

# Springer Texts in Statistics

*Series Editors:*

R. DeVeaux

S.E. Fienberg

I. Olkin

More information about this series at <http://www.springer.com/series/417>



David Ruppert • David S. Matteson

# Statistics and Data Analysis for Financial Engineering

with R examples

Second Edition

 Springer

David Ruppert  
Department of Statistical  
Science and School of ORIE  
Cornell University  
Ithaca, NY, USA

David S. Matteson  
Department of Statistical Science  
Department of Social Statistics  
Cornell University  
Ithaca, NY, USA

ISSN 1431-875X  
Springer Texts in Statistics  
ISBN 978-1-4939-2613-8  
DOI 10.1007/978-1-4939-2614-5

ISSN 2197-4136 (electronic)  
ISBN 978-1-4939-2614-5 (eBook)

Library of Congress Control Number: 2015935333

Springer New York Heidelberg Dordrecht London  
© Springer Science+Business Media New York 2011, 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*To Susan*

David Ruppert

*To my grandparents*

David S. Matteson



---

## Preface

The first edition of this book has received a very warm reception. A number of instructors have adopted this work as a textbook in their courses. Moreover, both novices and seasoned professionals have been using the book for self-study. The enthusiastic response to the book motivated a new edition. One major change is that there are now two authors. The second edition improves the book in several ways: all known errors have been corrected and changes in R have been addressed. Considerably more R code is now included. The GARCH chapter now uses the `rugarch` package, and in the Bayes chapter we now use JAGS in place of OpenBUGS.

The first edition was designed primarily as a textbook for use in university courses. Although there is an Instructor's Manual with solutions to all exercises and all problems in the R labs, this manual has been available only to instructors. No solutions have been available for readers engaged in self-study. To address this problem, the number of exercises and R lab problems has increased and the solutions to many of them are being placed on the book's web site.

Some data sets in the first edition were in R packages that are no longer available. These data sets are also on the web site. The web site also contains R scripts with the code used in the book.

We would like to thank Peter Dalgaard, Guy Yollin, and Aaron Fox for many helpful suggestions. We also thank numerous readers for pointing out errors in the first edition.

The book's web site is <http://people.orie.cornell.edu/davidr/SDAFE2/index.html>.

Ithaca, NY, USA  
Ithaca, NY, USA  
January 2015

David Ruppert  
David S. Matteson



---

## Preface to the First Edition

I developed this textbook while teaching the course *Statistics for Financial Engineering* to master's students in the financial engineering program at Cornell University. These students have already taken courses in portfolio management, fixed income securities, options, and stochastic calculus, so I concentrate on teaching statistics, data analysis, and the use of R, and I cover most sections of Chaps. 4–12 and 18–20. These chapters alone are more than enough to fill a one-semester course. I do not cover regression (Chaps. 9–11 and 21) or the more advanced time series topics in Chap. 13, since these topics are covered in other courses. In the past, I have not covered cointegration (Chap. 15), but I will in the future. The master's students spend much of the third semester working on projects with investment banks or hedge funds. As a faculty adviser for several projects, I have seen the importance of cointegration.

A number of different courses might be based on this book. A two-semester sequence could cover most of the material. A one-semester course with more emphasis on finance would include Chaps. 16 and 17 on portfolios and the CAPM and omit some of the chapters on statistics, for instance, Chaps. 8, 14, and 20 on copulas, GARCH models, and Bayesian statistics. The book could be used for courses at both the master's and Ph.D. levels.

Readers familiar with my textbook *Statistics and Finance: An Introduction* may wonder how that volume differs from this book. This book is at a somewhat more advanced level and has much broader coverage of topics in statistics compared to the earlier book. As the title of this volume suggests, there is more emphasis on data analysis and this book is intended to be more than just “an introduction.” Chapters 8, 15, and 20 on copulas, cointegration, and Bayesian statistics are new. Except for some figures borrowed from *Statistics and Finance*, in this book R is used exclusively for computations, data analysis, and graphing, whereas the earlier book used SAS and MATLAB. Nearly all of the examples in this book use data sets that are available in R, so readers can reproduce the results. In Chap. 20 on Bayesian statistics,

WinBUGS is used for Markov chain Monte Carlo and is called from R using the R2WinBUGS package. There is some overlap between the two books, and, in particular, a substantial amount of the material in Chaps. 2, 3, 9, 11–13, and 16 has been taken from the earlier book. Unlike *Statistics and Finance*, this volume does not cover options pricing and behavioral finance.

The prerequisites for reading this book are knowledge of calculus, vectors, and matrices; probability including stochastic processes; and statistics typical of third- or fourth-year undergraduates in engineering, mathematics, statistics, and related disciplines. There is an appendix that reviews probability and statistics, but it is intended for reference and is certainly not an introduction for readers with little or no prior exposure to these topics. Also, the reader should have some knowledge of computer programming. Some familiarity with the basic ideas of finance is helpful.

This book does not teach R programming, but each chapter has an “R lab” with data analysis and simulations. Students can learn R from these labs and by using R’s help or the manual *An Introduction to R* (available at the CRAN web site and R’s online help) to learn more about the functions used in the labs. Also, the text does indicate which R functions are used in the examples. Occasionally, R code is given to illustrate some process, for example, in Chap. 16 finding the tangency portfolio by quadratic programming. For readers wishing to use R, the bibliographical notes at the end of each chapter mention books that cover R programming and the book’s web site contains examples of the R and WinBUGS code used to produce this book. Students enter my course *Statistics for Financial Engineering* with quite disparate knowledge of R. Some are very accomplished R programmers, while others have no experience with R, although all have experience with some programming language. Students with no previous experience with R generally need assistance from the instructor to get started on the R labs. Readers using this book for self-study should learn R first before attempting the R labs.

Ithaca, NY, USA  
July 2010

David Ruppert

---

# Contents

<b>Notation</b> .....	xxv
<b>1 Introduction</b> .....	1
1.1 Bibliographic Notes .....	4
References .....	4
<b>2 Returns</b> .....	5
2.1 Introduction .....	5
2.1.1 Net Returns .....	5
2.1.2 Gross Returns .....	6
2.1.3 Log Returns .....	6
2.1.4 Adjustment for Dividends .....	7
2.2 The Random Walk Model .....	8
2.2.1 Random Walks .....	8
2.2.2 Geometric Random Walks .....	9
2.2.3 Are Log Prices a Lognormal Geometric Random Walk? .....	9
2.3 Bibliographic Notes .....	10
2.4 R Lab .....	11
2.4.1 Data Analysis .....	11
2.4.2 Simulations .....	13
2.4.3 Simulating a Geometric Random Walk .....	14
2.4.4 Let's Look at McDonald's Stock .....	15
2.5 Exercises .....	16
References .....	18
<b>3 Fixed Income Securities</b> .....	19
3.1 Introduction .....	19
3.2 Zero-Coupon Bonds .....	20
3.2.1 Price and Returns Fluctuate with the Interest Rate ...	20

3.3	Coupon Bonds . . . . .	22
3.3.1	A General Formula . . . . .	23
3.4	Yield to Maturity . . . . .	23
3.4.1	General Method for Yield to Maturity . . . . .	25
3.4.2	Spot Rates . . . . .	25
3.5	Term Structure . . . . .	26
3.5.1	Introduction: Interest Rates Depend Upon Maturity . . . . .	26
3.5.2	Describing the Term Structure . . . . .	27
3.6	Continuous Compounding . . . . .	32
3.7	Continuous Forward Rates . . . . .	33
3.8	Sensitivity of Price to Yield . . . . .	35
3.8.1	Duration of a Coupon Bond . . . . .	35
3.9	Bibliographic Notes . . . . .	36
3.10	R Lab . . . . .	37
3.10.1	Computing Yield to Maturity . . . . .	37
3.10.2	Graphing Yield Curves . . . . .	38
3.11	Exercises . . . . .	40
	References . . . . .	43
<b>4</b>	<b>Exploratory Data Analysis . . . . .</b>	<b>45</b>
4.1	Introduction . . . . .	45
4.2	Histograms and Kernel Density Estimation . . . . .	47
4.3	Order Statistics, the Sample CDF, and Sample Quantiles . . . . .	52
4.3.1	The Central Limit Theorem for Sample Quantiles . . . . .	54
4.3.2	Normal Probability Plots . . . . .	54
4.3.3	Half-Normal Plots . . . . .	58
4.3.4	Quantile–Quantile Plots . . . . .	61
4.4	Tests of Normality . . . . .	64
4.5	Boxplots . . . . .	65
4.6	Data Transformation . . . . .	67
4.7	The Geometry of Transformations . . . . .	71
4.8	Transformation Kernel Density Estimation . . . . .	75
4.9	Bibliographic Notes . . . . .	77
4.10	R Lab . . . . .	77
4.10.1	European Stock Indices . . . . .	77
4.10.2	McDonald’s Prices and Returns . . . . .	80
4.11	Exercises . . . . .	81
	References . . . . .	83
<b>5</b>	<b>Modeling Univariate Distributions . . . . .</b>	<b>85</b>
5.1	Introduction . . . . .	85
5.2	Parametric Models and Parsimony . . . . .	85
5.3	Location, Scale, and Shape Parameters . . . . .	86

5.4	Skewness, Kurtosis, and Moments . . . . .	87
5.4.1	The Jarque–Bera Test . . . . .	91
5.4.2	Moments . . . . .	92
5.5	Heavy-Tailed Distributions . . . . .	93
5.5.1	Exponential and Polynomial Tails . . . . .	93
5.5.2	$t$ -Distributions . . . . .	94
5.5.3	Mixture Models . . . . .	96
5.6	Generalized Error Distributions . . . . .	99
5.7	Creating Skewed from Symmetric Distributions . . . . .	101
5.8	Quantile-Based Location, Scale, and Shape Parameters . . . . .	103
5.9	Maximum Likelihood Estimation . . . . .	104
5.10	Fisher Information and the Central Limit Theorem for the MLE . . . . .	105
5.11	Likelihood Ratio Tests . . . . .	107
5.12	AIC and BIC . . . . .	109
5.13	Validation Data and Cross-Validation . . . . .	110
5.14	Fitting Distributions by Maximum Likelihood . . . . .	113
5.15	Profile Likelihood . . . . .	119
5.16	Robust Estimation . . . . .	121
5.17	Transformation Kernel Density Estimation with a Parametric Transformation . . . . .	123
5.18	Bibliographic Notes . . . . .	126
5.19	R Lab . . . . .	127
5.19.1	Earnings Data . . . . .	127
5.19.2	DAX Returns . . . . .	129
5.19.3	McDonald’s Returns . . . . .	130
5.20	Exercises . . . . .	131
	References . . . . .	134
<b>6</b>	<b>Resampling . . . . .</b>	<b>137</b>
6.1	Introduction . . . . .	137
6.2	Bootstrap Estimates of Bias, Standard Deviation, and MSE . . . . .	139
6.2.1	Bootstrapping the MLE of the $t$ -Distribution . . . . .	139
6.3	Bootstrap Confidence Intervals . . . . .	142
6.3.1	Normal Approximation Interval . . . . .	143
6.3.2	Bootstrap- $t$ Intervals . . . . .	143
6.3.3	Basic Bootstrap Interval . . . . .	146
6.3.4	Percentile Confidence Intervals . . . . .	146
6.4	Bibliographic Notes . . . . .	150
6.5	R Lab . . . . .	150
6.5.1	BMW Returns . . . . .	150
6.5.2	Simulation Study: Bootstrapping the Kurtosis . . . . .	152
6.6	Exercises . . . . .	154
	References . . . . .	156

<b>7</b>	<b>Multivariate Statistical Models</b> .....	157
7.1	Introduction .....	157
7.2	Covariance and Correlation Matrices .....	157
7.3	Linear Functions of Random Variables .....	159
7.3.1	Two or More Linear Combinations of Random Variables .....	161
7.3.2	Independence and Variances of Sums .....	162
7.4	Scatterplot Matrices .....	162
7.5	The Multivariate Normal Distribution .....	164
7.6	The Multivariate $t$ -Distribution .....	165
7.6.1	Using the $t$ -Distribution in Portfolio Analysis .....	167
7.7	Fitting the Multivariate $t$ -Distribution by Maximum Likelihood .....	168
7.8	Elliptically Contoured Densities .....	170
7.9	The Multivariate Skewed $t$ -Distributions .....	172
7.10	The Fisher Information Matrix .....	174
7.11	Bootstrapping Multivariate Data .....	175
7.12	Bibliographic Notes .....	177
7.13	R Lab .....	177
7.13.1	Equity Returns .....	177
7.13.2	Simulating Multivariate $t$ -Distributions .....	178
7.13.3	Fitting a Bivariate $t$ -Distribution .....	180
7.14	Exercises .....	181
	References .....	182
<b>8</b>	<b>Copulas</b> .....	183
8.1	Introduction .....	183
8.2	Special Copulas .....	185
8.3	Gaussian and $t$ -Copulas .....	186
8.4	Archimedean Copulas .....	187
8.4.1	Frank Copula .....	187
8.4.2	Clayton Copula .....	189
8.4.3	Gumbel Copula .....	191
8.4.4	Joe Copula .....	192
8.5	Rank Correlation .....	193
8.5.1	Kendall's Tau .....	194
8.5.2	Spearman's Rank Correlation Coefficient .....	195
8.6	Tail Dependence .....	196
8.7	Calibrating Copulas .....	198
8.7.1	Maximum Likelihood .....	199
8.7.2	Pseudo-Maximum Likelihood .....	199
8.7.3	Calibrating Meta-Gaussian and Meta- $t$ -Distributions ..	200
8.8	Bibliographic Notes .....	207

8.9	R Lab . . . . .	208
8.9.1	Simulating from Copula Models . . . . .	208
8.9.2	Fitting Copula Models to Bivariate Return Data . . . . .	210
8.10	Exercises . . . . .	213
	References . . . . .	214
<b>9</b>	<b>Regression: Basics</b> . . . . .	<b>217</b>
9.1	Introduction . . . . .	217
9.2	Straight-Line Regression . . . . .	218
9.2.1	Least-Squares Estimation . . . . .	218
9.2.2	Variance of $\hat{\beta}_1$ . . . . .	222
9.3	Multiple Linear Regression . . . . .	223
9.3.1	Standard Errors, $t$ -Values, and $p$ -Values . . . . .	225
9.4	Analysis of Variance, Sums of Squares, and $R^2$ . . . . .	227
9.4.1	ANOVA Table . . . . .	227
9.4.2	Degrees of Freedom (DF) . . . . .	229
9.4.3	Mean Sums of Squares (MS) and $F$ -Tests . . . . .	229
9.4.4	Adjusted $R^2$ . . . . .	231
9.5	Model Selection . . . . .	231
9.6	Collinearity and Variance Inflation . . . . .	233
9.7	Partial Residual Plots . . . . .	240
9.8	Centering the Predictors . . . . .	242
9.9	Orthogonal Polynomials . . . . .	243
9.10	Bibliographic Notes . . . . .	243
9.11	R Lab . . . . .	243
9.11.1	U.S. Macroeconomic Variables . . . . .	243
9.12	Exercises . . . . .	245
	References . . . . .	248
<b>10</b>	<b>Regression: Troubleshooting</b> . . . . .	<b>249</b>
10.1	Regression Diagnostics . . . . .	249
10.1.1	Leverages . . . . .	251
10.1.2	Residuals . . . . .	252
10.1.3	Cook's Distance . . . . .	253
10.2	Checking Model Assumptions . . . . .	255
10.2.1	Nonnormality . . . . .	256
10.2.2	Nonconstant Variance . . . . .	258
10.2.3	Nonlinearity . . . . .	259
10.3	Bibliographic Notes . . . . .	262
10.4	R Lab . . . . .	263
10.4.1	Current Population Survey Data . . . . .	263
10.5	Exercises . . . . .	265
	References . . . . .	268

<b>11 Regression: Advanced Topics</b> .....	269
11.1 The Theory Behind Linear Regression .....	269
11.1.1 Maximum Likelihood Estimation for Regression .....	270
11.2 Nonlinear Regression .....	271
11.3 Estimating Forward Rates from Zero-Coupon Bond Prices ....	276
11.4 Transform-Both-Sides Regression .....	281
11.4.1 How TBS Works .....	283
11.5 Transforming Only the Response .....	284
11.6 Binary Regression .....	286
11.7 Linearizing a Nonlinear Model .....	291
11.8 Robust Regression .....	293
11.9 Regression and Best Linear Prediction .....	295
11.9.1 Best Linear Prediction .....	295
11.9.2 Prediction Error in Best Linear Prediction .....	297
11.9.3 Regression Is Empirical Best Linear Prediction .....	298
11.9.4 Multivariate Linear Prediction .....	298
11.10 Regression Hedging .....	298
11.11 Bibliographic Notes .....	300
11.12 R Lab .....	300
11.12.1 Nonlinear Regression .....	300
11.12.2 Response Transformations .....	302
11.12.3 Binary Regression: Who Owns an Air Conditioner? ...	303
11.13 Exercises .....	304
References .....	305
<b>12 Time Series Models: Basics</b> .....	307
12.1 Time Series Data .....	307
12.2 Stationary Processes .....	307
12.2.1 White Noise .....	310
12.2.2 Predicting White Noise .....	311
12.3 Estimating Parameters of a Stationary Process .....	312
12.3.1 ACF Plots and the Ljung-Box Test .....	312
12.4 AR(1) Processes .....	314
12.4.1 Properties of a Stationary AR(1) Process .....	315
12.4.2 Convergence to the Stationary Distribution .....	316
12.4.3 Nonstationary AR(1) Processes .....	317
12.5 Estimation of AR(1) Processes .....	318
12.5.1 Residuals and Model Checking .....	318
12.5.2 Maximum Likelihood and Conditional Least-Squares ..	323
12.6 AR( $p$ ) Models .....	325
12.7 Moving Average (MA) Processes .....	328
12.7.1 MA(1) Processes .....	328
12.7.2 General MA Processes .....	330
12.8 ARMA Processes .....	331
12.8.1 The Backwards Operator .....	331

12.8.2	The ARMA Model . . . . .	332
12.8.3	ARMA(1,1) Processes . . . . .	332
12.8.4	Estimation of ARMA Parameters . . . . .	333
12.8.5	The Differencing Operator . . . . .	333
12.9	ARIMA Processes . . . . .	334
12.9.1	Drifts in ARIMA Processes . . . . .	337
12.10	Unit Root Tests . . . . .	338
12.10.1	How Do Unit Root Tests Work? . . . . .	341
12.11	Automatic Selection of an ARIMA Model . . . . .	342
12.12	Forecasting . . . . .	342
12.12.1	Forecast Errors and Prediction Intervals . . . . .	344
12.12.2	Computing Forecast Limits by Simulation . . . . .	346
12.13	Partial Autocorrelation Coefficients . . . . .	349
12.14	Bibliographic Notes . . . . .	352
12.15	R Lab . . . . .	352
12.15.1	T-bill Rates . . . . .	352
12.15.2	Forecasting . . . . .	355
12.16	Exercises . . . . .	356
	References . . . . .	360
<b>13</b>	<b>Time Series Models: Further Topics . . . . .</b>	<b>361</b>
13.1	Seasonal ARIMA Models . . . . .	361
13.1.1	Seasonal and Nonseasonal Differencing . . . . .	362
13.1.2	Multiplicative ARIMA Models . . . . .	362
13.2	Box-Cox Transformation for Time Series . . . . .	365
13.3	Time Series and Regression . . . . .	367
13.3.1	Residual Correlation and Spurious Regressions . . . . .	368
13.3.2	Heteroscedasticity and Autocorrelation Consistent (HAC) Standard Errors . . . . .	373
13.3.3	Linear Regression with ARMA Errors . . . . .	377
13.4	Multivariate Time Series . . . . .	380
13.4.1	The Cross-Correlation Function . . . . .	380
13.4.2	Multivariate White Noise . . . . .	382
13.4.3	Multivariate ACF Plots and the Multivariate Ljung-Box Test . . . . .	383
13.4.4	Multivariate ARMA Processes . . . . .	384
13.4.5	Prediction Using Multivariate AR Models . . . . .	387
13.5	Long-Memory Processes . . . . .	389
13.5.1	The Need for Long-Memory Stationary Models . . . . .	389
13.5.2	Fractional Differencing . . . . .	390
13.5.3	FARIMA Processes . . . . .	391
13.6	Bootstrapping Time Series . . . . .	394
13.7	Bibliographic Notes . . . . .	395
13.8	R Lab . . . . .	395
13.8.1	Seasonal ARIMA Models . . . . .	395
13.8.2	Regression with HAC Standard Errors . . . . .	396

13.8.3	Regression with ARMA Noise . . . . .	397
13.8.4	VAR Models . . . . .	397
13.8.5	Long-Memory Processes . . . . .	399
13.8.6	Model-Based Bootstrapping of an ARIMA Process . . . . .	400
13.9	Exercises . . . . .	401
	References . . . . .	403
<b>14</b>	<b>GARCH Models</b> . . . . .	<b>405</b>
14.1	Introduction . . . . .	405
14.2	Estimating Conditional Means and Variances . . . . .	406
14.3	ARCH(1) Processes . . . . .	407
14.4	The AR(1)+ARCH(1) Model . . . . .	409
14.5	ARCH( $p$ ) Models . . . . .	411
14.6	ARIMA( $p_M, d, q_M$ )+GARCH( $p_V, q_V$ ) Models . . . . .	411
14.6.1	Residuals for ARIMA( $p_M, d, q_M$ )+GARCH( $p_V, q_V$ ) Models . . . . .	412
14.7	GARCH Processes Have Heavy Tails . . . . .	413
14.8	Fitting ARMA+GARCH Models . . . . .	413
14.9	GARCH Models as ARMA Models . . . . .	418
14.10	GARCH(1,1) Processes . . . . .	419
14.11	APARCH Models . . . . .	421
14.12	Linear Regression with ARMA+GARCH Errors . . . . .	424
14.13	Forecasting ARMA+GARCH Processes . . . . .	426
14.14	Multivariate GARCH Processes . . . . .	428
14.14.1	Multivariate Conditional Heteroscedasticity . . . . .	428
14.14.2	Basic Setting . . . . .	431
14.14.3	Exponentially Weighted Moving Average (EWMA) Model . . . . .	432
14.14.4	Orthogonal GARCH Models . . . . .	433
14.14.5	Dynamic Orthogonal Component (DOC) Models . . . . .	436
14.14.6	Dynamic Conditional Correlation (DCC) Models . . . . .	439
14.14.7	Model Checking . . . . .	441
14.15	Bibliographic Notes . . . . .	443
14.16	R Lab . . . . .	443
14.16.1	Fitting GARCH Models . . . . .	443
14.16.2	The GARCH-in-Mean (GARCH-M) Model . . . . .	445
14.16.3	Fitting Multivariate GARCH Models . . . . .	445
14.17	Exercises . . . . .	447
	References . . . . .	451
<b>15</b>	<b>Cointegration</b> . . . . .	<b>453</b>
15.1	Introduction . . . . .	453
15.2	Vector Error Correction Models . . . . .	455
15.3	Trading Strategies . . . . .	459
15.4	Bibliographic Notes . . . . .	460

15.5	R Lab . . . . .	460
15.5.1	Cointegration Analysis of Midcap Prices . . . . .	460
15.5.2	Cointegration Analysis of Yields . . . . .	460
15.5.3	Cointegration Analysis of Daily Stock Prices . . . . .	461
15.5.4	Simulation . . . . .	462
15.6	Exercises . . . . .	462
	References . . . . .	463
<b>16</b>	<b>Portfolio Selection . . . . .</b>	<b>465</b>
16.1	Trading Off Expected Return and Risk . . . . .	465
16.2	One Risky Asset and One Risk-Free Asset . . . . .	465
16.2.1	Estimating $E(R)$ and $\sigma_R$ . . . . .	467
16.3	Two Risky Assets . . . . .	468
16.3.1	Risk Versus Expected Return . . . . .	468
16.4	Combining Two Risky Assets with a Risk-Free Asset . . . . .	469
16.4.1	Tangency Portfolio with Two Risky Assets . . . . .	469
16.4.2	Combining the Tangency Portfolio with the Risk-Free Asset . . . . .	471
16.4.3	Effect of $\rho_{12}$ . . . . .	472
16.5	Selling Short . . . . .	473
16.6	Risk-Efficient Portfolios with $N$ Risky Assets . . . . .	474
16.7	Resampling and Efficient Portfolios . . . . .	479
16.8	Utility . . . . .	484
16.9	Bibliographic Notes . . . . .	488
16.10	R Lab . . . . .	488
16.10.1	Efficient Equity Portfolios . . . . .	488
16.10.2	Efficient Portfolios with Apple, Exxon-Mobil, Target, and McDonald's Stock . . . . .	489
16.10.3	Finding the Set of Possible Expected Returns . . . . .	490
16.11	Exercises . . . . .	491
	References . . . . .	493
<b>17</b>	<b>The Capital Asset Pricing Model . . . . .</b>	<b>495</b>
17.1	Introduction to the CAPM . . . . .	495
17.2	The Capital Market Line (CML) . . . . .	496
17.3	Betas and the Security Market Line . . . . .	499
17.3.1	Examples of Betas . . . . .	500
17.3.2	Comparison of the CML with the SML . . . . .	500
17.4	The Security Characteristic Line . . . . .	501
17.4.1	Reducing Unique Risk by Diversification . . . . .	503
17.4.2	Are the Assumptions Sensible? . . . . .	504
17.5	Some More Portfolio Theory . . . . .	504
17.5.1	Contributions to the Market Portfolio's Risk . . . . .	505
17.5.2	Derivation of the SML . . . . .	505
17.6	Estimation of Beta and Testing the CAPM . . . . .	507

17.6.1	Estimation Using Regression .....	507
17.6.2	Testing the CAPM .....	509
17.6.3	Interpretation of Alpha .....	509
17.7	Using the CAPM in Portfolio Analysis .....	510
17.8	Bibliographic Notes .....	510
17.9	R Lab .....	510
17.9.1	Zero-beta Portfolios .....	512
17.10	Exercises .....	512
	References .....	515
<b>18</b>	<b>Factor Models and Principal Components .....</b>	<b>517</b>
18.1	Dimension Reduction .....	517
18.2	Principal Components Analysis .....	517
18.3	Factor Models .....	527
18.4	Fitting Factor Models by Time Series Regression .....	528
18.4.1	Fama and French Three-Factor Model .....	529
18.4.2	Estimating Expectations and Covariances of Asset Returns .....	534
18.5	Cross-Sectional Factor Models .....	538
18.6	Statistical Factor Models .....	540
18.6.1	Varimax Rotation of the Factors .....	545
18.7	Bibliographic Notes .....	546
18.8	R Lab .....	546
18.8.1	PCA .....	546
18.8.2	Fitting Factor Models by Time Series Regression .....	548
18.8.3	Statistical Factor Models .....	550
18.9	Exercises .....	551
	References .....	552
<b>19</b>	<b>Risk Management .....</b>	<b>553</b>
19.1	The Need for Risk Management .....	553
19.2	Estimating VaR and ES with One Asset .....	555
19.2.1	Nonparametric Estimation of VaR and ES .....	555
19.2.2	Parametric Estimation of VaR and ES .....	557
19.3	Bootstrap Confidence Intervals for VaR and ES .....	559
19.4	Estimating VaR and ES Using ARMA+GARCH Models .....	561
19.5	Estimating VaR and ES for a Portfolio of Assets .....	563
19.6	Estimation of VaR Assuming Polynomial Tails .....	565
19.6.1	Estimating the Tail Index .....	567
19.7	Pareto Distributions .....	571
19.8	Choosing the Horizon and Confidence Level .....	571
19.9	VaR and Diversification .....	573
19.10	Bibliographic Notes .....	575
19.11	R Lab .....	575
19.11.1	Univariate VaR and ES .....	575
19.11.2	VaR Using a Multivariate- <i>t</i> Model .....	576

19.12 Exercises .....	577
References .....	578
<b>20 Bayesian Data Analysis and MCMC .....</b>	<b>581</b>
20.1 Introduction .....	581
20.2 Bayes's Theorem .....	582
20.3 Prior and Posterior Distributions .....	584
20.4 Conjugate Priors .....	586
20.5 Central Limit Theorem for the Posterior .....	592
20.6 Posterior Intervals .....	593
20.7 Markov Chain Monte Carlo .....	595
20.7.1 Gibbs Sampling .....	596
20.7.2 Other Markov Chain Monte Carlo Samplers .....	597
20.7.3 Analysis of MCMC Output .....	597
20.7.4 JAGS .....	598
20.7.5 Monitoring MCMC Convergence and Mixing .....	602
20.7.6 DIC and $p_D$ for Model Comparisons .....	609
20.8 Hierarchical Priors .....	612
20.9 Bayesian Estimation of a Covariance Matrix .....	618
20.9.1 Estimating a Multivariate Gaussian Covariance Matrix .....	618
20.9.2 Estimating a Multivariate- $t$ Scale Matrix .....	620
20.9.3 Non-Wishart Priors for the Covariate Matrix .....	623
20.10 Stochastic Volatility Models .....	623
20.11 Fitting GARCH Models with MCMC .....	626
20.12 Fitting a Factor Model .....	629
20.13 Sampling a Stationary Process .....	632
20.14 Bibliographic Notes .....	635
20.15 R Lab .....	636
20.15.1 Fitting a $t$ -Distribution by MCMC .....	636
20.15.2 AR Models .....	639
20.15.3 MA Models .....	640
20.15.4 ARMA Models .....	641
20.16 Exercises .....	642
References .....	643
<b>21 Nonparametric Regression and Splines .....</b>	<b>645</b>
21.1 Introduction .....	645
21.2 Local Polynomial Regression .....	648
21.2.1 Lowess and Loess .....	652
21.3 Linear Smoothers .....	653
21.3.1 The Smoother Matrix and the Effective Degrees of Freedom .....	653
21.3.2 AIC, CV, and GCV .....	654
21.4 Polynomial Splines .....	654
21.4.1 Linear Splines with One Knot .....	655

21.4.2	Linear Splines with Many Knots . . . . .	656
21.4.3	Quadratic Splines . . . . .	656
21.4.4	$p$ th Degree Splines . . . . .	657
21.4.5	Other Spline Bases . . . . .	658
21.5	Penalized Splines . . . . .	658
21.5.1	Cubic Smoothing Splines . . . . .	659
21.5.2	Selecting the Amount of Penalization . . . . .	659
21.6	Bibliographic Notes . . . . .	664
21.7	R Lab . . . . .	664
21.7.1	Additive Model for Wages, Education, and Experience . . . . .	664
21.7.2	An Extended CKLS Model for the Short Rate . . . . .	665
21.8	Exercises . . . . .	666
	References . . . . .	667
<b>A</b>	<b>Facts from Probability, Statistics, and Algebra . . . . .</b>	<b>669</b>
A.1	Introduction . . . . .	669
A.2	Probability Distributions . . . . .	669
A.2.1	Cumulative Distribution Functions . . . . .	669
A.2.2	Quantiles and Percentiles . . . . .	670
A.2.3	Symmetry and Modes . . . . .	670
A.2.4	Support of a Distribution . . . . .	670
A.3	When Do Expected Values and Variances Exist? . . . . .	671
A.4	Monotonic Functions . . . . .	672
A.5	The Minimum, Maximum, Infimum, and Supremum of a Set . . . . .	672
A.6	Functions of Random Variables . . . . .	672
A.7	Random Samples . . . . .	673
A.8	The Binomial Distribution . . . . .	674
A.9	Some Common Continuous Distributions . . . . .	674
A.9.1	Uniform Distributions . . . . .	674
A.9.2	Transformation by the CDF and Inverse CDF . . . . .	675
A.9.3	Normal Distributions . . . . .	676
A.9.4	The Lognormal Distribution . . . . .	676
A.9.5	Exponential and Double-Exponential Distributions . . . . .	678
A.9.6	Gamma and Inverse-Gamma Distributions . . . . .	678
A.9.7	Beta Distributions . . . . .	679
A.9.8	Pareto Distributions . . . . .	680
A.10	Sampling a Normal Distribution . . . . .	681
A.10.1	Chi-Squared Distributions . . . . .	681
A.10.2	$F$ -Distributions . . . . .	681
A.11	Law of Large Numbers and the Central Limit Theorem for the Sample Mean . . . . .	682
A.12	Bivariate Distributions . . . . .	682

A.13	Correlation and Covariance . . . . .	683
	A.13.1 Normal Distributions: Conditional Expectations and Variance . . . . .	687
A.14	Multivariate Distributions . . . . .	687
	A.14.1 Conditional Densities . . . . .	688
A.15	Stochastic Processes . . . . .	688
A.16	Estimation . . . . .	689
	A.16.1 Introduction . . . . .	689
	A.16.2 Standard Errors . . . . .	689
A.17	Confidence Intervals . . . . .	690
	A.17.1 Confidence Interval for the Mean . . . . .	690
	A.17.2 Confidence Intervals for the Variance and Standard Deviation . . . . .	692
	A.17.3 Confidence Intervals Based on Standard Errors . . . . .	693
A.18	Hypothesis Testing . . . . .	693
	A.18.1 Hypotheses, Types of Errors, and Rejection Regions . .	693
	A.18.2 $p$ -Values . . . . .	693
	A.18.3 Two-Sample $t$ -Tests . . . . .	694
	A.18.4 Statistical Versus Practical Significance . . . . .	697
A.19	Prediction . . . . .	697
A.20	Facts About Vectors and Matrices . . . . .	698
A.21	Roots of Polynomials and Complex Numbers . . . . .	699
A.22	Bibliographic Notes . . . . .	700
	References . . . . .	700
<b>Index</b>	. . . . .	<b>703</b>



---

## Notation

The following conventions are observed as much as possible:

- Lowercase letters, e.g.,  $a$  and  $b$ , are used for nonrandom scalars.
- Lowercase boldface letters, e.g.,  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\boldsymbol{\theta}$ , are used for nonrandom vectors.
- Uppercase letters, e.g.,  $X$  and  $Y$ , are used for random variables.
- Uppercase bold letters either early in the Roman alphabet or in Greek without a “hat,” e.g.,  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\boldsymbol{\Omega}$ , are used for nonrandom matrices.
- A hat over a parameter or parameter vector, e.g.,  $\hat{\theta}$  and  $\hat{\boldsymbol{\theta}}$ , denotes an estimator of the corresponding parameter or parameter vector.
- $\mathbf{I}$  denotes the identity matrix with dimension appropriate for the context.
- $\text{diag}(d_1, \dots, d_p)$  is a diagonal matrix with diagonal elements  $d_1, \dots, d_p$ .
- Greek letters with a “hat” or uppercase bold letters later in the Roman alphabet, e.g.,  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\hat{\boldsymbol{\theta}}$ , will be used for random vectors.
- $\log(x)$  is the natural logarithm of  $x$  and  $\log_{10}(x)$  is the base-10 logarithm.
- $E(X)$  is the expected value of a random variable  $X$ .
- $\text{Var}(X)$  and  $\sigma_X^2$  are used to denote the variance of a random variable  $X$ .
- $\text{Cov}(X, Y)$  and  $\sigma_{XY}$  are used to denote the covariance between the random variables  $X$  and  $Y$ .
- $\text{Corr}(X, Y)$  and  $\rho_{XY}$  are used to denote the correlation between the random variables  $X$  and  $Y$ .
- $\text{COV}(\mathbf{X})$  is the covariance matrix of a random vector  $\mathbf{X}$ .
- $\text{CORR}(\mathbf{X})$  is the correlation matrix of a random vector  $\mathbf{X}$ .
- A Greek letter denotes a parameter, e.g.,  $\theta$ .
- A boldface Greek letter, e.g.,  $\boldsymbol{\theta}$ , denotes a vector of parameters.
- $\Re$  is the set of real numbers and  $\Re^p$  is the  $p$ -dimensional Euclidean space, the set of all real  $p$ -dimensional vectors.
- $A \cap B$  and  $A \cup B$  are, respectively, the intersection and union of the sets  $A$  and  $B$ .
- $\emptyset$  is the empty set.

- If  $A$  is some statement, then  $I\{A\}$  is called the indicator function of  $A$  and is equal to 1 if  $A$  is true and equal to 0 if  $A$  is false.
- If  $f_1$  and  $f_2$  are two functions of a variable  $x$ , then

$$f_1(x) \sim f_2(x) \text{ as } x \rightarrow x_0$$

means that

$$\lim_{x \rightarrow x_0} \frac{f_1(x)}{f_2(x)} = 1.$$

Similarly,

$$a_n \sim b_n$$

means that the sequences  $\{a_n\}$  and  $\{b_n\}$  are such that

$$\frac{a_n}{b_n} \rightarrow 1 \text{ as } n \rightarrow \infty.$$

- Vectors are column vectors and transposed vectors are rows, e.g.,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and

$$\mathbf{x}^\top = (x_1 \quad \cdots \quad x_n).$$

- $|\mathbf{A}|$  is the determinant of a square matrix  $\mathbf{A}$ .
- $\text{tr}(\mathbf{A})$  is the trace (sum of the diagonal elements) of a square matrix  $\mathbf{A}$ .
- $f(x) \propto g(x)$  means that  $f(x)$  is proportional to  $g(x)$ , that is,  $f(x) = ag(x)$  for some nonzero constant  $a$ .
- A word appearing in italic font is being defined or introduced in the text.