

Contents

I	Introducing Algorithmic Trading	1
1	Introduction to the Book	3
1.1	Introduction to QuantStart	3
1.2	What is this Book?	3
1.3	Who is this Book For?	3
1.4	What are the Prerequisites?	4
1.5	Software/Hardware Requirements	4
1.6	Book Structure	5
1.7	What the Book does not Cover	6
1.8	Where to Get Help	6
2	What Is Algorithmic Trading?	7
2.1	Overview	7
2.1.1	Advantages	7
2.1.2	Disadvantages	9
2.2	Scientific Method	10
2.3	Why Python?	11
2.4	Can Retail Traders Still Compete?	12
2.4.1	Trading Advantages	12
2.4.2	Risk Management	13
2.4.3	Investor Relations	13
2.4.4	Technology	14
II	Trading Systems	15
3	Successful Backtesting	17
3.1	Why Backtest Strategies?	17
3.1.1	Key Reasons to Backtest Strategies	18

3.2	Backtesting Biases	19
3.2.1	Optimisation Bias	19
3.2.2	Look-Ahead Bias	20
3.2.3	Survivorship Bias	20
3.2.4	Cognitive Bias	21
3.3	Exchange Issues	22
3.3.1	Order Types	22
3.3.2	Price Consolidation	22
3.3.3	Forex Trading and ECNs	23
3.3.4	Shorting Constraints	23
3.4	Transaction Costs	23
3.4.1	Commission	24
3.4.2	Slippage	24
3.4.3	Market Impact	24
3.5	Backtesting vs Reality	25
4	Automated Execution	27
4.1	Backtesting Platforms	27
4.1.1	Programming	29
4.1.2	Research Tools	29
4.1.3	Event-Driven Backtesting	30
4.1.4	Latency	30
4.1.5	Language Choices	31
4.1.6	Integrated Development Environments	33
4.2	Colocation	35
4.2.1	Home Desktop	35
4.2.2	VPS	36
4.2.3	Exchange	37
5	Sourcing Strategy Ideas	39
5.1	Identifying Your Own Personal Preferences for Trading	39
5.2	Sourcing Algorithmic Trading Ideas	41
5.2.1	Textbooks	41
5.2.2	The Internet	43
5.2.3	Journal Literature	44

5.2.4	Independent Research	45
5.3	Evaluating Trading Strategies	46
5.4	Obtaining Historical Data	49
III Data Platform Development		53
6	Software Installation	55
6.1	Operating System Choice	55
6.1.1	Microsoft Windows	55
6.1.2	Mac OSX	56
6.1.3	Linux	56
6.2	Installing a Python Environment on Ubuntu Linux	57
6.2.1	Python	57
6.2.2	NumPy, SciPy and Pandas	58
6.2.3	Statsmodels and Scikit-Learn	60
6.2.4	PyQt, IPython and Matplotlib	60
6.2.5	IbPy and Trader Workstation	62
7	Financial Data Storage	65
7.1	Securities Master Databases	65
7.2	Financial Datasets	66
7.3	Storage Formats	67
7.3.1	Flat-File Storage	67
7.3.2	Document Stores/NoSQL	68
7.3.3	Relational Database Management Systems	68
7.4	Historical Data Structure	69
7.5	Data Accuracy Evaluation	69
7.6	Automation	70
7.7	Data Availability	71
7.8	MySQL for Securities Masters	71
7.8.1	Installing MySQL	72
7.8.2	Configuring MySQL	72
7.8.3	Schema Design for EOD Equities	73
7.8.4	Connecting to the Database	76
7.8.5	Using an Object-Relational Mapper	76

7.9	Retrieving Data from the Securities Master	84
8	Processing Financial Data	87
8.1	Market and Instrument Classification	87
8.1.1	Markets	87
8.1.2	Instruments	88
8.1.3	Fundamental Data	88
8.1.4	Unstructured Data	89
8.2	Frequency of Data	90
8.2.1	Weekly and Monthly Data	90
8.2.2	Daily Data	90
8.2.3	Intraday Bars	90
8.2.4	Tick and Order Book Data	91
8.3	Sources of Data	91
8.3.1	Free Sources	92
8.3.2	Commercial Sources	93
8.4	Obtaining Data	94
8.4.1	Yahoo Finance and Pandas	94
8.4.2	Quandl and Pandas	96
8.4.3	DTN IQFeed	103
8.5	Cleaning Financial Data	105
8.5.1	Data Quality	106
8.5.2	Continuous Futures Contracts	106
IV	Modelling	113
9	Statistical Learning	115
9.1	What is Statistical Learning?	115
9.1.1	Prediction and Inference	116
9.1.2	Parametric and Non-Parametric Models	117
9.1.3	Supervised and Unsupervised Learning	118
9.2	Techniques	119
9.2.1	Regression	119
9.2.2	Classification	120
9.2.3	Time Series Models	120

10 Time Series Analysis	123
10.1 Testing for Mean Reversion	123
10.1.1 Augmented Dickey-Fuller (ADF) Test	124
10.2 Testing for Stationarity	127
10.2.1 Hurst Exponent	127
10.3 Cointegration	129
10.3.1 Cointegrated Augmented Dickey-Fuller Test	131
10.4 Why Statistical Testing?	136
11 Forecasting	139
11.1 Measuring Forecasting Accuracy	139
11.1.1 Hit Rate	140
11.1.2 Confusion Matrix	140
11.2 Factor Choice	141
11.2.1 Lagged Price Factors and Volume	141
11.2.2 External Factors	142
11.3 Classification Models	142
11.3.1 Logistic Regression	142
11.3.2 Discriminant Analysis	143
11.3.3 Support Vector Machines	144
11.3.4 Decision Trees and Random Forests	145
11.3.5 Principal Components Analysis	146
11.3.6 Which Forecaster?	146
11.4 Forecasting Stock Index Movement	148
11.4.1 Python Implementations	149
11.4.2 Results	153
V Performance and Risk Management	157
12 Performance Measurement	159
12.1 Trade Analysis	160
12.1.1 Summary Statistics	161
12.2 Strategy and Portfolio Analysis	162
12.2.1 Returns Analysis	162
12.2.2 Risk/Reward Analysis	164

12.2.3 Drawdown Analysis	171
13 Risk and Money Management	175
13.1 Sources of Risk	175
13.1.1 Strategy Risk	176
13.1.2 Portfolio Risk	176
13.1.3 Counterparty Risk	177
13.1.4 Operational Risk	177
13.2 Money Management	178
13.2.1 Kelly Criterion	179
13.3 Risk Management	182
13.3.1 Value-at-Risk	182
13.4 Advantages and Disadvantages	183
VI Automated Trading	187
14 Event-Driven Trading Engine Implementation	189
14.1 Event-Driven Software	189
14.1.1 Why An Event-Driven Backtester?	190
14.2 Component Objects	191
14.2.1 Events	192
14.2.2 Data Handler	198
14.2.3 Strategy	208
14.2.4 Portfolio	210
14.2.5 Execution Handler	223
14.2.6 Backtest	226
14.3 Event-Driven Execution	232
15 Trading Strategy Implementation	243
15.1 Moving Average Crossover Strategy	244
15.2 S&P500 Forecasting Trade	250
15.3 Mean-Reverting Equity Pairs Trade	257
15.4 Plotting Performance	269
16 Strategy Optimisation	271
16.1 Parameter Optimisation	271