

金融

时间序列预测

基于R语言的应用实践

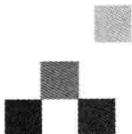
王乐 金珏 王水◎著 >>>>>

Forecasting Financial
Time Series



中国时代经济出版社

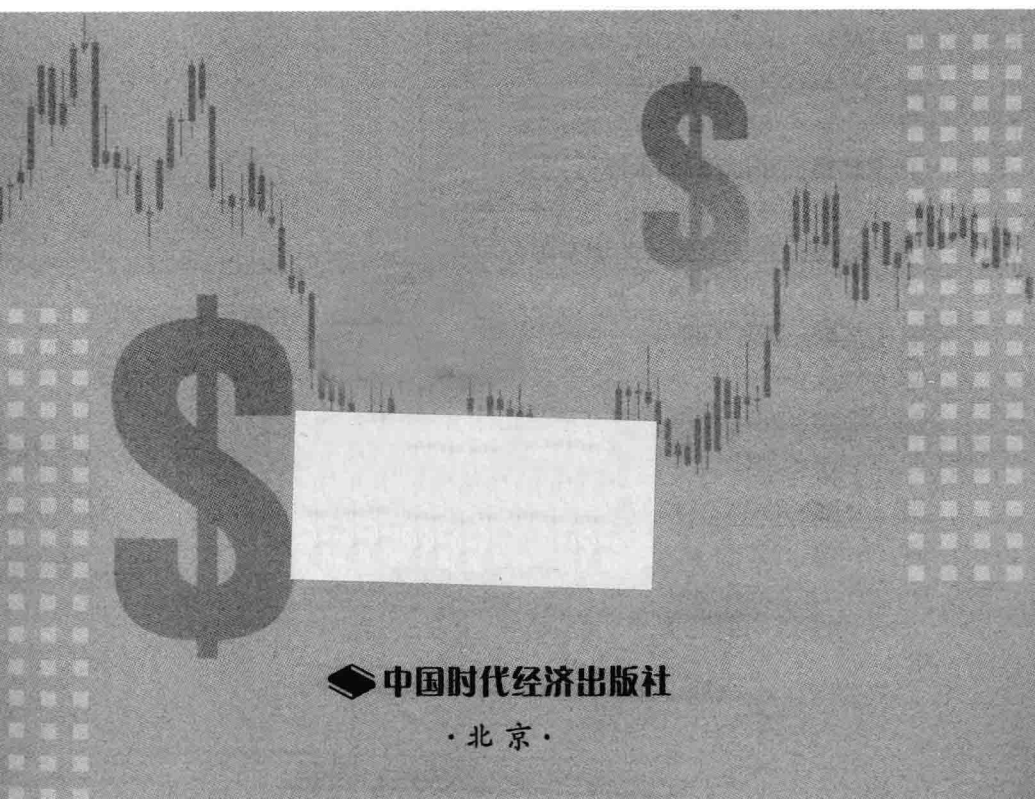
金融



时间序列预测

基于R语言的应用实践

王乐 金珏 王水◎著 >>>>>



 中国时代经济出版社

· 北京 ·

图书在版编目 (CIP) 数据

金融时间序列预测: 基于 R 语言的应用实践 / 王乐, 金珏, 王水著. —北京: 中国时代经济出版社, 2014. 9
ISBN 978-7-5119-0388-4

I. ①金… II. ①王… ②金… ③王… III. ①程序语言 - 应用 - 金融 - 时间序列分析 IV. ①F830 - 39

中国版本图书馆 CIP 数据核字 (2014) 第 218776 号

书 名: 金融时间序列预测: 基于 R 语言的应用实践
作 者: 王乐 金珏 王水

出版发行: 中国时代经济出版社

社 址: 北京市丰台区玉林里 25 号楼

邮政编码: 100069

发行热线: (010) 63508251 63508273

传 真: (010) 63508254 63508284

网 址: www.cmepub.com.cn

电子邮箱: zgsgdj@hotmail.com

经 销: 各地新华书店

印 刷: 北京紫瑞利印刷有限公司

开 本: 880 × 1230 1/32

字 数: 228 千字

印 张: 7

版 次: 2014 年 9 月第 1 版

印 次: 2014 年 9 月第 1 次印刷

书 号: ISBN 978-7-5119-0388-4

定 价: 29.00 元

本书如有破损、缺页、装订错误, 请与本社发行部联系更换
版权所有 侵权必究

前 言

随着金融自由化浪潮席卷全球，经济一体化趋势不断加强，金融创新达到了前所未有的高度，数量金融工程作为金融领域一门新兴学科得到迅速发展，而金融时间序列数据的分析则是经济决策的重要参考依据。但目前相关的中文书籍较少，部分专著又艰涩难懂，有曲高和寡之忧。由于这个原因，本书作者在攻读博士学位期间，以及在高校的科研和教学工作中，都深感不便；这也是撰写本书的主要动因之一。

由于本书独特的内容组织方式，没有 R 语言基础的读者也可以直接阅读，并进行实践和操作练习。当然如果读者曾经学习过某种编程语言，阅读起来会更为轻松。

本书以 R 语言计算环境为平台，从金融时间序列分析和预测的应用实践出发，以真实的金融数据为背景，对相关的 R 语言的语法和数据结构、时间序列数据的输入输出和可视化、回归方法、趋势分解、指数平滑方法、自回归移动平均方法等应用做了简明易懂的实例式讲述。其中，多数项目案例可以在数据分析实践中直接套用，对相关领域的工作人员和学者等，提供便捷的应用参考。

本书的第一章为 R 语言的快速入门，由于本章可以在极短的时间内让读者对 R 语言从陌生到熟悉建立起基本的应用概念，故称之为“闪电入门”。第二章为时间序列的 R 语言描述及常用处理，由于需要用到数据框等重要的数据结构，本章对列表、数据

框、矩阵、时间序列对象及相关 R 程序包等做了简要介绍。第三章为基本的时间序列分析方法，包括数据的获取、时间序列的分解和相关性分析等。第四章介绍基本预测方法及相关的概念，包括预处理、准确性度量，以及残差分析等。第五章为线性回归预测，主要是一元线性回归方法及分析。第六章是多重回归预测，以多只股票之间的关系为样例考察对象，介绍多元回归预测的基本方法。第七章为时间序列的分解和基于分解的预测。第八章是指数平滑方法及其包含趋势项和季节性组分的拓展形式。第九章为基于 ARIMA 模型的预测，包括季节性和非季节性 ARIMA 模型。第十章介绍使用技术分析指标的量化投资回测技术。

全书由王水提出创作动议及撰写“量化投资初步”，并负责通稿审校；王乐负责 R 语言入门、多元回归方法、时间序列分解、指数平滑和自回归移动平均方法相关章节；金珏负责时间序列数据处理、市场分析基本方法、简单预测和线性回归预测相关章节。

本书可作为金融数据分析领域的业务及研究人员参考使用，也可以作为金融、经济、管理等专业本科高年级学生和研究生教材，更可以直接作为普通投资者炒股理财的数据分析参考书。书中涉及的大部分操作的指令都极为简练，可以直接在 R 命令行输入运行；少数包含多条命令和函数的代码以 R 脚本的形式存放于 Google Code 可直接下载运行，其地址为：

<https://code.google.com/p/r-finance-forecast-practice-book/source/browse/>

由于作者的经验和水平所限，书中可能多有错谬之处，还望广大读者海涵并不吝赐教。E-mail: wangleboro@163.com。

2014 年 8 月于宁波大红鹰学院

目 录

第 1 章 R 语言的闪电入门 /1

1.1 R 简介.....	1
1.2 安装和配置 R 计算环境	4
1.3 十分钟的闪电教程	9
1.4 五分钟写两个 R 程序	16
1.5 R 大生境中的 SOS	20

第 2 章 金融时间序列的 R 表示 /22

2.1 时间序列数据的读入	22
2.2 列表 (list)	29
2.3 数据框：“列”的“列表”	33
2.4 矩阵 (matrix)	38
2.5 时间序列数据类型 ts	40

第 3 章 市场分析的基本方法 /45

3.1 读取在线股票数据	46
3.2 时间序列的分解	53

3.3 相关性分析	64
-----------------	----

第4章 股市的简单预测方法 /69

4.1 预测：能做到吗？	69
4.2 均值预测	73
4.3 单纯预测	76
4.4 预处理变换	81
4.5 衡量预测准确度	85
4.6 残差分析	90

第5章 线性回归预测 /96

5.1 线性回归	96
5.2 模型评价	102
5.3 R^2 指标	103
5.4 线性回归预测	104
5.5 拟合综述解释	107
5.6 虚假的回归？	109

第6章 多元回归预测 /113

6.1 多元线性回归的基本概念	113
6.2 残差分析	119
6.3 非线性回归	123
6.4 回归什么？预测什么？	131

第 7 章 季节和趋势：时间序列的分解 /133

7.1 序列分解的经典思路回顾	133
7.2 ts 数据类型	138
7.3 移动平均方法	140
7.4 经典分解法	143
7.5 STL 分解法	145
7.6 序列分解预测	147

第 8 章 指数平滑方法：原油价格预测 /150

8.1 简单指数平滑	151
8.2 Holt 线性趋势方法	156
8.3 阻尼趋势方法	159
8.4 Holt - Winters 季节方法	162
8.5 指数平滑组合模型	166

第 9 章 自回归移动平均 /175

9.1 平稳性和差分	175
9.2 回移算符	179
9.3 自回归模型	180
9.4 移动平均模型	182
9.5 非季节性 ARIMA 模型	183
9.6 ARIMA 模型预测的一般步骤	187
9.7 季节性 ARIMA	191
9.8 时间序列预测小结	194

第 10 章 R 量化投资初步 /197

10.1	回测	197
10.2	quantmod 包	198
10.3	技术指标	203
10.4	TTR 包	204
10.5	量化策略回测	209

第1章 R 语言的闪电入门

1.1 R 简介

简单地说（我们在本书中将一直坚持这个“简单”的风格），R 就是一个用于数据的统计处理的软件包，它支持使用一种简单的语言（就是所谓的 R 语言）来输入各种命令，例如：

```
>plot (1:10)
```

请注意，上面的“>”符号，是 R 的“提示符”，也就是说，所有命令都是在一个“>”符号后面输入的；因此，只有后面的“plot (1:10)”才是你输入的命令，而这个命令的含义就是：

“作一个图，里面包含从 1 到 10 共 10 个点”。

进一步我们可以这样理解：“作图”对应的英文单词就是“plot”，而“1:10”表示从 1 到 10 的是个整数。

上面的命令在 R 计算环境中将会产生如图 1-1 所示的图形。

如果引用维基百科（WikiPedia）的说法：

R 语言是一种自由软件编程语言与操作环境，主要用于统计分析、绘图、数据挖掘。R 本来是由来自新西兰奥克兰大学的 Ross Ihaka 和 Robert Gentleman 开发（也因此称为 R），现在由“R 开发核心团队”负责开发。R 是基于 S 语言的一个 GNU 计划

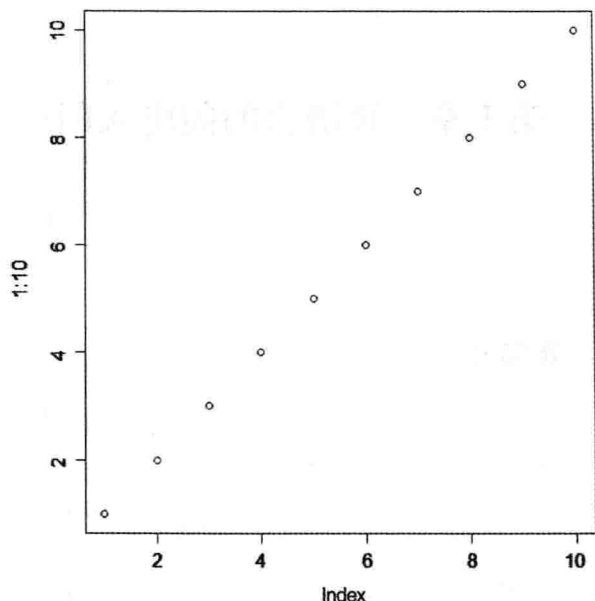


图 1-1 plot (1:10) 的作图效果

项目，所以也可以当作 S 语言的一种实现。

R 内置多种统计学及数字分析功能。R 的功能也可以通过安装包（Packages，用户撰写的功能）增强。

等等。

但是对于我们这些仅仅关注工具的使用的“俗人”来说，从上面以及其他一些官方的、板着脸的严肃说法中，能够领会出来意思可能只有下面 3 点：

第一：R 是免费的。

第二：R 最强大的用途就是数据处理和绘图。

第三：（也是 R 最为奇妙的特点）R 有无数个用途各异的“包（package）”；换句话说，只要你能想到的用途，极有可能在这个地球的某个角落已经有某个家伙把它写成了一个 R 包；而你

所要做的，就是找到这个 R 包，读读它的帮助文档，就可以着手开始你自己的工作了。

再次提请大家注意上面的第三点：是的，有点不可思议。我们不再需要抱着一本“R 语言设计大全”之类的教材，从第一页学到第 599，然后才能去开始我们的金融预测或基因分析之类与程序员八竿子打不着的工作，而是几乎立刻就可以开始自己真正关心的事情了——当然，前提是你很清晰地知道自己的任务。

实际上，现在 R 上至少已经有约 5000 个包，涵盖统计学、生物信息学、数据挖掘、人工智能、财经、生态、化学计量学、药物动力学、地理、高性能计算、基因学等几乎所有设计数值计算的学科，真真正正是“没有 R 做不到，只有你想不到”。当然，有些包因为是为特定的数据操作编写的，因此可能很快就会淡出 R 社区而被淘汰，但同时每天又会有新的 R 包在不断涌现。一些经典领域中的重要而常用的包，一般都有人在维护；关于这些领域的 R 包的简要介绍，可以在 R 的“任务视图 (Task Views)”网站找到，网站地址为：

<http://cran.r-project.org/web/views/>

我们关心的，主要是时间序列相关的分析和计算；该 Task View 的简介在：

时间序列：<http://cran.r-project.org/web/views/Time-Series.html>

当然，我们也可能需要另外两个任务集当中的一些功能，有兴趣的读者可以进一步阅读：

计量经济学：<http://cran.r-project.org/web/views/Econometrics.html>

金融学：<http://cran.r-project.org/web/views/Finance.html>

下面我们先看看如何在自己的电脑上安装 R 的运行环境。

1.2 安装和配置 R 计算环境

在你的电脑上安装 R 运行环境是一件极为简单的事情。但在安装之前，我们首先要知道，R 的运行环境是一套基本上不需要华丽人机界面的程序，因此，为了使用方便，我们还需要为 R 额外安装一个“外壳”，即图形界面程序，就像我们要用铁锹种庄稼，要在铁锹把上套一个橡胶套，并戴上手套一样。

因此，我们的安装过程可以由下面两个步骤实现：

- (1) 下载并安装 R（必需的）。
- (2) 下载并安装 RStudio（可选的，用着更顺手）。

首先第一步，安装 R：为简单起见，假定你的电脑使用的是 Windows 操作系统，到 R 的官方网站：

<http://www.r-project.org/>

找到“download R”的链接（<http://cran.r-project.org/mirrors.html>），你可以看到一串可以下载到 R 的镜像网站的列表（所谓镜像，就是内容一模一样的网站）。选一个和你离得比较近的镜像网站即可。例如，中国用户可以选择位于中国科技大学的镜像网站。

然后点击“Download R for Windows”，可以找到“install R for the first time”链接，这个链接中可以下载 R 的 Windows 版本。

需要说明的是，R 的 Windows 版本确实带有一个图形界面（GUI 界面），只是这个外壳有点简陋，如图 1-2 所示。

在“R Console”窗口的最下面，有一个小小的“>”，就是 R 的命令提示符，你可以在这个提示符后面键入你要 R 执行的命令。例如，看看它计算一个简单的正弦值的结果：

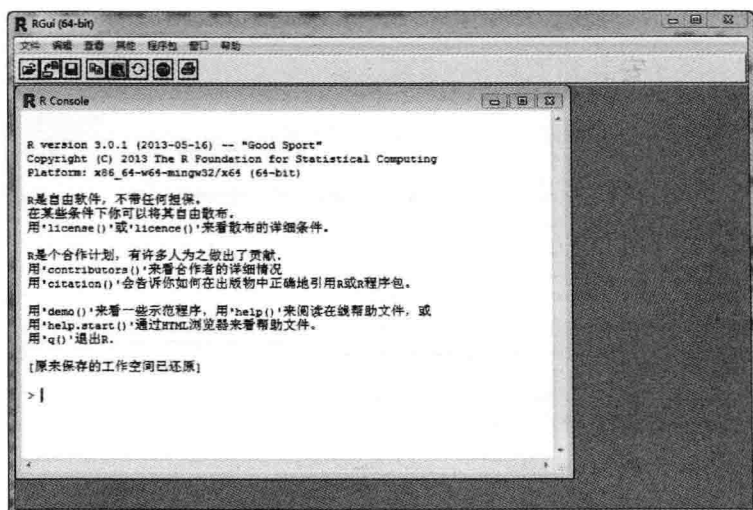


图 1-2 R for Windows 的运行界面

```
> sin(3.14)
[1] 0.001592653
```

上面显示的结果前面，有一个“[1]”，表示这是输出数据的第1项。这是R的特性之一：R总是将数据看成是一个系列数据的“列表”，哪怕你的结果只有一个数据也是如此。这样的列表，被称之为“向量”；向量的概念我们以后将经常遇到。

确认R安装完毕后，我们可以关闭上面的RGui窗口，然后到<http://www.rstudio.com/>下载并安装RStudio的Desktop版——这相当于我们前面说的“铁锹上的橡胶套”，用起来比直接使用RGui方便得多。注意：RStudio还有一个Server版，用于通过Web在服务器上执行R计算命令，不过这不在本书讨论范围之内。

RStudio 的运行界面如图 1-3 所示。

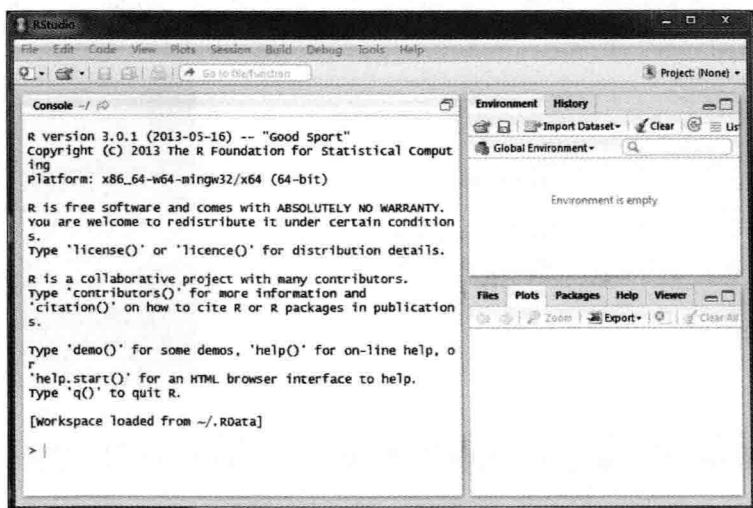


图 1-3 RStudio 的运行界面

和 RGui 类似，RStudio 也有一个带 “>” 提示符的 Console 窗口，可以直接输入命令。但是它还多了其他一些窗口，例如右面的 Environment 窗口用于显示当前系统内部的变量和数据，右下的 Plots 窗口用于绘制图形，Packages 窗口展示当前电脑中已经安装了的包，和当前运行的 RStudio 中已经“调入”了的包。

附注：所谓“调入”，指这个包的程序已经进入了你计算机的内存，随时可以使用了。仅仅安装一个包是不行的，就像一个游戏，仅仅拷贝到你的硬盘上还不能开始玩，只有调入内存后才可以玩。

除了使用 Packages 窗口调入你要用的包外，还可以使用 library 命令来调入，例如：

```
> library ("tseries")
```

调入已经安装过的“tseries”（时间序列分析）包。

安装一个程序包有两种方法。

第一种方法是使用 RStudio 的 Tools 菜单的 “Install Packages ...” 命令，该命令弹出如图 1-4 所示的对话框。

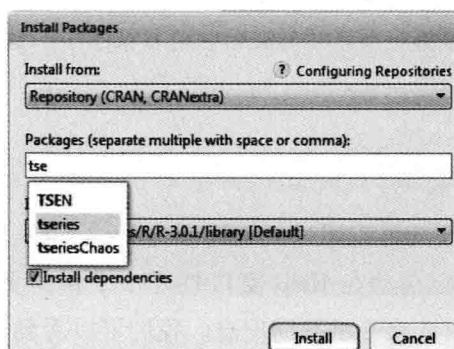


图 1-4 安装 R 程序包

在上面对话框中间的空白处，填上要安装的程序包的名称即可。该对话框有一个很好的特性，就是你在键入包名的时候，它可以随时给出提示，以免你键入名称的时候发生拼写错误，或者写错大小写等。

第二种安装程序包的方法是直接使用 `install.packages()` 命令。下面给出使用命令行来安装的具体输出结果。

```
> install.packages("tseries")
trying URL http://cran.rstudio.com/bin/windows/contrib/3.0/tseries_
0.10-32.zip
Content type 'application/zip' length 320827 bytes (313 Kb)
opened URL
downloaded 313 Kb
package 'tseries' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
```

```
C: \ Users \ Administrator \ AppData \ Local \ Temp \ RtmpMV5n9a \  
downloaded_ packages
```

注意：观察上面的输出信息，我们可以发现，系统将下载到的程序包存储到 C 盘一个非常隐蔽的临时目录下了。这显然不太适合我们以后的维护。

```
> ? install.packages
```

实际上，我们可以通过来查询 `install.packages()` 命令的使用帮助。系统会自动在 Help 窗口打开一个帮助页面，显示 `install.packages()` 命令的帮助信息。我们可以看到，该命令的使用语法是：

```
install.packages (pkgs, lib, repos = getOption ("repos"),  
  contriburl = contrib.url (repos, type),  
  method, available = NULL, destdir = NULL,  
  dependencies = NA, type = getOption ("pkgType"),  
  configure.args = getOption ("configure.args"),  
  configure.vars = getOption ("configure.vars"),  
  clean = FALSE, Ncpus = getOption ("Ncpus", 1L),  
  verbose = getOption ("verbose"),  
  libs_ only = FALSE, INSTALL_ opts, quiet = FALSE,  
  keep_ outputs = FALSE, ...)
```

上面语法中的大部分参数我们可以暂且不管（语法中的…表示还有其他的参数，被省略了；本书沿用这种表示法，在描述输入输出信息的时候，也用…表示被省略的信息），只需要知道，其中的 `lib` 参数指定了下载的程序包的存放位置；因此，将我们的安装命令改为：