Financial Econometrics
Models and Methods

This is a thorough exploration of the models and methods of financial econometrics by one of the world's leading financial econometricians and is for students in economics, finance, statistics, mathematics, and engineering who are interested in financial applications.

Based on courses taught around the world, the up-to-date content covers developments in econometrics and finance over the past twenty years while ensuring a solid grounding in the fundamental principles of the field.

Care has been taken to link theory and application to provide real-world context for students, worked exercises and empirical examples have also been included to make sure complicated concepts are solidly explained and understood.

Oliver Linton is a fellow of Trinity College and is Professor of Political Economy at Cambridge University. Formerly, Professor of Econometrics at the London School of Economics and Professor of Economics at Yale University. He obtained his PhD in Economics from the University of California at Berkeley in 1991. He has written more than a hundred articles on econometrics, statistics, and empirical finance. In 2015 he was a recipient of the Humboldt Research Award of the Alexander von Humboldt Foundation. He has been a Co-editor at the Journal of Econometrics since 2014. He is a Fellow of: the Econometric Society, the Institute of Mathematical Statistics, the Society for Financial Econometrics, the British Academy, and the International Foundation of Applied Econometrics. He was a lead expert in the U.K. Government Office for Science Foresight project: "The future of Computer Trading in Financial Markets", which published in 2012. He has appeared as an expert witness for the FSA and the FCA in several cases involving market manipulation.

Financial Econometrics
Models and Methods

OLIVER LINTON

University of Cambridge

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, 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

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.
It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107177154
DOI: 10.1017/9781316819302

First published 2019

Printed in the United Kingdom by TJ International Ltd, Padstow Cornwall

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

ISBN 978-1-107-17715-4 Hardback
ISBN 978-1-316-63033-4 Paperback

Additional resources for this publication at www.cambridge.org/linton

To my wife, Jianghong Song.

Short Contents

List of Figures — *page* xv
List of Tables — xix
Preface — xxi
Acknowledgments — xxv
Notation and Conventions — xxvii

1 Introduction and Background — 1
2 Econometric Background — 55
3 Return Predictability and the Efficient Markets Hypothesis — 75
4 Robust Tests and Tests of Nonlinear Predictability of Returns — 134
5 Empirical Market Microstructure — 152
6 Event Study Analysis — 201
7 Portfolio Choice and Testing the Capital Asset Pricing Model — 238
8 Multifactor Pricing Models — 279
9 Present Value Relations — 314
10 Intertemporal Equilibrium Pricing — 337
11 Volatility — 358
12 Continuous Time Processes — 422
13 Yield Curve — 463
14 Risk Management and Tail Estimation — 476
15 Exercises and Complements — 497
16 Appendix — 524

Bibliography — 533
Index — 553

Contents

List of Figures *page* xv
List of Tables xix
Preface xxi
Acknowledgments xxv
Notation and Conventions xxvii

1 Introduction and Background 1

1.1 Why Do We Have Financial Markets? 1
1.2 Classification of Financial Markets 3
1.3 Types of Markets and Trading 8
1.4 Financial Returns 12
1.5 Risk Aversion 27
1.6 Mean Variance Portfolio Analysis 39
1.7 Capital Asset Pricing Model 45
1.8 Arbitrage Pricing Theory 49
1.9 Appendix 53

2 Econometric Background 55

2.1 Linear Regression 55
2.2 Time Series 61

3 Return Predictability and the Efficient Markets Hypothesis 75

3.1 Efficient Markets Hypothesis 75
3.2 The Random Walk Model for Prices 81
3.3 Testing of Linear Weak Form Predictability 85
3.4 Testing under More General Conditions than rw1 106
3.5 Some Alternative Hypotheses 118
3.6 Empirical Evidence regarding Linear Predictability based on Variance Ratios 121
3.7 Trading Strategy Based Evidence 124
3.8 Regression Based Tests of Semi-Strong Form Predictability 129
3.9 Summary of Chapter 132

4 Robust Tests and Tests of Nonlinear Predictability of Returns 134

4.1 Robust Tests 134
4.2 Nonlinear Predictability and Nonparametric Autoregression 146
4.3 Further Empirical Evidence on Semistrong and Strong Form EMH 148
4.4 Explanations for Predictability 150
4.5 Summary of Chapter 151

5 Empirical Market Microstructure 152

5.1 Stale Prices 152
5.2 Discrete Prices and Quantities 164
5.3 Bid, Ask, and Transaction Prices 168
5.4 What Determines the Bid–Ask Spread? 173
5.5 Strategic Trade Models 183
5.6 Electronic Markets 187
5.7 Summary of Chapter 196
5.8 Appendix 196

6 Event Study Analysis 201

6.1 Some Applications 201
6.2 Basic Structure of an Event Study 202
6.3 Regression Framework 218
6.4 Nonparametric and Robust Tests 221
6.5 Cross-sectional Regressions 222
6.6 Time Series Heteroskedasticity 223
6.7 Panel Regression for Estimating Treatment Effects 224
6.8 Matching Approach 228
6.9 Stock Splits 229
6.10 Summary of Chapter 237
6.11 Appendix 237

7 Portfolio Choice and Testing the Capital Asset Pricing Model 238

7.1 Portfolio Choice 238
7.2 Testing the Capital Asset Pricing Model 241
7.3 Maximum Likelihood Estimation and Testing 246
7.4 Cross-sectional Regression Tests 259
7.5 Portfolio Grouping 264
7.6 Time Varying Model 268
7.7 Empirical Evidence on the CAPM 270
7.8 Summary of Chapter 273
7.9 Appendix 274

 Contents

8 Multifactor Pricing Models 279

8.1 Linear Factor Model 279
8.2 Diversification 280
8.3 Pervasive Factors 286
8.4 The Econometric Model 288
8.5 Multivariate Tests of the Multibeta Pricing Model with Observable Factors 290
8.6 Which Factors to Use? 293
8.7 Observable Characteristic Based Models 299
8.8 Statistical Factor Models 300
8.9 Testing the APT with Estimated Factors 312
8.10 The MacKinlay Critique 312
8.11 Summary of Chapter 312
8.12 Appendix 312

9 Present Value Relations 314

9.1 Fundamentals versus Bubbles 314
9.2 Present Value Relations 316
9.3 Rational Bubbles 319
9.4 Econometric Bubble Detection 321
9.5 Shiller Excess Volatility Tests 323
9.6 An Approximate Model of Log Returns 326
9.7 Predictive Regressions 329
9.8 Summary of Chapter 335
9.9 Appendix 336

10 Intertemporal Equilibrium Pricing 337

10.1 Dynamic Representative Agent Models 337
10.2 The Stochastic Discount Factor 338
10.3 The Consumption Capital Asset Pricing Model 339
10.4 The Equity Premium Puzzle and the Risk Free Rate Puzzle 345
10.5 Explanations for the Puzzles 346
10.6 Other Asset Pricing Approaches 353
10.7 Summary of Chapter 357

11 Volatility 358

11.1 Why is Volatility Important? 358
11.2 Implied Volatility from Option Prices 359
11.3 Intra Period Volatility 362
11.4 Cross-sectional Volatility 370
11.5 Empirical Studies 371

11.6 Discrete Time Series Models 374
11.7 Engle's ARCH Model 377
11.8 The GARCH Model 380
11.9 Asymmetric Volatility Models and Other Specifications 389
11.10 Mean and Variance Dynamics 392
11.11 Estimation of Parameters 395
11.12 Stochastic Volatility Models 402
11.13 Long Memory 404
11.14 Multivariate Models 407
11.15 Nonparametric and Semiparametric Models 412
11.16 Summary of Chapter 419
11.17 Appendix 419

12 Continuous Time Processes 422

12.1 Brownian Motion 422
12.2 Stochastic Integrals 427
12.3 Diffusion Processes 428
12.4 Estimation of Diffusion Models 436
12.5 Estimation of Quadratic Variation Volatility from High Frequency Data 450
12.6 Levy Processes 459
12.7 Summary of Chapter 462

13 Yield Curve 463

13.1 Discount Function, Yield Curve, and Forward Rates 463
13.2 Estimation of the Yield Curve from Coupon Bonds 464
13.3 Discrete Time Models of Bond Pricing 469
13.4 Arbitrage and Pricing Kernels 471
13.5 Summary of Chapter 475

14 Risk Management and Tail Estimation 476

14.1 Types of Risks 476
14.2 Value at Risk 477
14.3 Extreme Value Theory 479
14.4 A Semiparametric Model of Tail Thickness 482
14.5 Dynamic Models and VAR 487
14.6 The Multivariate Case 489
14.7 Coherent Risk Measures 492
14.8 Expected Shortfall 493
14.9 Black Swan Theory 494
14.10 Summary of Chapter 496

Cambridge University Press
978-1-107-17715-4 — Financial Econometrics
Oliver Linton
Frontmatter
More Information

15 Exercises and Complements 497

16 Appendix 524

16.1 Common Abbreviations 524
16.2 Two Inequalities 526
16.3 Signal Extraction 527
16.4 Lognormal Random Variables 528
16.5 Data Sources 529
16.6 A Short Introduction to Eviews 530

Bibliography 533
Index 553

List of Figures

1.1	Level of the S&P500 index	19
1.2	Daily return on the S&P500 daily index	23
1.3	Euro/dollar daily exchange rate	25
1.4	Return on the euro/dollar daily exchange rate	26
1.5	Daily Tbill Rates	26
1.6	Daily Federal Funds rate annualized	27
1.7	Price of West Texas Oil, monthly frequency	27
1.8	Daily return on West Texas Oil	28
3.1	Shows “Head and Shoulders” pattern in artificial dataset	79
3.2	Correlogram of S&P500 daily returns	92
3.3	Correlogram of S&P500 daily returns by decade	93
3.4	Correlogram of FTSE100 daily returns from 1984–2017	93
3.5	Correlogram of FTSE100 daily returns long horizon	94
3.6	ACF(1) of daily Dow stock returns against market capitalization	95
3.7	Average ACF of Dow stocks daily returns	97
3.8	Daily return on the Malaysian Ringgit	104
3.9	Variance ratio of S&P500 daily returns, 1950–2017	123
3.10	Variance ratio of FTSE100 daily returns, 1984–2017	123
3.11	ACF of the winner and loser stocks in sample	127
3.12	ACF of winner and loser stocks in and out of sample	128
3.13	ACF of the max versus max of the ACF	129
4.1	Cowles–Jones statistic with standard errors	139
4.2	Quantilogram of daily S&P500 returns	142
4.3	Fraction of positive returns on the stocks of the S&P500 index	143
4.4	The AD line on the daily S&P500 return	144
4.5	Length of daily runs on the S&P500	145
5.1	Prices and discretization	166
5.2	Histogram of integer remainders	166
5.3	Price changes on S&P500 Emini contract	170
5.4	Autocovariance of daily Dow	172
5.5	Efficient price volatility of the daily Dow	172
5.6	Price trajectory 1	180
5.7	Price trajectory 2	181
5.8	Daily Amihud illiquidity on S&P500	186
5.9	Daily Amihud illiquidity of S&P500 index since 2008	186
5.10	FTSE100 one minute data during nonfarm payroll announcement	191
5.11	Flash crash	192

5.12 Feedback loop 192
6.1 Shows examples of price trajectories around event at time 0 205
6.2 Dow Jones stock splits 232
6.3 Exxon stock splits CAR 233
6.4 Exxon splits average CAR 233
6.5 Exxon splits, longer window 234
6.6 Exxon splits average CAR over splits 234
6.7 Exxon splits, shorter window 235
6.8 Exxon splits, shorter window average CAR 235
7.1 Efficient frontier of Dow stocks 241
7.2 Skewness of Dow Jones returns by sampling frequency 245
7.3 Kurtosis of Dow Jones returns by frequency 245
7.4 Risk return relation 264
7.5 Time varying betas of IBM 269
7.6 Time varying alphas of IBM 269
7.7 SMB portfolio beta with market return 270
8.1 Quantiles of ordered (absolute) cross-correlations between S&P500 stocks 283
8.2 Quantiles of ordered (absolute) cross-correlations between S&P500 stocks idiosyncratic errors 284
8.3 Solnik curve of Dow stocks 285
8.4 GMV of Dow stocks 286
8.5 Rolling window correlation between the FTSE100 and FTSE250 287
8.6 Fama–French factors returns and implied prices 295
8.7 Variance ratios of Fama–French factors 297
8.8 US monthly CPI percentage change 298
8.9 Eigenvalues of $\hat{\Sigma}$ for daily S&P500 stocks, $N = 441$ and $T = 2732$ 309
8.10 Eigenvalues of $\hat{\Psi}$ for monthly S&P500 returns 309
8.11 Dominant principal component for monthly data 310
9.1 NASDAQ price level 315
9.2 Bitcoin price level from 2010–2017 315
9.3 Bubble time series 320
9.4 The logarithm of S&P500 index with trend line fitted from 1950 to 2017 325
9.5 Gross return on S&P500 333
9.6 Dividend yield on the S&P500 333
9.7 Rolling window ($\pm$20 years) R^2 of predictive regression 334
9.8 Rolling window ($\pm$20 years) slope coefficent of predictive regression 335
9.9 Overlapping daily 5 year returns on the S&P500 335
10.1 Growth rate of annual real PCE in 2009 dollars 346
10.2 Rolling window trailing 10 year gross nominal returns on the CRSP value weighted index 347
10.3 Distribution of the annual risk premium on the FF market factor from ten years of daily data 349
10.4 Quarterly US differenced log of PCE (seasonally adjusted) 349
10.5 Autocorrelation of quarterly US differenced log of PCE 350

10.6 Quarterly time series CAY_t 355
11.1 US commercial airline fatalities per mile 359
11.2 Daily VIX index 361
11.3 Histogram of the VIX index 361
11.4 ACF of the VIX 362
11.5 Annual volatility of the S&P500 364
11.6 Parkinson estimator of daily volatility 368
11.7 ACF of intraday volatility 368
11.8 Histogram of intraday volatility 369
11.9 Cross-section volatility 371
11.10 Correlogram of returns (first panel) and absolute returns (second panel) 375
11.11 Conditional standard deviation of daily returns 382
11.12 Standardized residuals 383
11.13 S&P500 daily return cross autocovariance $\text{cov}(Y_t^2, Y_{t-j}), j = -10, \ldots, 10$ 389
11.14 Comparison of the estimated news impact curves from GARCH(1,1) and GJR(1,1) for daily S&P500 returns 392
11.15 Density of S&P500 returns 413
11.16 Shows the nonparametrically estimated conditional comulants $cum_j(y_t|y_{t-k})$, for $j = 1, 2, 3, 4$ and $k = 1, \ldots, 50$ 415
11.17 Time varying cumulants of S&P500 returns 418
12.1 Crossing time density 425
12.2 ACF of daily federal funds rate 1954–2017 434
12.3 Conditional mean of exponential 435
12.4 Volatility signature plot 454
13.1 Time series of one month and ten year yields since 2000 470
13.2 Conditional cumulants of T-bill 474
13.3 Conditional cumulants of 10-year bonds 475
14.1 Estimated tail index of daily S&P500 returns 1950–2017 487
14.2 Expected shortfall 495
14.3 Daily S&P500 stock return events exceeding 6σ 496

List of Tables

1.1 Stock exchanges ranked by value traded 11
1.2 Descriptive statistics 1950–2017 23
1.3 Daily CRSP market returns descriptive statistics 24
1.4 Daily returns on individual stocks descriptive statistics, 1995–2017 25
3.1 Dow Jones stocks with market capitalization in January 2013 94
3.2 Parameter estimates of AR(22) model 104
3.3 Autocorrelation of squared returns and kurtosis of returns 113
3.4 The off-diagonal terms 113
3.5 Variance ratios for weekly small-size portfolio returns 114
3.6 Variance ratios for weekly medium-size portfolio returns 115
3.7 Variance ratios for weekly large-size portfolio returns 116
3.8 Day of the week effect 131
5.1 Tick size on the LSE 165
5.2a Limit order book example 188
5.2b Limit order book example 188
5.3 Round trip speed of the LSE 193
6.1 Recent stock splits on NASDAQ 230
6.2 Dow stocks split size distribution 232
7.1 Tangency and GMV portfolio weights for Dow Stocks 240
7.2 Market model estimates of Dow Stocks, daily data 247
7.3 Market model estimates of Dow Stocks, monthly data 248
8.1 Fama–French factors, mean and std 296
8.2 Fama–French factors correlation matrix 296
9.1 Dividend yield on the Dow stocks 336
10.1 Market risk premium 348
11.1 The FTSE100 top 20 most volatile days since 1984 369
11.2 The S&P500 top 20 most volatile days since 1960 370
11.3 Idiosyncratic volatility of Dow stocks 374
11.4 GARCH(1,1) parameter estimates 382
11.5 Estimated EGARCH model 391
11.6 Estimation of asymmetric GJR GARCH model 391
11.7 Estimates of GARCH in mean model 394
11.8 Correlation Matrix of estimated parameters from GARCH model 399
11.9 Daily GARCH in mean t-error 400
11.10 Estimated d by frequency 406
11.11 Estimated GARCH model by decade 417
13.1 Summary statistics of daily yields 471
13.2 Autocorrelation of daily yields 471

Preface

This work grew out of my teaching and research. It started from my considerable admiration of the seminal financial econometrics book Campbell, Lo, and MacKinlay (1997), henceforth CLM, and my teaching of that material to Master's students. I have kept along a similar line to that work in terms of selection of material and development, and have updated the material in several places. I have tried to adapt the material to a Master's audience by reducing the peerless literature review that is in that book and by amplifying some of the econometric discussions. I have included some theorems of varying degrees of rigor, meaning that I have not in every case specified all the required regularity conditions. I apologize for any upset this may cause, but the interests of time and space kept me from doing this. I hope the use of theorems can help to focus the material and make it easier to teach.

Financial econometrics has grown enormously as a discipline in the 20 years since CLM was published, and the range of authors engaged in its development has also increased. Computer scientists and so-called econo-physicists have taken an interest in the field and made major contributions to our understanding of financial markets. Mathematicians and statisticians have established rigorous proofs of many important results for our field and developed new tools for analyzing and understanding big financial data. The academic landscape has also become more international with a big expansion in the study of finance and financial statistics in China.

Data is the plural of anecdote, and happily there has recently been a massive increase in the amount of data, which has in itself stimulated research. Simultaneously computer power, both hardware and software, has increased substantially, allowing the analysis of much larger and more complex datasets than was previously possible. Econometric methodology has also expanded in many relevant areas, notably: volatility measurement and modelling; bounteous variate statistics where the size of the cross-section and time series is large; tools for extreme risk management; and quantifying causal effects. Despite the improvement in tools, the Global Financial Crisis following 2008, led to some skepticism about the value of economic theory and econometrics in predicting the armageddon that then ensued and in managing its aftermath, but this has in turn led to development of more relevant methodology, and hopefully some humility. Has all this attention and development improved our understanding of how financial markets work and how to change them for the better? I think so, but the subject is far from complete or satisfactory. The quantity and quality of empirical work and its presentation has improved substantially over time, but this has to some extent just made it harder for the reader to tell where the "bodies are buried" and what is the permanent value added of a particular study. Statistical methods are vital in this endeavor and in many cases their contribution is to provide a framework for acknowledging the limitations of our knowledge. Some of the empirical regularities that were cited in CLM have not stood the test of time, or at least they had to be qualified and their limitations acknowledged. This makes it hard to give a clear and simple picture that one could explain to a teenager.

This book is intended to be used as a text for advanced undergraduates and graduate students in economics, finance, statistics, mathematics, and engineering who are interested in financial applications and the methodology needed to understand and interpret those applications. I have taught part of this material at Yale University, the London School of Economics, the University of Cambridge, Humboldt University, Shandong University, SHUFE, and Renmin University. I provide two introductory chapters on financial institutions, financial economics, and econometrics, which provide some essential background that will be drawn on in later chapters. The main material begins with the efficient markets hypothesis and the predictability of asset returns, which is a central question of financial economics. I provide some updates on the empirical regularities found in CLM. I then provide a separate chapter on robust methods, which are important because large observations such as the October 1987 stock market crash can have an undue influence on estimation and hypothesis tests. I then cover some topics in market microstructure, which is a very active area struggling with new developments in market structure, technology, and regulation. I cover the classical topics of stale and discrete prices as well as the models for adverse selection and market impact that form the language of this area. I use some matrix algebra in Chapters 6 to 8; it is hard if not impossible to present this material cogently without these tools, and to understand big data without the basics of linear algebra is like trying to assemble an IKEA cupboard in the dark without an Allen key. Luckily there are many excellent books that provide suitable coverage of the necessary material, for example Linton (2016). The material on event studies has been expanded to include recent work coming from microeconometrics that can be used to evaluate the effects of policy changes on outcomes other than stock returns. I also include the standard methodology based on the market model but provide a more detailed discussion of the effects of overlapping estimation and event windows. I cover the CAPM next with some discussion of portfolio grouping methods and the two main testing methodologies. The chapter on multifactor models covers the main approaches including the Fama–French approach, which was still in its early days when CLM was published but is now one of the dominant methodologies. I also cover statistical factor models and characteristic based models. The next two chapters consider some intertemporal asset pricing material and the associated econometrics such as predictive regressions, volatility tests, and generalized method of moments (GMM). The chapter on volatility describes the three main approaches to measuring and defining volatility based on option prices, high frequency data, and dynamic time series modelling. The chapter on continuous time models develops some of this material further, but also introduces the models and methods widely used in this area. I cover some material on yield curve estimation and its application to pricing. The final chapter considers risk management including extreme value theory and dynamic modelling approaches. I use a number of datasets of different frequencies in the book to illustrate various points and to report on the main features of financial data. As usual, results can vary.

I have included short biographies of authors who have influenced me regarding financial econometrics particularly. My prediction is that at least one of them will win the Nobel Prize in economics.

The book contains many terms in bold face, which can then be investigated further by internet search. I have provided some computer code in different languages, such as `MATLAB`, `GAUSS`, and `R`, pertinent to various parts of the book. I am told that `STATA` can accomplish many things, but I have yet to see the light. A lot of analysis can be done by `EVIEWS`, and I provide a short introduction to its use in handling daily stock return data. I also provide a link to a number of data sources that can help with student projects and the like.

Acknowledgments

No man is an island, and I would like to thank the people who have had an influence on my research career. Even though they may not have been directly involved in the development of this book, I have drawn heavily on my interactions with them throughout this project. In roughly chronological order they are: Haya Friedman, Jan Magnus, Tom Rothenberg, Jens Perch Nielsen, Peter Bickel, Greg Connor, Peter Robinson, Wolfgang Härdle, Enno Mammen, Per Mykland, Miguel Delgado, Neil Shephard, Don Andrews, Peter Phillips, Xiaohong Chen, Arthur Lewbel, Yoon-Jae Whang, Zhijie Xiao, Christian Hafner, Frank Diebold, Eric Ghysels, Jean-Marie Dufour, Haim Levy, Andrew Patton, Jiti Gao, Jon Danielsson, Jean-Pierre Zigrand, Alexei Onatskiy, Andrew Harvey, Andrew Chesher, Hashem Pesaran, Richard Smith, and Mark Salmon. I would like also to thank my current and former PhD and Master's students who have contributed to the development of this book. I would like to thank Greg Connor, Katja Smetanina, and anonymous referees for comments.

Notation and Conventions

In this book I use the dating convention yyyymmddhhmmss. A visiting time traveller would surely prefer to know the year before the month or day, although he might ask why we have we have chosen 24 hours in a day and 60 minutes in an hour, etc. I use $\xrightarrow{P}$ to denote convergence in probability and $\Longrightarrow$ to denote convergence in distribution. $\log(x)$ is the natural logarithm unless otherwise stated. $\mathbb{R}$ is the set of real numbers, $'$ denotes differentiation, and $^{\intercal}$ denotes matrix transpose. We use $X_n = O(n)$ to mean that X_n/n is bounded for a deterministic sequence X_n. I use $\simeq$ to generically denote an approximation and $\sim$ to mean to have the same distribution as. I do not have a bracketing convention like some journals, but I do have a preference for round curved brackets over square ones. Dollars or $ are US unless otherwise specified.