

Quantitative Trading with R

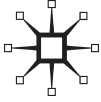
This page intentionally left blank

Quantitative Trading with R

**Understanding Mathematical and
Computational Tools from a
Quant's Perspective**

Harry Georgakopoulos

palgrave
macmillan



QUANTITATIVE TRADING WITH R

Copyright © Folk Creations, Inc., 2015.

Softcover reprint of the hardcover 1st edition 2015 978-1-137-35407-5

All right reserved.

First published in 2015 by

PALGRAVE MACMILLAN®

in the United States—a division of St. Martins Press LLC,

175 Fifth Avenue, New York, NY 10010.

Where this book is distributed in the UK, Europe and the rest of the world, this is by Palgrave Macmillan, a division of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978-1-349-46986-4 ISBN 978-1-137-43747-1 (eBook)

DOI 10.1057/9781137437471

Library of Congress Cataloging-in-Publication Data

Georgakopoulos, Harry.

Quantitative trading with R : understanding mathematical and computational tools from a quant's perspective / Harry Georgakopoulos.
pages cm

ISBN 978-1-349-46986-4

1. Stocks—Mathematical models. 2. Investment analysis—Mathematical models. 3. Corporations—Finance—Computer programs. 4. Commodity exchanges. I. Title.

HG4661.G46 2015

332.640285'5133—dc23

2014028408

A catalogue record of the book is available from the British Library.

Design by Newgen Knowledge Works (P) Ltd., Chennai, India.

First edition: January 2015

10 9 8 7 6 5 4 3 2 1

To Pinelopi, Maria, and Anastasia

This page intentionally left blank

Contents

<i>List of Figures</i>	xi
<i>List of Tables</i>	xv
<i>Acknowledgments</i>	xvii
1 An Overview	1
The mission statement	1
Financial markets and instruments	3
Trading strategies	5
High-frequency trading	5
About the orderbook	6
Trading automation	7
Where to get data from	8
Summary	9
2 Tools of the Trade	11
The R language	11
Getting started with R	14
The <code>c()</code> object	17
The <code>matrix()</code> object	19
The <code>data.frame()</code> object	21
The <code>list()</code> object	23
The <code>new.env()</code> object	25
Using the <code>plot()</code> function	26
Functional programming	30
Writing functions in R	31
Branching and looping	34
A recommended style guide	35
A pairwise correlation example	36
Summary	41
3 Working with Data	43
Getting data into R	43
Installing packages in R	45
Storing and transmitting data	46
Extracting data from a spreadsheet	48
Accessing a database	50
The dplyr package	53
Using the xts package	56

	Using the quantmod package	64
	Charting with quantmod	65
	Graphing with ggplot2	67
	Summary	70
4	Basic Statistics and Probability	71
	What is a statistic?	71
	Population versus sample	71
	Central Limit Theorem in R	74
	Unbiasedness and efficiency	76
	Probability basics	83
	Random variables	83
	Probabilities	84
	Probability distributions	85
	Bayes versus frequentist approach	87
	Simulations of coins	88
	On the use of RStan	90
	Summary	90
5	Intermediate Statistics and Probability	91
	Random process	91
	Stock price distributions	92
	Stationarity	95
	Determining stationarity with urca	98
	Assumptions of normality	99
	Correlation	103
	Filtering data	105
	R formulas	107
	The linear in linear regression	112
	Volatility	114
	Summary	117
6	Spreads, Betas and Risk	119
	Defining the stock spread	119
	Ordinary Least Squares versus Total Least Squares	121
	Constructing the spread	123
	Signal generation and validation	126
	Trading the spread	132
	Considering the risk	137
	More on the equity curve	138
	Strategy attributes	143
	Summary	144
7	Backtesting with Quantstrat	147
	Backtesting methodology	147
	About blotter and PerformanceAnalytics	149

Initial setup	149
The first strategy: A simple trend follower	151
Backtesting the first strategy	154
Evaluating the performance	158
The second strategy: Cumulative Connors RSI	162
Evaluating the mean-reverting strategy	168
Summary	175
8 High-Frequency Data	177
High-frequency quotes	178
Inter-quote arrival times	182
Identifying liquidity regimes	183
The micro-price	186
Distributions and autocorrelations	187
The highfrequency package	196
Summary	197
9 Options	199
Option theoretical value	199
A history of options	200
Valuation of options	200
Exploring options trade data	203
Implied volatility	208
Summary	214
10 Optimization	217
The motivating parabola	217
Newton's method	218
The brute-force approach	222
R optimization routines	224
A curve-fitting exercise	225
Portfolio optimization	230
Summary	236
11 Speed, Testing, and Reporting	237
Runtime execution improvements	237
Benchmarking R code	238
The Rcpp solution	239
Calling R from C++ with RInside	242
Writing unit tests with testthat	243
Using knitr for documentation	245
Summary	250
<i>Notes</i>	251
<i>References</i>	261
<i>Index</i>	269

This page intentionally left blank

Figures

1.1	Topic graph.	2
1.2	An orderbook.	6
2.1	R console.	14
2.2	Default plot of a vector.	26
2.3	Vector line plot.	27
2.4	A plot with extra attributes.	28
2.5	Four plots on a canvas.	29
2.6	Graph with legend and text.	30
2.7	If statement flow chart.	34
2.8	Sample stock price file.	38
2.9	Correlation matrix output.	40
2.10	Pairwise scatter plot.	41
3.1	Time series data of AAPL price.	44
3.2	Elementary stock line chart.	45
3.3	Sample signals sheet.	49
3.4	Sample strength sheet.	49
3.5	First xts plot.	58
3.6	An xts candle plot.	58
3.7	Subsetting an xts plot.	59
3.8	An xts plot with ablines.	61
3.9	Plots of time differences between trades.	64
3.10	AAPL data from quantmod	64
3.11	Chartseries of stock prices with indicators.	66
3.12	Chartseries of stock prices with indicators recharted.	67
3.13	Chartseries of stock prices with custom indicator.	68
3.14	Cuts applied to a vector.	68
3.15	Volume profile with ggplot2	69
4.1	Population versus sample.	72
4.2	Gaussian population.	73
4.3	Convergence with large N.	75
4.4	Convergence to a normal.	75
4.5	Nonnormal population.	76
4.6	Nonnormal sample distribution.	76
4.7	Bias versus efficiency.	77
4.8	Estimating variance.	79
4.9	Estimator consistency.	82
4.10	Random variable as a mapping.	83
4.11	Coin toss mass function.	86

4.12	The normal distribution.	86
4.13	Examples of continuous distributions.	87
4.14	Coin flip density.	88
4.15	Bayesian prior to posterior.	89
5.1	Random process.	91
5.2	Price distribution.	93
5.3	Four price distributions.	94
5.4	Stationary inputs to black box.	96
5.5	Stationary return distributions.	97
5.6	Stock prices and returns.	98
5.7	Normal histogram with density.	100
5.8	Leptokurtic and platykurtic.	101
5.9	Quantile-quantile plots.	102
5.10	Erroneous value.	103
5.11	Typical scatter plot of returns.	104
5.12	Outliers in VXX.	105
5.13	Removal of outliers.	106
5.14	Regression output on plot.	108
5.15	Diagnostics plots.	109
5.16	Scatter plot with lag.	111
5.17	Cross correlation of returns.	112
5.18	Line versus parabola.	113
5.19	Square root transformation.	114
5.20	Autocorrelation of returns.	116
5.21	Autocorrelation of theoretical squared returns.	116
5.22	Autocorrelation of actual squared returns.	117
6.1	Pepsi versus Coke price changes.	121
6.2	Ordinary least squares distance minimization.	122
6.3	Total least squares distance minimization.	122
6.4	Total least squares regression between SPY and AAPL.	123
6.5	The AAPL versus SPY spread.	126
6.6	Out of sample AAPL versus SPY spread.	127
6.7	Rolling beta spread.	129
6.8	Price difference spread for SPY versus AAPL.	130
6.9	Out of sample spread with bands.	131
6.10	Buy and sell signals superimposed.	132
6.11	The SPY versus AAPL equity curve.	136
6.12	Two equity curves with same ending value.	138
6.13	Two equity curves with different ending values.	139
6.14	Two equity curves with a different number of trades.	139
6.15	Drawdown curves.	141
6.16	Histogram of trade duration.	144
7.1	Clenow equity curve comparison.	161
7.2	Clenow XLB equity curve.	163

7.3	Connors equity curve comparison.	173
7.4	Connors XLB equity curve and indicators.	174
8.1	SPY bid price with outliers.	180
8.2	SPY bid price filtered.	181
8.3	SPY intraday bid ask spread.	184
8.4	SPY intraday spread histograms.	185
8.5	Micro-price between bid and ask.	187
8.6	Micro-price distribution.	188
8.7	Micro-price distribution with normal superimposed.	189
8.8	Micro-price autocorrelation of returns.	189
8.9	Trade price autocorrelation of returns.	191
8.10	Trade price distribution of returns.	192
8.11	Histogram of price levels.	194
8.12	Autocorrelation of trade prices.	195
9.1	Black Scholes inputs.	199
9.2	Option sensitivities.	204
9.3	Short dated option sensitivities.	204
9.4	Option traded volumes.	208
9.5	Spread of option quotes.	211
9.6	Call and put premiums.	213
9.7	Implied volatility skew.	214
10.1	Parabolic function and its derivative.	218
10.2	Rates and maturities graph.	226
10.3	Rates graph with polynomial fit.	227
10.4	Rates graph with 2 polynomial fits.	228
10.5	Rates graph with piecewise polynomial fits.	230
10.6	Rates graph with <code>loess</code> fits.	231
10.7	Equal and optimized weights.	236
11.1	Setting up knitr and pdflatex	247
11.2	Sample pdf document output.	249
11.3	A pdf with a ggplot2 plot.	250

This page intentionally left blank

Tables

1.1	Globex volume estimates by year	4
1.2	Popular stock exchanges	4
1.3	Popular futures exchanges	4
1.4	Active contract volumes	5
3.1	Popular relational databases	51
3.2	Sample data row from dplyr	55
3.3	Output of <code>summarise</code> for dplyr	56

This page intentionally left blank

Acknowledgments

You know that saying about standing on the shoulders of giants? Well, this book is dedicated to all those giants who, in one way or another, inspired and guided my work throughout the years. This book would not have been possible without the contribution of people like Jeff Ryan, Dirk Eddelbuettel, Ilya Kipnis, Hadley Wickham, Joshua Ulrich, Romain Francois, Guy Yollin, Bernhard Pfaff, Eric Zivot, Paul Teetor, Yihui Xie, Peter Carl, Jan Humme, Brian G. Peterson, Thomas Hutchinson, Steven Todd, Dimitrios Liakakos, Ed Zarek, and many others.

First and foremost, I would like to thank Ilya Kipnis for contributing excellent content on the backtesting of trading strategies via the use of the **quantstrat** package. Ilya maintains an insightful blog here:

<http://quantstrattrader.wordpress.com/>. He is also a prolific R developer, and his projects can be found on GitHub here: www.github.com/IlyaKipnis.

My gratitude and appreciation also go out to Tick Data, Inc. for graciously providing historical intraday data for use throughout this book. Tick Data, Inc. provides research-quality historical market data solutions to practitioners and academics. Their website is www.TickData.com. Readers of this book get a special 20 percent discount on all data purchases from Tick Data's online store by using promo-code: 13FC20.

Dirk Eddelbuettel (of **Rcpp**, **RProtoBuf** and **RQuantLib** fame) was gracious enough to provide guidance and insight during the beginning stages of this publication. I would like to thank him for this contribution of his, among those of many others within the R community.

A big thank you goes out to the graduate students of FINC 621 (Financial Mathematics and Modeling II) at Loyola University in Chicago for inspiring a lot of the content in this book.

Last, but not least, I am grateful to the R-core team, as well as the numerous third-party contributors for maintaining and improving the R language, which has become such an integral part of my daily work routine.