

# Contents

<b>I</b>	<b>Introducing Algorithmic Trading</b>	<b>1</b>
<b>1</b>	<b>Introduction to the Book</b>	<b>3</b>
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 Installation	4
1.5.1	Installing Python	4
1.5.2	Obtaining Financial Data	5
1.6	Book Structure	9
1.7	What the Book does not Cover	10
1.8	Where to Get Help	10
<b>2</b>	<b>What Is Algorithmic Trading?</b>	<b>11</b>
2.1	Overview	11
2.1.1	Advantages	11
2.1.2	Disadvantages	13
2.2	Scientific Method	14
2.3	Why Python?	15
2.4	Can Retail Traders Still Compete?	16
2.4.1	Trading Advantages	16
2.4.2	Risk Management	17
2.4.3	Investor Relations	17
2.4.4	Technology	18
<b>II</b>	<b>Trading Systems</b>	<b>19</b>
<b>3</b>	<b>Successful Backtesting</b>	<b>21</b>

---

3.1	Why Backtest Strategies? . . . . .	21
3.1.1	Key Reasons to Backtest Strategies . . . . .	22
3.2	Backtesting Biases . . . . .	23
3.2.1	Optimisation Bias . . . . .	23
3.2.2	Look-Ahead Bias . . . . .	24
3.2.3	Survivorship Bias . . . . .	24
3.2.4	Cognitive Bias . . . . .	25
3.3	Exchange Issues . . . . .	26
3.3.1	Order Types . . . . .	26
3.3.2	Price Consolidation . . . . .	26
3.3.3	Forex Trading and ECNs . . . . .	27
3.3.4	Shorting Constraints . . . . .	27
3.4	Transaction Costs . . . . .	27
3.4.1	Commission . . . . .	28
3.4.2	Slippage . . . . .	28
3.4.3	Market Impact . . . . .	28
3.5	Backtesting vs Reality . . . . .	29
<b>4</b>	<b>Automated Execution . . . . .</b>	<b>31</b>
4.1	Backtesting Platforms . . . . .	31
4.1.1	Programming . . . . .	33
4.1.2	Research Tools . . . . .	33
4.1.3	Event-Driven Backtesting . . . . .	34
4.1.4	Latency . . . . .	34
4.1.5	Language Choices . . . . .	35
4.1.6	Integrated Development Environments . . . . .	37
4.2	Colocation . . . . .	39
4.2.1	Home Desktop . . . . .	40
4.2.2	VPS . . . . .	40
4.2.3	Exchange . . . . .	41
<b>5</b>	<b>Sourcing Strategy Ideas . . . . .</b>	<b>43</b>
5.1	Identifying Your Own Personal Preferences for Trading . . . . .	43
5.2	Sourcing Algorithmic Trading Ideas . . . . .	45
5.2.1	Textbooks . . . . .	45

---

5.2.2	The Internet . . . . .	47
5.2.3	Journal Literature . . . . .	48
5.2.4	Independent Research . . . . .	49
5.3	Evaluating Trading Strategies . . . . .	50
5.4	Obtaining Historical Data . . . . .	53
<b>III Data Platform Development</b>		<b>57</b>
<b>6</b>	<b>Financial Data Storage . . . . .</b>	<b>59</b>
6.1	Securities Master Databases . . . . .	59
6.2	Financial Datasets . . . . .	60
6.3	Storage Formats . . . . .	61
6.3.1	Flat-File Storage . . . . .	61
6.3.2	Document Stores/NoSQL . . . . .	62
6.3.3	Relational Database Management Systems . . . . .	62
6.4	Historical Data Structure . . . . .	63
6.5	Data Accuracy Evaluation . . . . .	63
6.6	Automation . . . . .	64
6.7	Data Availability . . . . .	65
6.8	MySQL for Securities Masters . . . . .	65
6.8.1	Installing MySQL . . . . .	66
6.8.2	Configuring MySQL . . . . .	66
6.8.3	Schema Design for EOD Equities . . . . .	67
6.8.4	Connecting to the Database . . . . .	70
6.8.5	Using an Object-Relational Mapper . . . . .	70
6.9	Retrieving Data from the Securities Master . . . . .	79
<b>7</b>	<b>Processing Financial Data . . . . .</b>	<b>81</b>
7.1	Market and Instrument Classification . . . . .	81
7.1.1	Markets . . . . .	81
7.1.2	Instruments . . . . .	82
7.1.3	Fundamental Data . . . . .	82
7.1.4	Unstructured Data . . . . .	83
7.2	Frequency of Data . . . . .	84
7.2.1	Weekly and Monthly Data . . . . .	84

7.2.2	Daily Data . . . . .	84
7.2.3	Intraday Bars . . . . .	84
7.2.4	Tick and Order Book Data . . . . .	85
7.3	Sources of Data . . . . .	85
7.3.1	Free Sources . . . . .	86
7.3.2	Commercial Sources . . . . .	87
7.4	Obtaining Data . . . . .	88
7.4.1	Datareader and Pandas . . . . .	88
7.4.2	DTN IQFeed . . . . .	90
7.5	Cleaning Financial Data . . . . .	92
7.5.1	Data Quality . . . . .	92
7.5.2	Continuous Futures Contracts . . . . .	93
<b>IV Modelling</b>		<b>103</b>
<b>8</b>	<b>Statistical Learning . . . . .</b>	<b>105</b>
8.1	What is Statistical Learning? . . . . .	105
8.1.1	Prediction and Inference . . . . .	106
8.1.2	Parametric and Non-Parametric Models . . . . .	107
8.1.3	Supervised and Unsupervised Learning . . . . .	108
8.2	Techniques . . . . .	109
8.2.1	Regression . . . . .	109
8.2.2	Classification . . . . .	110
8.2.3	Time Series Models . . . . .	110
<b>9</b>	<b>Time Series Analysis . . . . .</b>	<b>113</b>
9.1	Testing for Mean Reversion . . . . .	113
9.1.1	Augmented Dickey-Fuller Test . . . . .	114
9.2	Testing for Stationarity . . . . .	117
9.2.1	Hurst Exponent . . . . .	118
9.3	Cointegration . . . . .	121
9.3.1	Cointegrated Augmented Dickey-Fuller Test . . . . .	123
9.4	Why Statistical Testing? . . . . .	130
<b>10</b>	<b>Forecasting . . . . .</b>	<b>131</b>

---

10.1	Measuring Forecasting Accuracy . . . . .	131
10.1.1	Hit Rate . . . . .	132
10.1.2	Confusion Matrix . . . . .	132
10.2	Factor Choice . . . . .	133
10.2.1	Lagged Price Factors and Volume . . . . .	133
10.2.2	External Factors . . . . .	134
10.3	Classification Models . . . . .	134
10.3.1	Logistic Regression . . . . .	134
10.3.2	Discriminant Analysis . . . . .	135
10.3.3	Support Vector Machines . . . . .	136
10.3.4	Decision Trees and Random Forests . . . . .	137
10.3.5	Principal Components Analysis . . . . .	138
10.3.6	Which Forecaster? . . . . .	138
10.4	Forecasting Stock Index Movement . . . . .	140
10.4.1	Python Implementations . . . . .	141
10.4.2	Results . . . . .	146
<b>V</b>	<b>Performance and Risk Management</b>	<b>149</b>
<b>11</b>	<b>Performance Measurement . . . . .</b>	<b>151</b>
11.1	Trade Analysis . . . . .	152
11.1.1	Summary Statistics . . . . .	153
11.2	Strategy and Portfolio Analysis . . . . .	154
11.2.1	Returns Analysis . . . . .	154
11.2.2	Risk/Reward Analysis . . . . .	156
11.2.3	Drawdown Analysis . . . . .	163
<b>12</b>	<b>Risk and Money Management . . . . .</b>	<b>167</b>
12.1	Sources of Risk . . . . .	167
12.1.1	Strategy Risk . . . . .	168
12.1.2	Portfolio Risk . . . . .	168
12.1.3	Counterparty Risk . . . . .	169
12.1.4	Operational Risk . . . . .	169
12.2	Money Management . . . . .	170
12.2.1	Kelly Criterion . . . . .	171

---

12.3 Risk Management . . . . .	174
12.3.1 Value-at-Risk . . . . .	174
12.4 Advantages and Disadvantages . . . . .	175
<b>VI Automated Trading</b>	<b>179</b>
<b>13 Event-Driven Trading Engine Implementation . . . . .</b>	<b>181</b>
13.1 Event-Driven Software . . . . .	181
13.1.1 Why An Event-Driven Backtester? . . . . .	182
13.2 Component Objects . . . . .	183
13.2.1 Events . . . . .	184
13.2.2 Data Handler . . . . .	190
13.2.3 Strategy . . . . .	200
13.2.4 Portfolio . . . . .	202
13.2.5 Execution Handler . . . . .	215
13.2.6 Backtest . . . . .	218
13.3 Event-Driven Execution . . . . .	223
<b>14 Trading Strategy Implementation . . . . .</b>	<b>233</b>
14.1 Moving Average Crossover Strategy . . . . .	234
14.2 S&P500 Forecasting Trade . . . . .	241
14.3 Mean-Reverting Equity Pairs Trade . . . . .	248
14.4 Plotting Performance . . . . .	260
<b>15 Strategy Optimisation . . . . .</b>	<b>263</b>
15.1 Parameter Optimisation . . . . .	263
15.1.1 Which Parameters to Optimise? . . . . .	264
15.1.2 Optimisation is Expensive . . . . .	265
15.1.3 Overfitting . . . . .	265
15.2 Model Selection . . . . .	266
15.2.1 Cross Validation . . . . .	266
15.2.2 Grid Search . . . . .	275
15.3 Optimising Strategies . . . . .	279
15.3.1 Intraday Mean Reverting Pairs . . . . .	279
15.3.2 Parameter Adjustment . . . . .	279

15.3.3 Visualisation . . . . .	284
--------------------------------	-----