

MARCOS LÓPEZ DE PRADO

ADVANCES
in
FINANCIAL
MACHINE
LEARNING

WILEY

Praise for *Advances in Financial Machine Learning*

In his new book *Advances in Financial Machine Learning*, noted financial scholar Marcos López de Prado strikes a well-aimed karate chop at the naive and often statistically overfit techniques that are so prevalent in the financial world today. He points out that not only are business-as-usual approaches largely impotent in today's high-tech finance, but in many cases they are actually prone to lose money. But López de Prado does more than just expose the mathematical and statistical sins of the finance world. Instead, he offers a technically sound roadmap for finance professionals to join the wave of machine learning. What is particularly refreshing is the author's empirical approach—his focus is on real-world data analysis, not on purely theoretical methods that may look pretty on paper but which, in many cases, are largely ineffective in practice. The book is geared to finance professionals who are already familiar with statistical data analysis techniques, but it is well worth the effort for those who want to do real state-of-the-art work in the field."

Dr. David H. Bailey, former Complex Systems Lead,
Lawrence Berkeley National Laboratory. Co-discoverer of the
BBP spigot algorithm

"Finance has evolved from a compendium of heuristics based on historical financial statements to a highly sophisticated scientific discipline relying on computer farms to analyze massive data streams in real time. The recent highly impressive advances in machine learning (ML) are fraught with both promise and peril when applied to modern finance. While finance offers up the nonlinearities and large data sets upon which ML thrives, it also offers up noisy data and the human element which presently lie beyond the scope of standard ML techniques. To err is human, but if you really want to f**k things up, use a computer. Against this background, Dr. López de Prado has written the first comprehensive book describing the application of modern ML to financial modeling. The book blends the latest technological developments in ML with critical life lessons learned from the author's decades of financial experience in leading academic and industrial institutions. I highly recommend this exciting book to both prospective students of financial ML and the professors and supervisors who teach and guide them."

Prof. Peter Carr, Chair of the Finance and Risk Engineering
Department, NYU Tandon School of Engineering

"Marcos is a visionary who works tirelessly to advance the finance field. His writing is comprehensive and masterfully connects the theory to the application. It is not often you find a book that can cross that divide. This book is an essential read for both practitioners and technologists working on solutions for the investment community."

Landon Downs, President and Cofounder, 1QBit

"Academics who want to understand modern investment management need to read this book. In it, Marcos López de Prado explains how portfolio managers use machine learning to derive, test, and employ trading strategies. He does this from a very unusual combination of an academic perspective and extensive experience in industry, allowing him to both explain in detail what happens in industry and to explain

how it works. I suspect that some readers will find parts of the book that they do not understand or that they disagree with, but everyone interested in understanding the application of machine learning to finance will benefit from reading this book.”

Prof. David Easley, Cornell University. Chair of the
NASDAQ-OMX Economic Advisory Board

“For many decades, finance has relied on overly simplistic statistical techniques to identify patterns in data. Machine learning promises to change that by allowing researchers to use modern nonlinear and highly dimensional techniques, similar to those used in scientific fields like DNA analysis and astrophysics. At the same time, applying those machine learning algorithms to model financial problems would be dangerous. Financial problems require very distinct machine learning solutions. Dr. López de Prado’s book is the first one to characterize what makes standard machine learning tools fail when applied to the field of finance, and the first one to provide practical solutions to unique challenges faced by asset managers. Everyone who wants to understand the future of finance should read this book.”

Prof. Frank Fabozzi, EDHEC Business School. Editor of
The Journal of Portfolio Management

“This is a welcome departure from the knowledge hoarding that plagues quantitative finance. López de Prado defines for all readers the next era of finance: industrial scale scientific research powered by machines.”

John Fawcett, Founder and CEO, Quantopian

“Marcos has assembled in one place an invaluable set of lessons and techniques for practitioners seeking to deploy machine learning techniques in finance. If machine learning is a new and potentially powerful weapon in the arsenal of quantitative finance, Marcos’s insightful book is laden with useful advice to help keep a curious practitioner from going down any number of blind alleys, or shooting oneself in the foot.”

Ross Garon, Head of Cubist Systematic Strategies. Managing
Director, Point72 Asset Management

“The first wave of quantitative innovation in finance was led by Markowitz optimization. Machine Learning is the second wave, and it will touch every aspect of finance. López de Prado’s *Advances in Financial Machine Learning* is essential for readers who want to be ahead of the technology rather than being replaced by it.”

Prof. Campbell Harvey, Duke University. Former President of
the American Finance Association

“The complexity inherent to financial systems justifies the application of sophisticated mathematical techniques. *Advances in Financial Machine Learning* is an exciting book that unravels a complex subject in clear terms. I wholeheartedly recommend this book to anyone interested in the future of quantitative investments.”

Prof. John C. Hull, University of Toronto. Author of
Options, Futures, and other Derivatives

“Prado’s book clearly illustrates how fast this world is moving, and how deep you need to dive if you are to excel and deliver top of the range solutions and above the curve performing algorithms... Prado’s book is clearly at the bleeding edge of the machine learning world.”

Irish Tech News

“Financial data is special for a key reason: The markets have only one past. There is no ‘control group’, and you have to wait for true out-of-sample data. Consequently, it is easy to fool yourself, and with the march of Moore’s Law and the new machine learning, it’s easier than ever. López de Prado explains how to avoid falling for these common mistakes. This is an excellent book for anyone working, or hoping to work, in computerized investment and trading.”

Dr. David J. Leinweber, Former Managing Director, First Quadrant.
Author of *Nerds on Wall Street: Math, Machines and Wired Markets*

“In his new book, Dr. López de Prado demonstrates that financial machine learning is more than standard machine learning applied to financial datasets. It is an important field of research in its own right. It requires the development of new mathematical tools and approaches, needed to address the nuances of financial datasets. I strongly recommend this book to anyone who wishes to move beyond the standard Econometric toolkit.”

Dr. Richard R. Lindsey, Managing Partner, Windham Capital Management.
Former Chief Economist, U.S. Securities and Exchange Commission

“Dr. Lopez de Prado, a well-known scholar and an accomplished portfolio manager who has made several important contributions to the literature on machine learning (ML) in finance, has produced a comprehensive and innovative book on the subject. He has illuminated numerous pitfalls awaiting anyone who wishes to use ML in earnest, and he has provided much needed blueprints for doing it successfully. This timely book, offering a good balance of theoretical and applied findings, is a must for academics and practitioners alike.”

Prof. Alexander Lipton, Connection Science Fellow, Massachusetts Institute of Technology. *Risk’s* Quant of the Year (2000)

“How does one make sense of today’s financial markets in which complex algorithms route orders, financial data is voluminous, and trading speeds are measured in nanoseconds? In this important book, Marcos López de Prado sets out a new paradigm for investment management built on machine learning. Far from being a “black box” technique, this book clearly explains the tools and process of financial machine learning. For academics and practitioners alike, this book fills an important gap in our understanding of investment management in the machine age.”

Prof. Maureen O’Hara, Cornell University. Former President of the American Finance Association

“Marcos López de Prado has produced an extremely timely and important book on machine learning. The author’s academic and professional first-rate credentials shine through the pages of this book—indeed, I could think of few, if any, authors better suited to explaining both the theoretical and the practical aspects of this new and (for most) unfamiliar subject. Both novices and experienced professionals will find insightful ideas, and will understand how the subject can be applied in novel and useful ways. The Python code will give the novice readers a running start and will allow them to gain quickly a hands-on appreciation of the subject. Destined to become a classic in this rapidly burgeoning field.”

Prof. Riccardo Rebonato, EDHEC Business School. Former
Global Head of Rates and FX Analytics at PIMCO

“A tour de force on practical aspects of machine learning in finance, brimming with ideas on how to employ cutting-edge techniques, such as fractional differentiation and quantum computers, to gain insight and competitive advantage. A useful volume for finance and machine learning practitioners alike.”

Dr. Collin P. Williams, Head of Research, D-Wave Systems

Advances in Financial Machine Learning

Advances in Financial Machine Learning

MARCOS LÓPEZ DE PRADO

WILEY

Cover image: © Erikona/Getty Images

Cover design: Wiley

Copyright © 2018 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. The views expressed in this book are the author's and do not necessarily reflect those of the organizations he is affiliated with.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993, or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

ISBN 978-1-119-48208-6 (Hardcover)

ISBN 978-1-119-48211-6 (ePDF)

ISBN 978-1-119-48210-9 (ePub)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Dedicated to the memory of my coauthor and friend,
Professor Jonathan M. Borwein, FRSC, FAAAS,
FBAS, FAustMS, FAA, FAMS, FRSNSW
(1951–2016)

There are very few things which we know, which are not capable of being reduced to a mathematical reasoning. And when they cannot, it's a sign our knowledge of them is very small and confused. Where a mathematical reasoning can be had, it's as great a folly to make use of any other, as to grope for a thing in the dark, when you have a candle standing by you.

—*Of the Laws of Chance*, Preface (1692)
John Arbuthnot (1667–1735)

Contents

About the Author	xxi
PREAMBLE	1
1 Financial Machine Learning as a Distinct Subject	3
1.1 Motivation, 3	
1.2 The Main Reason Financial Machine Learning Projects Usually Fail, 4	
1.2.1 The Sisyphus Paradigm, 4	
1.2.2 The Meta-Strategy Paradigm, 5	
1.3 Book Structure, 6	
1.3.1 Structure by Production Chain, 6	
1.3.2 Structure by Strategy Component, 9	
1.3.3 Structure by Common Pitfall, 12	
1.4 Target Audience, 12	
1.5 Requisites, 13	
1.6 FAQs, 14	
1.7 Acknowledgments, 18	
Exercises, 19	
References, 20	
Bibliography, 20	
PART 1 DATA ANALYSIS	21
2 Financial Data Structures	23
2.1 Motivation, 23	

2.2	Essential Types of Financial Data, 23
2.2.1	Fundamental Data, 23
2.2.2	Market Data, 24
2.2.3	Analytics, 25
2.2.4	Alternative Data, 25
2.3	Bars, 25
2.3.1	Standard Bars, 26
2.3.2	Information-Driven Bars, 29
2.4	Dealing with Multi-Product Series, 32
2.4.1	The ETF Trick, 33
2.4.2	PCA Weights, 35
2.4.3	Single Future Roll, 36
2.5	Sampling Features, 38
2.5.1	Sampling for Reduction, 38
2.5.2	Event-Based Sampling, 38
	Exercises, 40
	References, 41

3 Labeling 43

3.1	Motivation, 43
3.2	The Fixed-Time Horizon Method, 43
3.3	Computing Dynamic Thresholds, 44
3.4	The Triple-Barrier Method, 45
3.5	Learning Side and Size, 48
3.6	Meta-Labeling, 50
3.7	How to Use Meta-Labeling, 51
3.8	The Quantamental Way, 53
3.9	Dropping Unnecessary Labels, 54
	Exercises, 55
	Bibliography, 56

4 Sample Weights 59

4.1	Motivation, 59
4.2	Overlapping Outcomes, 59
4.3	Number of Concurrent Labels, 60
4.4	Average Uniqueness of a Label, 61
4.5	Bagging Classifiers and Uniqueness, 62
4.5.1	Sequential Bootstrap, 63
4.5.2	Implementation of Sequential Bootstrap, 64

4.5.3	A Numerical Example, 65	
4.5.4	Monte Carlo Experiments, 66	
4.6	Return Attribution, 68	
4.7	Time Decay, 70	
4.8	Class Weights, 71	
	Exercises, 72	
	References, 73	
	Bibliography, 73	
5	Fractionally Differentiated Features	75
5.1	Motivation, 75	
5.2	The Stationarity vs. Memory Dilemma, 75	
5.3	Literature Review, 76	
5.4	The Method, 77	
5.4.1	Long Memory, 77	
5.4.2	Iterative Estimation, 78	
5.4.3	Convergence, 80	
5.5	Implementation, 80	
5.5.1	Expanding Window, 80	
5.5.2	Fixed-Width Window Fracdiff, 82	
5.6	Stationarity with Maximum Memory Preservation, 84	
5.7	Conclusion, 88	
	Exercises, 88	
	References, 89	
	Bibliography, 89	
PART 2	MODELLING	91
6	Ensemble Methods	93
6.1	Motivation, 93	
6.2	The Three Sources of Errors, 93	
6.3	Bootstrap Aggregation, 94	
6.3.1	Variance Reduction, 94	
6.3.2	Improved Accuracy, 96	
6.3.3	Observation Redundancy, 97	
6.4	Random Forest, 98	
6.5	Boosting, 99	

6.6	Bagging vs. Boosting in Finance, 100	
6.7	Bagging for Scalability, 101	
	Exercises, 101	
	References, 102	
	Bibliography, 102	
7	Cross-Validation in Finance	103
7.1	Motivation, 103	
7.2	The Goal of Cross-Validation, 103	
7.3	Why K-Fold CV Fails in Finance, 104	
7.4	A Solution: Purged K-Fold CV, 105	
7.4.1	Purging the Training Set, 105	
7.4.2	Embargo, 107	
7.4.3	The Purged K-Fold Class, 108	
7.5	Bugs in Sklearn's Cross-Validation, 109	
	Exercises, 110	
	Bibliography, 111	
8	Feature Importance	113
8.1	Motivation, 113	
8.2	The Importance of Feature Importance, 113	
8.3	Feature Importance with Substitution Effects, 114	
8.3.1	Mean Decrease Impurity, 114	
8.3.2	Mean Decrease Accuracy, 116	
8.4	Feature Importance without Substitution Effects, 117	
8.4.1	Single Feature Importance, 117	
8.4.2	Orthogonal Features, 118	
8.5	Parallelized vs. Stacked Feature Importance, 121	
8.6	Experiments with Synthetic Data, 122	
	Exercises, 127	
	References, 127	
9	Hyper-Parameter Tuning with Cross-Validation	129
9.1	Motivation, 129	
9.2	Grid Search Cross-Validation, 129	
9.3	Randomized Search Cross-Validation, 131	
9.3.1	Log-Uniform Distribution, 132	
9.4	Scoring and Hyper-parameter Tuning, 134	

Exercises, 135
References, 136
Bibliography, 137

PART 3 BACKTESTING **139**

10 Bet Sizing **141**

10.1 Motivation, 141
10.2 Strategy-Independent Bet Sizing Approaches, 141
10.3 Bet Sizing from Predicted Probabilities, 142
10.4 Averaging Active Bets, 144
10.5 Size Discretization, 144
10.6 Dynamic Bet Sizes and Limit Prices, 145
Exercises, 148
References, 149
Bibliography, 149

11 The Dangers of Backtesting **151**

11.1 Motivation, 151
11.2 Mission Impossible: The Flawless Backtest, 151
11.3 Even If Your Backtest Is Flawless, It Is Probably Wrong, 152
11.4 Backtesting Is Not a Research Tool, 153
11.5 A Few General Recommendations, 153
11.6 Strategy Selection, 155
Exercises, 158
References, 158
Bibliography, 159

12 Backtesting through Cross-Validation **161**

12.1 Motivation, 161
12.2 The Walk-Forward Method, 161
 12.2.1 Pitfalls of the Walk-Forward Method, 162
12.3 The Cross-Validation Method, 162
12.4 The Combinatorial Purged Cross-Validation Method, 163
 12.4.1 Combinatorial Splits, 164
 12.4.2 The Combinatorial Purged Cross-Validation
 Backtesting Algorithm, 165
 12.4.3 A Few Examples, 165

12.5 How Combinatorial Purged Cross-Validation Addresses Backtest Overfitting, 166

Exercises, 167

References, 168

13 Backtesting on Synthetic Data 169

13.1 Motivation, 169

13.2 Trading Rules, 169

13.3 The Problem, 170

13.4 Our Framework, 172

13.5 Numerical Determination of Optimal Trading Rules, 173

13.5.1 The Algorithm, 173

13.5.2 Implementation, 174

13.6 Experimental Results, 176

13.6.1 Cases with Zero Long-Run Equilibrium, 177

13.6.2 Cases with Positive Long-Run Equilibrium, 180

13.6.3 Cases with Negative Long-Run Equilibrium, 182

13.7 Conclusion, 192

Exercises, 192

References, 193

14 Backtest Statistics 195

14.1 Motivation, 195

14.2 Types of Backtest Statistics, 195

14.3 General Characteristics, 196

14.4 Performance, 198

14.4.1 Time-Weighted Rate of Return, 198

14.5 Runs, 199

14.5.1 Returns Concentration, 199

14.5.2 Drawdown and Time under Water, 201

14.5.3 Runs Statistics for Performance Evaluation, 201

14.6 Implementation Shortfall, 202

14.7 Efficiency, 203

14.7.1 The Sharpe Ratio, 203

14.7.2 The Probabilistic Sharpe Ratio, 203

14.7.3 The Deflated Sharpe Ratio, 204

14.7.4 Efficiency Statistics, 205

14.8 Classification Scores, 206

14.9 Attribution, 207

Exercises, 208
References, 209
Bibliography, 209

15 Understanding Strategy Risk 211

15.1 Motivation, 211
15.2 Symmetric Payouts, 211
15.3 Asymmetric Payouts, 213
15.4 The Probability of Strategy Failure, 216
 15.4.1 Algorithm, 217
 15.4.2 Implementation, 217
Exercises, 219
References, 220

16 Machine Learning Asset Allocation 221

16.1 Motivation, 221
16.2 The Problem with Convex Portfolio Optimization, 221
16.3 Markowitz's Curse, 222
16.4 From Geometric to Hierarchical Relationships, 223
 16.4.1 Tree Clustering, 224
 16.4.2 Quasi-Diagonalization, 229
 16.4.3 Recursive Bisection, 229
16.5 A Numerical Example, 231
16.6 Out-of-Sample Monte Carlo Simulations, 234
16.7 Further Research, 236
16.8 Conclusion, 238
Appendices, 239
 16.A.1 Correlation-based Metric, 239
 16.A.2 Inverse Variance Allocation, 239
 16.A.3 Reproducing the Numerical Example, 240
 16.A.4 Reproducing the Monte Carlo Experiment, 242
Exercises, 244
References, 245

PART 4 USEFUL FINANCIAL FEATURES 247

17 Structural Breaks 249

17.1 Motivation, 249
17.2 Types of Structural Break Tests, 249

- 17.3 CUSUM Tests, 250
 - 17.3.1 Brown-Durbin-Evans CUSUM Test on Recursive Residuals, 250
 - 17.3.2 Chu-Stinchcombe-White CUSUM Test on Levels, 251
- 17.4 Explosiveness Tests, 251
 - 17.4.1 Chow-Type Dickey-Fuller Test, 251
 - 17.4.2 Supremum Augmented Dickey-Fuller, 252
 - 17.4.3 Sub- and Super-Martingale Tests, 259
- Exercises, 261
- References, 261

18 Entropy Features 263

- 18.1 Motivation, 263
- 18.2 Shannon's Entropy, 263
- 18.3 The Plug-in (or Maximum Likelihood) Estimator, 264
- 18.4 Lempel-Ziv Estimators, 265
- 18.5 Encoding Schemes, 269
 - 18.5.1 Binary Encoding, 270
 - 18.5.2 Quantile Encoding, 270
 - 18.5.3 Sigma Encoding, 270
- 18.6 Entropy of a Gaussian Process, 271
- 18.7 Entropy and the Generalized Mean, 271
- 18.8 A Few Financial Applications of Entropy, 275
 - 18.8.1 Market Efficiency, 275
 - 18.8.2 Maximum Entropy Generation, 275
 - 18.8.3 Portfolio Concentration, 275
 - 18.8.4 Market Microstructure, 276
- Exercises, 277
- References, 278
- Bibliography, 279

19 Microstructural Features 281

- 19.1 Motivation, 281
- 19.2 Review of the Literature, 281
- 19.3 First Generation: Price Sequences, 282
 - 19.3.1 The Tick Rule, 282
 - 19.3.2 The Roll Model, 282

- 19.3.3 High-Low Volatility Estimator, 283
- 19.3.4 Corwin and Schultz, 284
- 19.4 Second Generation: Strategic Trade Models, 286
 - 19.4.1 Kyle's Lambda, 286
 - 19.4.2 Amihud's Lambda, 288
 - 19.4.3 Hasbrouck's Lambda, 289
- 19.5 Third Generation: Sequential Trade Models, 290
 - 19.5.1 Probability of Information-based Trading, 290
 - 19.5.2 Volume-Synchronized Probability of Informed Trading, 292
- 19.6 Additional Features from Microstructural Datasets, 293
 - 19.6.1 Distribution of Order Sizes, 293
 - 19.6.2 Cancellation Rates, Limit Orders, Market Orders, 293
 - 19.6.3 Time-Weighted Average Price Execution Algorithms, 294
 - 19.6.4 Options Markets, 295
 - 19.6.5 Serial Correlation of Signed Order Flow, 295
- 19.7 What Is Microstructural Information?, 295
- Exercises, 296
- References, 298

PART 5 HIGH-PERFORMANCE COMPUTING RECIPES 301

20 Multiprocessing and Vectorization 303

- 20.1 Motivation, 303
- 20.2 Vectorization Example, 303
- 20.3 Single-Thread vs. Multithreading vs. Multiprocessing, 304
- 20.4 Atoms and Molecules, 306
 - 20.4.1 Linear Partitions, 306
 - 20.4.2 Two-Nested Loops Partitions, 307
- 20.5 Multiprocessing Engines, 309
 - 20.5.1 Preparing the Jobs, 309
 - 20.5.2 Asynchronous Calls, 311
 - 20.5.3 Unwrapping the Callback, 312
 - 20.5.4 Pickle/Unpickle Objects, 313
 - 20.5.5 Output Reduction, 313
- 20.6 Multiprocessing Example, 315
- Exercises, 316

Reference, 317
Bibliography, 317

21 Brute Force and Quantum Computers 319

21.1 Motivation, 319
21.2 Combinatorial Optimization, 319
21.3 The Objective Function, 320
21.4 The Problem, 321
21.5 An Integer Optimization Approach, 321
 21.5.1 Pigeonhole Partitions, 321
 21.5.2 Feasible Static Solutions, 323
 21.5.3 Evaluating Trajectories, 323
21.6 A Numerical Example, 325
 21.6.1 Random Matrices, 325
 21.6.2 Static Solution, 326
 21.6.3 Dynamic Solution, 327
Exercises, 327
References, 328

22 High-Performance Computational Intelligence and Forecasting Technologies 329

Kesheng Wu and Horst D. Simon

22.1 Motivation, 329
22.2 Regulatory Response to the Flash Crash of 2010, 329
22.3 Background, 330
22.4 HPC Hardware, 331
22.5 HPC Software, 335
 22.5.1 Message Passing Interface, 335
 22.5.2 Hierarchical Data Format 5, 336
 22.5.3 *In Situ* Processing, 336
 22.5.4 Convergence, 337
22.6 Use Cases, 337
 22.6.1 Supernova Hunting, 337
 22.6.2 Blobs in Fusion Plasma, 338
 22.6.3 Intraday Peak Electricity Usage, 340
 22.6.4 The Flash Crash of 2010, 341
 22.6.5 Volume-synchronized Probability of Informed Trading Calibration, 346

22.6.6	Revealing High Frequency Events with Non-uniform Fast Fourier Transform, 347
22.7	Summary and Call for Participation, 349
22.8	Acknowledgments, 350
	References, 350

About the Author

Prof. Marcos López de Prado is the CIO of True Positive Technologies (TPT), and Professor of Practice at Cornell University's School of Engineering. He has over 20 years of experience developing investment strategies with the help of machine learning algorithms and supercomputers. Marcos launched TPT after he sold some of his patents to AQR Capital Management, where he was a principal and AQR's first head of machine learning. He also founded and led Guggenheim Partners' Quantitative Investment Strategies business, where he managed up to \$13 billion in assets, and delivered an audited risk-adjusted return (information ratio) of 2.3.

Concurrently with the management of investments, between 2011 and 2018 Marcos was a research fellow at Lawrence Berkeley National Laboratory (U.S. Department of Energy, Office of Science). He has published dozens of scientific articles on machine learning and supercomputing in the leading academic journals, is a founding co-editor of The Journal of Financial Data Science, and SSRN ranks him as the most-read author in economics.

Marcos earned a PhD in financial economics (2003), a second PhD in mathematical finance (2011) from Universidad Complutense de Madrid, and is a recipient of Spain's National Award for Academic Excellence (1999). He completed his post-doctoral research at Harvard University and Cornell University, where he is a faculty member. In 2019, Marcos received the 'Quant of the Year Award' from The Journal of Portfolio Management.

For more information, visit www.QuantResearch.org

Preamble

Chapter 1: Financial Machine Learning as a Distinct Subject, 3

