

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?	3
1.5	Software/Hardware Requirements	4
1.6	Book Structure	4
1.7	What the Book does not Cover	5
1.8	Where to Get Help	5
2	What Is Algorithmic Trading?	7
2.1	Overview	7
2.1.1	Advantages	7
2.1.2	Disadvantages	8
2.2	Scientific Method	9
2.3	Why Python?	9
2.4	Can Retail Traders Still Compete?	10
2.4.1	Trading Advantages	10
2.4.2	Risk Management	11
2.4.3	Investor Relations	11
2.4.4	Technology	11
II	Trading Systems	13
3	Successful Backtesting	15
3.1	Why Backtest Strategies?	15
3.2	Backtesting Biases	16
3.2.1	Optimisation Bias	16
3.2.2	Look-Ahead Bias	16
3.2.3	Survivorship Bias	17
3.2.4	Cognitive Bias	17
3.3	Exchange Issues	18
3.3.1	Order Types	18
3.3.2	Price Consolidation	18
3.3.3	Forex Trading and ECNs	19
3.3.4	Shorting Constraints	19
3.4	Transaction Costs	19
3.4.1	Commission	19
3.4.2	Slippage	19
3.4.3	Market Impact	20
3.5	Backtesting vs Reality	20
4	Automated Execution	21
4.1	Backtesting Platforms	21

4.1.1	Programming	22
4.1.2	Research Tools	22
4.1.3	Event-Driven Backtesting	23
4.1.4	Latency	23
4.1.5	Language Choices	23
4.1.6	Integrated Development Environments	24
4.2	Colocation	26
4.2.1	Home Desktop	26
4.2.2	VPS	27
4.2.3	Exchange	27
5	Sourcing Strategy Ideas	29
5.1	Identifying Your Own Personal Preferences for Trading	29
5.2	Sourcing Algorithmic Trading Ideas	30
5.2.1	Textbooks	30
5.2.2	The Internet	31
5.2.3	Journal Literature	33
5.2.4	Independent Research	33
5.3	Evaluating Trading Strategies	34
5.4	Obtaining Historical Data	35
III	Data Platform Development	39
6	Software Installation	41
6.1	Operating System Choice	41
6.1.1	Microsoft Windows	41
6.1.2	Mac OSX	41
6.1.3	Linux	42
6.2	Installing a Python Environment on Ubuntu Linux	42
6.2.1	Python	43
6.2.2	NumPy, SciPy and Pandas	43
6.2.3	Statsmodels and Scikit-Learn	44
6.2.4	PyQt, IPython and Matplotlib	44
6.2.5	IbPy and Trader Workstation	45
7	Financial Data Storage	47
7.1	Securities Master Databases	47
7.2	Financial Datasets	48
7.3	Storage Formats	48
7.3.1	Flat-File Storage	48
7.3.2	Document Stores/NoSQL	49
7.3.3	Relational Database Management Systems	49
7.4	Historical Data Structure	49
7.5	Data Accuracy Evaluation	50
7.6	Automation	51
7.7	Data Availability	51
7.8	MySQL for Securities Masters	51
7.8.1	Installing MySQL	51
7.8.2	Configuring MySQL	51
7.8.3	Schema Design for EOD Equities	52
7.8.4	Connecting to the Database	54
7.8.5	Using an Object-Relational Mapper	54
7.9	Retrieving Data from the Securities Master	58
8	Processing Financial Data	61
8.1	Market and Instrument Classification	61

8.1.1	Markets	61
8.1.2	Instruments	61
8.1.3	Fundamental Data	62
8.1.4	Unstructured Data	62
8.2	Frequency of Data	63
8.2.1	Weekly and Monthly Data	63
8.2.2	Daily Data	63
8.2.3	Intraday Bars	63
8.2.4	Tick and Order Book Data	63
8.3	Sources of Data	64
8.3.1	Free Sources	64
8.3.2	Commercial Sources	65
8.4	Obtaining Data	66
8.4.1	Yahoo Finance and Pandas	66
8.4.2	Quandl and Pandas	66
8.4.3	DTN IQFeed	71
8.5	Cleaning Financial Data	73
8.5.1	Data Quality	73
8.5.2	Continuous Futures Contracts	73
IV Modelling		77
9	Statistical Learning	79
9.1	What is Statistical Learning?	79
9.1.1	Prediction and Inference	79
9.1.2	Parametric and Non-Parametric Models	80
9.1.3	Supervised and Unsupervised Learning	81
9.2	Techniques	81
9.2.1	Regression	81
9.2.2	Classification	82
9.2.3	Time Series Models	82
10	Time Series Analysis	85
10.1	Testing for Mean Reversion	85
10.1.1	Augmented Dickey-Fuller (ADF) Test	86
10.2	Testing for Stationarity	87
10.2.1	Hurst Exponent	87
10.3	Cointegration	89
10.3.1	Cointegrated Augmented Dickey-Fuller Test	90
10.4	Why Statistical Testing?	93
11	Forecasting	95
11.1	Measuring Forecasting Accuracy	95
11.1.1	Hit Rate	95
11.1.2	Confusion Matrix	96
11.2	Factor Choice	96
11.2.1	Lagged Price Factors and Volume	96
11.2.2	External Factors	97
11.3	Classification Models	97
11.3.1	Logistic Regression	97
11.3.2	Discriminant Analysis	98
11.3.3	Support Vector Machines	98
11.3.4	Decision Trees and Random Forests	99
11.3.5	Principal Components Analysis	99
11.3.6	Which Forecaster?	99

11.4 Forecasting Stock Index Movement	101
11.4.1 Python Implementations	101
11.4.2 Results	104
V Performance and Risk Management	105
12 Performance Measurement	107
12.1 Trade Analysis	108
12.1.1 Summary Statistics	108
12.2 Strategy and Portfolio Analysis	109
12.2.1 Returns Analysis	109
12.2.2 Risk/Reward Analysis	110
12.2.3 Drawdown Analysis	115
13 Risk and Money Management	117
13.1 Sources of Risk	117
13.1.1 Strategy Risk	117
13.1.2 Portfolio Risk	118
13.1.3 Counterparty Risk	118
13.1.4 Operational Risk	118
13.2 Money Management	119
13.2.1 Kelly Criterion	119
13.3 Risk Management	121
13.3.1 Value-at-Risk	121
13.4 Advantages and Disadvantages	122
VI Automated Trading	125
14 Event-Driven Trading Engine Implementation	127
14.1 Event-Driven Software	127
14.1.1 Why An Event-Driven Backtester?	128
14.2 Component Objects	128
14.2.1 Events	129
14.2.2 Data Handler	132
14.2.3 Strategy	138
14.2.4 Portfolio	139
14.2.5 Execution Handler	147
14.2.6 Backtest	149
14.3 Event-Driven Execution	152
15 Trading Strategy Implementation	159
15.1 Moving Average Crossover Strategy	159
15.2 S&P500 Forecasting Trade	162
15.3 Mean-Reverting Equity Pairs Trade	167
16 Strategy Optimisation	175
16.1 Parameter Optimisation	175
16.1.1 Which Parameters to Optimise?	175
16.1.2 Optimisation is Expensive	176
16.1.3 Overfitting	176
16.2 Model Selection	177
16.2.1 Cross Validation	177
16.2.2 Grid Search	182
16.3 Optimising Strategies	184
16.3.1 Intraday Mean Reverting Pairs	185

16.3.2	Parameter Adjustment	185
16.3.3	Visualisation	188