

HZ BOOKS
华章教育

WILEY

统计学精品译丛

金融数据分析导论

基于R语言

An Introduction to Analysis of Financial Data with R



(美) Ruey S. Tsay 著
芝加哥大学

李洪成 尚秀芬 郝瑞丽 译



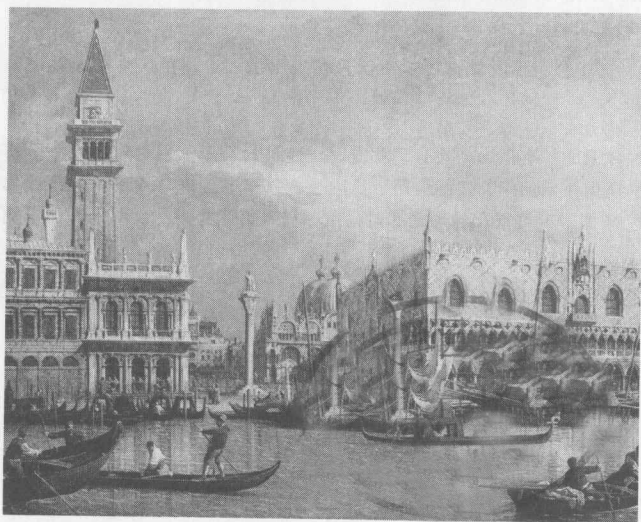
机械工业出版社
China Machine Press

统计学精品译丛

金融数据分析导论

基于R语言

An Introduction to Analysis of Financial Data with R



(美) Ruey S. Tsay 著
芝加哥大学

李洪成 尚秀芬 郝瑞丽 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

金融数据分析导论: 基于 R 语言 / (美) 蔡瑞胸著; 李洪成, 尚秀芬, 郝瑞丽译. —北京: 机械工业出版社, 2013. 10

(统计学精品译丛)

书名原文: An Introduction to Analysis of Financial Data with R

ISBN 978-7-111-43506-8

I. 金… II. ①蔡… ②李… ③尚… ④郝… III. 金融-数据-分析-教材 IV. F830.41

中国版本图书馆 CIP 数据核字 (2013) 第 176528 号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2013-1416

All Rights Reserved. This translation published under license. Authorized translation from the English language edition, entitled *An Introduction to Analysis of Financial Data with R*, ISBN 978-0-470-89081-3, by Ruey S. Tsay, Published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

本书中文简体字版由约翰·威利父子公司授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

本书向读者展示了可视化金融数据的基本概念, 共有 7 章内容, 涉及 R 软件、线性时间序列分析、资产波动率的不同计算方法、波动率模型在金融中的实际应用、高频金融数据的处理、用于风险管理的量化方法等。贯通全书, 作者都是通过 R 图形以可视化的形式把讨论主题展现给读者, 并以两个详细案例展示了金融中统计学的应用。

本书是高年级本科生或研究生阶段学习时间序列和商务统计学的优秀教材。对于希望进一步加强对金融数据和当今金融市场理解的研究人员以及商业、金融和经济领域的从业者, 该书也是极佳的选择。

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 盛思源

北京市荣盛彩色印刷有限公司印刷

2013 年 10 月第 1 版第 1 次印刷

186mm×240mm·19.75 印张

标准书号: ISBN 978-7-111-43506-8

定 价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

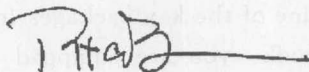
读者信箱: hzjsj@hzbook.com

推 荐 序

伟大的统计学家 George Box 有一句名言：“所有的模型都是错误的，但其中有一些是有用的。”这句话给出了现实中统计模型的现状。现实中不确定性无处不在，确定性“正确”模型是不存在的。如果 Box 在现在这个时代，他可能会增加一句“并且其中一些模型是危险的”，尤其对于金融领域的一些模型。2008 年的金融危机，在某种程度上是由错误的金融模型造成的，既有模型过于简单的原因又有模型过于复杂的原因。房地产经纪人和买家依赖于一个隐式模型，它表明价格已经在高位，并且还会继续上涨。贷款人使用统计模型来对打包的按揭产品进行分析设计，这似乎可以奇迹般地降低风险。然而最后的结果是灾难性的，在 5 年之后仍然可以感受到房地产泡沫的影响。

因此，如何进行有用的并且没有危险的金融分析呢？首先应该对金融数据有一个基本的理解，尤其是时间序列数据。因为不确定性是主要的影响因素，例如，可以应用概率模型来描述资产收益率的频率分布。本书给出了大量的时间序列模型，它们可以对数据进行描述、平滑和季节调整。

要成为一个统计分析专家，没有对实际数据的分析经验是绝不可能的。蔡瑞胸教授在本书中给出了进行实际数据分析所需的数据和统计工具。这里的统计分析工具是 R 软件，它是一款开源的统计软件包，可以和现在的商业软件包媲美。R 软件功能强大，免费并且有数千个用于完成特定任务的添加包。蔡教授在书中给出了金融数据分析的几个关键添加包。通过本书提供的数据集，你就可以应用 R 软件来学习金融时间序列的实际应用。



2013 年 6 月 17 日

美国统计教育学院，Statistics.com 在线课程网站总裁 Peter Bruce

Recommendation Preface

The great statistician George Box famously said “All models are wrong, but some are useful.” This statement embodies the statistical approach to reality, in which uncertainty is everpresent, and a deterministic “truth” does not exist. Box might now add “and some models are dangerous” particularly those in the financial arena. The financial collapse of 2008 was caused, in part, by faulty financial models, both at a simple and a complex level. Real estate brokers and buyers relied upon an implicit model that said that prices have been going up, so they will continue to go up. Lenders used statistical models to engineer bundled mortgage products that magically seemed to reduce risk. The result was a disastrous real estate bubble whose effects were still being felt half a decade later.

So how do you conduct a financial analysis that is useful, and not dangerous? The first thing to focus on is a basic understanding of financial data—particularly time series data. Since uncertainty plays a large role, probabilistic models are used, for example, to portray the frequency distribution of asset returns. There are a number of time series models that describe, smooth, and seasonally adjust the data.

Expertise in statistical analysis cannot be obtained without getting the hands dirty in actual data and analysis. Ruey Tsay provides both the data and the statistical tool you will need to get your hands dirty. The tool is R—the popular open-source statistical software that is now on a par with commercial programs. It is powerful, free and has spawned the creation of thousands of R “packages” to accomplish specific tasks. Tsay provides pointers to some of the key packages for financial analysis. Together with the datasets provided in the book, you are equipped to learn the practical application of financial time series using R.



Statistics.com

The Institute for Statistics Education

译者序

本书是原作者继《Analysis of Financial Time Series》之后的又一本力作。由于金融计量学和金融统计变得越来越重要，所以读者很迫切地需要这方面的资料。而国内相关的教材目前还比较少，机械工业出版社及时引进了《金融数据分析导论》的中文版权，使得国内读者能够几乎和国外同步地学习该书的内容。

本书应用开源的 R 软件，结合具体的金融数据来讲解金融数据的分析方法和模型。第 1 章首先向读者展示了可视化金融数据的基本概念、金融计量模型和它们在现实世界中的丰富应用。第 2 章从线性时间序列模型开始，介绍常见的金融数据分析模型。第 3 章介绍了 3 个金融数据分析的实际案例，读者可以从中学习线性时间序列模型在实际中的应用。第 4 章介绍了波动率模型。第 5 章介绍波动率模型在金融中的实际应用。第 6 章介绍如何处理高频金融数据。第 7 章介绍用量化方法进行风险管理，包括风险值和条件风险值。

对于实际的金融数据案例，本书应用免费的 R 软件给出了具体的分析过程和 R 代码。读者可以按照书中的讲授，一步一步地进行实践，加深对本书内容的理解。

在本书的翻译过程中，我们得到了蔡瑞胸教授的帮助，他帮助译者澄清了一些问题。中文版对原书中发现的错误进行了相应的更正。本书第 1 章、第 3 章和第 5 章由尚秀芬翻译，第 2 章、第 4 章、第 6 章、第 7 章由李洪成完成，其中郝瑞丽协助翻译了第 2 章的部分内容。全书由李洪成进行统一定稿。本书的责任编辑盛思源为本书的出版付出了大量的劳动，由于她的认真校对和修订，本书才能及时得以出版。另外，特别感谢明永玲编辑对本书翻译工作的大力支持和帮助。本书的翻译还得到了上海金融学院的领导和同事的帮助和支持，在此一并表示衷心的感谢。由于时间仓促，加之本人的精力和水平有限，翻译不当之处在所难免，请读者和同行指正。

李洪成

2013 年 6 月

前 言

我经常会被读者问到有关金融时间序列分析的问题：你能否让金融数据分析更简单些？我也经常被我的学生问道：如何简化实证研究工作，统计学和金融学究竟存在怎样的关系？这些都是很重要的问题，它们促使我编写这本入门教材。

为了简化实证数据分析，我决定使用 R 软件进行所有的分析。这样做是基于下面几个原因。第一，R 软件在大多数平台上都能免费使用。第二，已有许多研究人员针对金融数据开发出很好的软件包，尤其是 RMetrics 提供了许多有用的软件包。第三，R 软件添加包的功能正在飞速增加，并且这种趋势还会继续下去。第四，我写过一些简单的 R 代码，用以在某些场合展示某些具体分析方法。这些代码起两个作用。其一，它们能满足我展示概念和方法的特殊需要。其二，更重要的是，读者只要稍有一些 R 软件的使用经验就可以简单地修改这些代码和命令以适应他们的需要，从而简化他们的金融数据分析任务。

为了使计量经济学和统计学的理论更简单易懂，我尝试以一种简洁的方法来讲解，并应用大量实际案例来阐释这些内容。本书共有 7 章，其中两章为案例分析。这两章的案例揭示了统计学与金融学之间的联系。其余章节用来帮助读者理解概念，获取金融数据分析的经验。第 1 章介绍金融数据，并讨论了汇总统计量和数据可视化的知识。此外，还介绍了 R 软件，这样读者就可以应用它来探索金融数据。第 2 章介绍线性时间序列分析的基本内容，涵盖了商业、金融和经济学中常用的一些简单计量经济模型。在保持行文简洁的同时，我尝试使该章内容尽量全面。该章还介绍了指数平滑预测法和模型比较方法。第 3 章介绍了 3 个案例学习 (case study)。其中使用的模型并不简单，但是它们可以帮助读者理解线性时间序列模型在实际应用中的价值和局限性。第 4 章介绍资产波动率的不同计算方法，以及多种波动率模型。其中讨论的方法包括应用日开盘价、收盘价、最高价和最低价的方法。同样，我尝试使该章的内容尽量全面，同时避免过多的理论细节。第 5 章介绍波动率模型在金融中的实际应用。该章旨在帮助读者更好地理解波动率的期限结构，以及波动率在金融实践中的应用。第 6 章介绍如何处理高频金融数据，包括价格变动、交易强度以及实际波动率的简单模型。最后，第 7 章介绍用量化方法进行风险管理，包括风险值和条件风险值。该章涵盖风险评估的重要计量经济学方法和统计学方法，包含基于极值理论和分位数回归的一些方法。

本书包含许多图表和示例，它们旨在简化金融数据分析的过程，使结果容易理解。限于篇幅，不可避免地省略了一些重要内容。本书涵盖的内容与《Analysis of Financial Time Series》一书有部分重叠，但是本书所用到的案例都是全新的。

借此机会，我感谢我的妻子。如果没有她的关心和帮助，本书就无法完成。同时我

也想感谢我的孩子，他们给予我很多灵感，帮助我编辑部分章节。另外还有许多读者和学生不断给予我反馈和建议，他们的意见是弥足珍贵的。最后，我想感谢 Steve Quigley、Jacqueline Palmieri 和他们的 Wiley 团队给予我的支持和鼓励。

本书网站：<http://faculty.chicagobooth.edu/ruey.tsay/teaching/introTS>.

R. S. T

伊利诺伊州芝加哥市

2012 年 10 月

目 录

推荐序	
译者序	
前言	
第 1 章 金融数据及其特征	1
1.1 资产收益率	1
1.2 债券收益和价格	5
1.3 隐含波动率	7
1.4 R 软件包及其演示	8
1.4.1 R 软件包的安装	9
1.4.2 Quantmod 软件包	9
1.4.3 R 的基本命令	11
1.5 金融数据的例子	12
1.6 收益率的分布性质	14
1.7 金融数据的可视化	19
1.8 一些统计分布	23
1.8.1 正态分布	23
1.8.2 对数正态分布	23
1.8.3 稳态分布	24
1.8.4 正态分布的尺度混合	24
1.8.5 多元收益率	25
习题	27
参考文献	27
第 2 章 金融时间序列的线性模型	28
2.1 平稳性	30
2.2 相关系数和自相关函数	31
2.3 白噪声和线性时间序列	36
2.4 简单自回归模型	37
2.4.1 AR 模型的性质	38
2.4.2 实践中 AR 模型的识别	44
2.4.3 拟合优度	49
2.4.4 预测	50
2.5 简单移动平均模型	52
2.5.1 MA 模型的性质	53
2.5.2 MA 模型定阶	54
2.5.3 模型估计	55
2.5.4 用 MA 模型预测	55
2.6 简单 ARMA 模型	58
2.6.1 ARMA (1, 1) 模型的性质	59
2.6.2 一般 ARMA 模型	60
2.6.3 ARMA 模型的识别	60
2.6.4 用 ARMA 模型进行预测	63
2.6.5 ARMA 模型的三种表示方式	63
2.7 单位根非平稳性	65
2.7.1 随机游动	65
2.7.2 带漂移的随机游动	66
2.7.3 趋势平稳时间序列	68
2.7.4 一般单位根非平稳模型	68
2.7.5 单位根检验	69
2.8 指数平滑	72
2.9 季节模型	74
2.9.1 季节差分	75
2.9.2 多重季节模型	77
2.9.3 季节哑变量	82
2.10 带时间序列误差的回归模型	84
2.11 长记忆模型	89
2.12 模型比较和平均	92
2.12.1 样本内比较	92
2.12.2 样本外比较	92
2.12.3 模型平均	96
习题	96
参考文献	97
第 3 章 线性时间序列分析案例学习	99
3.1 每周普通汽油价格	99
3.1.1 纯时间序列模型	100

3.1.2	原油价格的使用	102	4.9.2	模型的另一种形式	170
3.1.3	应用滞后期的原油价格数据	103	4.9.3	第二个示例	170
3.1.4	样本外预测	104	4.9.4	用 EGARCH 模型进行预测	172
3.2	全球温度异常值	108	4.10	门限 GARCH 模型	173
3.2.1	单位根平稳	109	4.11	APARCH 模型	175
3.2.2	趋势非平稳	112	4.12	非对称 GARCH 模型	177
3.2.3	模型比较	114	4.13	随机波动率模型	179
3.2.4	长期预测	116	4.14	长记忆随机波动率模型	180
3.2.5	讨论	117	4.15	另一种方法	181
3.3	美国月失业率	121	4.15.1	高频数据的应用	181
3.3.1	单变量时间序列模型	121	4.15.2	应用日开盘价、最高价、 最低价和收盘价	183
3.3.2	一个替代模型	125	习题		187
3.3.3	模型比较	128	参考文献		188
3.3.4	使用首次申请失业救济金 人数	128	第 5 章	波动率模型的应用	190
3.3.5	模型比较	135	5.1	GARCH 波动率期限结构	190
习题		135	5.2	期权定价和对冲	194
参考文献		136	5.3	随时间变化的协方差和 β 值	196
第 4 章	资产波动率及其模型	137	5.4	最小方差投资组合	203
4.1	波动率的特征	137	5.5	预测	207
4.2	模型的结构	138	习题		214
4.3	模型的建立	140	参考文献		214
4.4	ARCH 效应的检验	141	第 6 章	高频金融数据	215
4.5	ARCH 模型	143	6.1	非同步交易	215
4.5.1	ARCH 模型的性质	144	6.2	交易价格的买卖报价差	218
4.5.2	ARCH 模型的优点与缺点	145	6.3	交易数据的经验特征	220
4.5.3	ARCH 模型的建立	145	6.4	价格变化模型	224
4.5.4	例子	149	6.4.1	顺序概率值模型	224
4.6	GARCH 模型	154	6.4.2	分解模型	228
4.6.1	实例说明	156	6.5	持续期模型	232
4.6.2	预测的评估	163	6.5.1	日模式的成分	233
4.6.3	两步估计方法	164	6.5.2	ACD 模型	235
4.7	求和 GARCH 模型	164	6.5.3	估计	237
4.8	GARCH-M 模型	166	6.6	实际波动率	241
4.9	指数 GARCH 模型	168	6.6.1	处理市场微结构噪声	247
4.9.1	第一个示例	169	6.6.2	讨论	249

附录 A 概率分布概览	251	7.6 极值理论	280
附录 B 危险率函数	253	7.6.1 极值理论概览	280
习题	254	7.6.2 经验估计	282
参考文献	255	7.6.3 股票收益率的应用	284
第 7 章 极值理论、分位数估计 与 VaR	257	7.7 极值在 VaR 中的应用	288
7.1 风险测度和一致性	257	7.7.1 讨论	289
7.1.1 风险值	258	7.7.2 多期 VaR	290
7.1.2 期望损失	262	7.7.3 收益率水平	290
7.2 计算风险度量的注记	263	7.8 超出门限的峰值	291
7.3 风险度量制	264	7.8.1 统计理论	292
7.3.1 讨论	267	7.8.2 超额均值函数	293
7.3.2 多个头寸	268	7.8.3 估计	294
7.4 VaR 计算的计量经济学方法	270	7.8.4 另外一种参数化方法	296
7.5 分位数估计	275	7.9 平稳损失过程	298
7.5.1 分位数与次序统计量	276	习题	299
7.5.2 分位数回归	277	参考文献	300
		索引	302

第 1 章 金融数据及其特征

近年来,数量分析方法在商业和金融市场上的重要性持续增加,因为我们有丰富的数据环境,经济和金融市场的数据库比以前更加综合和完整。在许多国家,成百上千个变量的数据可以更系统、更精确地搜集,计算处理上的便利和统计软件包的使用使得对复杂的高维金融数据的分析成为可能。通过互联网,人们可以很容易地应用开源软件包下载公开的金融数据,比如 R 软件。所有这些软件的特性和功能免费公开,因而被广泛使用。

该书的目的是提供金融时间序列分析的基本知识,介绍分析金融数据的各种统计工具,获得各种计量经济学方法的金融数据应用经验。第 1 章,我们首先分析了本书有关的基本概念,通过实例引入了 R 软件。接下来,我们讨论 R 软件中用不同的方法可视化这些金融数据。第 2 章回顾了非线性时间序列的基本概念,如平稳性和自相关函数,介绍处理数据中序列依存关系的简单的线性模型。随后,讨论具有时间序列误差的回归模型、季节性调整以及单位根的非平稳性和长期记忆的过程。本章还讨论了用指数平滑法进行预测以及不同模型之间的比较。第 3 章对第 2 章介绍的模型进行具体应用,给出一些案例学习,有助于帮助读者更好地了解和金融数据、实证模型以及进行统计推断,同时也指出了线性时间序列模型在长期预测中的局限性。第 4 章主要讨论条件异方差模型的建模(即资产收益率的条件方差)。首先介绍刻画资产波动率随时间演变的计量经济模型。本章还讨论了不同的波动率模型的其他替代方法,包括使用资产的每日最高及最低价格来建模。第 5 章演示了不同波动率模型的应用,使用一些案例学习,给出波动率模型的建模步骤和过程,并对各种波动率模型的优缺点进行讨论,包括扩散限制的连续时间模型。第 6 章是关于高频金融数据的分析的。首先给出了高频数据的特性,分析高频数据相应的模型和方法,应用非同步交易和竞价反弹策略来研究股票收益率之间的序列相关问题。此外,还研究了交易间隔时间的动态性和分析交易数据的计量经济模型。特别地,我们讨论了应用逻辑线性回归模型和概率模型(probit model)来分析股票价格在连续交易中的运动过程。最后,使用日内对数收益率数据来研究实际波动率。第 7 章讨论金融头寸的风险度量及其在风险管理中的应用,引入了风险值和条件风险值来量化在金融头寸持有期间的各种风险,包括 RiskMetrics 模型、极值理论、分位数回归以及阈值峰值理论。

1

本书非常重视模型的应用和实证分析,每章都包含了很多实际的例子。在很多时候,金融时间序列的经验特征是计量经济模型发展的动因。对某些案例,本书网站上给出了某些具体分析简单 R 脚本。每章后面的习题都给出了一些真实的数据可供使用。

1.1 资产收益率

大多数金融研究都是针对资产收益率,而不是资产价格。Campbell 等(1997)给出了使用资产收益率的两个主要原因。首先,对于一个普通的投资者来说,资产收益率代表一个完全的、尺度自由的投资机会的总结和概括。其次,资产收益率序列比价格序列更容易处

理,前者有更好的统计特性.然而,资产收益率有多种不同的定义.

设 P_t 是 t 时刻的资产价格,下面给出本书中用到的一些资产收益率的定义.暂时假定该过程资产不支付红利.

2 单期简单收益率

假设投资者在一个周期内拥有某种资产,从第 $t-1$ 天到第 t 天,其简单毛收益率为:

$$1 + R_t = \frac{P_t}{P_{t-1}} \text{ 或 } P_t = P_{t-1}(1 + R_t) \quad (1-1)$$

相对应的单期简单净收益率(simple net return)或简单收益率(simple return)为:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1-2)$$

表 1-1 给出了 2011 年 12 月苹果公司股票每日收盘价.从表 1-1 中可知,从 2011 年 12 月 8 日到 12 月 9 日,持有苹果公司股票每日的总收益率为 $1 + R_t = 393.62/390.66 \approx 1.0076$.相应的每日简单收益率为 0.76%,即 $(393.62 - 390.66)/390.66 = 0.76\%$.

多期简单收益率

假设从第 $t-k$ 天到第 t 天,这 k 个周期内持有某种资产,则 k 期简单毛收益率为:

$$\begin{aligned} 1 + R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j}) \end{aligned}$$

这样, k 期简单毛收益率是其包含的这 k 个单期简单毛收益率的乘积,称为复合收益率(compound return). k 期简单净收益率为 $R_t[k] = (P_t - P_{t-k})/P_{t-k}$.

为了说明这一点,表 1-1 给出了苹果公司股票的每日收盘价.由于 2011 年 12 月 2 日和 9 日是星期五,所以每周的股票简单毛收益率为 $1 + R_t[5] = 393.62/389.70 \approx 1.0101$,即每周的简单收益率为 1.01%.

表 1-1 苹果公司股票从 2011 年 12 月 2 日到 12 月 9 日的每日收盘价

日期	12/02	12/05	12/06	12/07	12/08	12/09
价格(\$)	389.70	393.01	390.95	389.09	390.66	393.62

在实际中,确切的时间区间对讨论和比较收益率是非常重要的(例如月收益率还是年收益率).若时间区间没有给出,这里隐含的假定时间区间为一年.如果持有资产的期限为 k 年,则(平均)年度化收益率定义为

$$\text{年度化的 } \{R_t[k]\} = \left[\prod_{j=0}^{k-1} (1 + R_{t-j}) \right]^{1/k} - 1$$

这是由它所包含的 k 个单期简单毛收益率几何平均得到的,可用下式计算:

$$\text{年度化的 } \{R_t[k]\} = \exp \left[\frac{1}{k} \sum_{j=0}^{k-1} \ln(1 + R_{t-j}) \right] - 1$$

其中 $\exp(x)$ 表示指数函数, $\ln(x)$ 是正数 x 的自然对数. 因为算术平均值比几何平均值计算起来容易, 并且单期收益率一般很小, 所以我们可用一阶泰勒 (Taylor) 展开来近似表示年度化的收益率, 则有

$$\text{年度化的 } \{R_t[k]\} \approx \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j} \quad (1-3)$$

然而, 在有些应用中, 式(1-3)的近似精确度可能不够.

连续复利收益率

在引进连续复利收益率之前, 先讨论复利收益率的效果. 假定银行存款的年利率为 10%, 最初的存款为 1 美元. 如果银行每年支付一次利息, 则一年后存款的净值变为 $1 \times (1+0.1) = 1.1$ 美元. 如果该银行半年付息一次, 则六个月的利息率为 $10\%/2 = 5\%$, 一年后净值为 $1 \times (1+0.1/2)^2 = 1.1025$ 美元. 通常, 如果银行一年付息 m 次, 则每次支付的利率为 $10\%/m$, 一年后存款的净值变成了 $1 \times (1+0.1/m)^m$ 美元. 表 1-2 给出了年利率为 10% 时, 一些常用的时间间隔下存款 1 美元的净值结果. 尤其是, 净值 $\exp(0.1) \approx 1.1052$ 美元, 这个值是连续复利的结果 (result of continuous compounding). 我们可以清楚地看到复利的效果.

表 1-2 收益率复利效果的演示 (期限为 1 年, 年利率为 10%)

类型	支付次数	每期利率	净值
一年	1	0.1	\$ 1.100 00
半年	2	0.05	\$ 1.102 50
季度	4	0.025	\$ 1.103 81
每月	12	0.0083	\$ 1.104 71
每周	52	0.1/52	\$ 1.105 06
每天	365	0.1/365	\$ 1.105 16
连续	∞		\$ 1.105 17

通常, 连续复利的资产净值 A 可以表示为

$$A = C \exp(r \times n) \quad (1-4)$$

其中 r 是年利率, C 是初始资本, n 是持有资产的年数. 由式(1-4), 则有

$$C = A \exp(-r \times n) \quad (1-5)$$

这里, C 是从现在开始 n 年后价值为 A 的资产的现值. 假定 r 是用连续复利表示的年利率.

连续复合收益率

资产的简单毛收益率的自然对数称为连续复合收益率或对数收益率(log-return):

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1} \quad (1-6)$$

其中, $p_t = \ln(P_t)$, 与简单净收益率 R_t 相比, 连续复合收益率 r_t 有很多优点. 首先, 对多期收益率, 我们有

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) = \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \cdots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$

这样, 连续复合多期收益率是它所包含的连续复合单期收益率之和. 其次, 对数收益率具有更容易处理的统计特性.

为了说明这一点, 我们再次考虑表 1-1 中苹果公司股票的日收盘价格. 从 12 月 8 日到

12月9日的日对数收益率为 $r_t = \log(393.62) - \log(390.66) \approx 0.75\%$, 从12月2日到12月9日的周对数收益率为 $r_t[5] = \log(393.62) - \log(389.70) \approx 1.00\%$. 显而易见, 周对数收益率是该周的5个日对数收益率之和.

资产组合收益率

若一个资产组合由 N 项资产组成, 则该资产组合的简单净收益率是它所包含的各项资产的简单净收益率的加权平均, 其中每个资产的权重是该资产的价值占资产组合总价值的百分比. 设 p 是一个资产组合, 它在第 i 项资产上的权重为 w_i , 那么 p 在 t 时刻的简单收益率为 $R_{p,t} = \sum_{i=1}^N w_i R_{i,t}$, 其中 $R_{i,t}$ 是组合中第 i 个资产的简单收益率.

然而, 资产组合的连续复合收益率没有上述方便的性质. 如果简单收益率 $R_{i,t}$ 的绝对值很小, 我们有 $r_{p,t} \approx \sum_{i=1}^N w_i r_{i,t}$, 其中 $r_{p,t}$ 是该组合在 t 时刻的连续复合收益率. 经常用这种近似来研究资产组合的收益率.

红利支付

如果一项资产周期性地支付红利, 则需要修改资产收益率的定义. 设 D_t 是一项资产在第 $t-1$ 天和第 t 天之间支付的红利, P_t 是该资产在第 t 个周期末的价格. 这样, 红利并没有包含在 P_t 中. 则 t 时刻的简单净收益率和连续复合收益率分别为

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, \quad r_t = \ln(P_t + D_t) - \ln(P_{t-1})$$

超额收益率

一项资产在 t 时刻的超额收益率是该项资产的收益率与某项参照资产的收益率之差. 参照资产通常是无风险的, 如美国短期国债收益率. 简单超额收益率和对数超额收益率分别定义为

$$Z_t = R_t - R_{0,t}, \quad z_t = r_t - r_{0,t} \quad (1-7)$$

其中, $R_{0,t}$ 和 $r_{0,t}$ 分别是该参照资产的简单收益率和对数收益率. 在金融学文献中, 超额收益率被认为是一个套利投资组合的盈利. 在这个投资组合中, 对某资产持有多头头寸而对其参照资产持有空头头寸, 且初始净投资额为零.

注记: 多头金融头寸是指持有某资产. 空头头寸是指卖出不属于自己的资产, 这需要通过从已购买该资产的投资者那里借入资产来实现. 随后的某天, 卖空者有义务买进与借入时完全相同数量的份额来偿还借出者. 偿还时要求相等数量的份额, 而不是相等数量的美元, 因此卖空者会由于该资产价格的下跌而获利. 如果在空头持有期间该资产支付现金红利, 这些红利将支付给做空合约的买方. 卖空者必须从自己的资源里支付相应的现金红利来补偿借出者. 换句话说, 空头方有义务支付所借资产的现金红利给资产的借出方. ■

关系小结

简单收益率 R_t 与连续复合对数收益率 r_t 的关系是

$$r_t = \ln(1 + R_t), \quad R_t = e^{r_t} - 1$$

如果收益率 R_t 与 r_t 用百分数表示, 则有

$$r_t = 100 \ln \left(1 + \frac{R_t}{100} \right), \quad R_t = 100 (e^{r_t/100} - 1)$$

把收益率进行时间累加, 则有

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})$$

$$r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k+1}$$

如果连续复合年利率为 r , 则资产现值与资产的未来价值之间的关系为

$$A = C \exp(r \times n), \quad C = A \exp(-r \times n)$$

例 1.1 若某项资产的月对数收益率为 4.46%, 则相应的月简单收益率是 $100[\exp(4.46/100) - 1] = 4.56\%$. 同样, 若某项资产在一个季度内的月对数收益率分别为 4.46%、-7.34% 和 10.77%, 则该资产季度的对数收益率为 $(4.46 - 7.34 + 10.77)\% = 7.89\%$. ■

图 1-1 显示了 IBM 股票的日简单收益率及其对数收益率的时序图, 从 2001 年 1 月 2 日到 2010 年 12 月 31 日, 共有 2515 个观测值. 从图 1-1 可知, 和简单收益率相比, 对数收益率更简单. 事实上, 简单收益率和对数收益率之间的相关系数为 0.9997. 这是可以理解的, 因为当 x 接近零时, $\log(1+x) \approx x$. 在样本持续期内, IBM 股票的日简单收益率的数额很小.

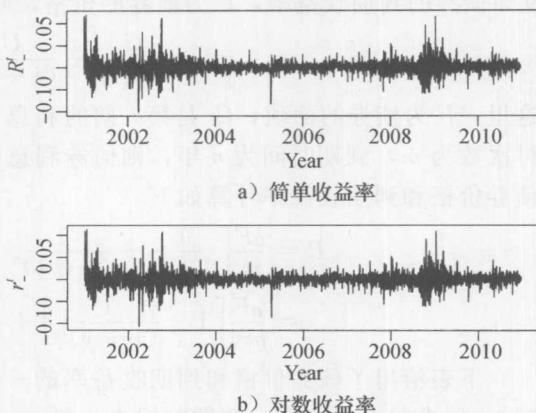


图 1-1 IBM 股票从 2001 年 1 月 2 日到 2010 年 12 月 31 日的日收益率时序图

1.2 债券收益和价格

债券是一种金融工具, 在到期日, 向债券的持有者支付票面价值(或面值). 有些债券根据票面利率(coupon payment)定期支付利息. 而零息债券则不需要定期支付利息. 债券收益率是持有债券至到期日, 投资者将收取的回报. 在金融上, 有几种类型的常用债券. 这些债券的收益率是指当期收益率和到期收益率(Yield To Maturity, YTM).

当期收益率

当期收益率是指每年支付给投资者回报的百分数. 从数学上, 我们有

$$\text{当期收益率} = \frac{\text{支付的年度利息额}}{\text{债券的市场价格}} \times 100\%$$

例如, 一个投资者购买债券支付 90 美元, 债券面值(也称为面值(par value))为 100 元, 债券的票面利率为每年 5%, 则该债券当期收益率为

$$c_t = (0.05 \times 100) / 90 \times 100\% = 5.56\%$$

这里, 我们用下标 t 表示收益率通常随时间变化. 从定义可知, 当期收益率不包含任何投资过程中的资本收益或损失. 对于零息债券, 收益率的计算方法如下:

$$\text{当期收益率} = \left(\frac{\text{面值}}{\text{购买价格}} \right)^{1/k} - 1$$

这里, k 指用年度表示的到期时间. 举例来说, 如果投资者购买零息债券, 支付 90 美元, 其面值为 100 美元, 该债券在 2 年内到期, 则该债券的收益率为 $c_t = (100/90)^{1/2} - 1 = 5.41\%$.

到期收益率

当期收益率没有考虑货币的时间价值, 因为它没有考虑投资者未来收到的债券利息的当前价值. 因此, 一个较为常用的债券投资收益是到期收益率(YTM). 然而, 到期收益率的计算更复杂. 简单地说, 到期收益率是通过将债券的价格与将来所有的现金流入的现值相等而计算出的收益率. 在购买日和到期日之间, 假设该债券的投资者收到 k 期利息支付. y 为债券的到期收益率, P 为债券的价格, 则有

$$P = \frac{C_1}{1+y} + \frac{C_2}{(1+y)^2} + \dots + \frac{C_k + F}{(1+y)^k}$$

这里, F 为债券的面值, C_i 是第 i 期的利息支付. 假定每年的债券票面利率为 α , 每年支付次数为 m , 到期时间为 n 年, 则债券利息每期的现金流入为 $F\alpha/m$, 支付期数为 $k = mn$. 债券价格和到期收益率计算如下

$$\begin{aligned} P &= \frac{\alpha F}{m} \left[\frac{1}{1+y} + \frac{1}{(1+y)^2} + \dots + \frac{1}{(1+y)^k} \right] + \frac{F}{(1+y)^k} \\ &= \frac{\alpha F}{my} \left[1 - \frac{1}{(1+y)^k} \right] + \frac{F}{(1+y)^k} \end{aligned}$$

下表给出了债券价格和到期收益率的一些结果, 这里假设 $F=100$ 美元, 票面利率为 5%, 每半年支付一次, 到期时间为 3 年.

到期收益率(%)	半年收益率(%)	债券价格(美元)
6	3.0	97.29
7	3.5	94.67
8	4.0	92.14
9	4.5	89.68
10	5.0	87.31

从表中可以看到, 随着债券到期收益率的增加, 债券价格在下降. 换句话说, 到期收益率和债券价格成反比. 在实践中, 通过观测到的债券价格, 可以计算出到期收益率. 一般不容易得到精确的解, 但我们可以获得一些精确的近似解. 例如, 在前面的表中, 购买债券的投资者支付的价格为 94 美元, 则到期收益率在 7%~8% 之间. 通过试验-更正错误的试验, 得到投资者每年到期收益率约为 7.26%. 许多金融机构提供在线程序来计算债券到期收益率及其价格, 例如, Fidelity 投资公司.

到期收益率(%)	半年收益率(%)	债券价格(美元)
7.1	3.55	94.41
7.2	3.6	94.16
7.3	3.65	93.90
7.25	3.625	94.03
7.26	3.63	94.00