

O'REILLY®



Python

金融大数据分析

Python for Finance

[德] Yves Hilpisch 著
姚军 译

 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS

O'REILLY®

Python 金融大数据分析

[德] Yves Hilpisch 著

姚 军 译

人民邮电出版社
北京

图书在版编目 (C I P) 数据

Python金融大数据分析 / (德) 希尔皮斯科著 ; 姚军译. — 北京 : 人民邮电出版社, 2015. 12
ISBN 978-7-115-40445-9

I. ①P… II. ①希… ②姚… III. ①软件工具—程序设计—应用—金融—分析 IV. ①F830.41-39

中国版本图书馆CIP数据核字(2015)第244879号

版权声明

Copyright © 2015 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2015. Authorized translation of the English edition, 2015 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体字版由 O'Reilly Media, Inc. 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复制或抄袭。

版权所有, 侵权必究。

◆ 著 [德] Yves Hilpisch

译 姚 军

责任编辑 傅道坤

责任印制 张佳莹 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>

三河市海波印务有限公司印刷

◆ 开本: 787×1000 1/16

印张: 33

字数: 687 千字

2015 年 12 月第 1 版

印数: 1-2 500 册

2015 年 12 月河北第 1 次印刷

著作权合同登记号 图字: 01-2014-7519 号

定价: 99.00 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

Python 凭借其简单、易读、可扩展性以及拥有巨大而活跃的科学计算社区，在需要分析、处理大量数据的金融行业得到了广泛而迅速的应用，并且成为该行业开发核心应用的首选编程语言。本书提供了使用 Python 进行数据分析，以及开发相关应用程序的技巧和工具。

本书总计分为 3 部分，共 19 章，第 1 部分介绍了 Python 在金融学中的应用，其内容涵盖了 Python 用于金融行业的原因、Python 的基础架构和工具，以及 Python 在计量金融学中的一些具体入门实例；第 2 部分介绍了金融分析和应用程序开发中最重要的 Python 库、技术和方法，其内容涵盖了 Python 的数据类型和结构、用 matplotlib 进行数据可视化、金融时间序列数据处理、高性能输入/输出操作、高性能的 Python 技术和库、金融学中需要的多种数学工具、随机数生成和随机过程模拟、Python 统计学应用、Python 和 Excel 的集成、Python 面向对象编程和 GUI 的开发、Python 与 Web 技术的集成，以及基于 Web 应用和 Web 服务的开发；第 3 部分关注的是蒙特卡洛模拟期权与衍生品定价实际应用的开发工作，其内容涵盖了估值框架的介绍、金融模型的模拟、衍生品的估值、投资组合的估值、波动率期权等知识。

本书适合对使用 Python 进行大数据分析、处理感兴趣的金融行业开发人员阅读。

O'Reilly Media, Inc. 介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

— Wired

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

— Business 2.0

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

— CRN

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

— Irish Times

“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路遇到岔路口，走小路（岔路）。’回顾过去 Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

— Linux Journal

前言

不久以前，在金融行业，Python 作为一种编程语言和平台技术还被视为异端。相比之下，2014 年有许多大型金融机构——如美国银行、美林证券的“石英”项目或者摩根大通的“雅典娜”项目——战略性地使用了 Python 和其他既定的技术，构建、改进和维护其核心 IT 系统。众多大大小小的对冲基金也大量使用 Python 的功能，进行高效的金融应用程序开发和金融分析工作。

同样，当今许多金融工程硕士课程（或者授予类似学位的课程）也使用 Python 作为核心语言之一，教授计量金融理论与可执行计算机代码之间的转换方法。针对金融专业人士的教育项目和培训也越来越多地在课程中加入 Python。有些课程将它作为主要实现语言。

Python 最近取得这样的成功，而且在未来似乎还会继续下去，这有许多原因。其中包括它的语法、Python 开发人员可用的科学生态系统和数据分析库、易于和几乎所有其他技术集成，以及其开源地位（更多这方面的深入探讨请参见第 1 章）。

因此，有许多好的书籍，从不同角度和焦点传授 Python。本书是最先介绍和传授 Python 金融应用的书籍之一，特别是将 Python 用于计量金融学和金融分析。书中采用的方法很实用，实现和说明先于理论细节，通常将焦点更多地放在大局上，而非某些类或者函数晦涩难懂的参数化选项。

本书的大部分是在基于浏览器的强大交互式环境 IPython Notebook（在第 2 章中有更详细的介绍）中编写的，因此有可能为读者提供本书中几乎所有例子的可执行、交互式版本。

希望立即开始使用完备的交互式 Python（以及 R 和 Julia）金融分析环境的读者，应该前往 <http://oreilly.quant-platform.com>，尝试 Python Quant 平台（结合本书提供的 IPython Notebook 文件）。你还应该关注基于 Python 的金融分析库 DX analytics（<http://dx-analytics.com>）。我的另一本书《Derivatives Analytics with Python》（Wiley Finance）更详细地介绍高级衍生品分析的理论 and 数值方法，书中也提供了丰富而易用的 Python 代码。进一步的材料，特别是有关 Python 计量金融学应用的幻灯片及视频，可以在我的私人网站上找到（<http://hilpisch.com>）。

如果你想参加 Python 计量金融学应用社区的活动，在世界上的金融中心有各种各样的机会。例如，我自己在伦敦（http://www.meetup.com/Python-for-Quant_finance-London/）和纽约（http://www.meetup.com/Python-for-Quant_finance-NYC/）组织以此为焦点的讨

论组。每年还有多次 For Python Quants 会议和研讨 (<http://forpythonquants.com> 和 <http://pythonquants.com>)。

对于 Python 确立金融行业中重要技术地位这一事实，我确实感到很兴奋。我敢肯定，它在未来将会起到更重要的作用，例如在衍生品和风险分析或者高性能计算领域中。我希望本书能够帮助专业人士、研究人员和学生在面对这一迷人领域中的挑战时，最大限度地利用 Python。

本书的惯例



提示

提示这个图标用来强调一个提示、建议或一般说明。



警告

警告这个图标用来表示一个警告或注意事项。

代码示例的使用

补充材料（特别是 IPython Notebooks 和 Python 脚本/模块）可以从 <http://oreilly.quantplatform.com> 下载。

本书的目的是为了帮助读者完成工作。一般而言，你可以在你的程序和文档中使用本书中的代码，而且也没有必要取得我们的许可。但是，如果你要复制的是核心代码，则需要和我们打个招呼。例如，你可以在无需获取我们许可的情况下，在程序中使用本书中的多个代码块。但是，销售或分发 O'Reilly 图书中的代码光盘则需要取得我们的许可。通过引用本书中的示例代码来回答问题时，不需要事先获得我们的许可。但是，如果你的产品文档中融合了本书中的大量示例代码，则需要取得我们的许可。

在引用本书中的代码示例时，如果能列出本书的属性信息是最好不过。一个属性信息通常包括书名、作者、出版社和 ISBN。例如：“*Python for Finance* by Yves Hilpisch (O'Reilly). Copyright 2015 Yves Hilpisch, 978-1-491-94528-5.”

在使用书中的代码时，如果不确定是否属于正常使用，或是否超出了我们的许可，请通过 permissions@oreilly.com 与我们联系。

联系方式

如果你想就本书发表评论或有任何疑问，敬请联系出版社：

美国：

O'Reilly Media Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室（100035）

奥莱利技术咨询（北京）有限公司

我们还为本书建立了一个网页，其中包含了勘误表、示例和其他额外的信息。你可以通过如下地址访问该网页：

<http://bit.ly/python-finance>

关于本书的技术性问题或建议，请发邮件到：

bookquestions@oreilly.com

欢迎登录我们的网站（<http://www.oreilly.com>），查看更多我们的书籍、课程、会议和最新动态等信息。

Facebook: <http://facebook.com/oreilly>

Twitter: <http://twitter.com/oreillymedia>

YouTube: <http://www.youtube.com/oreillymedia>

Safari® 在线图书

Safari 在线图书是一个按需订阅的数字图书馆。它有不少于 7500 本技术和创意相关的书籍和视频供你参考和搜索。

通过订阅，你可以在线阅读任何页面或任何视频，甚至可以从手机或移动设备上在线阅读。你可以在书籍出版前访问到它们，并给读者发送反馈。其他功能还包括：复制和粘贴代码、组织收藏夹、下载和标记章节、做笔记、打印等。

O'Reilly Media 已经将本书英文版上传到 Safari 在线图书服务。在 <http://my.safaribooksonline.com> 上免费注册，你就可以访问本书所有章节以及类似主题的书籍。

致谢

我要感谢所有为本书的实现提供帮助的人，特别是认真反馈甚至完全理解书中例子的人，例如 Ben Lerner、James Powell、Michael Schwed、Thomas Wiecki 或者 Felix

Zumstein。同样，我要感谢审稿人 Hugh Brown、Jennifer Pierce、Kevin Sheppard 和 Galen Wilkerson。本书从他们宝贵的反馈和许多建议中受益良多。

本书还从我在 2013 年和 2014 年出席的会议和研讨活动的参与者们那里得到很多有益的反馈，这些会议包括：PyData、For Python Quants、Big Data in Quant Finance、EuroPython、EuroScipy、PyCon DE、PyCon Ireland、Parallel Data Analysis、Budapest BI Forum 和 CodeJam。我还在柏林、伦敦、纽约的 Python 讨论组上的演讲期间得到了宝贵的反馈。

最后，我要感谢家人，他们完全接受我集中地做自己最热爱的事。完成这样篇幅的书籍需要在一年之中、在我已经安排得很满的工作负荷和行程基础上投入大量的时间——我也必须花费更多的时间，独自坐在计算机前。因此，感谢 Sandra、Lilli 和 Henry 的理解与支持。我将本书献给可爱的妻子 Sandra，她是我们全家的核心。

Yves

德国萨尔州，2014 年 11 月

目录

第1部分 Python 与金融

第1章	为什么将 Python 用于金融	3
1.1	Python 是什么	3
1.1.1	Python 简史	5
1.1.2	Python 生态系统	5
1.1.3	Python 用户谱系	7
1.1.4	科学栈	7
1.2	金融中的科技	8
1.2.1	科技开销	9
1.2.2	作为业务引擎的科技	9
1.2.3	作为进入门槛的科技和人才	9
1.2.4	不断提高的速度、频率、数据量	10
1.2.5	实时分析的兴起	11
1.3	用于金融的 Python	12
1.3.1	金融和 Python 语法	12
1.3.2	Python 的效率和生产率	15
1.3.3	从原型化到生产	19
1.4	结语	20
1.5	延伸阅读	20
第2章	基础架构和工具	21
2.1	Python 部署	22
2.1.1	Anaconda	22
2.1.2	Python Quant Platform	27
2.1.3	工具	30
2.1.4	Python	30
2.1.5	IPython	30
2.1.6	Spyder	40
2.2	结语	42
2.3	延伸阅读	43
第3章	入门示例	45
3.1	隐含波动率	46

3.2	蒙特卡洛模拟	54
3.2.1	纯 Python	56
3.2.2	用 NumPy 向量化	57
3.2.3	利用对数欧拉方法实现全向量化	59
3.2.4	图形化分析	60
3.2.5	技术分析	62
3.3	结语	67
3.4	延伸阅读	68

第 2 部分 金融分析和开发

第 4 章	数据类型和结构	71
4.1	基本数据类型	72
4.1.1	整数	72
4.1.2	浮点数	73
4.1.3	字符串	75
4.2	基本数据结构	77
4.2.1	元组	77
4.2.2	列表	78
4.2.3	离题: 控制结构	80
4.2.4	离题: 函数式编程	81
4.2.5	字典	82
4.2.6	集合	84
4.3	NumPy 数据结构	85
4.3.1	用 Python 列表形成数组	85
4.3.2	常规 NumPy 数组	87
4.3.3	结构数组	90
4.4	代码向量化	91
4.5	内存布局	93
4.6	结语	95
4.7	延伸阅读	95
第 5 章	数据可视化	97
5.1	二维绘图	97
5.1.1	一维数据集	98
5.1.2	二维数据集	103
5.1.3	其他绘图样式	109
5.2	金融学图表	116
5.3	3D 绘图	119

5.4	结语	122
5.5	延伸阅读	122
第 6 章	金融时间序列	123
6.1	pandas 基础	124
6.1.1	使用 DataFrame 类的第一步	124
6.1.2	使用 DataFrame 类的第二步	127
6.1.3	基本分析	131
6.1.4	Series 类	134
6.1.5	GroupBy 操作	135
6.2	金融数据	136
6.3	回归分析	142
6.4	高频数据	150
6.5	结语	154
6.6	延伸阅读	154
第 7 章	输入/输出操作	155
7.1	Python 基本 I/O	156
7.1.1	将对象写入磁盘	156
7.1.2	读写文本文件	159
7.1.3	SQL 数据库	160
7.1.4	读写 NumPy 数组	162
7.2	Pandas 的 I/O	164
7.2.1	SQL 数据库	165
7.2.2	从 SQL 到 pandas	166
7.2.3	CSV 文件数据	168
7.2.4	Excel 文件数据	169
7.3	PyTables 的快速 I/O	170
7.3.1	使用表	170
7.3.2	使用压缩表	175
7.3.3	使用数组	176
7.3.4	内存外计算	177
7.4	结语	179
7.5	延伸阅读	180
第 8 章	高性能的 Python	181
8.1	Python 范型与性能	182
8.2	内存布局与性能	184
8.3	并行计算	186
8.3.1	蒙特卡洛算法	186

8.3.2	顺序化计算	187
8.3.3	并行计算	188
8.3.4	性能比较	191
8.4	多处理	191
8.5	动态编译	193
8.5.1	介绍性示例	193
8.5.2	二项式期权定价方法	195
8.6	用 Cython 进行静态编译	199
8.7	在 GPU 上生成随机数	201
8.8	结语	205
8.9	延伸阅读	205
第 9 章	数学工具	207
9.1	逼近法	208
9.1.1	回归	208
9.1.2	插值	218
9.2	凸优化	221
9.2.1	全局优化	222
9.2.2	局部优化	223
9.2.3	有约束优化	224
9.3	积分	226
9.3.1	数值积分	228
9.3.2	通过模拟求取积分	228
9.4	符号计算	229
9.4.1	基本知识	229
9.4.2	方程式	230
9.4.3	积分	231
9.4.4	微分	232
9.5	结语	233
9.6	延伸阅读	233
第 10 章	推断统计学	235
10.1	随机数	236
10.2	模拟	241
10.2.1	随机变量	241
10.2.2	随机过程	244
10.2.3	方差缩减	256
10.3	估值	259

10.3.1	欧式期权	259
10.3.2	美式期权	263
10.4	风险测度	266
10.4.1	风险价值	266
10.4.2	信用价值调整	270
10.5	结语	272
10.6	延伸阅读	273
第 11 章	统计学	275
11.1	正态性检验	276
11.1.1	基准案例	277
11.1.2	现实世界的数据	284
11.2	投资组合优化	289
11.2.1	数据	290
11.2.2	基本理论	291
11.2.3	投资组合优化	294
11.2.4	有效边界	296
11.2.5	资本市场线	297
11.3	主成分分析	300
11.3.1	DAX 指数和 30 种成分股	301
11.3.2	应用 PCA	301
11.3.3	构造 PCA 指数	302
11.4	贝叶斯回归	305
11.4.1	贝叶斯公式	305
11.4.2	PyMC3	306
11.4.3	介绍性示例	307
11.4.4	真实数据	310
11.5	结语	318
11.6	延伸阅读	318
第 12 章	Excel 集成	321
12.1	基本电子表格交互	322
12.1.1	生成工作簿 (.xls)	323
12.1.2	生成工作簿 (.xlsx)	324
12.1.3	从工作簿中读取	326
12.1.4	使用 OpenPyxl	328
12.1.5	使用 pandas 读写	329
12.2	用 Python 编写 Excel 脚本	332

12.2.1	安装 DataNitro	333
12.2.2	使用 DataNitro	333
12.3	xlwings	342
12.4	结语	342
12.5	延伸阅读	343
第 13 章	面向对象和图形用户界面	345
13.1	面向对象	345
13.1.1	Python 类基础知识	346
13.1.2	简单的短期利率类	350
13.1.3	现金流序列类	354
13.2	图形用户界面	356
13.2.1	带 GUI 的短期利率类	356
13.2.2	值的更新	358
13.2.3	带 GUI 的现金流序列类	360
13.3	结语	362
13.4	延伸阅读	362
第 14 章	Web 集成	365
14.1	Web 基础知识	366
14.1.1	ftplib	366
14.1.2	httplib	368
14.1.3	urllib	369
14.2	Web 图表绘制	372
14.2.1	静态图表绘制	372
14.2.2	交互式图表绘制	374
14.2.3	实时图表绘制	375
14.3	快速 Web 应用	383
14.3.1	交易者的聊天室	384
14.3.2	数据建模	384
14.3.3	Python 代码	385
14.3.4	模板	391
14.3.5	样式化	396
14.4	Web 服务	397
14.4.1	金融模型	399
14.4.2	实现	400
14.5	结语	406
14.6	延伸阅读	406

第3部分 衍生品分析库

第15章 估值框架	409
15.1 资产定价基本定理	409
15.1.1 简单示例	409
15.1.2 一般结果	410
15.2 风险中立折现	412
15.2.1 日期建模和处理	412
15.2.2 固定短期利率	413
15.3 市场环境	415
15.4 结语	418
15.5 延伸阅读	419
第16章 金融模型的模拟	421
16.1 随机数生成	422
16.2 泛型模拟类	423
16.3 几何布朗运动	427
16.3.1 模拟类	427
16.3.2 用例	429
16.4 跳跃扩散	431
16.4.1 模拟类	431
16.4.2 用例	434
16.5 平方根扩散	435
16.5.1 模拟类	435
16.5.2 用例	437
16.6 结语	438
16.7 延伸阅读	440
第17章 衍生品估值	441
17.1 泛型估值类	441
17.2 欧式行权	445
17.3 估值类	445
17.4 美式行权	451
17.4.1 最小二乘蒙特卡洛方法	451
17.4.2 估值类	453
17.4.3 用例	454
17.5 结语	457
17.6 延伸阅读	458

第 18 章 投资组合估值	459
18.1 衍生品头寸	460
18.1.1 类	460
18.1.2 用例	462
18.2 衍生品投资组合	463
18.2.1 类	463
18.2.2 用例	467
18.3 结语	472
18.4 延伸阅读	474
第 19 章 波动率期权	475
19.1 VSTOXX 数据	476
19.1.1 VSTOXX 指数数据	476
19.1.2 VSTOXX 期货数据	477
19.1.3 VSTOXX 期权数据	479
19.2 模型检验	480
19.2.1 相关市场数据	480
19.2.2 期权建模	481
19.2.3 检验过程	483
19.3 基于 VSTOXX 的美式期权	487
19.3.1 期权头寸建模	487
19.3.2 期权投资组合	488
19.4 结语	489
19.5 延伸阅读	490
附录 A 精选的最佳实践	491
附录 B 看涨期权类	499
附录 C 日期和时间	503