

量化交易学习指南 ——基于R语言

Learning Quantitative Finance with R

[印度] 帕勒姆·吉特 (Param Jeet) 普拉桑特·瓦次 (Prashant Vats) 著

曾永艺 许健男 译

学习机器学习、时间序列分析等技术，解决实际的量化交易难题

Packt>



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

量化交易学习指南

——基于R语言

Learning Quantitative Finance with R

[印度] 帕勒姆·吉特 (Param Jeet) 普拉桑特·瓦次 (Prashant Vats) 著

曾永艺 许健男 译

人民邮电出版社
北 京

图书在版编目 (C I P) 数据

量化交易学习指南：基于R语言 / (印) 帕勒姆·吉特 (Param Jeet), (印) 普拉桑特·瓦次 (Prashant Vats) 著；曾永艺, 许健男译. — 北京：人民邮电出版社, 2019. 1
ISBN 978-7-115-49874-8

I. ①量… II. ①帕… ②普… ③曾… ④许… III. ①程序语言—应用—金融—经济数学 IV. ①F830

中国版本图书馆CIP数据核字(2018)第267654号

版 权 声 明

Copyright ©2017 Packt Publishing. First published in the English language under the title *Learning Quantitative Finance with R*.

All rights reserved.

本书由英国 Packt Publishing 公司授权人民邮电出版社出版。未经出版者书面许可，对本书的任何部分不得以任何方式或任何手段复制和传播。

版权所有，侵权必究。

-
- ◆ 著 [印度] 帕勒姆·吉特 (Param Jeet)
[印度] 普拉桑特·瓦次 (Prashant Vats)
译 曾永艺 许健男
责任编辑 胡俊英
责任印制 焦志炜
- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
天津翔远印刷有限公司印刷
- ◆ 开本：800×1000 1/16
印张：14.5
字数：279 千字 2019 年 1 月第 1 版
印数：1—2 400 册 2019 年 1 月天津第 1 次印刷
- 著作权合同登记号 图字：01-2018-3478 号
-

定价：69.00 元

读者服务热线：(010)81055410 印装质量热线：(010)81055316
反盗版热线：(010)81055315

内容提要

R 既是统计、挖掘、计算、分析、制图等方面的优秀工具，也是一个强大的开发与应用平台。在大数据时代，任何与数据相关的难题，都可以借助 R 语言来解决。而金融领域正是与数据密切相关的行业，可以通过 R 这一工具来实现量化金融建模与量化交易。

本书包括 9 章内容，书中包含诸多真实的金融案例，旨在通过循序渐进的讲解帮助读者了解 R，并学会在量化金融与量化交易中使用 R。本书还介绍了有关统计建模、计量分析与小波分析、时间序列建模、算法交易、基于机器学习的交易、风险管理、最优化、衍生品定价等重要内容。

本书适合对 R 及其应用感兴趣的读者阅读，尤其适合想要在量化交易中使用 R 的读者学习。本书并不要求读者具备 R 编程的知识，但希望读者对数学分析有一些了解。

作者简介

Param Jeet 从印度理工学院马德拉斯分校 (IITM) 获得数学博士学位。Param Jeet 博士在多个国际刊物上发表多篇数学方面的研究论文。过去几年, Param Jeet 博士进入数据分析行业, 并以数据科学家的身份为多家顶级的跨国公司工作或提供咨询服务。

感谢我的父母亲 S. Dhayan Singh 和 Jeet Kaur, 在我生命中的每个阶段给我提供一贯的支持; 感谢我的妻子 Manpreet Kaur, 总是把我摆在第一位, 全力支持并鼓励我撰写本书; 感谢我的孩子 Kavan Singh, 他天真无邪的笑容让我倍加珍爱家人并努力工作。还要感谢我的博士论文指导老师 Satyajit Roy 教授这么多年来给我指导。感谢同事和朋友, 如果没有他们的帮助, 这本书绝不可能完成。我希望能将我的知识分享给那些迫切想要用 R 学习量化金融的读者。

Prashant Vats 从印度理工学院获得数学硕士学位。Prashant 进入数据分析行业已有 10 年时间, 作为数据科学家供职于多家顶级的跨国公司, 提供多领域的咨询服务。

感谢我的父母亲 Devendra K. Singh 和 Sushila Sinha, 允许我追逐梦想并在我的职业生涯中给我以支持; 感谢我的妻子 Namrata, 在人生的每个阶段总陪在我身边, 支持我撰写本书; 感谢我的孩子 Aahan Vats, 他的微笑总是给我以灵感。还要感谢我的导师、同事和朋友, 如果没有他们的帮助, 这本书不可能完成。我希望能将自己的知识分享给更多有需要的人。

审稿人简介

Manuel Amunategui 是一位应用数据科学家，为多个行业（包括医疗、金融和销售）开发企业级的预测分析解决方案。在此之前，他作为量化开发人员在华尔街为一家大规模的股票和期权做市商工作 6 年，并曾作为软件开发人员在微软公司工作 4 年。他拥有美国西北大学预测分析专业的硕士学位以及国际训练学院国际管理专业的硕士学位。目前他是 SpringML 公司的数据科学副总裁。这是一家提供先进的预测性 CRM 分析建议、仪表盘和自动报告工具的创业公司。SpringML 公司的客户包括谷歌云平台、雪佛兰、雅马哈、Tesoro 和 Salesforce。他还是数据科学的倡导者、博主/视频博主（amunategui.github.io）以及 Udemy.com 和 O'Reilly Media 的培训师。

前言

本书借助统计语言 R 来阐释量化金融领域的实用例子。写作本书的目的在于向读者介绍如何利用 R 来学习量化金融。本书涉及从基础到高级的众多话题。具体而言，我们会介绍统计学、时间序列分析、小波分析以及它们在算法交易中的应用。在本书中，我们还会尽己所能地解释机器学习、风险管理、最优化、期权定价等领域知识的应用。

本书内容

第 1 章“R 语言入门”介绍 R 语言相关的基础知识，逐一说明 R 和 R 程序包的安装、数据类型、数据框、循环等知识。本章还说明如何编写和调用函数以及如何从不同格式的文档导入数据到 R。本章的目的在于向读者提供关于 R 语言的基础性说明。

第 2 章“统计建模”介绍如何进行探索性分析（如常用的概率分布、相关系数、趋中性度量、异常值侦测等）以更好地理解数据。本章还介绍数据抽样和数据标准化/归一化的技术，这有助于我们更好地为分析步骤准备数据。在本章的最后，我们还说明如何进行假设检验和参数估计。

第 3 章“计量分析与小波分析”解释简单线性回归模型和多元线性回归模型，它们是构成很多统计分析技术的支柱。本章还介绍 ANOVA 和特征选择的相关知识。最后，我们利用小波分析构建若干模型。

在第 4 章“时间序列建模”中，我们借助例子展示如何使用 `ts()`、`zoo()` 和 `xts()` 等函数将数据转换为时间序列数据，这是时间序列预测建模的基础性工作。接着，我们讨论不同的预测技术，如 AR、ARIMA、GARCH 和 VGARCH 等，并使用例子说明它们在 R

中如何实现。

第 5 章“算法交易”包含多个算法交易领域的生动示例，涵盖动量交易以及使用多种方法的配对交易等。本章还涉及 CAPM、多因子模型、投资组合构建等方面的知识。

第 6 章“基于机器学习的交易”展示如何使用资本市场的数据构建机器学习算法模型，具体包括监督型学习算法和非监督型学习算法。

第 7 章“风险管理”讨论度量市场风险和投资组合风险的技术，并解释计算 VAR 的常见方法。本章还给出银行业内度量信用风险的最佳实践示例。

第 8 章“最优化”展示金融领域应用最优化技术的不同例子，如动态再平衡、前行测试、网格测试、基因算法等。

第 9 章“衍生品定价”介绍如何用 R 进行衍生品定价。本章涵盖普通期权定价、奇异期权定价、债券定价、信用利差和信用违约互换等知识。本章相对复杂，要求读者具备关于衍生品的基础知识。

准备工作

首先，确保电脑上已经安装好 R 软件。书中的例子全部都用 R 来实现，相关代码可在 R 编辑器或控制台上直接执行。R 是开源软件，可免费从官网下载并安装到任何类型的操作系统中。在该网站上可找到安装指南。一旦安装好 R 软件后，你就可以从第 1 章开始学习。每个章节都会说明本章所需的 R 程序包以及如何把它们安装并加载到 R 的工作环境中。

目标读者

写作本书的目的在于向目标读者对学习 R 语言及其应用感兴趣的读者介绍知识。书中所用例子来自金融领域。这些例子有的比较基础，有的相对复杂，所用到的 R 代码的复杂性也有所不同。本书并不要求读者已经掌握 R 语言编程的知识，但读者要具备一些数学分析的概念。即使读者已经能够熟练使用 R，本书中用到的众多来自数据分析领域（特别是资本市场）的生动例子对这些读者肯定也会大有裨益。

排版惯例

在本书中，你会发现多种用来区分不同类型信息的文本格式。以下是一些文本格式的示例以及相应说明。

正文中的代码、数据表名、文件夹名、文件名、文件扩展名、路径、URL、用户输入、推特标签等使用如下格式：“我们会多次用到 `quantmod` 程序包。”

代码块示例如下：

```
getSymbols("^DJI", src = "yahoo")
dji <- DJI[, "DJI.Close"]
```

当我们需要你关注代码块的某个特定部分时，相应的代码行会加粗显示：

```
corr <- rollapply(data, 252, correlation, by.column = FALSE)
```

为了方便读者直接复制并执行相应代码，本书中输出的代码行前没有 R 控制台的命令提示符 `>`，长表达式的换行也不注明 `+`，而执行代码的输出结果前有两个注释符号，即 `##`。

新术语和**关键词**以粗体表示，例如以读者在屏幕上、菜单栏或对话框中看到的关键词在文中格式如下：“点击 **Next** 按钮进入下一界面。”

资源与支持

本书由异步社区出品，社区（<https://www.epubit.com/>）为您提供相关资源和后续服务。

配套资源

本书提供配套源代码，要获得该配套资源，请在异步社区本书页面中单击 **配套资源**，跳转到下载界面，按提示进行操作即可。注意：为保证购书读者的权益，该操作会给出相关提示，要求输入提取码进行验证。

如果您是教师，希望获得教学配套资源，请在社区本书页面中直接联系本书的责任编辑。

提交勘误

作者和编辑尽最大努力来确保书中内容的准确性，但难免会存在疏漏。欢迎您将发现的问题反馈给我们，帮助我们提升图书的质量。

当您发现错误时，请登录异步社区，按书名搜索，进入本书页面，单击“提交勘误”，输入勘误信息，单击“提交”按钮即可。本书的作者和编辑会对您提交的勘误进行审核，确认并接受后，您将获赠异步社区的 100 积分。积分可用于在异步社区兑换优惠券、样书或奖品。



The screenshot shows a web form titled "提交勘误" (Submit勘误) with three tabs: "详细报错" (Detailed Error Report), "写书评" (Write Book Review), and "提交勘误" (Submit勘误). The form contains three input fields: "页码:" (Page Number), "页内位置 (行号):" (Position within page (Line Number)), and "勘误内容:" (勘误内容:). Below these fields is a large text area for the error description. At the bottom right, there is a "提交" (Submit) button.

扫码关注本书

扫描下方二维码，您将会在异步社区微信服务号中看到本书信息及相关的服务提示。



与我们联系

我们的联系邮箱是 contact@epubit.com.cn。

如果您对本书有任何疑问或建议，请您发邮件给我们，并请在邮件标题中注明本书书名，以便我们更高效地做出反馈。

如果您有兴趣出版图书、录制教学视频，或者参与图书翻译、技术审校等工作，可以发邮件给我们；有意出版图书的作者也可以到异步社区在线提交投稿（直接访问 www.epubit.com/selfpublish/submission 即可）。

如果您是学校、培训机构或企业，想批量购买本书或异步社区出版的其他图书，也可以发邮件给我们。

如果您在网上发现有针对异步社区出版图书的各种形式的盗版行为，包括对图书全部或部分内容的非授权传播，请您将怀疑有侵权行为的链接发邮件给我们。您的这一举动是对作者权益的保护，也是我们持续为您提供有价值的内容的动力之源。

关于异步社区和异步图书

“异步社区”是人民邮电出版社旗下 IT 专业图书社区，致力于出版精品 IT 技术图书和相关学习产品，为作译者提供优质出版服务。异步社区创办于 2015 年 8 月，提供大量精品 IT 技术图书和电子书，以及高品质技术文章和视频课程。更多详情请访问异步社区官网 <https://www.epubit.com>。

“异步图书”是由异步社区编辑团队策划出版的精品 IT 专业图书的品牌，依托于人民邮电出版社近 30 年的计算机图书出版积累和专业编辑团队，相关图书在封面上印有异步图书的 LOGO。异步图书的出版领域包括软件开发、大数据、AI、测试、前端、网络技术等。



异步社区



微信服务号

目录

第 1 章 R 语言入门	1
1.1 为什么要用 R	1
1.2 下载并安装 R 软件	2
1.3 安装 R 程序包	3
1.3.1 直接从 CRAN 安装	3
1.3.2 手动安装	4
1.4 数据类型	4
1.4.1 向量	5
1.4.2 列表	6
1.4.3 矩阵	7
1.4.4 数组	7
1.4.5 因子	7
1.4.6 数据框	8
1.5 导入和输出不同类型的数据	8
1.5.1 读取和存储 CSV 格式文档	9
1.5.2 XLSX 文档	10
1.5.3 网络数据或在线数据资源	11
1.5.4 数据库	12
1.6 编写表达式	13
1.6.1 表达式	13
1.6.2 符号和赋值	15

1.6.3	关键词	15
1.6.4	变量命名	15
1.7	函数	16
1.7.1	调用函数（无需参数）	17
1.7.2	调用函数（需要参数）	17
1.8	如何执行 R 程序	18
1.8.1	在 R 代码窗中运行存盘的脚本文档	18
1.8.2	加载 R 脚本文档	18
1.9	循环与条件	19
1.9.1	if 语句	19
1.9.2	if ... else 语句	20
1.9.3	for 循环	20
1.9.4	while 循环	21
1.9.5	apply()函数	21
1.9.6	sapply()函数	21
1.10	循环控制语句	22
1.10.1	break 语句	22
1.10.2	next 语句	22
1.11	问题	23
1.12	小结	23
第 2 章	统计建模	25
2.1	概率分布	25
2.1.1	正态分布	26
2.1.2	对数正态分布	28
2.1.3	泊松分布	30
2.1.4	均匀分布	31
2.1.5	极值理论	31
2.2	抽样	33
2.2.1	随机抽样	33
2.2.2	分层抽样	35

2.3 统计量	36
2.3.1 均值	36
2.3.2 中位数	36
2.3.3 众数	36
2.3.4 概览	37
2.3.5 矩	37
2.3.6 峰度	37
2.3.7 偏度	38
2.4 相关性	38
2.4.1 自相关性	38
2.4.2 偏自相关性	39
2.4.3 交叉相关性	40
2.5 假设检验	40
2.5.1 总体均值的下侧单尾检验（方差已知）	40
2.5.2 总体均值的上侧单尾检验（方差已知）	41
2.5.3 总体均值的双尾检验（方差已知）	42
2.5.4 总体均值的下侧单尾检验（方差未知）	43
2.5.5 总体均值的上侧单尾检验（方差未知）	44
2.5.6 总体均值的双尾检验（方差未知）	44
2.6 参数估计	45
2.6.1 极大似然估计	45
2.6.2 线性模型	47
2.7 异常值侦测	48
2.7.1 箱形图	48
2.7.2 LOF 算法	48
2.8 标准化	49
2.9 归一化	50
2.10 问题	50
2.11 小结	51

第3章 计量分析与小波分析	52
3.1 简单线性回归	53
3.1.1 散点图	53
3.1.2 可决系数	55
3.1.3 显著性检验	55
3.1.4 置信区间	56
3.1.5 残差图	56
3.1.6 误差项正态分布	57
3.2 多元线性回归	58
3.2.1 可决系数	59
3.2.2 置信区间	60
3.2.3 多重共线性	60
3.3 ANOVA	60
3.4 特征选择	61
3.4.1 考察相关系数矩阵	62
3.4.2 逐步法变量选择	63
3.4.3 用分类技术进行变量选择	63
3.4.4 变量排序	64
3.5 小波分析	64
3.6 快速傅里叶变换	70
3.7 希尔伯特变换	71
3.8 问题	73
3.9 小结	74
第4章 时间序列建模	75
4.1 时间序列概述	76
4.2 将数据转化为时间序列	76
4.3 zoo 包	78
4.3.1 构建 zoo 对象	78
4.3.2 用 zoo 包读入外部数据	79
4.3.3 zoo 对象的优点	79

4.3.4 zoo 对象的缺点	81
4.4 xts 包	81
4.4.1 调用 as.xts() 构建 xts 对象	81
4.4.2 从头开始构建 xts 对象	82
4.5 线性过滤器	83
4.6 AR 模型	84
4.7 MA 模型	85
4.8 ARIMA 模型	86
4.9 GARCH 模型	90
4.10 EGARCH 模型	91
4.11 VGARCH 模型	92
4.12 动态条件相关性模型	94
4.13 问题	96
4.14 小结	96
第 5 章 算法交易	97
5.1 动量或方向性交易	98
5.2 配对交易	106
5.2.1 基于距离的配对交易	106
5.2.2 基于相关性的配对交易	112
5.2.3 基于协整关系的配对交易	115
5.3 资本资产定价模型	118
5.4 多因子模型	120
5.5 构建投资组合	126
5.6 问题	130
5.7 小结	130
第 6 章 基于机器学习的交易	131
6.1 逻辑回归	132
6.2 神经网络	138
6.3 深度神经网络	146
6.4 K 均值算法	148
6.5 K 近邻法	151

6.6	支持向量机	154
6.7	决策树	156
6.8	随机森林	157
6.9	问题	160
6.10	小结	161
第7章	风险管理	162
7.1	市场风险	162
7.2	投资组合风险	164
7.3	风险值	167
7.3.1	参数 VaR	167
7.3.2	历史 VaR	168
7.4	蒙特卡罗模拟	170
7.5	风险对冲	171
7.6	巴塞尔监管	172
7.7	信用风险	173
7.8	欺诈识别	177
7.9	负债管理	180
7.10	问题	180
7.11	小结	181
第8章	最优化	182
8.1	动态再平衡	182
8.2	前行测试	185
8.3	网格测试	185
8.4	遗传算法	188
8.5	问题	192
8.6	小结	192
第9章	衍生品定价	193
9.1	期权定价	193
9.1.1	Black-Scholes 模型	194
9.1.2	Cox-Ross-Rubinstein 模型	195