

金融信息 分析



Financial Information
Analysis

林建忠 编



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

金融信息分析

林建忠 编

上海交通大学出版社

内容提要

本书介绍了金融信息分析中统计分析方法的基本知识,内容包括金融时间序列的基本概念和基本统计特征、平稳金融时间序列模型、条件异方差模型、风险值及分位数估计、神经网络和支持向量机分类方法及其在股市涨跌研判中的应用、生存数据与变量类型、生存分析的基本函数和参数模型、估计基本特征函数的非参数方法、比较生存函数的非参数方法和比例危险率模型。

本书可作为普通高等院校数学与应用数学专业本科生、应用统计专业硕士学位研究生教材,也可作为统计学和计量经济学等专业方向的研究生的教学参考书以及金融界高级从业人员的参考用书。

图书在版编目(CIP)数据

金融信息分析 / 林建忠编. — 上海: 上海交通大学出版社, 2015
ISBN 978-7-313-12714-3

I. ①金… II. ①林… III. ①金融—信息资源—分析
IV. ①F830

中国版本图书馆 CIP 数据核字 (2015) 第 042059 号

金融信息分析

编 者: 林建忠

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 韩建民

印 制: 上海天地海设计印刷有限公司

开 本: 787 mm×960 mm 1/16

字 数: 371 千字

版 次: 2015 年 4 月第 1 版

书 号: ISBN 978-7-313-12714-3/F

定 价: 39.80 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 21

印 次: 2015 年 4 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 021-64835344

前言

金融信息是指在组织和管理货币流通、各种金融证券交易、信用活动以及资金结算过程中产生的信号、指令、数据、情况、消息,其内容包括政府金融信息、银行信息、证券市场信息、客户资信信息等。金融市场的信息往往波诡云谲,对这些信息进行及时的整理、分析并从中提炼出有商业参考价值的本质特征,一直是金融业从业人员十分关心的问题。业界和学术界研究人员对此类问题已有大量的研究,积累了丰富的成果,形成了繁多的学术流派。从统计学的观点对金融信息进行分析是一种重要且颇具影响的流派,本书旨在介绍这一流派的基本思想和基本方法。

全书分三个部分共 12 章。第一部分讲解金融时间序列统计建模理论与分析方法,以及 Eviews 软件的基本操作,其中第 1 章介绍金融时间序列的基本概念和基本统计特征,第 2 章介绍平稳金融时间序列模型,第 3 章和第 4 章介绍几类重要的条件异方差模型,第 5 章概要介绍了风险值及分位数估计。第二部分讲解神经网络和支持向量机分类判别方法,以及应用 MatLab 软件相应模块研判股市涨跌,其中第 6 章介绍神经网络分类判别方法,第 7 章介绍支持向量机分类判别方法。第三部分讲解生存分析,以及应用 SAS 软件进行信用分析,其中第 8 章介绍生存数据与变量类型,第 9 章介绍生存分析的基本函数和参数模型,第 10 章介绍估计基本特征函数的非参数方法,第 11 章介绍比较生存函数的非参数方法,第 12 章介绍极其重

要的比例危险率模型。

金融信息分析的内容十分丰富,本书旨在介绍这一领域最基本的知识,使读者对金融信息的统计分析方法的轮廓和相应的统计软件有一基本的了解,以便为今后的深入学习打下必要的基础。正因为此,读者只要掌握了大学的微积分、线性代数和初等概率统计知识,就可以顺利阅读本书的绝大部分内容。本书可以作为数学系大四学生和应用统计专业硕士研究生的选修课程,同时也可以作为统计学和计量经济学等专业方向的研究生的教学参考书以及金融界高级从业人员的参考用书。

在本书结稿之际,我要感谢所有关心和支持我写作和出版此书的人们。我的博士导师范剑青教授把我引入了非参数统计建模领域,上海交通大学数学系叶中行教授和上海交通大学图像处理与模式识别研究所杨杰教授加深了我对统计学习理论的理解。本教材的大部分内容在上海交通大学数学系的《金融信息分析》课程中讲授过,参加过课程学习的同学对课程的成功和本书的成书作出了贡献,其中张轶、张肖、刘忞超、张翼飞、雪洋、李丹、罗兰和张译文等同学提供了部分实证例题。本书的写作和出版得到了上海交通大学 985 三期 2014 年校级教材建设项目的重点资助,以及国家自然科学基金项目“引入背景风险理论的主权财富基金战略资产配置”项目(No. 71271135(2013 - 2016))的资助。

本书相关数据及课件存放在公共邮箱 jrxxfx@aliyun.com 中,登录密码为 jrxxfx2015。

由于作者水平所限,书中存在的缺点或错误恳请同行和广大读者批评指正。

林建忠

上海交通大学数学系

2015 年 1 月

目 录

1 金融时间序列及其特征	001
1.1 资产收益率.....	001
1.2 收益率的分布性质.....	006
1.2.1 统计分布及其矩的回顾	006
1.2.2 收益率的分布	013
1.2.3 收益率的经验性质	017
1.3 Eviews 软件相关操作	018
1.3.1 简介	018
1.3.2 启动软件包	019
1.3.3 创建工作文件	021
1.3.4 输入和编辑数据	022
1.3.5 查看序列的数据特征	024
1.4 习题.....	025
2 线性时间序列分析及其应用	026
2.1 平稳性.....	026
2.2 自相关函数.....	027
2.3 自回归模型.....	032
2.3.1 AR 模型及性质	032
2.3.2 实际中怎样识别 AR 模型	042
2.3.3 拟合优度	049

2.3.4	预测	049
2.4	移动平均模型	053
2.4.1	MA 模型的性质	053
2.4.2	识别 MA 的阶	054
2.4.3	估计	055
2.4.4	用 MA 模型预测	056
2.5	ARMA 模型	057
2.5.1	ARMA(1, 1)模型的性质	057
2.5.2	一般的 ARMA 模型	059
2.5.3	识别 ARMA 模型	059
2.5.4	用 ARMA 模型预测	062
2.5.5	ARMA 模型的三种表示	063
2.6	单位根非平稳时间序列	064
2.6.1	随机游动	064
2.6.2	带漂移的随机游动	065
2.6.3	带趋势项的时间序列	066
2.6.4	单整与单位根非平稳模型	066
2.6.5	非平稳序列的单位根检验	067
2.6.6	DGP 识别	071
2.6.7	Eviews 相关操作	073
2.7	带时间序列误差的回归模型	075
2.8	异方差性和自相关一致协方差估计	078
2.9	习题	081

3	条件异方差模型	082
3.1	波动率的特征与模型的结构	082
3.2	ARCH 模型	086
3.2.1	ARCH 模型的结构	086
3.2.2	ARCH 模型的性质	088
3.2.3	ARCH 效应的检验	090
3.2.4	ARCH 模型的建立	090
3.2.5	例子	093

3.2.6	ARCH 模型的缺点	096
3.3	GARCH 模型	096
3.3.1	模型的结构与性质	096
3.3.2	模型的预测	097
3.3.3	例子	098
3.4	求和 GARCH 模型	101
3.5	GARCH - M 模型	103
3.6	指数 GARCH 模型	104
3.7	门限 GARCH 模型	107
3.8	习题	108
4	非线性模型及其应用	110
4.1	非线性模型	110
4.1.1	门限自回归模型	110
4.1.2	平滑转移 AR(STAR)模型	113
4.1.3	马尔科夫转换模型	114
4.1.4	非参数方法	116
4.2	非线性检验	124
4.2.1	非参数检验	124
4.2.2	参数检验	126
4.3	建模与预测	128
4.3.1	建模	128
4.3.2	参数自助法	128
4.3.3	预测的评估	129
4.4	习题	131
5	风险值与分位数估计	132
5.1	风险值	132
5.2	风险度量制	133
5.2.1	讨论	135
5.2.2	多个头寸	136

5.2.3	预期损失	136
5.3	VaR 计算的计量经济方法	137
5.4	分位数估计	142
5.4.1	分位数与次序统计量	142
5.4.2	分位数回归	143
5.5	习题	146
6	神经网络	148
6.1	神经网络的基本特征和通有性质	149
6.1.1	神经网络的形式化描述	149
6.1.2	神经网络的互连结构形态	152
6.1.3	前向神经网络的符号表示	153
6.2	MP 模型和 Hebb 学习规则	155
6.2.1	MP 模型	155
6.2.2	Hebb 学习规则	156
6.3	感知器	156
6.3.1	线性阈值单元	156
6.3.2	单层感知器	157
6.3.3	感知器学习算法	160
6.3.4	多层感知器	164
6.3.5	最小均方(LMS)算法	166
6.4	BP 网络及 BP 算法(反向传播算法)	171
6.4.1	BP 网络概述	171
6.4.2	反向传播算法	174
6.4.3	BP 网络的训练函数	179
6.4.4	算法的 Matlab 实现	184
6.4.5	BP 网络的设计原则	186
6.5	径向基函数网络	189
6.5.1	RBF 网络结构和工作原理	189
6.5.2	函数逼近与内插	190
6.5.3	Tikhonov 正规化理论	191
6.5.4	径向基网络的学习	195

6.5.5	径向基网络设计的基本方法	196
6.6	基于神经网络的股价对数收益率模型	197
6.7	基于 BP 神经网络和 RBF 神经网络的期权 定价	199
6.7.1	基于神经网络的期权定价 模型	199
6.7.2	基于 B-S 模型和神经网络模型的 实证分析及比较	202
6.8	习题	205
7	支持向量机	207
7.1	分离超平面	207
7.2	支持向量分类器	210
7.3	支持向量机	212
7.3.1	支持向量机原理	212
7.3.2	计算分类的 SVM	213
7.4	MatLab 上支持向量机的使用及其在股市 预测中的应用	216
7.4.1	MatLab 上支持向量机的使用 ...	216
7.4.2	数据预处理与指标选取	220
7.4.3	模型的训练和验证方法	223
7.4.4	模型训练和预测分析	224
7.5	习题	228
8	生存数据与变量类型	230
8.1	生存数据与删失性	230
8.1.1	生存数据	230
8.1.2	右删失	232
8.1.3	左删失和区间删失	237
8.2	习题	238

9 基本函数和参数模型	239
9.1 生存分析的基本函数	239
9.1.1 生存函数	239
9.1.2 危险函数(Hazard Function)	241
9.1.3 平均剩余寿命函数和中位寿命	244
9.1.4 几个函数间的关系	245
9.2 生存数据建模常用的参数模型	247
9.2.1 指数分布	247
9.2.2 威布尔分布(Weibull Distribution)	247
9.2.3 对数正态分布	248
9.2.4 对数 Logistic 分布	250
9.2.5 Gamma 分布	252
9.2.6 广义 Gamma 分布	254
9.3 次序统计量的分布与极值分布	254
9.3.1 次序统计量的分布	254
9.3.2 极值分布	257
9.4 删失数据似然函数的构造	259
9.4.1 I 型删失数据的似然函数	259
9.4.2 II 型删失数据的似然函数	261
9.4.3 III 型(随机)删失数据的似然函数	263
9.5 习题	264
10 估计基本特征函数的非参数方法	265
10.1 寿命表法(life table method)	265
10.2 右删失数据的生存函数、累积危险函数的估计	267
10.2.1 单变量数据的生存函数的估计(乘积限估计)	267

10.2.2	累积还款率和违约率的估计	271
10.3	生存函数点估计的置信区间	273
10.3.1	生存函数的置信区间	273
10.3.2	累积还款力函数、累积违约力的 置信区间	274
10.4	生存时间均值和中位数的点估计与区间 估计	275
10.4.1	生存函数均值的估计	275
10.4.2	生存时间分位点的估计	276
10.5	双删失数据的生存函数估计(特恩伯 估计)	277
10.5.1	非分组数据情形的方法	277
10.5.2	分组数据情形的方法和理 论	283
10.6	习题	286
11	比较生存函数的非参数方法	287
11.1	两个生存函数的比较	287
11.1.1	Gehan-Wilcoxon 检验	287
11.1.2	Cox-Mantel 检验	291
11.1.3	对数秩检验	292
11.2	分层情形下的 Mantel-Haenszel 检 验	295
11.3	M 个样本情形的比较($M \geq 2$)	297
11.3.1	完全数据情形下的检验方 法	297
11.3.2	基于 Kruskal-Wallis 检验的多重 比较	299
11.3.3	含有删失数据时的检验方 法	301
11.4	习题	305

12 比例危险率模型	307
12.1 参数型比例危险率模型	307
12.2 Cox 半参数比例危险率模型	308
12.3 习题	313
参考文献	314

1 金融时间序列及其特征

金融时间序列分析考虑的是资产价值随时间演变的理论与实践。它是一个带有高度经验性的学科,但也像其他学科领域一样,理论是形成分析推断的基础。然而,金融时间序列分析有一个区别于其他时间序列分析的主要特点:金融理论及其经验的时间序列都包含不确定性。例如,资产波动率有各种不同的定义,对一个股票收益率序列,波动率是不能直接观察到的。正因为带有不确定性,统计的理论和方法在金融时间序列分析中起重要作用。

1.1 资产收益率

资产(asset)是指企业(自然人、国家)过去的交易或事项形成的,由企业(自然人、国家)拥有或控制的,预期会给企业(自然人、国家)带来能以货币衡量的经济利益的资源。本书讨论的股票、债权等各种有价证券都是资产。图 1-1(a)是上证指数从 1991 年 12 月到 2011 年 12 月的月指数走势。虽然资产价格序列是金融研究的自然对象,然而本书以后介绍的实证研究表明,将资产价格作为直接的研究对象很难形成满意的统计理论框架。目前,多数金融研究以资产收益率(rate of return, yield)序列作为金融研究的直接对象。正如 Campbell, Lo 和 MacKinlay (1997)指出的,对普通的投资者来说,资产收益率不仅完全体现了该资产的投资机会且与其投资规模无关,更重要的是以资产收益率序列作为直接研究对象能够形成比较完整的统计分析框架。

1) 单周期收益率

设资产在 t 时刻的价格为 P_t ,且假设资产不支付红利。现假设以 1 天作为一个周期,若从第 $t-1$ 天到第 t 天持有某种资产,定义单周期总收益率为相隔一天的资产价格比,即

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

而相隔一天的资产收益率(或称回报率、涨跌幅)定义为

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

图 1-1(b)是上证指数从 1991 年 12 月到 2011 年 12 月月指数的收益率。

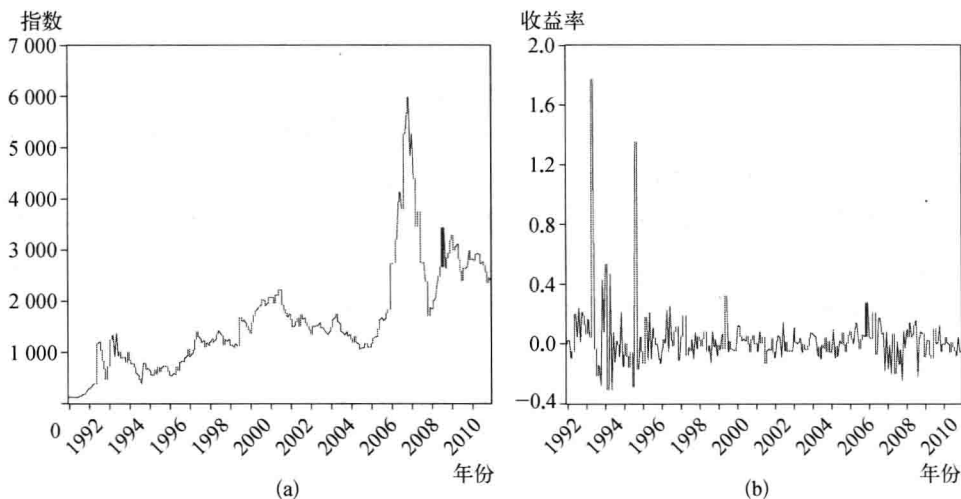


图 1-1 上证月指数

(a) 走势 (b) 收益率

2) 多周期收益率

若从第 $t-k$ 天到第 t 天(k 个周期)持有某种资产,定义 k 周期总收益率为相隔 k 天的资产价格比,即

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

这样, k 期资产总收益率就是其所包含的这 k 个单周期资产总收益率的乘积。相应地, k 期收益率定义为

$$R_t[k] = \frac{P_t}{P_{t-k}} - 1 = \frac{P_t - P_{t-k}}{P_{t-k}}。$$

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。如果时间区间没有给出,那么就隐含地假定时间区间为 1 年。如果持有资产的期限为 k 年,则(平均的)年化收益率(annualized rate of return)定义为

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

这是由它所包含的 k 个单期总收益率的几何平均减去 1 所得到的。因为计算算术平均值比计算几何平均值容易,上式也可写为

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

如果单期收益率或周期很小,可以用一阶泰勒(Taylor)展开来近似表示年度化的收益率,得到近似计算公式:

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

3) 连续复合

假定银行存款的年利率为 10%,最初存款为 1 元人民币(RMB)。如果该银行每年支付一次利息,那么 1 年之后存款的总价值变为 1 元 $\times (1 + 0.1) = 1.1$ 元。如果该银行半年付息一次,6 个月的利息率是 $10\%/2 = 5\%$,第 1 年之后总价值是 1 元 $\times (1 + 0.1/2)^2 = 1.1025$ 元。一般地,如果银行 1 年付息 m 次,那么每次支付的利息率为 $10\%/m$,第 1 年之后总价值是 1 元 $\times (1 + 0.1/m)^m$ 元。表 1-1 给出了年利率为 10% 时一些常用的时间间隔下存款 1 元人民币的结果。

表 1-1 复合效果的展示(期限为 1 年,年利率为 10%)

类 型	支付次数	每期的利率	总价值(元)
一年	1	0.1	1.1
半年	2	0.05	1.1025
季度	4	0.025	1.10381
月	12	0.0083	1.10471
周	52	0.1/52	1.10506
天	365	0.1/365	1.10516
连续地	∞		1.10517

特别地,总价值趋于 1.1052 元 $\approx \exp(0.1)$ 元,这个值就是连续复合的

结果。

一般地,连续复合的资产总价值 A 为

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

式中, r 是年利率, C 是初始资本, n 是年数。由式(1-4),有

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

C 称为 n 年后价值为 A 的资产的现值,这里假定连续复合的年利率为 r 。

4) 连续复合收益率

资产的总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):

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

式中, $p_t = \ln P_t$ 。与收益率 R_t 相比,对数收益率 r_t 有一些优点。首先,对多期对数收益率,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,多期对数收益率就是它所包含的单期对数收益率之和,这一性质使得对数收益率更容易进行统计分析。图 1-2 是上证指数从 1991 年 12 月到 2011 年 12 月的月指数的对数收益率图。

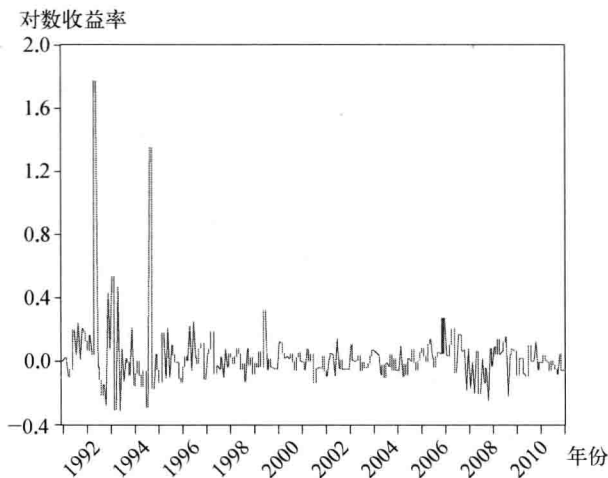


图 1-2 上证月指数对数收益率图