



量化投资 以 Python为工具

蔡立崙 著

Quantitative
Investment Using Python

国内率先面世基于Python的量化投资专著

源自作者多年研究与实战
有技术，有案例，有工具，有总结

 中国工信出版集团

 电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



量化投资 以 Python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

本书主要讲解量化投资思想和策略，并借助 Python 语言进行实战。本书一共分为 5 部分，第 1 部分是 Python 入门，第 2 部分是统计学基础，第 3 部分是金融理论、投资组合与量化选股，第 4 部分是时间序列简介与配对交易，第 5 部分是技术指标与量化投资。本书首先对 Python 编程语言进行介绍，通过学习，读者可以迅速掌握用 Python 语言处理数据的方法，并灵活运用 Python 解决实际金融问题；其次，向读者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和类型等方面；最后讲述如何在 Python 语言中构建量化投资策略。

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 Python 为工具 / 蔡立嵩著. — 北京：电子工业出版社，2017.2
(金融科技丛书)

ISBN 978-7-121-30514-6

I. ①量…II. ①蔡…III. ①投资—软件工具 IV. ①F830.59-39

中国版本图书馆 CIP 数据核字 (2016) 第 288007 号

策划编辑：高洪霞

责任编辑：黄爱萍

印 刷：涿州市京南印刷厂

装 订：涿州市京南印刷厂

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：34.5 字数：938 千字

版 次：2017 年 2 月第 1 版

印 次：2017 年 2 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, 88258888。

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。

本书咨询联系方式：(010) 51260888-819, faq@phei.com.cn。

序 言

过去十年，一股“量化投资”的热潮在中国悄然掀起。最近这两年，投资人对量化的关注更是到达了前所未有的地步。除了业界到处寻找量化团队以外，各种量化基金如雨后春笋般出现，各个大学校园也开始举办一场又一场的量化讲座、研讨会等。量化投资一时蔚为风行，产官学共襄盛举。

这么受人瞩目的议题，到底它的内涵是什么呢？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）。在资讯工具不发达的年代，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伤害投资绩效。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。正因为如此，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及的优势。若再善用计算机技术，量化交易者可以处理的资讯量更让主观交易者望尘莫及。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，虽然很重要，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做广泛与初步的介绍，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讯，可发邮件至 service@baoquant.com 索取，来信请在邮件标题中写明书名：《量化投资：以 Python 为工具》。

目 录

第 1 部分 Python 入门	1
第 1 章 Python 简介与安装使用	2
1.1 Python 概述	2
1.2 Python 的安装	3
1.2.1 下载安装 Python 执行文件	3
1.2.2 下载安装 Anaconda	4
1.2.3 多种 Python 版本并存	6
1.3 Python 的简单使用	7
1.4 交互对话环境 IPython	8
1.4.1 IPython 的安装	8
1.4.2 IPython 的使用	9
1.4.3 IPython 功能介绍	10
第 2 章 Python 代码的编写与执行	14
2.1 创建 Python 脚本文件	15
2.1.1 记事本	15
2.1.2 Python 默认的 IDLE 环境	15
2.1.3 专门的程序编辑器	15
2.2 执行 .py 文件	17
2.2.1 IDLE 环境自动执行	17
2.2.2 在控制台 cmd 中执行	18
2.2.3 在 Anaconda Prompt 中执行	19
2.3 Python 编程小技巧	20
2.3.1 Python 行	20
2.3.2 Python 缩进	21
第 3 章 Python 对象类型初探	23
3.1 Python 对象	23
3.2 变量命名规则	24

3.3	数值类型	25
3.3.1	整数	25
3.3.2	浮点数	26
3.3.3	布尔类型	26
3.3.4	复数	27
3.4	字符串	28
3.5	列表	29
3.6	可变与不可变	30
3.7	元组	32
3.8	字典	33
3.9	集合	33
第 4 章	Python 集成开发环境: Spyder 介绍	36
4.1	代码编辑器	37
4.2	代码执行 Console	39
4.3	变量查看与编辑	40
4.4	当前工作路径与文件管理	41
4.5	帮助文档与在线帮助	42
4.6	其他功能	43
第 5 章	Python 运算符与使用	44
5.1	常用运算符	44
5.1.1	算术运算符	45
5.1.2	赋值运算符	46
5.1.3	比较运算符	47
5.1.4	逻辑运算符	48
5.1.5	身份运算符	49
5.1.6	成员运算符	51
5.1.7	运算符的优先级	52
5.2	具有运算功能的内置函数	52
第 6 章	Python 常用语句	55
6.1	赋值语句	55
6.1.1	赋值含义与简单赋值	55
6.1.2	多重赋值	57

6.1.3	多元赋值	58
6.1.4	增强赋值	58
6.2	条件语句	59
6.3	循环语句	60
6.3.1	for 循环	60
6.3.2	while 循环	61
6.3.3	嵌套循环	62
6.3.4	break、continue 等语句	62
第 7 章	函数	66
7.1	函数的定义与调用	66
7.2	函数的参数	68
7.3	匿名函数	71
7.4	作用域	72
第 8 章	面向对象	75
8.1	类	75
8.2	封装	77
8.3	继承 (Inheritance)	79
第 9 章	Python 标准库与数据操作	82
9.1	模块、包和库	82
9.1.1	模块	82
9.1.2	包	84
9.1.3	库	85
9.2	Python 标准库介绍	85
9.3	Python 内置数据类型与操作	91
9.3.1	序列类型数据操作	91
9.3.1.1	list 类型与操作	91
9.3.1.2	tuple 类型与操作	95
9.3.1.3	range 类型与操作	97
9.3.1.4	字符串操作	98
9.3.2	字典类型操作	103
9.3.3	集合操作	106

第 10 章 常用第三方库：Numpy 库与多维数组	111
10.1 NumPy 库	111
10.2 创建数组	111
10.3 数组元素索引与切片	115
10.4 数组运算	118
第 11 章 常用第三方库：Pandas 与数据处理	120
11.1 Series 类型数据	120
11.1.1 Series 对象的创建	120
11.1.2 Series 对象的元素提取与切片	122
11.1.2.1 调用方法提取元素	122
11.1.2.2 利用位置或标签提取元素与切片	123
11.1.3 时间序列	124
11.2 DataFrame 类型数据	128
11.2.1 创建 DataFrame 对象	128
11.2.2 查看 DataFrame 对象	130
11.2.3 DataFrame 对象的索引与切片	131
11.2.4 DataFrame 的操作	135
11.2.5 DataFrame 的运算	139
11.3 数据规整化	142
11.3.1 缺失值的处理	142
11.3.1.1 缺失值的判断	142
11.3.1.2 选出不是缺失值的数据	143
11.3.2 缺失值的填充	143
11.3.3 缺失值的选择删除	145
11.3.4 删除重复数据	146
第 12 章 常用第三方库：Matplotlib 库与数据可视化	149
12.1 Matplotlib 简介	149
12.2 修改图像属性	152
12.2.1 坐标	152
12.2.1.1 更改坐标轴范围	152
12.2.1.2 设定坐标标签与显示角度	153
12.2.2 添加文本	155
12.2.2.1 添加标题	155

12.2.2.2	中文显示问题	157
12.2.2.3	设定坐标轴标签	159
12.2.2.4	增加图形背景 grid	160
12.2.2.5	增加图例	161
12.2.3	多种线条属性	162
12.2.3.1	线条的类型	162
12.2.3.2	图形的颜色	163
12.2.3.3	点的形状类型	164
12.2.3.4	线条宽度	166
12.3	常见图形的绘制	167
12.3.1	柱状图 (Bar charts)	167
12.3.2	直方图	170
12.3.3	饼图	172
12.3.4	箱线图	172
12.4	Figure、Axes 对象与多图绘制	173
12.4.1	Figure、Axes 对象	174
12.4.2	多图绘制	176
12.4.2.1	多个子图绘制	176
12.4.2.2	一个图中多条曲线绘制	178

第 2 部分 统计学基础 180

第 13 章 描述性统计 181

13.1	数据类型	182
13.2	图表	182
13.2.1	频数分布表	182
13.2.2	直方图	183
13.3	数据的位置	184
13.4	数据的离散度	186

第 14 章 随机变量简介 190

14.1	概率与概率分布	190
14.1.1	离散型随机变量	190
14.1.2	连续型随机变量	192
14.2	期望值与方差	193

14.3	二项分布	194
14.4	正态分布	197
14.5	其他连续分布	199
14.5.1	卡方分布	199
14.5.2	t 分布	199
14.5.3	F 分布	200
14.6	变量的关系	202
14.6.1	联合概率分布	202
14.6.2	变量的独立性	203
14.6.3	变量的相关性	203
14.6.4	上证综指与深证综指的相关性分析	205
第 15 章	推断统计	208
15.1	参数估计	208
15.1.1	点估计	209
15.1.2	区间估计	209
15.2	案例分析	212
15.3	假设检验	213
15.3.1	两类错误	214
15.3.2	显著性水平与 p 值	215
15.3.3	确定小概率事件	215
15.4	t 检验	216
15.4.1	单样本 t 检验	216
15.4.2	独立样本 t 检验	217
15.4.3	配对样本 t 统计量的构造	218
第 16 章	方差分析	221
16.1	方差分析之思想	221
16.2	方差分析之原理	222
16.2.1	离差平方和	223
16.2.2	自由度	224
16.2.3	显著性检验	225
16.3	方差分析之 Python 实现	226
16.3.1	单因素方差分析	227
16.3.2	多因素方差分析	228

16.3.3	析因方差分析	228
第 17 章	回归分析	231
17.1	一元线性回归模型	231
17.1.1	一元线性回归模型	231
17.1.2	最小平方法	232
17.2	模型拟合度	233
17.3	古典假设条件下 $\hat{\alpha}$ 、 $\hat{\beta}$ 之统计性质	234
17.4	显著性检验	235
17.5	上证综指与深证成指的回归分析与 Python 实践	236
17.5.1	Python 拟合回归函数	236
17.5.2	绘制回归诊断图	238
17.6	多元线性回归模型	240
17.7	多元线性回归案例分析	241
17.7.1	价格水平对 GDP 的影响	241
17.7.2	考量自变量共线性因素的新模型	243
第 3 部分	金融理论、投资组合与量化选股	246
第 18 章	资产收益率和风险	247
18.1	单期与多期简单收益率	248
18.1.1	单期简单收益率	248
18.1.2	多期简单收益率	249
18.1.3	Python 函数计算简单收益率	252
18.1.4	单期与多期简单收益率的关系	252
18.1.5	年化收益率	254
18.1.6	考虑股利分红的简单收益率	256
18.2	连续复利收益率	259
18.2.1	多期连续复利收益率	260
18.2.2	单期与多期连续复利收益率的关系	262
18.3	绘制收益图	263
18.4	资产风险的来源	264
18.4.1	市场风险	264
18.4.2	利率风险	264
18.4.3	汇率风险	265

18.4.4	流动性风险	265
18.4.5	信用风险	265
18.4.6	通货膨胀风险	266
18.4.7	营运风险	266
18.5	资产风险的测度	266
18.5.1	方差	266
18.5.2	下行风险	268
18.5.3	风险价值	269
18.5.4	期望亏空	271
18.5.5	最大回撤	271
第 19 章	投资组合理论及其拓展	276
19.1	投资组合的收益率与风险	276
19.2	Markowitz 均值-方差模型	280
19.3	Markowitz 模型之 Python 实现	285
19.4	Black-Litterman 模型	289
第 20 章	资本资产定价模型 (CAPM)	298
20.1	资本资产定价模型的核心思想	298
20.2	CAPM 模型的应用	299
20.3	Python 计算单资产 CAPM 实例	301
20.4	CAPM 模型的评价	305
第 21 章	Fama-French 三因子模型	308
21.1	Fama-French 三因子模型的基本思想	308
21.2	三因子模型之 Python 实现	310
21.3	三因子模型的评价	315
第 4 部分	时间序列简介与配对交易	317
第 22 章	时间序列基本概念	318
22.1	认识时间序列	318
22.2	Python 中的时间序列数据	320
22.3	选取特定日期的时间序列数据	321
22.4	时间序列数据描述性统计	323

第 23 章 时间序列的基本性质	326
23.1 自相关性	326
23.1.1 自协方差	327
23.1.2 自相关系数	327
23.1.3 偏自相关系数	327
23.1.4 $acf(\)$ 函数与 $pacf(\)$ 函数	328
23.1.5 上证综指的收益率指数的自相关性判断	328
23.2 平稳性	331
23.2.1 强平稳	331
23.2.2 弱平稳	332
23.2.3 强平稳与弱平稳的区别	332
23.3 上证综指的平稳性检验	333
23.3.1 观察时间序列图	333
23.3.2 观察序列的自相关图和偏自相关图	333
23.3.3 单位根检验	336
23.4 白噪声	340
23.4.1 白噪声	340
23.4.2 白噪声检验——Ljung-Box 检验	341
23.4.3 上证综合指数的白噪声检验	343
第 24 章 时间序列预测	345
24.1 移动平均预测	345
24.1.1 简单移动平均	345
24.1.2 加权移动平均	346
24.1.3 指数加权移动平均	346
24.2 ARMA 模型预测	347
24.2.1 自回归模型	348
24.2.2 移动平均模型	350
24.3 自回归移动平均模型	350
24.4 ARMA 模型的建模过程	351
24.5 CPI 数据的 ARMA 短期预测	351
24.5.1 序列识别	351
24.5.2 模型识别与估计	354
24.5.3 模型诊断	356
24.5.4 运用模型进行预测	359

24.6 股票收益率的平稳时间序列建模	359
第 25 章 GARCH 模型	364
25.1 资产收益率的波动率与 ARCH 效应	364
25.2 ARCH 模型和 GARCH 模型	365
25.2.1 ARCH 模型	365
25.2.2 GARCH 模型	366
25.3 ARCH 效应检验	368
25.4 GARCH 模型构建	370
第 26 章 配对交易策略	372
26.1 什么是配对交易	372
26.2 配对交易的思想	373
26.3 配对交易的步骤	374
26.3.1 股票对的选择	374
26.3.2 配对交易策略的制定	383
26.4 构建 PairTrading 类	387
26.5 Python 实测配对交易交易策略	391
第 5 部分 技术指标与量化投资	399
第 27 章 K 线图	400
27.1 K 线图简介	400
27.2 Python 绘制上证综指 K 线图	403
27.3 Python 捕捉 K 线图的形态	405
27.3.1 Python 捕捉“早晨之星”	406
27.3.2 Python 语言捕捉“乌云盖顶”形态	410
第 28 章 动量交易策略	416
28.1 动量概念介绍	416
28.2 动量效应产生的原因	416
28.3 价格动量的计算公式	417
28.3.1 作差法求动量值	417
28.3.2 做除法求动量值	418
28.4 编写动量函数 momentum()	420

28.5	万科股票 2015 年走势及 35 日动量线	420
28.6	动量交易策略的一般思路	423
第 29 章	RSI 相对强弱指标	429
29.1	RSI 基本概念	429
29.2	Python 计算 RSI 值	429
29.3	Python 编写 rsi() 函数	434
29.4	RSI 天数的差异	435
29.5	RSI 指标判断股票超买和超卖状态	436
29.6	RSI 的“黄金交叉”与“死亡交叉”	437
29.7	交通银行股票 RSI 指标交易实测	438
29.7.1	RSI 捕捉交通银行股票买卖点	438
29.7.2	RSI 交易策略执行及回测	440
第 30 章	均线系统策略	446
30.1	简单移动平均	446
30.1.1	简单移动平均数	446
30.1.2	简单移动平均函数	448
30.1.3	期数选择	449
30.2	加权移动平均	449
30.2.1	加权移动平均数	449
30.2.2	加权移动平均函数	451
30.3	指数加权移动平均	452
30.3.1	指数加权移动平均数	452
30.3.2	指数加权移动平均函数	454
30.4	创建 movingAverage 模组	454
30.5	常用平均方法的比较	455
30.6	中国银行股价数据与均线分析	456
30.7	均线时间跨度	458
30.8	中国银行股票均线系统交易	459
30.8.1	简单移动均线制定中国银行股票的买卖点	459
30.8.2	双均线交叉捕捉中国银行股票的买卖点	462
30.9	异同移动均线 (MACD)	464
30.9.1	MACD 的求值过程	464
30.9.2	异同均线 (MACD) 捕捉中国银行股票的买卖点	466

30.10 多种均线指标综合运用模拟实测	468
第 31 章 通道突破策略	473
31.1 通道突破简介	473
31.2 唐奇安通道	473
31.2.1 唐奇安通道刻画	473
31.2.2 Python 捕捉唐奇安通道突破	476
31.3 布林带通道	478
31.4 布林带通道与市场风险	481
31.5 通道突破交易策略的制定	484
31.5.1 一般布林带上下通道突破策略	484
31.5.2 特殊布林带通道突破策略	485
第 32 章 随机指标交易策略	489
32.1 什么是随机指标 (KDJ)	489
32.2 随机指标的原理	489
32.3 KDJ 指标的计算公式	490
32.3.1 未成熟随机指标 RSV	490
32.3.2 K、D 指标计算	495
32.3.3 J 指标计算	497
32.3.4 KDJ 指标简要分析	498
32.4 KDJ 指标的交易策略	499
32.5 KDJ 指标交易实测	499
32.5.1 KD 指标交易策略	499
32.5.2 KDJ 指标交易策略	503
32.5.3 K 线、D 线“金叉”与“死叉”	504
第 33 章 量价关系分析	509
33.1 量价关系概述	509
33.2 量价关系分析	509
33.2.1 价涨量增	510
33.2.2 价涨量平	512
33.2.3 价涨量缩	512
33.2.4 价平量增	513
33.2.5 价平量缩	514