146
8.8 Risk aversion of market makers 147
8.9 Conclusion and further remarks 149

9 Deciphering How Investors' Daily Flows are Forming Prices
Daniel Giamouridis, Georgios V. Papaioannou and Brice Rosenzweig 153
9.1 Introduction 153
9.2 Data description and exploratory statistics 156
9.3 Modeling and methodology 160
9.4 Empirical results 162

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

9.4.1 Aggregate flow and price formation 162
9.4.2 Participant type flow imbalance and price formation 163
9.4.3 Co-Impact 168
9.5 Summary and Conclusions 169

TOWARDS BETTER RISK INTERMEDIATION 173

Part III High Frequency Finance 175

10 Introduction to Part III
Robert Almgren 177
10.1 Chapters in this Part 177
10.2 State of the field and future prospects 179
10.2.1 Data needs and simulation 179
10.2.2 Game formulation 180
10.2.3 Conclusion 181

11 Reinforcement Learning Methods in Algorithmic Trading
Olivier Guéant 182
11.1 Introduction 182
11.1.1 The recent successes of reinforcement learning 182
11.1.2 Finance, it might be your go 183
11.2 A brief introduction to reinforcement learning 184
11.2.1 Markov Decision Processes and optimization problems 185
11.2.2 Basic concepts 186
11.2.3 Main RL methods 188
11.3 Finance is not a game 192
11.3.1 States and actions 192
11.3.2 The role of models 192
11.3.3 The question of risk 194
11.3.4 The question of time steps 194
11.4 A review of existing works 194
11.4.1 Statistical arbitrage 195
11.4.2 Optimal execution 196
11.4.3 Market making 198
11.5 Conclusion and perspectives for the future 199

12 Stochastic Approximation Applied to Optimal Execution: Learning by Trading
Sophie Laruelle 205
12.1 Introduction 205
12.2 Stochastic approximation: results on a.s. convergence and its rate 207
12.2.1 Back to deterministic recursive methods 207
12.2.2 Stochastic recursive methods 208
12.3 Applications to optimal execution 216
12.3.1 Optimal split of an order across liquidity pools 216
12.3.2 Optimal posting price of limit orders 221

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

13 Reinforcement Learning for Algorithmic Trading
Álvaro Cartea, Sebastian Jaimungal and Leandro Sánchez-Betancourt 230
13.1 Learning in financial markets 230
13.2 Statistical arbitrage: trading an FX triplet 232
13.2.1 Market model 234
13.3 The reinforcement learning paradigm 235
13.3.1 Deep Q-learning (DQN) 237
13.3.2 Reinforced deep Markov models 239
13.3.3 Implementation of RDMM 244
13.4 Optimal trading in triplet 245
13.4.1 Remarks on RDMM versus DDQN 248
13.5 Conclusions and future work 249

Part IV Advanced Optimization Techniques 251

14 Introduction to Part IV. *Advanced Optimization Techniques for Banks and Asset Managers*
Paul Bilokon, Matthew F. Dixon and Igor Halperin 253
14.1 Introduction 253
14.1.1 Pelger's asset pricing model 257
14.2 Data wins the center stage 259
14.2.1 Deep hedging vs. reinforcement learning for option pricing 259
14.2.2 Market simulators 261
14.3 Stratified models for portfolio construction 262
14.4 Summary 263

15 Harnessing Quantitative Finance by Data-Centric Methods
Blanka Horvath, Aitor Muguruza Gonzalez and Mikko S. Pakkanen 265
15.1 Data-centric methods in quantitative finance 265
15.2 Pricing and calibration by supervised learning 269
15.2.1 Model calibration framework 269
15.2.2 Pricing and calibration aided by deep neural networks 270
15.2.3 An example where deep pricing makes a difference: the rough Bergomi model 272
15.2.4 Choosing the feature set 273
15.2.5 Supervised learning approaches, and their ability to loosen limitations of the tractability mantra 274
15.3 Pricing and hedging by unsupervised deep learning 275
15.3.1 Deep hedging 275
15.3.2 Utility indifference pricing 278
15.3.3 Numerical illustration 279
15.4 Market generators 284
15.4.1 The case for more flexible data-driven market models 285
15.4.2 The case for classical models 288
15.4.3 Synergies between the classical and modern approaches, and further risk management considerations 288
15.5 Outlook and challenges with data at centre stage 288

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

16 Asset Pricing and Investment with Big Data 293
Markus Pelger
16.1 Overview 293
16.2 No-arbitrage pricing and investment 294
16.3 Factor models 296
16.4 Deep learning in asset pricing 300
16.4.1 Forecasting 300
16.4.2 No-arbitrage model 301
16.4.3 Economic dynamics 303
16.4.4 Model architecture 305
16.4.5 Empirical results 305
16.5 Decision trees in asset pricing 308
16.5.1 SDF recovery as a mean-variance optimization problem 310
16.5.2 Empirical results 312
16.6 Directions for future research 314

17 Portfolio Construction Using Stratified Models 317
Jonathan Tuck, Shane Barratt and Stephen Boyd
17.1 Introduction 317
17.1.1 Related work 319
17.2 Laplacian regularized stratified models 319
17.3 Dataset 320
17.4 Stratified market conditions 322
17.5 Stratified return model 324
17.5.1 Hyper-parameter search 324
17.5.2 Final stratified return model 325
17.6 Stratified risk model 326
17.6.1 Hyper-parameter search 327
17.6.2 Final stratified risk model 327
17.7 Trading policy and backtest 328
17.7.1 Trading policy 328
17.7.2 Backtests 330
17.7.3 Hyper-parameter selection 331
17.7.4 Final trading policy results 332
17.8 Extensions and variations 335
17.9 Conclusions 337

Part V New Frontiers for Stochastic Control in Finance 341

18 Introduction to Part V. *Machine Learning and Applied Mathematics: a Game of Hide-and-Seek?* 343
Gilles Pagès

19 The Curse of Optimality, and How to Break it? 354
Xun Yu Zhou
19.1 Introduction 354
19.2 Entropy-regularized exploratory formulation 357
19.2.1 Classical stochastic control 357

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

19.2.2 Exploratory formulation 359
19.2.3 Entropy regularization 360
19.3 Optimal distributional policies 361
19.4 Non-convex optimization and Langevin diffusions 363
19.5 Algorithmic considerations for RL 365
19.6 Conclusion 367

20 Deep Learning for Mean Field Games and Mean Field Control with Applications to Finance
René Carmona and Mathieu Laurière 369
20.1 Introduction 369
20.1.1 Literature review 370
20.1.2 Definition of the problems 372
20.2 Direct method for MKV Control 373
20.2.1 Description of the method 374
20.2.2 Numerical illustration: a price impact model 376
20.3 Deep BSDE method for MKV FBSDEs 379
20.3.1 Description of the method 379
20.3.2 Numerical illustration: a toy model of systemic risk 381
20.4 DGM method for mean field PDEs 382
20.4.1 Description of the method 383
20.4.2 Numerical illustration: a crowded trade model 386
20.5 Conclusion 389

21 Reinforcement Learning for Mean Field Games, with Applications to Economics
Andrea Angiuli, Jean-Pierre Fouque and Mathieu Laurière 393
21.1 Introduction 393
21.2 Finite horizon mean field problems 396
21.2.1 Mean field games 397
21.2.2 Mean field control 397
21.3 Two-timescale approach 398
21.3.1 Discrete formulation 398
21.3.2 Action-value function 400
21.3.3 Unification through a two-timescale approach 401
21.4 Reinforcement learning algorithm 404
21.4.1 Reinforcement learning 404
21.4.2 Algorithm 404
21.4.3 Learning rates 406
21.4.4 Application to continuous problems 407
21.5 A mean field accumulation problem 407
21.5.1 Description of the problem 407
21.5.2 Solution of the MFG 409
21.5.3 Solution of the MFC 410
21.5.4 Numerical results 411
21.6 A mean field execution problem 413
21.6.1 The MFG trader problem 415
21.6.2 Solution of the MFG problem 416
21.6.3 The MFC trader problem 416

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

21.6.4 Solution of the MFC problem 417
21.6.5 Numerical results 418
21.7 Conclusion 421

22 Neural Networks-Based Algorithms for Stochastic Control and PDEs in Finance
Maximilien Germain, Huyên Pham and Xavier Warin 426
22.1 Breakthrough in the resolution of high-dimensional nonlinear problems 426
22.2 Deep learning approach to stochastic control 427
22.2.1 Global approach 428
22.2.2 Backward dynamic programming approach 429
22.3 Machine learning algorithms for nonlinear PDEs 430
22.3.1 Deterministic approach by neural networks 431
22.3.2 Probabilistic approach by neural networks 432
22.3.3 Case of fully nonlinear PDEs 436
22.3.4 Limitations of the machine learning approach 440
22.4 Numerical applications 440
22.4.1 Numerical tests on credit valuation adjustment pricing 441
22.4.2 Portfolio allocation in stochastic volatility models 443
22.5 Extensions and perspectives 448

23 Generative Adversarial Networks: Some Analytical Perspectives
Haoyang Cao and Xin Guo 453
23.1 Introduction 453
23.2 Basics of GANs: an analytical view 455
23.3 GANs Training 460
23.4 Applications of GANs 467
23.4.1 Computing MFGs via GANs 467
23.4.2 GANs in Mathematical Finance 471
23.5 Conclusion and Discussion 475

CONNECTIONS WITH THE REAL ECONOMY 479

Part VI Nowcasting with Alternative Data 481

24 Introduction to Part VI. *Nowcasting is Coming*
Michael Recce 483
24.1 Micro before macro 483
24.2 Advance driven by Moore's law 484
24.3 The CEO dashboard 485
24.4 Internet companies led progress in nowcasting 486
24.5 CEO dashboard from alternative data 486
24.6 Nowcasting with alternative data 487

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

25 Data Preselection in Machine Learning Methods: An Application to Macroeconomic Nowcasting with Google Search Data 490
Laurent Ferrara and Anna Simoni
25.1 Introduction 490
25.2 Structure of Google search database 493
25.3 The nowcasting approach 494
25.3.1 Linear Bridge equation 494
25.3.2 Preselection of Google search variables 495
25.4 Factor models 496
25.5 Methods based on regularisation: Ridge 499
25.6 Nowcasting euro area GDP growth: Empirical eesults 500
25.7 Conclusions 505

26 Alternative data and ML for macro nowcasting 507
Apurv Jain
26.1 The fundamental problems of macro data 507
26.2 High-dimensionality problem 510
26.3 Nowcasting the big and jagged data 512
26.3.1 Nonlinearity and ML for nowcasting 516
26.4 Dimensions of alternative data quality 519
26.4.1 A crowd-sourced experiment 519
26.4.2 The need for a hypothesis 519
26.5 Non-farm payrolls and web search case study 522
26.5.1 Background and related work 522
26.5.2 Government non-farm payrolls data overview 526
26.5.3 Information content of NFP revisions 528
26.5.4 Web search data 530
26.5.5 Search and NFP correlation 531
26.5.6 Regression results 533
26.5.7 Robustness 535
26.5.8 Machine learning for NFP revisions 536
26.6 Conclusion and future work 538

27 Nowcasting Corporate Financials and Consumer Baskets with Alternative Data 545
Michael Fleder and Devavrat Shah
27.1 Quant for alt data 546
27.2 Nowcasting company financials 547
27.2.1 Problem statement and model 547
27.2.2 Contributions 549
27.2.3 Theoretical results 551
27.2.4 Experiments 552
27.3 Inferring products in anonymized transactions 554
27.3.1 Problem statement and model 554
27.3.2 Contributions 554
27.3.3 Algorithm 556
27.3.4 Main results 557
27.3.5 Experiments 558
27.4 Conclusion 561
27.5 Relevant literature 561

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

28 NLP in Finance
Prabhanjan Kambadur, Gideon Mann and Amanda Stent 563
28.1 Core NLP techniques 564
28.1.1 Basic language analytics 565
28.1.2 Higher-level linguistic analysis 566
28.2 Mathematics for NLP 568
28.2.1 Introduction to supervised learning 569
28.2.2 Machine learning methods for NLP 571
28.3 Applications 575
28.3.1 Information extraction 576
28.3.2 NSTM: Identifying key themes in news 580
28.3.3 Market sentiment analysis 584
28.4 Conclusion 586

29 The Exploitation of Recurrent Satellite Imaging for the Fine-Scale Observation of Human Activity
Carlo de Franchis, Sébastien Drouyer, Gabriele Facciolo, Rafael Grompone von Gioi, Charles Hessel and Jean-Michel Morel 593
29.1 Introduction 593
29.2 What is recurrent satellite imaging? 594
29.2.1 A landscape of satellites images 595
29.3 3D monitoring from space 599
29.4 The SAR revolution of oil storage 604
29.4.1 Overview 605
29.4.2 Detailed description 605
29.5 Creating a movie of the earth 610
29.5.1 Overview of the existing methods 610
29.5.2 Algorithms 612
29.6 Ground visibility and cloud detection 613
29.7 Detecting and counting cars from space 616
29.7.1 Vehicle detection on high resolution satellite images 618
29.7.2 Parking occupancy estimation on PlanetScope satellite images 620

Part VII Biases and Model Risks of Data-Driven Learning 629

30 Introduction to Part VII. *Towards the Ideal Mix between Data and Models*
Mathieu Rosenbaum 631
30.1 What are we exactly talking about? 631
30.2 What is a good model? 631
30.3 Simulating what has never been observed 632
30.4 Being pragmatic 632

31 Generative Pricing Model Complexity: The Case for Volatility-Managed Portfolios
Brian Clark, Akhtar Siddique and Majeed Simaan 634
31.1 Introduction 634
31.2 Quantifying model complexity 640
31.2.1 The main idea 640

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51619-5 — Machine Learning and Data Sciences for Financial Markets
Edited by Agostino Capponi, Charles-Albert Lehalle
Table of Contents
More Information

	31.2.2 ALE functions	640
	31.2.3 Interaction strength (IAS)	642
	31.2.4 Main effect complexity (MEC)	644
31.3	Portfolio problem	647
	31.3.1 Volatility managed portfolio	647
	31.3.2 The appeal of ML in portfolio	648
31.4	Empirical investigation	649
	31.4.1 Data	649
	31.4.2 Training and testing	650
	31.4.3 Performance	651
	31.4.4 Results and discussion	652
	31.4.5 Additional results	654
31.5	Concluding remarks	657
32	**Bayesian Deep Fundamental Factor Models**	
	Matthew F. Dixon and Nicholas G. Polson	661
32.1	Introduction	661
	32.1.1 Why Bayesian deep learning?	662
	32.1.2 Connection with fundamental factor models	663
	32.1.3 Overview	664
32.2	Barra fundamental factor models	665
	32.2.1 Prediction with the Barra model	666
32.3	Bayesian deep learning	667
	32.3.1 Deep probabilistic models	667
	32.3.2 Variational Approximation	669
	32.3.3 Bayes by backprop	670
32.4	Bayesian deep fundamental factor models	671
	32.4.1 Probabilistic prediction	671
32.5	Deep network interpretability	672
	32.5.1 Factor sensitivities	673
32.6	Applications: Russell 1000-factor modeling	675
32.7	Discussion	680
32.A	Appendix: Gradients of two-layer feedforward networks	684
32.B	Description of Russell 1000-factor model	685
33	**Black-Box Model Risk in Finance**	
	Samuel N. Cohen, Derek Snow and Lukasz Szpruch	687
33.1	Introduction	687
33.2	A practical application of machine learning	689
	33.2.1 How to use neural networks for derivative modelling	689
	33.2.2 Black-box trade-offs	692
33.3	The role of data	693
	33.3.1 Data risks	693
	33.3.2 Data solutions	695
33.4	The role of models	699
	33.4.1 Model risks	699
	33.4.2 Model solutions	705
Index		718

© in this web service Cambridge University Press & Assessment www.cambridge.org