

量化投资

以R语言为工具

蔡立崙 / 著





量化投资

以R语言为工具

蔡立崙 著

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

主要讲解量化投资思想和策略，并借助 R 语言进行实战。由以下几部分组成：

首先，是对 R 编程语言的介绍，通过学习，读者可以迅速掌握用 R 语言处理数据的方法，灵活运用 R 语言解决实际金融问题；其次，向读者介绍统计学的基础知识；再次介绍量化投资的理论知识，主要讲解量化投资所需的数量基础与量化投资的类型等方面；最后，将以上两部分内容结合起来，讲述如何在 R 语言中构建量化投资策略。

本书适合从事量化投资、数据分析等工作的专业人士；金融、经济、管理、统计学、计算机相关专业的学生和教师，以及希望学习 R 语言进行量化投资的广大学者、科研人员学习和参考。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目 (CIP) 数据

量化投资：以 R 语言为工具 / 蔡立嵩著. —北京：电子工业出版社，2016.1

(Broadview 金融工程)

ISBN 978-7-121-27585-2

I. ①量… II. ①蔡… III. ①程序语言—程序设计—应用—投资 IV. ①F830.59-39

中国版本图书馆 CIP 数据核字 (2015) 第 273740 号

责任编辑：高洪霞

印 刷：北京京师印务有限公司

装 订：北京京师印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×1092 1/16 印张：34.75 字数：890 千字

版 次：2016 年 1 月第 1 版

印 次：2016 年 1 月第 1 次印刷

定 价：99.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

序 言

过去十年，一股“量化投资”的热潮在中国悄然掀起。到了最近这一两年，读者不难发现，投资人对量化的关注达到了前所未有的地步。业界到处寻找量化团队，各种量化基金如雨后春笋般出现，学校里也开始举办一场又一场的量化讲座、研讨会等。量化投资可以说一时蔚为风行，产官学共襄盛举。

这么受人瞩目的议题，到底它的含义是什么呢？为了了解量化投资这个概念，我们先回顾一下投资分析与决策过程。在投资分析与实战中，虽然个中滋味如人饮水，个中细节一言难尽，但“投资”大致上会有如下几个阶段：首先，投资人利用各种工具与分析方法，建构模型（系统）来验证买卖标的、时点、价位等的有效性。第二阶段则筛选经过分析与验证得到的结论，实际应用于交易。一个严谨的投资人，通常还会有第三阶段，即在实际投资的过程中，不断地修正与完善自己的模型（系统）。

在信息工具不发达的年代，这些过程往往以质化为主。例如，基金经理人会研究上市公司财务报表，拜访公司高层，以经验判断技术指标的趋势与形态，做出投资的买卖决策。这种做法带有很大的主观性，因此又被称为“主观交易”。主观交易的流弊，在于决策基础多源于“大胆假设”而缺乏科学方法“小心求证”的过程。更有甚者，行为金融学指出，投资人的行为往往易受各种心理认知谬误的影响而影响投资绩效。除此之外，在瞬息万变的金融市场中，主观交易者若要处变不惊地坚守操作纪律，同时眼明手快地捕捉稍纵即逝的机会，也常有“力不从心”之叹。

相较于主观交易所遭遇的问题，量化投资则在上述投资的各个阶段，利用数学、统计、计算机等分析工具来建立模型，据以客观地分析数据，按事先设定好的投资逻辑来进行投资决策，在理想状况下自动化执行下单。正因如此，量化投资拥有可验证性、纪律性与即时性等许多主观交易不可企及的优势。若再善用计算机技术，量化交易者可以处理的信息量更让主观交易者难以望其项背。如此说来，采用量化技术岂非在投资上立于不败之地？

读者只要稍加思考即可发现，量化投资的模型很容易因建模者的能力不同而良莠不齐。此外，绝大多数模型的核心思想在于“以史为鉴”；在对历史数据依赖度高的前提下，一旦遇到新兴的金融市场或历史上不曾出现的事件，量化投资者也只能徒呼负负。既然主观交易有诸多限制，量化交易看来又并非万能，那么，对投资绩效念兹在兹的投资者，究竟该何去何从呢？我们要提醒读者的是，编程语言、统计、金融、技术指标等量化投资常用的知识，只是工具！它们就像武侠小说中的宝剑与武功秘籍，固然重要，却不是笑傲江湖的保证。“宝剑锋从磨砺出”，只有勤练武艺，在实战中积累经验，才能审时度势，百战不殆。

本书旨在对量化投资作广泛与初步的介绍，希望能引领读者进入这个引人入胜的学术与实务领域。囿于笔者的学养见识，书中内容或有疏漏谬误之处，尚祈先进专家能不吝指正。最后，谨以此书表达对热血投资大众的献曝之忱。

若读者需要书中的习题解答、代码、数据、勘误补充及量化相关资料，可发邮件至 service@jadebol.com 索取，来信请在邮件标题写明书名《量化投资：以 R 语言为工具》。

目 录

第 1 部分 熟悉 R 语言	1
第 1 章 R 的简介与安装	2
1.1 R 语言简介	2
1.2 RGui 的下载和安装	2
1.3 RGui 使用简要介绍	4
1.4 统计功能 Gui: R Commander	6
1.4.1 R Commander 的安装与加载	6
1.4.2 R Commander 简单操作	8
第 2 章 R 使用入门	13
2.1 R 代码编写	13
2.2 R 代码执行与脚本	14
2.3 R 脚本的保存与工作空间管理	15
2.3.1 R 脚本的保存	15
2.3.2 R 工作空间与工作目录	16
2.4 R 的帮助系统	17
2.4.1 单击“帮助”标签获取资源	17
2.4.2 R 函数获取帮助	18
第 3 章 R 包简介	22
3.1 包的安装与加载	22
3.1.1 单击下载安装包	22
3.1.2 函数下载安装包	23
3.1.3 本地安装包	23
3.2 包的加载	24
3.3 R 基础包	24
3.4 常用扩展包	25
第 4 章 RStudio 使用	27
4.1 RStudio 的下载和安装	27
4.2 Rstudio 的界面介绍	27
4.3 RStudio 的使用入门	28

4.3.1	自动补全功能	28
4.3.2	历史查询功能	29
4.3.3	其他标签的功能	30
4.3.4	RStudio 中脚本文件的使用	32
第 5 章	R 语言数据类型	34
5.1	几种常见的数据类型	34
5.2	数据类型的识别	36
5.3	数据类型的转换	36
第 6 章	R 语言数据结构	39
6.1	数据结构	39
6.2	向量	39
6.2.1	创建向量	39
6.2.2	向量元素的索引	42
6.3	矩阵	43
6.3.1	创建新矩阵	43
6.3.2	矩阵元素索引	44
6.4	数组	45
6.4.1	数组的创建	45
6.4.2	数组元素的索引	47
6.5	向量、矩阵、数组的联系与区别	48
6.5.1	向量和矩阵、数组的区别	49
6.5.2	矩阵与数组的联系与区别	51
6.6	因子	52
6.6.1	创建因子	52
6.6.2	选取因子中元素	54
6.7	数据框	54
6.7.1	创建数据框	55
6.7.2	访问数据框	56
6.8	列表	57
6.8.1	列表的创建	57
6.8.2	访问列表	58
6.9	变量的查看与删除	59
6.9.1	变量的查看	59
6.9.2	变量的删除	62

第 7 章 数据导入和导出	64
7.1 数据导入	64
7.1.1 read.table() 函数	64
7.1.2 读取 Excel 文件	65
7.1.3 读取 Stata、SAS 与 SPSS 的数据文件	66
7.1.4 读取网页数据	66
7.1.5 连接数据库	67
7.2 数据导出	68
第 8 章 数据编辑	70
8.1 编辑方式	70
8.2 变量命名	72
8.3 索引	73
8.4 数据结构转换	75
8.5 缺失值处理	75
第 9 章 数据整合	78
9.1 变量合并	78
9.2 列联表	79
9.3 reshape2 包	82
第 10 章 R 语言编程	85
10.1 流程控制	85
10.1.1 循环语句	85
10.1.2 条件语句	86
10.2 自编函数	87
10.3 数据操作	88
10.3.1 数学运算符	88
10.3.2 基本数据操作函数	89
10.3.3 字符型数据操作	92
10.4 apply 函数族	93
10.4.1 apply() 函数	94
10.4.2 tapply() 函数	94
10.4.3 lapply() 函数	95
第 11 章 R 语言绘图基础	97
11.1 一个简单的例子	97
11.2 修改图形属性	98
11.2.1 图形类型	98

11.2.2	颜色	99
11.2.3	大小	104
11.2.4	文本	105
11.2.5	par()	108
11.3	常见图形类型	109
11.3.1	柱状图	109
11.3.2	直方图与密度曲线图	112
11.3.3	饼图	113
11.3.4	箱线图	114
11.3.5	时间序列图	115
11.4	绘图窗口	116
11.4.1	绘图窗口	116
11.4.2	窗口分割	117
第 12 章	绘图系统 ggplot2	119
12.1	简介	119
12.2	使用 qplot() 作图	119
12.2.1	一个小例子	119
12.2.2	修改图形属性	121
12.2.3	绘制常见图形	123
12.2.4	分面	126
12.3	基本语法	127
12.3.1	数据和映射	128
12.3.2	标尺	129
12.3.3	统计变换和几何对象	130
12.4	使用 ggplot 作图	131
12.4.1	构建图层	131
12.4.2	映射函数	133
12.4.3	几何对象函数和统计变换函数	134
12.4.4	标尺函数	136
12.4.5	分面函数和坐标系统函数	139
12.4.6	图形输出	140
第 2 部分	统计学基础	142
第 13 章	描述性统计	143
13.1	数据类型	144
13.2	图表	144

13.2.1 频数分布表	144
13.2.2 直方图	145
13.3 数据的位置	145
13.4 数据的离散度	148
第 14 章 随机变量简介	152
14.1 概率与概率分布	152
14.1.1 离散型随机变量	152
14.1.2 连续型随机变量	153
14.2 期望值与方差	154
14.3 二项分布	155
14.4 正态分布 (Normal Distribution)	158
14.5 其他连续分布	160
14.5.1 卡方分布	160
14.5.2 t 分布	161
14.5.3 F 分布	162
14.6 变量的关系	163
14.6.1 联合概率分布	163
14.6.2 变量的独立性	164
14.6.3 变量的相关性	164
14.6.4 上证综指与深证综指的相关性分析	165
第 15 章 推断统计	169
15.1 参数估计	169
15.1.1 点估计	170
15.1.2 区间估计	170
15.2 案例分析	172
15.3 假设检验	175
15.3.1 两类错误	176
15.3.2 显著性水平与 p 值	176
15.3.3 确定小概率事件	177
15.4 t 检验	177
15.4.1 单样本 t 检验	178
15.4.2 独立样本 t 检验	179
15.4.3 配对样本 t 统计量的构造	180

第 16 章 方差分析	183
16.1 方差分析之思想	183
16.2 方差分析之原理	184
16.2.1 离差平方和	185
16.2.2 自由度	186
16.2.3 显著性检验	187
16.3 方差分析之 R 语言实现	188
16.3.1 单因素方差分析	188
16.3.2 多因素方差分析	189
16.3.3 析因方差分析	191
第 17 章 回归分析	193
17.1 一元线性回归模型	193
17.1.1 一元线性回归模型	193
17.1.2 最小平方方法	194
17.2 模型拟合度	195
17.3 古典假设条件下 $\hat{\alpha}$ 、 $\hat{\beta}$ 的统计性质	195
17.4 显著性检验	197
17.5 上证综指与深证成指的回归分析与 R 语言	197
17.5.1 R 语言拟合回归函数	198
17.5.2 R 语言回归诊断函数	199
17.6 多元线性回归模型	201
17.6.1 多元线性回归模型	202
17.7 多元线性回归案例分析	203
第 3 部分 金融基础、投资组合与量化选股	207
第 18 章 资产收益率和风险	208
18.1 单期与多期简单收益率	209
18.1.1 单期简单收益率	209
18.1.2 多期简单收益率	209
18.1.3 R 函数计算简单收益率	212
18.1.4 单期与多期简单收益率的关系	214
18.1.5 年化收益率	216
18.1.6 考虑股利分红的简单收益率	218
18.2 连续复利收益率	220
18.2.1 多期连续复利收益率	223
18.2.2 单期与多期连续复利收益率的关系	224

18.3 绘制收益图	225
18.4 资产风险的来源	226
18.4.1 市场风险	226
18.4.2 利率风险	227
18.4.3 汇率风险	227
18.4.4 流动性风险	227
18.4.5 信用风险	228
18.4.6 通货膨胀风险	228
18.4.7 营运风险	228
18.5 资产风险的测度	228
18.5.1 方差	228
18.5.2 下行风险	230
18.5.3 风险价值	231
18.5.4 期望亏空	233
18.5.5 最大回撤	233
第 19 章 投资组合理论及其拓展	239
19.1 投资组合的收益率与风险	239
19.2 Markowitz 均值-方差模型	243
19.3 Markowitz 模型之 R 语言实现	247
19.3.1 数据读取与整理	247
19.4 Black-Litterman 模型	252
第 20 章 资本资产定价模型	260
20.1 资本资产定价模型的核心思想	260
20.2 CAPM 模型的应用	261
20.3 R 语言计算单资产 CAPM 实例	263
20.4 CAPM 模型的评价	266
第 21 章 Fama-French 三因子模型	269
21.1 Fama-French 三因子模型的基本思想	269
21.2 三因子模型之 R 语言实现	271
21.3 三因子模型的评价	276
第 4 部分 时间序列基础与配对交易	278
第 22 章 时间序列基本概念	279
22.1 认识时间序列	279

22.2 R 中的时间序列分析包	280
22.3 时间序列数据处理函数	283
22.4 选取特定日期的时间序列数据	284
22.5 时间序列数据描述性统计	286
第 23 章 时间序列的基本性质	289
23.1 自相关性	289
23.1.1 自协方差	290
23.1.2 自相关系数	290
23.1.3 偏自相关系数	290
23.1.4 $acf()$ 函数与 $pacf()$ 函数	291
23.1.5 上证综指的收益率指数的自相关性判断	291
23.2 平稳性	295
23.2.1 强平稳	295
23.2.2 弱平稳	295
23.2.3 强平稳与弱平稳的区别	296
23.3 上证综指的平稳性检验	297
23.3.1 观察时间序列图	297
23.3.2 观察序列的自相关图和偏自相关图	298
23.3.3 单位根检验	299
23.4 白噪声	304
23.4.1 白噪声	304
23.4.2 白噪声检验——Ljung-Box 检验	305
23.4.3 上证综合指数的白噪声检验	307
第 24 章 时间序列预测	309
24.1 移动平均预测	309
24.1.1 简单移动平均	309
24.1.2 加权移动平均	310
24.1.3 指数加权移动平均	310
24.2 ARMA 模型预测	310
24.2.1 自回归模型	311
24.2.2 移动平均模型	313
24.3 自回归移动平均模型	314
24.4 ARMA 模型的建模过程	314
24.5 CPI 数据的 ARMA 短期预测	315
24.6 上证指数的平稳时间序列建模	322

第 25 章 GARCH 模型	327
25.1 资产收益率的波动率与 ARCH 效应	327
25.2 ARCH 模型和 GARCH 模型	327
25.2.1 ARCH 模型	327
25.2.2 GARCH 模型	329
25.3 ARCH 效应检验	330
25.4 GARCH 模型构建	332
25.5 GARCH 模型之 VaR 应用	336
第 26 章 配对交易策略	341
26.1 什么是配对交易?	341
26.2 配对交易的思想	342
26.3 配对交易的步骤	343
26.3.1 股票对的选择	343
26.3.2 配对交易策略的制定	355
26.3.3 多空股票的仓位配比	359
26.4 配对交易与 R 语言	360
26.4.1 PairTrading 包	360
26.4.2 R 语言实测配对交易策略	365
第 5 部分 技术指标与量化投资	377
第 27 章 K 线图	378
27.1 K 线图简介	378
27.2 R 绘制上证综指 K 线图	380
27.3 R 捕捉 K 线图的形态	384
27.3.1 R 语言捕捉“早晨之星”	384
27.3.2 R 语言捕捉“乌云盖顶”形态	389
第 28 章 动量交易策略	396
28.1 动量概念介绍	396
28.2 动量效应产生原因	396
28.3 价格动量的计算公式	397
28.3.1 作差法求动量值	397
28.3.2 作除法求动量值	399
28.4 R 中的动量相关函数	400
28.4.1 momentum() 函数	400
28.4.2 ROC() 函数	401

28.5 万科股票 2015 年走势及动量线	402
28.6 动量交易策略的一般思路	403
28.6.1 运用动量指标交易万科股票	403
第 29 章 RSI 相对强弱指标	410
29.1 RSI 基本概念	410
29.2 R 语言计算 RSI 值	410
29.3 TTR 包中的 RSI() 函数	417
29.4 RSI 天数的差异	418
29.5 RSI 指标判断股票超买和超卖状态	419
29.6 RSI 的“黄金交叉”与“死亡交叉”	420
29.7 交通银行股票 RSI 指标交易实测	421
29.7.1 RSI 捕捉交通银行股票买卖点	422
29.7.2 RSI 交易策略执行及回测	426
第 30 章 均线系统策略	431
30.1 简单移动平均	431
30.1.1 简单移动平均数	431
30.1.2 简单移动平均函数	434
30.1.3 期数选择	435
30.2 加权移动平均	435
30.2.1 加权移动平均数	435
30.2.2 加权移动平均函数	438
30.3 指数加权移动平均	438
30.3.1 指数加权移动平均数	438
30.3.2 指数加权移动平均函数	441
30.4 常用平均方法的比较	442
30.5 TTR 包中的平均函数	442
30.6 中国银行股价数据与均线分析	443
30.7 均线时间跨度	447
30.8 中国银行股票均线系统交易	448
30.8.1 简单移动均线制定中国银行股票的买卖点	448
30.8.2 双均线交叉捕捉中国银行股票的买卖点	452
30.9 异同移动均线 (MACD)	457
30.9.1 MACD 的求值过程	457
30.9.2 TTR 包中的 MACD() 函数	459
30.9.3 异同均线 (MACD) 捕捉中国银行股票的买卖点	460
30.10 多种均线指标综合运用模拟实测	463

第 31 章 通道突破策略	470
31.1 通道突破简介	470
31.2 唐奇安通道 (Donchian Channel)	470
31.2.1 唐奇安通道刻画	470
31.2.2 R 语言捕捉唐奇安通道突破	474
31.3 布林带 (Bollinger Band) 通道	478
31.3.1 布林带通道的计算方式	479
31.3.2 通道突破 BBands() 函数	481
31.4 布林带通道与市场风险	483
31.5 通道突破交易策略的制定	486
31.5.1 布林带上下通道突破策略	486
31.5.2 另一种布林带通道突破策略	488
第 32 章 随机指标 (KDJ) 交易策略	491
32.1 什么是随机指标 (KDJ)	491
32.2 随机指标 (KDJ) 的原理	491
32.3 KDJ 指标的计算公式	492
32.3.1 未成熟随机指标 RSV	492
32.3.2 K、D 指标计算	497
32.3.3 J 指标计算	501
32.3.4 KDJ 指标简要分析	502
32.4 KDJ 指标的交易策略	504
32.5 R 语言 KDJ 指标交易实测	504
32.5.1 KD 指标交易策略	504
32.5.2 KDJ 指标交易策略	508
32.5.3 K 线、D 线“金叉”与“死叉”	510
第 33 章 量价关系分析	516
33.1 量价关系概述	516
33.2 量价关系分析	516
33.2.1 价涨量增	516
33.2.2 价涨量平	518
33.2.3 价涨量缩	519
33.2.4 价平量增	520
33.2.5 价平量缩	520
33.2.6 价跌量增	520
33.2.7 价跌量平	521
33.2.8 价跌量缩	521

33.3 不同价格段位的成交量与 R 语言	522
33.4 成交量与均线思想结合制定交易策略	524
第 34 章 OBV 指标交易策略	532
34.1 OBV 指标概念	532
34.2 OBV 指标计算方法	532
34.3 OBV 指标的理论依据	536
34.4 OBV 指标的交易策略制定	536
34.5 OBV 指标交易策略的 R 语言实测	536
34.6 OBV 指标的应用原则	540