
Contents

Part I DATA EXPLORATION, ESTIMATION AND SIMULATION	1
1 UNIVARIATE DATA DISTRIBUTIONS	3
1.1 Probability Distributions and Their Parameters	3
1.1.1 Standard Probability Distribution Families	3
1.1.2 Estimation from Empirical Data	18
1.1.3 Quantiles and Q-Q Plots	22
1.2 Observations and Nonparametric Density Estimation	31
1.2.1 Sample Data	31
1.2.2 Nonparametric Estimation	35
1.2.3 Histograms	39
1.2.4 Kernel Density Estimation	41
1.2.5 Empirical Q-Q Plots	48
1.3 Monte Carlo Computations	49
1.3.1 Generating Random Samples in R	49
1.3.2 Limit Theorems and Monte Carlo Computations	51
1.3.3 Home Grown Random Samples	60
Problems	62
Notes & Complements	67
2 HEAVY TAIL DISTRIBUTIONS	69
2.1 A Primer on Extreme Value Theory	69
2.1.1 Empirical Evidence of Extreme Events	69
2.1.2 Pareto Distributions	71
2.1.3 Tidbits of Extreme Value Theory	75
2.2 GEV & GPD Parameter Estimation	83
2.2.1 The Method of L-Moments	83
2.2.2 Maximum Likelihood Estimation	90

- 2.2.3 An Example Chosen for Pedagogical Reasons 93
 - 2.2.4 Implementation of the Block-Maxima Method 95
- 2.3 Semi Parametric Estimation 97
 - 2.3.1 Threshold Exceedances 97
 - 2.3.2 Semi Parametric Estimation 100
 - 2.3.3 The Example of the PCS Index Revisited 102
 - 2.3.4 The Example of the Weekly S&P Returns 106
- Appendix: Risk Measures: Why and What For? 109
- Problems 114
- Notes & Complements 118

3 DEPENDENCE & MULTIVARIATE DATA EXPLORATION 121

- 3.1 Multivariate Data and First Measure of Dependence 121
 - 3.1.1 Density Estimation 123
 - 3.1.2 The Correlation Coefficient 126
- 3.2 The Multivariate Normal Distribution 128
 - 3.2.1 Important Remark about Independence 130
 - 3.2.2 Simulation of Random Samples 130
 - 3.2.3 The Bivariate Case 131
 - 3.2.4 A Simulation Example 132
 - 3.2.5 Let’s Have Some Coffee 133
 - 3.2.6 Is the Joint Distribution Normal? 134
- 3.3 Marginals and More Measures of Dependence 135
 - 3.3.1 Estimation of the Coffee Log-Return Distributions 136
 - 3.3.2 More Measures of Dependence 142
- 3.4 Copulas 144
 - 3.4.1 Definitions and First Properties 145
 - 3.4.2 Examples of Copula Families 146
 - 3.4.3 Copulas and General Bivariate Distributions 148
 - 3.4.4 Fitting Copulas 151
 - 3.4.5 Monte Carlo Simulations with Copulas 152
 - 3.4.6 A Risk Management Example 154
 - 3.4.7 A First Example from the Credit Markets 157
 - 3.4.8 Higher Dimensional Copulas 158
 - 3.4.9 Multi Name Credit Derivatives and CDOs 164
- 3.5 Principal Component Analysis 171
 - 3.5.1 Identification of the Principal Components of a Data Set . . . 172
 - 3.5.2 PCA with R 175
 - 3.5.3 Effective Dimension of the Space of Yield Curves 175
 - 3.5.4 Swap Rate Curves 178
- Appendix 1: Calculus with Random Vectors and Matrices 180
- Appendix 2: Families of Copulas 183
- Problems 186
- Notes & Complements 195

Part II REGRESSION

197

4	PARAMETRIC REGRESSION	199
4.1	Simple Linear Regression	199
4.1.1	Getting the Data	200
4.1.2	First Plots	201
4.1.3	Regression Set-up	202
4.1.4	Simple Linear Regression	205
4.1.5	Cost Minimizations	208
4.1.6	Regression as a Minimization Problem	209
4.2	Regression for Prediction & Sensitivities	211
4.2.1	Prediction	211
4.2.2	Introductory Discussion of Sensitivity and Robustness	213
4.2.3	Comparing L2 and L1 Regressions	214
4.2.4	Taking Another Look at the Coffee Data	216
4.3	Smoothing Versus Distribution Theory	217
4.3.1	Regression and Conditional Expectation	218
4.3.2	Maximum Likelihood Approach	219
4.4	Multiple Regression	224
4.4.1	Notation	224
4.4.2	The R Function <code>lm</code>	225
4.4.3	R^2 as a Regression Diagnostic	226
4.5	Matrix Formulation and Linear Models	228
4.5.1	Linear Models	228
4.5.2	Least Squares (Linear) Regression Revisited	229
4.5.3	Confidence and Prediction Intervals	235
4.5.4	First Extensions	236
4.5.5	Testing the CAPM	239
4.6	Polynomial Regression	243
4.6.1	Polynomial Regression as a Linear Model	243
4.6.2	Example of R Commands	243
4.6.3	Important Remark	244
4.6.4	Prediction with Polynomial Regression	245
4.6.5	Choice of the Degree p	248
4.7	Nonlinear Regression	249
4.7.1	A First Model	249
4.7.2	Transformation of the Variables	251
4.7.3	A Second Model	252
4.8	Term Structure of Interest Rates: A Crash Course	252
4.8.1	Zero Coupon Bonds	253
4.8.2	Coupon Bearing Bonds	254
4.8.3	Constructing the Term Structure by Linear Regression?	255
4.8.4	Clean Prices & Duration	256
4.8.5	The Three Different Forms of Term Structure	257

4.9	Parametric Yield Curve Estimation	258
4.9.1	Estimation Procedures	259
4.9.2	Practical Implementation	260
4.9.3	R Experiments	261
4.9.4	Concluding Remarks	263
	Appendix: Cautionary Notes on Some R Idiosyncracies	264
	Problems	268
	Notes & Complements	276
5	LOCAL AND NONPARAMETRIC REGRESSION	277
5.1	Review of the Regression Setup	277
5.2	Basis Expansion Regression	279
5.2.1	Natural Splines as Local Smoothers	279
5.2.2	Feature Function Expansions	280
5.3	Nonparametric Scatterplot Smoothers	283
5.3.1	Smoothing Splines	283
5.3.2	Locally Weighted Regression	285
5.3.3	A Robust Smoother	286
5.3.4	The Super Smoother	287
5.3.5	The Kernel Smoother	287
5.4	More Yield Curve Estimation	291
5.4.1	A First Estimation Method	291
5.4.2	A Direct Application of Smoothing Splines	292
5.4.3	US and Japanese Instantaneous Forward Rates	292
5.5	Multivariate Kernel Regression	293
5.5.1	Running the Kernel in R	296
5.5.2	An Example Involving the June 1998 S&P Futures Contract	297
5.6	Projection Pursuit Regression	303
5.6.1	The R Function ppr	304
5.6.2	ppr Prediction of the S&P Afternoon Indicators	306
5.6.3	More on the Comparison of the Two Methods	310
5.7	Nonparametric Option Pricing	311
5.7.1	Nonparametric Pricing Alternatives	311
5.7.2	Description of the Data	312
5.7.3	The Actual Experiment	313
5.7.4	Numerical Results	319
	Appendix: Kernel Density Estimation & Kernel Regression	322
	Problems	324
	Notes & Complements	338

Part III TIME SERIES & STATE SPACE MODELS 343

6	TIME SERIES MODELS: AR, MA, ARMA, & ALL THAT	345
6.1	Notation and First Definitions	345
6.1.1	Notation	346
6.1.2	Regular Time Series and Signals	346
6.1.3	Calendar and Irregular Time Series	348
6.1.4	Creating and Plotting <code>timeSeries</code> Objects in R	349
6.1.5	High Frequency Data	349
6.1.6	<code>TimeDate</code> Manipulations	353
6.2	Time Dependent Statistics and Stationarity	355
6.2.1	Statistical Moments	356
6.2.2	The Notion of Stationarity	356
6.2.3	The Search for Stationarity	361
6.2.4	The Example of the CO_2 Concentrations	363
6.3	First Examples of Models	367
6.3.1	White Noise	367
6.3.2	Random Walk	371
6.3.3	Auto Regressive Time Series	372
6.3.4	Moving Average Time Series	376
6.3.5	Using the Backward Shift Operator B	379
6.3.6	Linear Processes	380
6.3.7	Causality, Stationarity and Invertibility	381
6.3.8	ARMA Time Series	385
6.3.9	ARIMA Models	386
6.4	Fitting Models to Data	386
6.4.1	Practical Steps	387
6.4.2	R Implementation	388
6.5	Putting a Price on Temperature	398
6.5.1	Generalities on Degree Days	398
6.5.2	Temperature Options	399
6.5.3	Statistical Analysis of Temperature Historical Data	402
	Problems	415
	Notes & Complements	420
7	MULTIVARIATE TIME SERIES, LINEAR SYSTEMS AND KALMAN FILTERING	423
7.1	Multivariate Time Series	423
7.1.1	Stationarity and Auto-Covariance Functions	424
7.1.2	Multivariate White Noise	424
7.1.3	Multivariate AR Models	425
7.1.4	Back to Temperature Options	429
7.1.5	Multivariate MA & ARIMA Models	432
7.1.6	Cointegration	433
7.2	State Space Models: Mathematical Set Up	435

7.3	Factor Models as Hidden Markov Processes	437
7.3.1	Factor Models	437
7.3.2	Assumptions of the Model	438
7.3.3	Dynamics of the Factors	439
7.4	Kalman Filtering of Linear Systems	439
7.4.1	One-Step-Ahead Prediction	440
7.4.2	Derivation of the Recursive Filtering Equations	440
7.4.3	Filtering	444
7.4.4	More Predictions	445
7.4.5	Estimation of the Parameters	447
7.5	Applications to Linear Models	448
7.5.1	State Space Representation of Linear Models	448
7.5.2	Linear Models with Time Varying Coefficients	449
7.5.3	CAPM with Time Varying β 's	450
7.6	State Space Representation of Time Series	453
7.6.1	The Case of AR Series	453
7.6.2	The General Case of ARMA Series	455
7.6.3	Fitting ARMA Models by Maximum Likelihood	456
7.7	Example: Prediction of Quarterly Earnings	457
	Problems	461
	Notes & Complements	471

8 NONLINEAR TIME SERIES: MODELS AND SIMULATION 473

8.1	First Nonlinear Time Series Models	473
8.1.1	Fractional Time Series	474
8.1.2	Nonlinear Auto-Regressive Series	475
8.1.3	Statistical Estimation	476
8.2	More Nonlinear Models: ARCH, GARCH & All That	478
8.2.1	Motivation	478
8.2.2	ARCH Models	479
8.2.3	GARCH Models	481
8.2.4	Summary	481
8.2.5	R Commands	482
8.2.6	Fitting a GARCH Model to Real Data	483
8.2.7	Generalizations	493
8.3	Stochastic Volatility Models	495
8.3.1	Information Structure	495
8.3.2	State Space Formulation	496
8.3.3	Excess Kurtosis	496
8.3.4	Leverage Effect	497
8.3.5	Comparison with ARCH and GARCH Models	498
8.3.6	The Smile Effect	499

8.4	Discretization of Stochastic Differential Equations	500
8.4.1	Discretization Schemes	501
8.4.2	Monte Carlo Simulations: A First Example.....	503
8.5	Random Simulation and Scenario Generation	505
8.5.1	A Simple Model for the S&P 500 Index	505
8.5.2	Modeling the Short Interest Rate	508
8.5.3	Modeling the Spread	509
8.5.4	Putting Everything Together	511
8.6	Filtering of Nonlinear Systems	513
8.6.1	Hidden Markov Models	513
8.6.2	General Filtering Approach.....	514
8.6.3	Particle Filter Approximations	515
8.6.4	Filtering in Finance? Statistical Issues	519
8.6.5	Application: Tracking Volatility	520
	Problems	526
	Notes & Complements	531

Part IV BACKGROUND MATERIAL 535

9 APPENDICES 537

9.1	Appendix A: A Quick Introduction to R	537
9.1.1	Starting R Under Mac OS, Windows and Under UNIX....	537
9.1.2	Creating R Objects	538
9.1.3	Random Generation and White Noise	540
9.1.4	More Functions and for Loops	542
9.1.5	Importing Data	544
9.1.6	Programming in R: A First Function	549
9.2	Appendix B: A Crash Course on Black-Scholes Option Pricing	551
9.2.1	Generalities on Option Pricing	551
	Notes & Complements	558

References 559

Notation Index 565

Data Set Index 569

R Index 571

Author Index 575

Subject Index 579