Machine Learning and Data Sciences for Financial Markets

Leveraging the research efforts of more than 60 experts in the area, this book reviews cutting-edge practices in machine learning for financial markets. Instead of seeing machine learning as a new field, the authors explore the connection between knowledge developed in quantitative finance over the past 40 years and modern techniques generated by the current revolution in data sciences and artificial intelligence.

The text is structured around three main areas: "Interacting with investors and asset owners," which covers robo-advisors and price formation; "Towards better risk intermediation," which discusses derivative hedging, portfolio construction, and machine learning for dynamic optimization; and "Connections with the real economy," which explores nowcasting, alternative data, and ethics of algorithms.

Accessible to a wide audience, this invaluable resource will allow practitioners to include machine learning driven techniques in their day-to-day quantitative practices, while students will build intuition and come to appreciate the technical tools and motivation behind the theory.

AGOSTINO CAPPONI is Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. He conducts research in financial technology and market microstructure. His work has been recognized with the NSF CAREER Award, and a JP Morgan AI Research award. Capponi is a co-editor of *Management Science and Mathematics and Financial Economics*. He is a Council member of the Bachelier Financial Society, and recently served as Chair of the SIAM-FME and INFORMS Finance.

CHARLES-ALBERT LEHALLE is Global Head of Quantitative R&D at Abu Dhabi Investment Authority and Visiting Professor at Imperial College London. He has a PhD in machine learning, was previously Head of Data Analytics at CFM, and held different Global Head positions at Crédit Agricole CIB. Recognized as an expert in market microstructure, Lehalle is often invited to present to regulators and policy-makers.

"Agostino Capponi and Charles-Albert Lehalle have edited an excellent book tackling the most important topics associated with the application of machine learning and data science techniques to the challenging field of finance, including robo-advisory, high-frequency trading, nowcasting, and alternative data. I highly recommend this book to any reader interested in our field, regardless of experience or background."
— Marcos López de Prado, *Abu Dhabi Investment Authority & Cornell University*

"Beginning with the 1973 publication of the Black–Scholes formula, mathematical models coupled with computing revolutionized finance. We are now witnessing a second revolution as larger-scale computing makes data science and machine learning methods feasible. This book demonstrates that the second revolution is not a departure from, but rather a continuation of, the first revolution. It will be essential reading for researchers in quantitative finance, whether they were participants in the first revolution or are only now joining the fray."
— Steven E. Shreve, *Carnegie Mellon University*

"[This book] comes at a critical time in the financial markets. The amount of machine-readable data available to practitioners, the power of the statistical models they can build, and the computational power available to train them keeps growing exponentially. AI and machine learning are increasingly embedded into every aspect of the investing process. The common curriculum, however, both in finance and in applications of machine learning, lags behind. This book provides an excellent and very thorough overview of the state of the art in the field, with contributions by key researchers and practitioners. The monumental work done by the editors and reviewers shows in the wide diversity of current topics covered – from deep learning for solving partial differential equations to transformative breakthroughs in NLP. This book, which I cannot recommend highly enough, will be useful to any practitioner or student who wishes to familiarize themselves with the current state of the art and build their careers and research on a solid foundation."
— Gary Kazantsev, *Bloomberg & Columbia University*

Machine Learning and Data Sciences for Financial Markets

A Guide to Contemporary Practices

Edited by

Agostino Capponi
Columbia University, New York

Charles-Albert Lehalle
Abu Dhabi Investment Authority

CAMBRIDGE UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316516195
DOI: 10.1017/9781009028943

First published 2023

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

ISBN 978-1-316-51619-5 Hardback

Contents

Contributors

Robert Almgren *Quantitative Brokers, New York; and Princeton University.*
Andrea Angiuli *Department of Statistics and Applied Probability, University of California, Santa Barbara.*
Shane Barratt *Stanford University, Department of Electrical Engineering.*
Milo Bianchi *Toulouse School of Economics, TSM; and IUF, University of Toulouse Capitole.*
Paul Bilokon *Department of Mathematics, Imperial College, London.*
Jean-Philippe Bouchaud *Capital Fund Management, Paris.*
Stephen Boyd *Stanford University, Department of Electrical Engineering.*
Haoyang Cao *CMAP, École Polytechnique.*
Marie Brière *Amundi, Paris Dauphine University, and Université Libre de Bruxelles.*
Luca Capriotti *Department of Mathematics, University College London; and New York University, Tandon School of Engineering.*
René Carmona *Department of Operations Research and Financial Engineering, Princeton University.*
Álvaro Cartea *Oxford University, Mathematical Institute, and Oxford–Man Institute of Quantitative Finance.*
Umut Çetin *London School of Economics, Department of Statistics.*
Brian Clark *Lally School of Management, Rensselaer Polytechnic Institute.*
Samuel N. Cohen *Mathematical Institute, University of Oxford.*
Francesco D'Acunto *Carroll School of Management, Boston College.*
Carlo de Franchis *ENS Paris-Saclay, CNRS; and Kayrros, Paris.*
Matthew F. Dixon *Department of Applied Mathematics, Illinois Institute of Technology.*
Sébastien Drouyer *ENS Paris-Saclay, CNRS.*
Gabriele Facciolo *ENS Paris-Saclay, CNRS.*
Laurent Ferrara *Skema Business School, University Côte d'Azur; and QuantCube Technology.*
Michael Fleder *Massachusetts Institute of Technology; and Covariance Labs, New York.*

Jean-Pierre Fouque *Department of Statistics and Applied Probability, University of California, Santa Barbara.*
Maximilien Germain *LPSM, Université de Paris.*
Aitor Muguruza Gonzalez *Imperial College London and Kaiju Capital Management.*
Daniel Giamouridis *Bank of America, Data and Innovation Group, London.*
Adam Grealish *Altruist, Los Angeles.*
Rafael Grompone von Gioi *ENS Paris-Saclay, CNRS.*
Olivier Guéant *Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne.*
Xin Guo *University of California, Berkeley, Department of Industrial Engineering and Operations Research.*
Igor Halperin *AI Research, Fidelity Investments, Boston.*
Artur Henrykowski *Department of Mathematics, University College London.*
Charles Hessel *ENS Paris-Saclay, CNRS; and Kayrros, Paris.*
Blanka Horvath *Technical University of Munich; Munich Data Science Institute; King's College London; and The Alan Turing Institute.*
Lisa L. Huang *Head of AI Investment Management and Planning, Fidelity.*
Sebastian Jaimungal *University of Toronto, Statistical Sciences.*
Apurv Jain *MacroXStudio, San Francisco.*
Prabhanjan Kambadur *Bloomberg, New York.*
Petter N. Kolm *Courant Institute of Mathematical Sciences, New York University.*
Sophie Laruelle *Université Paris Est Creteil, CNRS, LAMA; and Université Gustave Eiffel, LAMA, Marne-la-Vallée.*
Mathieu Laurière *Department of Operations Research and Financial Engineering, Princeton University.*
Jacky Lee *Department of Mathematics, University College London.*
Fabrizio Lillo *University of Bologna; and Scuola Normale Superiore.*
Gideon Mann *Bloomberg, New York.*
Alberto G. Rossi *McDonough School of Business, Georgetown University.*
Jean-Michel Morel *ENS Paris-Saclay, CNRS.*
Gilles Pagès *LPSM, Sorbonne-Université.*
Mikko S. Pakkanen *Imperial College London.*
Georgios V. Papaioannou *Bank of America, Data and Innovation Group, London.*
Markus Pelger *Stanford University, Department of Management Science & Engineering.*
Huyên Pham *LPSM, Université de Paris.*
Nicholas G. Polson *ChicagoBooth, University of Chicago.*
Michael Recce *CEO, AlphaROC Inc., New York.*
Mathieu Rosenbaum *CMAP, École Polytechnique.*
Brice Rosenzweig *Bank of America, Data and Innovation Group, London.*
Leandro Sánchez-Betancourt *Oxford University, Mathematical Institute.*
Devavrat Shah *Massachusetts Institute of Technology.*

Akhtar Siddique *Economics Department, Office of the Comptroller of the Currency.*
Majeed Simaan *School of Business, Stevens Institute of Technology.*
Anna Simoni *CREST, CNRS, ENSAE, École Polytechnique, Institut Polytechnique de Paris.*
Derek Snow *The Alan Turing Institute.*
Amanda Stent *Colby College.*
Lukasz Szpruch *School of Mathematics, University of Edinburgh.*
Jonathan Tuck *Stanford University, Department of Electrical Engineering.*
Xavier Warin *EDF R&D.*
Dominic Wright *Department of Mathematics, University College London.*
Xun Yu Zhou *Columbia University, Department of Industrial Engineering and Operations Research; and The Data Science Institute, New York.*

Preface

Machine learning, Artificial Intelligence (AI), and data science pervade every aspect of our everyday life. Many of the techniques developed by the Computer Science community are becoming increasingly used in the area of financial engineering, ranging from the use of deep learning methods for hedging and risk management through the exploitation of AI techniques for investment or design of trading systems. These techniques are also having enormous implications on the operations of financial markets. It is thus not surprising to see increasingly the proliferation of AI research groups or recently created "AI Labs" at major banks, centered around topics of key relevance to financial services. Those include, among others, explainable AI, human-machine interaction, and DS methods for extracting information from data and using it to support investment decisions. The integration of AI methods in the decision making process may also have unintended or unanticipated consequences especially in a sector like finance, where bad intermediation of risk can spread over the whole economy. Many of the ethical issues expected from AI systems, including privacy, data manipulation, opacity, and discrimination, can be detrimental to financial markets. For example, data leakage is a key concern for banks; regulatory authorities need to deal with it, and so is fairness in the distribution of debt and issuance of loans. In asset management, the question of bias introduced by a dataset and its stationarity has been known for a long time; the more data dominate decisions, the more important they are. All those issues are getting increasing consideration from major regulatory bodies worldwide.

We should mention that if we come back to the early age of machine learning, the techniques and tools used to provide theoretical grounds to the process of learning from data share their roots with the ones that gave birth to online optimization and stochastic control. They are based on asymptotics of discrete stochastic processes and on stochastic algorithms that support frameworks in which the learned parameters, like the weights of a neural network, are seen as controls that evolve during the learning process. These parameters start at an arbitrary point (they are often randomly initialized) and are meant to follow flows which minimize a criterion usually referred to as a loss function: they are "controls" driving the neural network from a random state to a state where a target task can be performed. These technical tools, designed to capture the behavior of a stochastic system that is driven to a specific state in a noisy environment,

evolved in parallel to address important problems arising in financial markets. A prominent example is "hedging", where one needs to hedge a portfolio of derivatives by replicating the risks embedded into the derivative constituents of the portfolio. In such a case, this portfolio is a control driving the balance sheet of an institution towards a state with minimal unhedged risk. Other business needs require the design of a portfolio that captures investment goals stated in a more generic way (with no specification in terms of tradable instruments). Hence financial engineering has exploited these tools from the 1980s and contributed to their improvement. This community did it independently from the machine learning community, which also contributed to improving these tools mostly from an algorithmic perspective. In recent years, the disciplines of data science and AI have started to be seriously involved in the analysis of financial markets. It is important to not forget what academics and practitioners understood about these tools, and especially the way they can improve risk management in markets. Since the dream of replacing reasoning and modelling by data and black boxes is dangerous in the non stationary environment of financial markets, it is important to integrate machine learning practices with the structural knowledge developed by quantitative finance during the last 40 years. "Old" knowledge and new approaches should cross-fertilize, injecting the structural nonlinearities of learning machines and their capability to extract structures from data exactly where more formal methods had a lack of adaptiveness.

Inspired by these considerations, we have decided to collect the most relevant sample of cutting edge research developed in the fields of Machine learning, Data Sciences, and AI with application to finance into a book. Our book project has been strongly supported by the academic community. We have invited active researchers with demonstrated expertise and leadership in their own areas of relevance to contribute a chapter to the book. They have enthusiastically responded to our call, and submitted high quality chapters. Their chapters have been reviewed by a team of qualified referees, who have carefully processed the content and provided excellent feedback for improvement. Our project has also received strong support by the Cambridge University Press (CUP), which has kindly agreed to publish the volume. This book follows the tradition of the financial engineering community started in the last decade to spotlight topics of increasing importance for the community and the broad society overall, and culminating then into the *Handbook on Systemic Risk* published by CUP. The topics of the chapters are highly reflective of the research agenda of the two most prominent financial engineering societies, namely the SIAM-FM Activity group currently chaired by Agostino Capponi, and the Finance and Insurance Reloaded program (FaIR) within the Institute of Louis Bachelier Paris, which Charles-Albert Lehalle started a few years ago. The last two biennial meetings of the SIAM-FM group, held in 2019 and 2021, featured many plenary talks, invited minisymposia, and tutorials in the area of machine learning and data science. Talks given by a mix of academics and industry practitioners, reflected both an algorithmic technical perspective and the integration of ML methodologies

against financial markets data. Relatedly, the FaIR transverse program has been a unique occasion to meet researchers involved in the use of new technologies for financial markets. The series of thematic workshops organized by FaIR and the ACPR (French regulator for banking), as well as its kick-off workshop at the Collège de France, have specially been places of intense thinking and brainstorming on how machine learning would influence these industries.

Since starting our effort, it has been our intention to structure the book around three main areas of interest: "Interactions with investors and asset owners" which mainly covers robo-advisors and price formation; "Risk intermediation" which covers portfolio construction, and machine learning for dynamic optimization, including optimal trading; and "Connections with the real economy" covers nowcasting, alternative data and ethics of algorithms. This structure offers a comprehensive and easy to read perspective on the areas of machine learning, AI and data science in financial markets.

We believe that now, more than ever, is now a good time to collect the various efforts made by leading and high profile researchers, including academics, practitioners and policy makers, into a book. We have developed this book with the idea that it becomes a key reference in the field. It will serve as the main reference for experienced researchers with training in quantitative methods, who want to increase their awareness of the cutting edge research being done in the area. We have also paid attention to a pedagogic component, and strived to make each chapter comprehensive enough and understandable by advanced graduate students. Those in search of a new topic to explore for their dissertation at the intersection of machine learning, data science, and finance will be inspired by the methodologies and applications presented in the book.

We expect the handbook to be received well beyond the academic community. Financial institutions and policy makers wishing to bring rigor to their business will be able to leverage upon the methodologies discussed in the book, and integrate them with data. As a result, the book will have a high potential of increasing the collaborations of the academia with the public and private sector, and to educate new generations of scientists who will build the new AI technologies in the financial sector.

The editors, Agostino Capponi and Charles-Albert Lehalle
New York and Abu Dhabi

Acknowledgments of referees.

The editors and contributors would like to thank the referees who took time to read and comment the contributions of this book:

- Agustin Lifschitz, Capital Fund Management, Paris, France.
- Amine Raboun, Euronext Paris, Courbevoie, France.
- Andrea Angiuli, Department of Statistics and Applied Probability, University of California, Santa Barbara.
- Bobby Shackelton, Head of Geospatial, Bloomberg LP.
- Emmanuel Sérié, Capital Fund Management, Paris, France.
- Frederic Bucci,
- Haoran Wang, CAI Data Science and Machine Learning, The Vanguard Group, Inc., Malvern, PA, USA.
- Haoyang Cao, The Alan Turing Institute.
- Harvey Stein, Head, Quantitative Risk Analytics, Bloomberg and Adjunct Professor, Mathematics Department, Columbia University.
- Ibrahim Ekren, Florida State University, Department of Mathematics, Tallahassee, FL.
- Iuliia Manziuk, Engineers Gate, Quantitative Researcher, London.
- Jiacheng Zhang, Department of Operations Research and Financial Engineering, Princeton University.
- Matthew Dixon, Illinois Institute of Technology, Department of Applied Mathematics.
- Michael Fleder, Massachusetts Institute of Technology and Covariance.AI.
- Michael Reher, University of California San Diego, Rady School of Management.
- Noufel Frikha, Université de Paris, Laboratoire de Probabilités, Statistiques et Modélisation.
- Othmane Mounjid, University of California, Berkeley (IEOR department).
- Renyuan Xu, Industrial and Systems Engineering, University of Southern California.
- Ruimeng Hu, Department of Mathematics, Department of Statistics and Applied Probability, University of California, Santa Barbara.
- Shihao Gu, Booth School of Business, University of Chicago.
- Sveinn Olafsson, Stevens Institute of Technology.
- Sylvain Champonnois, Capital Fund Management, Paris, France.
- Symeon Chouvardas, Independent Researcher.
- Zhaoyu Zhang, Department of Mathematics, USC.