

“十二五”经济管理类实验系列规划教材

金融定量分析 与 S-Plus 运用

Financial Quantitative Analysis
& its Applications in S-Plus

朱 敏 王 翔 编著



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

上海市金融保险本科教育高地建设项目

金融定量分析与S-Plus 运用

朱 敏 王 翔 编著

上海交通大学出版社

内 容 提 要

随着现代金融朝着数理化、微观化和工程化的趋势发展,运用模型量化分析金融数据已成为学术研究和实践运用的主流模式,掌握分析技术原理和实现方法已成为金融方向研究人员或金融从业技术人员的基本素质,也是目前高校金融、经济类专业人才培养的必修内容。

目前相关书籍可以归为两类:一类属“理论类”,侧重于理论,强调严谨的数学推导;另一类属“应用类”,侧重于运用,强调方法通过软件实现。本书力求找到理论与运用的平衡。为了说明原理,有严谨的推导过程,但尽量简洁。在此基础上,重点解析了对应方法在软件中的实现过程。

本书适合金融专业师生、金融方向研究人员以及金融从业技术人员参考阅读。

图书在版编目(CIP)数据

金融定量分析与 S-Plus 运用/朱敏,王翔编著. —上海:
上海交通大学出版社,2013

ISBN 978-7-313-08962-5

I. 金... II. ①朱... ②王... III. 金融—定量分
析—应用软件—高等学校—教材 IV. F83-39

中国版本图书馆 CIP 数据核字(2012)第 199599 号

金融定量分析与 S-Plus 运用

朱 敏 王 翔 编著

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

上海颀辉印刷厂印刷 全国新华书店经销

开本:787mm×960mm 1/16 印张:18.75 字数:354 千字

2013 年 1 月第 1 版 2013 年 1 月第 1 次印刷

ISBN 978-7-313-08962-5/F 定价:42.00 元

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:021-57602918

前 言

随着现代金融朝着数理化、微观化和工程化的趋势发展,运用模型量化分析金融数据已成为学术研究和实践运用的主流模式,掌握分析技术原理和实现方法已成为金融方向研究人员或金融从业技术人员的基本素质,也是目前高校金融、经济类专业人才培养的必修内容。

金融定量研究的对象基本都是时间序列数据,因而主要运用经济计量学的专门方法即“时间序列分析”来分析处理数据,这也使得“时间序列分析”成为金融定量分析技术学习的主要内容。由于较多运用了矩阵分析和随机分析等数学内容,使得非数学专业学生在学习这部分知识时有较大困难。

环顾目前相关教材或参考书籍,大体可以归为两类:一类属“理论类”书籍,相对“高深”,侧重于理论,强调严谨的数学推导;另一类属“应用类”书籍,相对“浅显”,侧重于运用,强调方法通过软件实现。对于初学者而言,大部分人都会选择“应用类”书籍,既有案例又有软件操作过程,掌握起来“短平快”,能很快上手,学以致用。但这样的学习过程存在一定的“隐患”,知道技术的使用用法,但不清楚其内在原理,往往会误用甚至用错。笔者在指导学生论文、查阅某些期刊论文过程中发现过定量分析中出现的各类问题。究其原因,就是出在了知其然而不知其所以然。

解决的措施就是由浅入深,进一步“啃”理论类书籍。这是一个痛苦的过程,特别是对非数学类专业的学生,数字符号过于抽象,即使能够推导一遍,也难以与实际问题联系起来。这也是笔者在高年级本科生以及研究生教学过程中遇到的一个难题。笔者采取的一个解决措施就是通过软件尽量把抽象的内容具体化,如通过模拟生成特定类型数据、通过作图展示数据特征、通过分步运算展示实现细节等。通过多轮的教学尝试,这一措施获得了学生普遍的好评,教学材料也积累了许多,由此萌生把教学讲义付之桑梓,为学习者提供有益参考的想法。

本书力求找到理论与运用的平衡。为了说明原理,有严谨的推导过程,但尽量

简洁。在此基础上,重点解析了对应方法在软件中的实现过程。与大部分“应用类”书籍不同,本书不仅仅介绍软件中实现方法的具体函数,而是打开函数这一“黑箱”,把实现的程序详细讲解。

本书由朱敏提出编写框架,并撰写了第 1、2、4、5、6、7、8 章,完成了对全书的统稿;王翔撰写了第 3 章,帮助完成了部分推导的校对;研究生帮助完成了部分公式的编辑转换。由于金融定量分析的内容深入而广泛,笔者对某些细节的理解相对有限,因而书中可能存在差错和遗漏,希望读者对本书提出宝贵的批评和建议。

感谢上海师范大学金融工程重点学科(DZW912)对本著作出版的资金支持。

朱 敏 王 翔

2012 年 6 月于学思湖畔

目 录

第 1 章 S-Plus 的基本使用方法	1
1.1 S-Plus 软件简介	1
1.2 S 语言的基本对象	2
1.2.1 向量	3
1.2.2 矩阵	6
1.2.3 因子	10
1.2.4 列表	12
1.2.5 数据框	14
1.3 S 语言的基本语法	16
1.3.1 分支语句	16
1.3.2 循环语句	18
1.4 S 语言的自编函数	21
1.4.1 工作空间管理	21
1.4.2 参数(自变量)	23
1.4.3 作用域	24
1.5 S-Plus 的绘图	26
1.5.1 散点图	26
1.5.2 密度函数的绘制	28
1.5.3 多组数据的图形绘制	29
1.5.4 多图控制	30
1.6 用 S-Plus 读写数据文件	32
1.6.1 文本格式文件的读写方式	32
1.6.2 特定数据源数据的读写方式	33
1.6.3 利用菜单读写数据	35
1.7 数据的基本处理方法	37
1.7.1 特定数据的选择	37
1.7.2 数据的合并	38

1.7.3	数据的排序	41
1.7.4	数据的筛选	42
第 2 章	金融时间序列数据的处理方法	44
2.1	时间序列数据	44
2.1.1	“timeSeries”对象	44
2.1.2	“timeSeries”的基本操作	47
2.1.3	“timeDate”对象	49
2.1.4	“timeDate”对象的基本操作	50
2.1.5	递增时间序列的生成方法	54
2.1.6	生成“timeSeries”对象	59
2.1.7	时间序列数据的分组统计	61
2.1.8	时间序列数据的合并	65
2.1.9	缺失值的处理方法	67
2.2	在 S-Plus 中对时间序列的处理	67
2.2.1	生成滞后和差分	68
2.2.2	收益率的定义	70
2.2.3	收益率的计算	72
2.2.4	收益率的特征	74
第 3 章	时间序列数据的回归分析	78
3.1	最小二乘估计量	78
3.1.1	线性回归模型的矩阵方法;OLS 的数值性质	78
3.1.2	回归模型的经典假设;OLS 的统计性质	81
3.2	OLS 估计量的有限样本性质	83
3.2.1	估计量 $\hat{\beta}$ 的有限样本性质	83
3.2.2	随机扰动项的方差估计量 $\hat{\sigma}^2$ 的有限样本性质	84
3.2.3	时间序列的有限样本性质	84
3.3	假设检验	87
3.3.1	个别回归系数的假设检验	87
3.3.2	线性约束检验	88
3.3.3	不包括常数项的所有系数的统计显著性; F 统计量	90
3.4	违背经典假设数据的问题与处理	90

3.4.1	异方差	90
3.4.2	自相关	94
3.5	使用 S+Finmetrics 函数“OLS”进行时间序列回归	97
第 4 章	平稳时间序列模型	105
4.1	平稳时间序列的几个重要概念	105
4.1.1	平稳性	105
4.1.2	自相关	105
4.1.3	偏自相关	106
4.2	自回归模型	106
4.2.1	AR(1)模型	107
4.2.2	AR(1)模型的识别特征	108
4.2.3	AR(1)模型的生成程序	109
4.2.4	AR(1)模型的估计	112
4.2.5	AR(2)模型	116
4.2.6	AR(p)模型	118
4.2.7	自回归模型的建模法则	119
4.2.8	自回归模型的预测	123
4.3	移动平均模型	125
4.3.1	MA(1)模型	126
4.3.2	MA(1)模型的识别特征	127
4.3.3	MA(1)模型的生成程序	128
4.3.4	MA(1)模型的估计	129
4.3.5	MA(2)模型	130
4.3.6	MA(q)模型	133
4.3.7	移动平均模型的建模法则	133
4.3.8	移动平均模型的预测	135
4.4	自回归移动平均模型	137
4.4.1	ARMA(1,1)模型	137
4.4.2	ARMA(1,1)的识别特征	138
4.4.3	ARMA(1,1)的生成程序	139
4.4.4	ARMA 模型的建模法则	141
4.4.5	自回归移动平均模型的预测	143

第 5 章 波动率建模	145
5.1 ARCH 模型	146
5.1.1 ARCH 模型的结构	146
5.1.2 ARCH 模型的性质	146
5.1.3 ARCH 模型的诊断	148
5.1.4 ARCH 模型的估计	150
5.1.5 ARCH 模型的预测	154
5.2 GARCH 模型	155
5.2.1 GARCH 模型的结构	155
5.2.2 GARCH 模型的性质	156
5.2.3 GARCH 模型的估计方法	159
5.2.4 GARCH 模型的优选	161
5.2.5 GARCH 模型的预测	162
5.3 风险效应模型	163
5.3.1 风险效应	163
5.3.2 GARCH-M 模型	164
5.3.3 GARCH-M 模型的估计	164
5.3.4 GARCH-M 模型的预测	165
5.4 杠杆效应模型	166
5.4.1 波动的非对称性	166
5.4.2 杠杆效应的检验	166
5.4.3 EGARCH 模型	168
5.4.4 EGARCH 模型的估计	169
5.4.5 EGARCH 模型的预测	172
5.4.6 TGARCH 模型	174
5.4.7 TGARCH 模型的估计	175
5.5 含有外生变量的 GARCH 模型	176
5.5.1 外生变量与 GARCH 模型	177
5.5.2 含有外生变量的 GARCH 模型估计	177
第 6 章 非平稳序列与单位根	182
6.1 非平稳序列	182

6.1.1	随机游走	182
6.1.2	带漂移的随机游走	184
6.2	伪回归	185
6.3	DF 检验	188
6.3.1	DF 检验的基本思想	188
6.3.2	DF 统计量的构造	188
6.3.3	蒙特卡洛模拟与 DF 分布的获取	191
6.3.4	S-Plus 中 DF 检验方法	197
6.4	ADF 检验	198
6.4.1	ADF 检验的基本概念	198
6.4.2	ADF 检验的经验法则	199
6.5	单位根检验的一个实际例子	202
第 7 章	向量自回归	209
7.1	向量自回归(VAR)模型的构造	209
7.2	VAR 模型的残差滞后项表达式	210
7.2.1	VAR 模型的估计方法	213
7.2.2	VAR 模型滞后阶数的选择	215
7.2.3	VAR 模型的稳定性条件	217
7.3	格兰杰因果检验	221
7.3.1	格兰杰因果检验的定义	221
7.3.2	F 检验	222
7.3.3	Wald 检验	225
7.4	VAR 模型的脉冲响应分析	229
7.5	VAR 模型预测误差的方差分解	235
7.6	结构向量自回归	241
7.6.1	结构模型还原的识别条件	242
7.6.2	SVAR 模型的一般形式	244
7.6.3	SVAR 模型的似然估计	245
7.6.4	长期累积效应约束与 BQ 分解	247
7.6.5	SVAR 模型的脉冲响应函数	250
7.7	结构向量自回归的实例	253
7.7.1	基于 VAR(2) 的模型	253

7.7.2 基于 SVAR(2)的模型	255
第 8 章 协整与误差修正模型	260
8.1 协整的概念	260
8.1.1 单整	260
8.1.2 协整的定义	261
8.2 两个变量的协整分析	264
8.2.1 协整检验的基本思想	264
8.2.2 误差修正模型	266
8.2.3 E-G 两部法	268
8.3 多个变量的协整分析	272
8.3.1 向量的协整	275
8.3.2 Johanson 检验	278
参考文献	289

第 1 章 S-Plus 的基本使用方法

随着信息技术的发展,现代金融的理论研究和实践运用,其微观化、数量化的特征越来越明显。处理的数据数以亿计,运用的模型精密复杂,仅靠人工手算已经无法胜任。因此,在熟悉基本金融理论的基础上,掌握一个专业的金融运算软件已成为金融理论研究人员亦或金融业界从业人员必不可少的技能。S-Plus 作为与 SAS、SPSS 齐名的优秀统计软件,专为金融研究与运用开发了强大的 Finmetric 模块。只要掌握 S-Plus 的基本语言和金融定量分析的基本原理,我们就能轻松面对海量的金融数据,通过强大的软件对纷繁的信息抽丝剥茧,发现现象背后的本质规律。

1.1 S-Plus 软件简介

1. 软件功能特点

S-Plus 是基于 S 语言发展而成的,目前由 MathSoft 公司的统计科学部进一步完善。作为统计学家及一般研究人员的通用方法工具箱,S-Plus 强调演示图形、探索性数据分析、统计方法、开发新统计工具的计算方法以及可扩展性。S-Plus 是一个商业软件,可以直接用来进行标准的统计分析得到所需结果,但是它的主要的特点是交互地从各个方面去发现数据中的信息,并可以很容易地实现一个新的统计方法。另外,Auckland 大学的 Robert Gentleman 和 Rossihaka 及其他志愿人员开发了一个 R 系统,其语法形式与 S 语言基本相同,但实现不同,两种语言的程序有一定的兼容性。

在你数据存取更方便、分析更自由的时候,能够容易、有效地存取数据是非常重要的。S-Plus 有流畅、可视化的用户界面及广泛的输入输出功能,不论你的数据在哪儿、数据的格式如何,都可以及时存取。S-Plus 提供 Microsoft Office 的外观,让你方便地存取及分析数据。整合了 Microsoft Excel,确切说,Excel 完全成为 S-Plus 的一个组成部分,你可以在 S-Plus 中使用 Excel 的所有功能,直到你遇到 Excel 的极限或不能做的工作,这时你自然用到 S-Plus。

S-Plus 提供了强大的数据接口,几乎可以存取任何格式的数据,包括 Excel、Access、Oracle、Sybase、SQLServer 与许多财经专业数据库。S-Plus 可以帮你将数据从一种格式转到另一种格式,S-Plus 具有独一无二的可视化交互式图形显示,其

中包括著名的 Trellis 图形分析在做数据探索及分享结果时,强大的图形显示扮演关键性的角色。当更深入地探索数据时,你常常会迷失在许多简要的统计数值及文字报表中,而不能即时发现隐藏在数据中的趋势及相关性,但你运用 S-Plus 的探索式图形技术,就可以透视及展现隐藏在数据中的模式,获得新的发现。

S-Plus 的图形是面向对象的,因此你可以修改每个细节,来产生适合报告用的图形;此外 S-Plus 提供广泛的 2D 及 3D 图形种类,包含非常广泛的图形库。你可以轻易地调整图形的属性,包括线条、颜色、字型等。创新的 S-Plus Graphlets 技术可以产生互动式的图形,让使用者能够对图形逐层下探观察数据,或是连接到 Web 网页,让读者可以即时地与你的数据互动。S-Plus 拥有最广泛、最全面的统计分析与数值计算方法,S-Plus 提供你强大的分析能力和尖端的统计、数据挖掘、预测分析的方法,让你可以深入地了解你的资料。对数据越深入的了解,就会产生越佳的决策,在充满竞争的时代,这将会是企业关键性的战略优势。

S-Plus 提供超过 4 200 种包含现代及传统技巧的数据分析函数。方便的菜单、工具列和对话框,让你轻松地分析数据。S-Plus 的所有函数、数据及模型都被视为对象,你可以使用传统及现代的不同方法来符合不同的模型,并选择最佳结果的模型。S-Plus 提供强大的程序语言,你可以自创或延伸分析方法。当你的分析变得很复杂时,S-Plus 的弹性扩展功能可以满足你的挑战。S-Plus 是数据分析软件领域的 Linux,其开放性与广泛使用使得全球的众多数据分析专家首先应用 S-Plus 函数及程序或 S-语言来实现新的分析算法。这些新的算法库可以在专业网站自由下载。

2. 开发公司背景

Insightful 公司是一家具备丰富数据分析经验的专业公司,其旗舰产品 S-Plus 在各行业已经拥有 18 年的使用历史。作为世界上最为流行的统计分析软件之一,它曾获得著名的“美国计算机协会优秀软件奖”,而其他获此殊荣的是大名鼎鼎的 TCP/IP、UNIX 和网上浏览器的创始者 MOSAIC。S 语言是由 AT&T 贝尔实验室开发的一种用来进行数据探索、统计分析、作图的解释型语言。它的丰富的数据类型(向量、数组、列表、对象等)特别有利于实现新的统计算法,其交互式运行方式及强大的图形及交互图形功能使得我们可以方便地探索数据。目前 S 语言的实现版本主要就是 S-Plus。

1.2 S 语言的基本对象

S-Plus 中出现的任何内容其实质都是对象(object),而任何一个对象都对应唯一的数据类型。

1.2.1 向量

最简单的结构就是由一系列有序数值构成的数值向量(vector)。向量中的元素允许数值型、逻辑性、字符型等各类数据类型,但向量不允许元素有不同的数据类型,因此当赋值中存在不同的数据类型时,系统就把输入值强制作为字符型处理。如果要创建一个名称为 x 的向量,那么相应的命令应为:

代码 1.1

创建数值型向量

```
> x=c(2.4,2.5,5,6)
> x
[1] 2.4 2.5 5.0 6.0
```

创建逻辑型向量

```
> x=c(T,F)
> x
[1] T F
```

创建字符型向量

```
> x=c("yes", "no")
> x
[1] "yes" "no"
```

输入不同数据类型的值

```
> x=c(2.3, "yes")
> x
[1] "2.3" "yes"
```

用函数 `c()` 完成赋值语句。这里的函数 `c()` 可以有任意多个参数,而它的值则是一个把这些参数首尾相连形成的一个向量。其中,“=”是赋值符号,把等式右边具体的数值赋予等式左边的 x 。这样一个包含有 4 个数值的向量就创建成功了。对向量的引用,只需要键入向量名称 x 。例如:

代码 1.2

求 x 的倒数

```
> 1/x
[1] 0.4166667 0.4000000 0.2000000 0.1666667
```

更复杂一些的运算,对 x 减 5 后取自然数

```
> exp(x-5)
[1] 0.0742736 0.0820850 1.0000000 2.7182818
```

S-Plus 的一大特点在于可以对向量进行数学运算。使用特定数学函数,可以一次对向量中的每个元素进行复杂运算,最后结果还是一个向量。此外,也可以使用统计函数对向量进行统计运算,计算的结果则转化为一个数值。常用的数学运算函数和统计运算函数如表 1-1 所示:

表 1-1 S-Plus 中常用函数表

数学函数		统计函数	
$\exp(x)$	求 e^x 的值	$\min(x)$	求 x 中的最小值
$\log(x)$	求 $\ln x$ 的值	$\max(x)$	求 x 中的最大值
$\sqrt{x}$	求 $\sqrt{x}$ 的值	$\text{mean}(x)$	求 x 的均值
$\text{abs}(x)$	求 x 的绝对值	$\text{var}(x)$	求 x 的方差
$\sin(x)$	求 $\sin(x)$ 的值	$\text{sd}(x)$	求 x 的标准差
$\cos(x)$	求 $\cos(x)$ 的值	$\text{sum}(x)$	求 x 之和
$\tan(x)$	求 $\tan(x)$ 的值	$\text{rang}(x)$	求 x 之全距
$\text{diff}(x)$	求 x 的差分	$\text{media}(x)$	求 x 之中位数
$\text{cumsum}(x)$	求 x 的向上累计值	$\text{cor}(x, y)$	求 x 与 y 的相关系数

例:对 10 名成年人的身高(单位:cm)进行抽样调查,得到的具体数据如下所列:166,169,172,177,180,170,172,174,168,173

现利用 S-Plus 对成年人的身高做一定的统计分析:

代码 1.3

```
> height=c(166,169,172,177,180,170,172,174,168,173)
```

```
> height
```

```
[1] 166 169 172 177 180 170 172 174 168 173
```

利用 $\text{mean}()$ 、 $\min()$ 、 $\max()$ 和 $\text{sd}()$ 等函数计算统计量

```
> mean(height)
```

```
[1] 172.1
```

```
> max(height)
```

```
[1] 180
```

```
> min(height)
```

```
[1] 166
```

```
> sd(height)
```

```
[1] 4.201851
```

运用函数 $\text{length}()$ 能够获得向量的长度(向量里元素的个数),而运用函数

`class()` 可以获得向量中元素的数据类型。例如：

代码 1.4

获得向量的长度

```
> length(x)
```

```
[1] 4
```

获得变量的类型

```
> class(x)
```

```
[1] "numeric"
```

运用符号“:”可以生成等差数列。例如：

代码 1.5

1 到 5 的递增数列

```
> 1:5
```

```
[1] 1 2 3 4 5
```

5 到 1 的递减数列

```
> 5:1
```

```
[1] 5 4 3 2 1
```

符号“*”的运算级别最高

```
> 2*1:5
```

```
[1] 2 4 6 8 10
```

如果想要生成更为复杂的数列,可以使用 `seq()` 函数,其语法结构为:

```
seq(from,to,by,length)
```

其中,from 为序列的起始值,to 是序列的最终值,by 是步长。有时候我们关注生成数列的个数,而不在乎序列的终值是多少,可以通过设置 `length` 实现。例如:

代码 1.6

如果省略步长,则系统默认为 1

```
> seq(1,5)
```

```
[1] 1 2 3 4 5
```

生成公差为 2 的等差序列

```
> seq(1,5,2)
```

```
[1] 1 3 5
```

步长为 2,元素个数为 4 的序列

```
> seq(1,,2,length=4)
```



```
[1] 1 3 5 7
```

上面我们空缺参数 to, 否则系统把第二个位置的读为终值, 取 1 与 2 之间等长位置上的 4 个值

```
> seq(1,2,length=4)
```

```
[1] 1.000000 1.333333 1.666667 2.000000
```

运用函数 rep() 能重复特定次数的同一对象。其中, 参数 value 是要创建的值, 参数 times 是重复的次数。例如:

代码 1.7

重复五次 2

```
> rep(2,5)
```

```
[1] 2 2 2 2 2
```

重复 2,3 两次

```
> rep(c(2,3),2)
```

```
[1] 2 3 2 3
```

重复 2 两次, 3 两次

```
> rep(c(2,3),each=2)
```

```
[1] 2 2 3 3
```

重复 1 到 3 两次

```
> rep(1:3,2)
```

```
[1] 1 2 3 1 2 3
```

1.2.2 矩阵

矩阵(matrix)是将数据用行和列排列的长方形表格, 它是二维的数组, 其单元必须是相同的数据类型。S-Plus 中生成矩阵的函数是 matrix(), 其语法为:

```
matrix(x,nrow,ncol,byrow=T)
```

其中, x 是需要输入的矩阵元素, nrow 设置矩阵的行数, ncol 设置矩阵的列数。byrow 设置矩阵元素输入的顺序, 如果为“T”, 输入依照行顺序完成, 即输完第一行再从第二行开始输入。如果为“F”, 输入依照列顺序完成, 即输完第一列再从第二列开始输入。系统默认的 byrow 参数为“F”。

例如, 要生成形如:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}$$

可以输入如下命令: