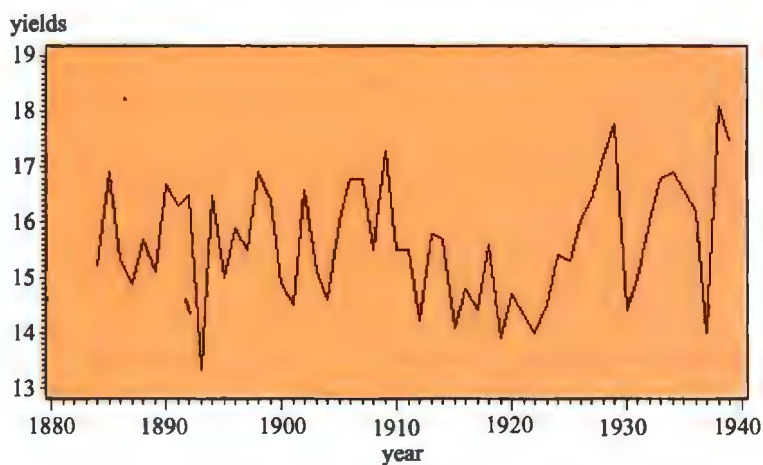


时间序列分析

—— 基于R

Time Series Analysis with R

王 燕 编著





基于R应用的统计学丛书

统计学——基于R

多元统计分析——基于R

● 时间序列分析——基于R

回归分析——基于R

贾俊平

费宇

王燕

吴喜之



时间序列分析

——基于R

Time Series Analysis with R

人大经管图书在线 www.rdjg.com.cn
了解图书出版信息 下载教学辅助资料

策划编辑 王伟娟
责任编辑 王前 张佳佳
封面设计 贾俊平 拾光书坊

ISBN 978-7-300-20939-5



定价：35.00元



基于R应用的统计学丛书

时间序列分析

—— 基于R

Time Series Analysis with R

王 燕 编著

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

时间序列分析：基于 R / 王燕编著. — 北京：中国人民大学出版社，2015. 3
(基于 R 应用的统计学丛书)
ISBN 978-7-300-20939-5

I. ①时… II. ①王… III. ①时间序列分析－高等学校－教材 IV. ①O211.61

中国版本图书馆 CIP 数据核字 (2015) 第 039238 号

基于 R 应用的统计学丛书
时间序列分析——基于 R
王 燕 编著
Shijianxulie Fenxi: Jiyu R

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511770 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	三河市汇鑫印务有限公司		
规 格	185 mm × 260 mm 16 开本	版 次	2015 年 9 月第 1 版
印 张	16.75 插页 1	印 次	2015 年 9 月第 1 次印刷
字 数	368 000	定 价	35.00 元

版权所有 侵权必究 印装差错 负责调换

前 言

时间序列分析是统计学科的一个重要分支. 它主要研究随着时间的变化事物发生、发展的过程, 寻找事物发展变化的规律, 并预测未来的走势. 在日常生产、生活中, 时间序列比比皆是, 所以目前时间序列分析方法广泛应用于经济、金融、天文、气象、海洋、物理、化学、医学、质量控制等诸多领域, 成为众多行业经常使用的统计方法.

目前, 国内有关时间序列分析的著作和教材有很多, 但主要是基于 SAS 软件或 EViews 软件编写的. 近几年来, R 语言开始崛起, 它是一个可进行交互式数据分析的强大平台. R 语言的以下几个特征让它在学界和业界都受到很大的重视.

首先, R 语言是自由的开源软件, 也就是说, 它是全球免费的. 正版的专业统计软件通常要收取高额的版权使用费, 这使很多高校的教师和学生无法获得正版的专业统计软件, 而 R 语言没有版权使用费这个障碍, 人们可以光明正大、放心大胆地使用.

其次, R 语言不仅是一款统计软件, 还是一个可以进行交互式数据分析和探索的强大平台, 金融、经济、医疗、数据挖掘等诸多领域都基于 R 在研发它们的分析方法. 在这个平台上, 时间序列分析方法可以非常便捷地嵌入其他领域的研究中, 成为各行各业实务分析的基本方法.

最重要的一点是, 由于 R 语言的开放性和资源共享性, 使得它可以汇集全球 R 用户的智慧和创造力, 以惊人的速度发展. 新方法的更新速度是以周来计算的, 这是传统统计软件所无法比拟的. R 具有自由广阔的发展前景, 可以预期, 它很有可能会打破专业统计软件的功能边界, 与时俱进, 不断拓宽应用领域, 不断创造出更多的功能和解法. 所以, 我们需要学习并共同发展 R 语言.

基于 R 语言的这些特点, 我们推出教材《时间序列分析——基于 R》.

需要说明的是, 本书直接使用了笔者编著的 21 世纪统计学系列教材《应用时间序列分析》的结构和内容. 这样安排主要是为了便于部分教师和读者在原有 SAS 软件的基础上, 迅速掌握 R 语言的应用.

最后, 感谢所有使用这本教材的朋友们. 尽管笔者本着认真的态度编写, 但是水平

有限, 书中谬误之处在所难免, 欢迎大家批评指正.

王 燕

教师教学服务说明

中国人民大学出版社工商管理分社以出版经典、高品质的工商管理、财务会计、统计、市场营销、人力资源管理、运营管理、物流管理、旅游管理等领域的各层次教材为宗旨。

为了更好地为一线教师服务，近年来工商管理分社着力建设了一批数字化、立体化的网络教学资源。教师可以通过以下方式获得免费下载教学资源的权限：

在“人大经管图书在线”（www.rdjg.com.cn）注册，下载“教师服务登记表”，或直接填写下面的“教师服务登记表”，加盖院系公章，然后邮寄或传真给我们。我们收到表格后将在一个工作日内为您开通相关资源的下载权限。

如您需要帮助，请随时与我们联系：

中国人民大学出版社工商管理分社

联系电话：010－62515735，62515749，62515987

传 真：010－62515732，62514775 电子邮箱：rdcbjsjg@crup.com.cn

通讯地址：北京市海淀区中关村大街甲 59 号文化大厦 1501 室（100872）

教师服务登记表

姓 名	☐ 先生 ☐ 女士		职 称		
座机/手机				电子邮箱	
通讯地址				邮 编	
任教学校				所在院系	
所授课程	课程名称	现用教材名称	出版社	对象（本科生/研究生/MBA/其他）	学生人数
需要哪本教材的配套资源					
人大经管图书在线用户名					
院/系领导（签字）： 院/系办公室盖章					

目 录

第 1 章 时间序列分析简介	1
1.1 引 言	1
1.2 时间序列的定义	1
1.3 时间序列分析方法	2
1.3.1 描述性时序分析	2
1.3.2 统计时序分析	4
1.4 R 简介	6
1.4.1 R 的特点	6
1.4.2 R 的安装	7
1.4.3 R 语言基本规则	8
1.4.4 生成时间序列数据	11
1.4.5 时间序列数据的处理	14
1.4.6 时间序列数据导出	16
1.5 习 题	16
第 2 章 时间序列的预处理	17
2.1 平稳时间序列	17
2.1.1 特征统计量	17
2.1.2 平稳时间序列的定义	19
2.1.3 平稳时间序列的统计性质	20
2.1.4 平稳时间序列的意义	21
2.2 时序图与自相关图	23
2.2.1 时序图	23

2.2.2 绘制序列自相关图	29
2.3 平稳性的检验	30
2.3.1 时序图检验	30
2.3.2 自相关图检验	32
2.4 纯随机性检验	33
2.4.1 纯随机序列的定义	34
2.4.2 白噪声序列的性质	35
2.4.3 纯随机性检验	36
2.5 习 题	41
第 3 章 平稳时间序列分析	44
3.1 方法性工具	44
3.1.1 差分运算	44
3.1.2 延迟算子	45
3.1.3 线性差分方程	45
3.2 ARMA 模型的性质	47
3.2.1 AR 模型	47
3.2.2 MA 模型	62
3.2.3 ARMA 模型	69
3.3 平稳序列建模	72
3.3.1 建模步骤	72
3.3.2 样本自相关系数与偏自相关系数	73
3.3.3 模型识别	73
3.3.4 参数估计	81
3.3.5 模型检验	87
3.3.6 模型优化	91
3.4 序列预测	96
3.4.1 线性预测函数	96
3.4.2 预测方差最小原则	97
3.4.3 线性最小方差预测的性质	98
3.5 习 题	105
第 4 章 非平稳序列的确定性分析	109
4.1 时间序列的分解	109
4.1.1 Wold 分解定理	109
4.1.2 Cramer 分解定理	110

4.2	确定性因素分解	111
4.3	趋势分析	112
4.3.1	趋势拟合法	112
4.3.2	平滑法	117
4.4	季节效应分析	125
4.5	综合分析	127
4.6	习 题	133
第 5 章	非平稳序列的随机分析	136
5.1	差分运算	136
5.1.1	差分运算的实质	136
5.1.2	差分方式的选择	137
5.1.3	过差分	141
5.2	ARIMA 模型	142
5.2.1	ARIMA 模型的结构	142
5.2.2	ARIMA 模型的性质	143
5.2.3	ARIMA 模型建模	145
5.2.4	ARIMA 模型预测	147
5.2.5	疏系数模型	150
5.2.6	季节模型	154
5.3	残差自回归模型	161
5.3.1	模型结构	162
5.3.2	残差自相关检验	165
5.3.3	残差自相关模型拟合	168
5.4	异方差的性质	170
5.4.1	异方差的影响	170
5.4.2	异方差的直观诊断	171
5.5	方差齐性变换	173
5.6	条件异方差模型	175
5.6.1	ARCH 模型	176
5.6.2	GARCH 模型	184
5.6.3	GARCH 的衍生模型	191
5.7	习 题	193
第 6 章	多元时间序列分析	198
6.1	平稳多元序列建模	198
6.2	虚假回归	203

6.3 单位根检验	205
6.3.1 DF 检验	205
6.3.2 ADF 检验	214
6.4 协 整	219
6.4.1 单整与协整	219
6.4.2 协整检验	220
6.5 误差修正模型	223
6.6 习 题	225
附 录	228
参考文献	258

1.1 引言

最早的时间序列分析可以追溯到 7 000 年前的古埃及. 当时, 为了发展农业生产, 古埃及人一直在密切关注尼罗河泛滥的规律. 把尼罗河涨落的情况逐天记录下来, 就构成了所谓的时间序列. 对这个时间序列长期的观察使他们发现尼罗河的涨落非常有规律. 天狼星第一次和太阳同时升起的那一天之后, 再过 200 天左右, 尼罗河就开始泛滥, 泛滥期将持续七八十天, 洪水过后, 土地肥沃, 随意播种就会有丰厚的收成. 由于掌握了尼罗河泛滥的规律, 古埃及的农业迅速发展, 解放出大批的劳动力去从事非农业生产, 从而创建了古埃及灿烂的史前文明.

像古埃及人一样, 按照时间的顺序把随机事件变化发展的过程记录下来就构成了一个时间序列. 对时间序列进行观察、研究, 寻找它变化发展的规律, 预测它将来的走势就是时间序列分析.

1.2 时间序列的定义

在统计研究中, 常用按时间顺序排列的一组随机变量

$$X_1, X_2, \dots, X_t, \dots \quad (1.1)$$

来表示一个随机事件的时间序列, 简记为 $\{X_t, t \in T\}$ 或 $\{X_t\}$.

用

$$x_1, x_2, \dots, x_n \quad (1.2)$$

或 $\{x_t, t = 1, 2, \dots, n\}$ 表示该随机序列的 n 个有序观察值, 称为序列长度为 n 的观察值序列, 有时也称式 (1.2) 为式 (1.1) 的一个实现.

在日常生产生活中, 观察值序列比比皆是. 比如把全国 1999—2008 年普通高等学校每年的招生人数按照时间顺序记录下来, 就构成了一个序列长度为 10 的全国普通高等学校招生人数时间序列 (单位: 万人):

159.7, 220.6, 268.3, 320.5, 382.2, 447.3, 504.5, 546.1, 565.9, 607.7

我们进行时序研究的目的是想揭示随机时序 $\{X_t\}$ 的性质, 而要实现这个目标就要分析它的观察值序列 $\{x_t\}$ 的性质, 由观察值序列的性质来推断随机时序 $\{X_t\}$ 的性质.

1.3 时间序列分析方法

1.3.1 描述性时序分析

早期的时序分析通常都是通过直观的数据比较或绘图观测, 寻找序列中蕴涵的发展规律, 这种分析方法就称为描述性时序分析. 古埃及人发现尼罗河泛滥的规律就是依靠这种分析方法. 而在天文、物理、海洋学等自然科学领域, 这种简单的描述性时序分析方法也常常能使人们发现意想不到的规律.

比如根据《史记·货殖列传》记载, 早在春秋战国时期, 范蠡和计然就提出我国农业生产具有“六岁穰, 六岁旱, 十二岁一大饥”的自然规律.《越绝书·计倪内经》则描述得更加详细:“太阴三岁处金则穰, 三岁处水则毁, 三岁处木则康, 三岁处火则旱……天下六岁一穰, 六岁一康, 凡十二岁一饥.”

用现代汉语来表述就是:“木星绕天空运行, 运行三年, 如果处于金位, 则该年为大丰收年; 如果处于水位, 则该年为大灾年; 再运行三年, 如果处于木位, 则该年为小丰收年; 如果处于火位, 则该年为小灾年, 所以天下平均六年一大丰收, 六年一小丰收, 十二年一大饥荒.”这是 2 500 多年前, 我国对农业生产具有 3 年一小波动、12 年左右一个大周期的记录, 是一个典型的描述性时间序列分析.

描述性时序分析方法是人们在认识自然、改造自然的过程中发现的实用方法. 对于很多自然现象, 只要人们观察时间足够长, 就能运用描述性时序分析发现蕴涵在时间里的自然规律. 根据自然规律, 做恰当的政策安排, 就能有利于社会的发展和进步.

比如范蠡根据“六岁穰, 六岁旱, 十二岁一大饥”的自然规律提出:“夫粢, 二十病农, 九十病末. 末病则财不出, 农病则草不辟矣. 上不过八十, 下不减三十, 则农末俱利, 平粢齐物, 关市不乏, 治国之道也.”这段话的意思是: 如果丰收年粮食贱卖, 会挫伤农民种粮的积极性; 如果大灾年粮价高涨, 会危及老百姓的生存, 所以要实行“平粢”法. 政府应该在粮食丰收时以高于最低价的价格购买粮食进行储备, 以保护农民的利益; 在粮食短缺时, 将储备的粮食投放市场, 以稳定粮价, 确保百姓的生存. 这是对农民和百姓都有利的政策, 是一个国家的治国之道.

在范蠡故去 2 000 年之后, 欧洲经济学家在研究欧洲各地粮食产量时发现了类似规律. 比如 19 世纪末 20 世纪初英格兰和威尔士的小麦平均亩产序列就具有这种规律 (数据见表 A1—1), 如图 1—1 所示.

小麦的产量直接影响到小麦的价格, 丰收时价格便宜, 价格指数就偏低; 歉收时, 价格上涨, 价格指数就偏高. 在时间序列领域里, 有一个非常著名的序列叫贝弗里奇 (Beveridge) 小麦价格指数序列, 它由 1500—1869 年的小麦价格构成 (数据见表 A1—2). 1971 年 Granger 和 Hughes 分析该序列, 发现该序列有一个 13 年左右的周期. 部

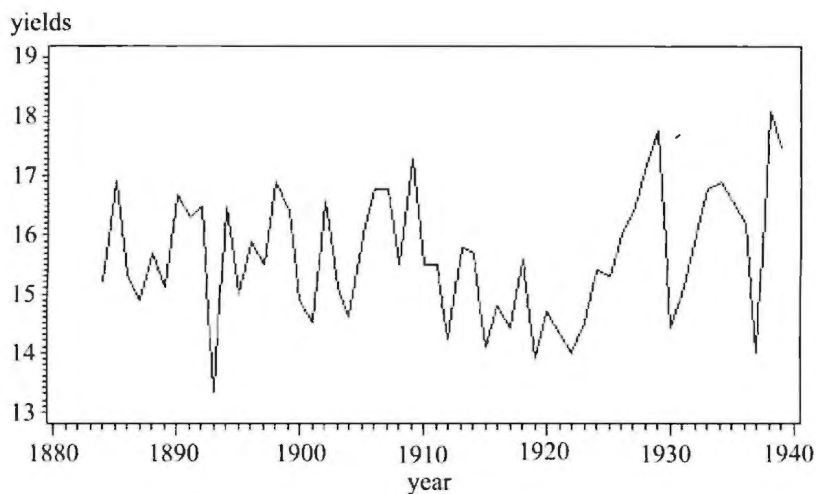


图 1—1 1884—1939 年英格兰与威尔士每亩小麦产量时序图

分 Beveridge 小麦价格指数序列走势如图 1—2 所示.

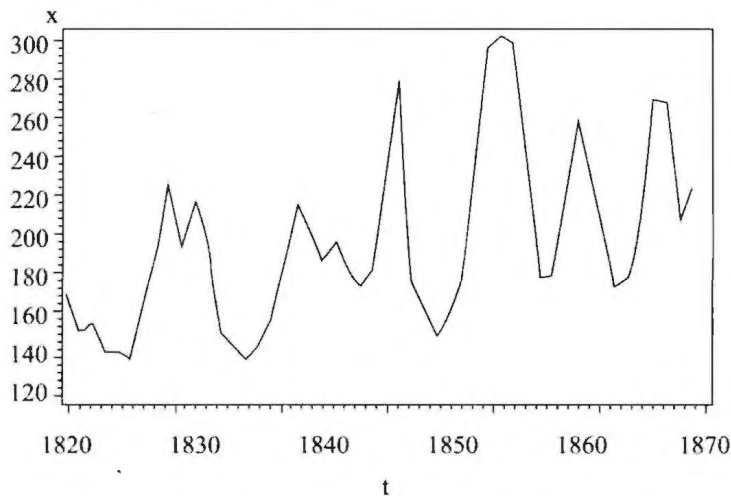


图 1—2 部分 Beveridge 小麦价格指数时序图

西方学者致力于研究为什么粮食产量会有这样的周期波动. 19 世纪中后期, 德国药剂师、业余天文学家 S. H. Schwabe 经过几十年不间断的观察、记录, 发现了太阳黑子的活动具有 11~12 年的周期 (数据见表 A1—3), 如图 1—3 所示. 太阳黑子的运动周期和农业生产的周期长度非常接近, 这引起了英国天文学家、天王星的发现者 F. W. Herschel 的关注. 最后他发现, 当太阳黑子变少时, 地球上的雨量也会减少. 所以在没有良好人工灌溉技术的时代, 农业生产会有和太阳黑子近似的变化周期.

人们没有采用任何复杂的模型或分析方法, 仅仅是按照时间顺序收集数据, 描述和呈现序列的波动, 就了解到小麦产量的周期波动特征、产生该周期波动的气候原因以及该周期波动对价格的影响. 操作简单、直观有效是描述性时间序列分析方法的突出特点. 它通常也是人们进行统计时序分析的第一步, 通过图示的方法直观地反映出序列的波动特征.

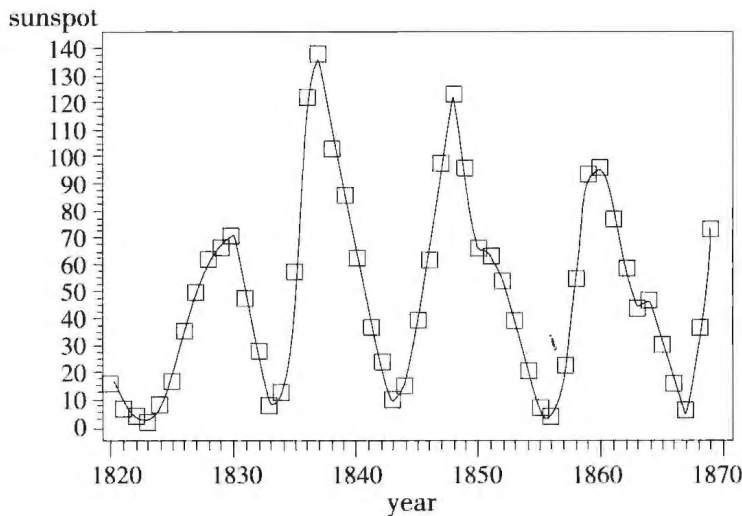


图 1—3 1820—1869 年太阳黑子年度数据时序图

1.3.2 统计时序分析

随着研究领域的不断拓展,人们发现单纯的描述性时序分析有很大的局限性.在金融、保险、法律、人口、心理学等社会科学研究领域,随机变量的发展通常会呈现出非常强的随机性,想通过对序列简单的观察和描述总结出随机变量发展变化的规律,并准确预测出它们将来的走势通常是非常困难的.

为了更准确地估计随机序列发展变化的规律,从 20 世纪 20 年代开始,学术界利用数理统计学原理分析时间序列.研究的重心从总结表面现象转移到分析序列值内在的相关关系上,由此开辟了一门应用统计学科——时间序列分析.

纵观时间序列分析方法的发展历史可以将时间序列分析方法分为两大类.

一、频域分析方法

频域 (frequency domain) 分析方法也称为“频谱分析”或“谱分析” (spectral analysis) 方法.

早期的频域分析方法假设任何一种无趋势的时间序列都可以分解成若干不同频率的周期波动,借助富里埃分析从频率的角度揭示时间序列的规律,后来又借助傅里叶变换,用正弦、余弦项之和来逼近某个函数.20 世纪 60 年代, Burg 在分析地震信号时提出最大熵谱估计理论,该理论克服了传统谱分析所固有的分辨率不高和频率漏泄等缺点,使谱分析进入一个新阶段,称为现代谱分析阶段.

目前谱分析方法主要运用于电力工程、信息工程、物理学、天文学、海洋学和气象科学等领域,它是一种非常有用的纵向数据分析方法.但是由于谱分析过程一般都比较复杂,研究人员通常要具有很强的数学基础才能熟练使用,同时它的分析结果也比较抽象,不易于进行直观解释,所以谱分析方法的使用具有很大的局限性.

二、时域分析方法

时域 (time domain) 分析方法主要是从序列自相关的角度揭示时间序列的发展规律. 相对于谱分析方法, 它具有理论基础扎实、操作步骤规范、分析结果易于解释等优点. 目前它已广泛应用于自然科学和社会科学的各个领域, 成为时间序列分析的主流方法. 本书主要就是介绍时域分析方法.

时域分析方法的基本思想是事件的发展通常都具有一定的惯性, 这种惯性用统计的语言来描述就是序列值之间存在一定的相关关系, 而且这种相关关系具有某种统计规律. 我们分析的重点就是寻找这种规律, 并拟合出适当的数学模型来描述这种规律, 进而利用这个拟合模型来预测序列未来的走势.

时域分析方法具有相对固定的分析套路, 通常都遵循如下分析步骤:

第一步: 考察观察值序列的特征.

第二步: 根据序列的特征选择适当的拟合模型.

第三步: 根据序列的观察数据确定模型的口径.

第四步: 检验模型, 优化模型.

第五步: 利用拟合好的模型来推断序列其他的统计性质或预测序列将来的发展.

时域分析方法的产生最早可以追溯到 1927 年, 英国统计学家 G. U. Yule (1871—1951) 提出自回归 (autoregressive, AR) 模型. 不久之后, 英国数学家、天文学家 G. T. Walker 爵士在分析印度大气规律时使用了移动平均 (moving average, MA) 模型和自回归移动平均 (autoregressive moving average, ARMA) 模型 (1931 年). 这些模型奠定了时间序列时域分析方法的基础.

1970 年, 美国统计学家 G. E. P. Box 和英国统计学家 G. M. Jenkins 联合出版了 *Time Series Analysis: Forecasting and Control* 一书. 在书中, Box 和 Jenkins 在总结前人研究的基础上, 系统地阐述了对求和自回归移动平均 (autoregressive integrated moving average, ARIMA) 模型的识别、估计、检验及预测的原理和方法. 这些知识现在称为经典时间序列分析方法, 是时域分析方法的核心内容. 为了纪念 Box 和 Jenkins 对时间序列发展的特殊贡献, 现在人们也常把 ARIMA 模型称为 Box-Jenkins 模型.

Box-Jenkins 模型实际上是主要运用于单变量、同方差场合的线性模型. 随着人们对各领域时间序列的深入研究, 发现该经典模型在理论和应用上都还存在许多局限性. 所以从 20 世纪 80 年代以来, 统计学家纷纷转向多变量场合、异方差场合和非线性场合的时间序列分析方法的研究, 并取得了突破性的进展.

在异方差场合, 美国统计学家、计量经济学家 Robert F. Engle 在 1982 年提出了自回归条件异方差 (ARCH) 模型, 用以研究英国通货膨胀率的建模问题. 为了进一步放宽 ARCH 模型的约束条件, Bollerslov 在 1985 年提出了广义自回归条件异方差 (GARCH) 模型. 随后 Nelson 等人又提出了指数广义自回归条件异方差 (EGARCH) 模型、方差无穷广义自回归条件异方差 (IGARCH) 模型和依均值广义自回归条件异方差 (GARCH-M) 模型等限制条件更为宽松的异方差模型. 这些异方差模型是对经典的 ARIMA 模型很好的补充. 它们比传统的方差齐性模型更准确地刻画了金融市场风

险的变化过程, 因此 ARCH 模型及其衍生出的一系列拓展模型在计量经济学领域有着广泛的应用. Engle 也因此获得 2003 年诺贝尔经济学奖.

在多变量场合, Box 和 Jenkins 在 *Time Series Analysis: Forecasting and Control* 一书中研究过平稳多变量序列的建模, Box 和 Tiao 在 1970 年前后讨论过带干扰变量的时间序列分析. 这些研究实际上是把对随机事件的横向研究和纵向研究有机地融合在一起, 提高了对随机事件分析和预测的精度. 1987 年, 英国统计学家、计量经济学家 C. Granger 提出了协整 (co-integration) 理论, 进一步为多变量时间序列建模松绑. 有了协整的概念之后, 在多变量时间序列建模过程中“变量是平稳的”不再是必需条件, 而只要求它们的某种线性组合平稳. 协整概念的提出极大地促进了多变量时间序列分析方法的发展, Granger 因此与 Engle 一起获得了 2003 年诺贝尔经济学奖.

在非线性场合, 各种新的模型层出不穷. Granger 和 Andersen 在 1978 年提出了双线性模型, Howell Tong 于 1978 年提出了门限自回归模型, Priestley 于 1980 年提出了状态相依模型, Hamilton 于 1989 年提出了马尔科夫转移模型, Lewis 和 Stevens 于 1991 年提出了多元适应回归样条方法, Carlin 等人于 1992 年提出了非线性状态空间建模的方法, Chen 和 Tsay 于 1993 年提出了非线性可加自回归模型. 非线性是一个异常广阔的研究空间, 在非线性的模型构造、参数估计、参数检验等各方面都有大量的研究工作需要完成.

1.4 R 简介

1.4.1 R 的特点

R 是 S 语言的一种实现. 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. 1995 年新西兰奥克兰大学的 Ross Ihaka 和 Robert Gentleman 在 S 语言的基础上开发了一个新系统, 由于这两位科学家的名字的首字母都是 R, 所以该系统软件被取名为 “R”.

R 从开发伊始就定位为开源软件, 所以它所有的代码和帮助文件都免费向全球用户开放. 目前 R 语言的维护和开发由 R 核心开发小组 (R Development Core Team) 负责, 核心团队秉承兴趣与共享的原则, 义务而高效地为全球 R 用户服务.

R 是一个交互式的平台系统. 一个标配的 R 语言 (Base R) 里面只安装了 25 个标准包, 给用户提供最基本的数据存储和处理功能. 很多特殊的功能由一系列程序包 (package) 提供, 目前 R 可以提供 6 000 多个程序包供用户使用. 用户可以根据自己的需要从全球 100 多个 R 语言镜像站下载不同的程序包, 以满足自己对特殊功能的需要.

目前, R 语言已经有若干个专门用于时间序列分析的程序包. 借助这些程序包, 我们能够完成序列读入、绘图、识别、建模、预测等一系列分析工作. R 语言编程简洁、绘图功能强大、分析结果准确, 是进行时间序列分析与预测的常用软件之一.

1.4.2 R 的安装

一、R 软件的下载安装

登录 CRAN (Comprehensive R Archive Network) 的官网 (<http://cran.r-project.org>), 根据自己电脑的操作系统 (Windows, Linux 或者 Mac OS X), 选择适当的 R 版本免费下载。下载之后, 根据提示安装 R 软件。安装结束后, 桌面上会出现 R 快捷图标。双击 R 图标就可以进入 R 操作系统了。

二、程序包的下载安装

R 语言提供多种下载与安装程序包的方式, 我们在此介绍最简单的两种。

1. 通过菜单栏选项下载安装

下面以下载时间序列分析程序包 `tseries` 为例, 介绍菜单栏下载步骤。

第一步: 双击 R 图标进入 R (3.10 版本) 操作系统, 点击界面最上面菜单栏“程序包”按钮, 再点击“安装程序包”(见图 1—4)。

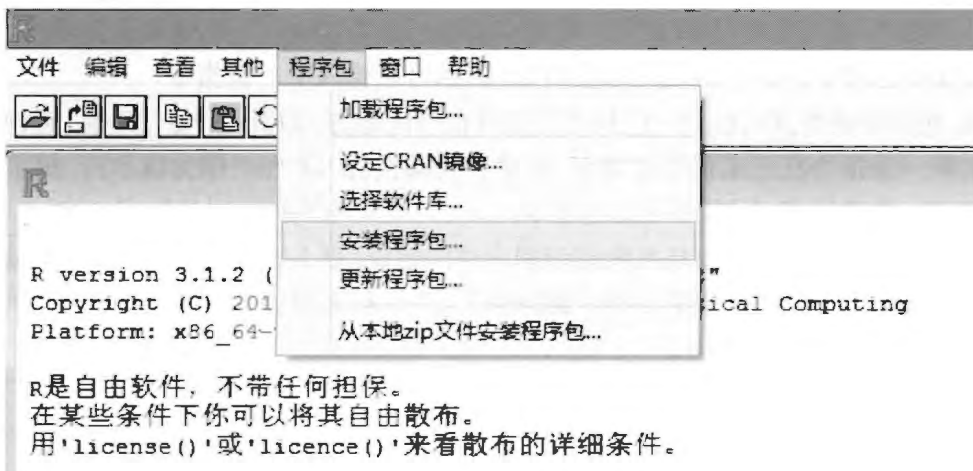


图 1—4 选择安装程序包选项

第二步: 选择本次下载的镜像站。全球一共有 100 多个镜像站, 中国境内有 4 个: 北京 2 个, 合肥 1 个, 厦门 1 个, 使用者可以根据自己的网络情况, 任意选择一个自己可以访问的镜像站作为程序包的下载平台。

第三步: 选择下载程序包。选择某个镜像站之后, 就进入程序包选择下载界面。根据程序包名字的字母检索, 选中本次想要下载的程序包 `tseries`, 点击“确定”即可。

2. 输入指令下载安装

在计算机联网状态下, 直接在对话窗口输入安装 `tseries` 程序包的指令:

```
> install.packages("tseries")
```


并按回车键,系统会自动弹出镜像站选择界面,任意选择一个自己可以访问的镜像站作为自己程序包的下载平台即可。

三、程序包的加载调用

原则上讲,用户可以下载安装所有的 R 语言程序包。如果每次启动 R 语言,所有下载的程序包都调入内存备用,那么 R 语言将非常庞大,并占用很多内存资源。而实际上我们通常只用有限的几个 R 程序包就可以解决所研究的问题。所以 R 语言规定,每次进入 R 语言之后,下载的 R 程序包需要通过 `library` 命令进行加载,才可以在本次 R 语言运行中被调用。加载 `tseries` 程序包的指令为:

```
> library(tseries)
```

1.4.3 R 语言基本规则

一、输入指令的规则

进入 R 语言对话窗口,红色的“>”后面光标在闪烁,提示在这个位置可以输入指令。一条指令输入完毕,回车之后,系统立刻执行该条指令,无论结果是否输出,下一行起首还是红色的“>”,后面光标在闪烁,提示可以输入下一条指令。

R 允许对指令进行注释。注释可以被放在任何地方,以井号“#”开始,到行末结束。如果一条指令在行末仍没有结束,R 将会自动换行,以“+”作为提示符,提示继续书写注释。注释写完,回车即可运行指令。

假如指令写得不对,而且系统能识别出你的错误,那么在指令输完回车后,系统会立刻提示你错误信息,这条错误指令就结束了,下一行起首还是红色的“>”,提示可以输入新的指令了。

```
> sun(1,2,3)           #输入错误函数名
错误: 没有"sun"这个函数  #错误提示
>
```

假如指令写得不对,而且系统不能识别出你的错误,那么下一行起始符号就是“+”,这表示系统依然在等你输入一个完整的正确的指令。如果在“+”后面输入右括号并回车,系统会自动将两部分的指令合并在一起,给出结果,而且下一行起始符号为“>”,系统在等待我们输入新的指令。

```
> sum(1,2,3           #求和函数没有编辑右括号
+)                   #补右括号
[1]6                 #给出求和结果
>
```

如果系统无法识别我们的错误,我们自己也没办法在“+”后面完善这条指令,因

此希望放弃执行这条指令,则可以按电脑上的 ESC 功能键,退出这一条指令编辑,这时下一行起始符号又会变为“>”,我们就可以输入新的指令了。

如果一行内要编辑多条指令,可以用分号来分隔不同的指令,回车后就可以将多条指令一次性执行了。

```
> x=1; y=2; x+y      #把 1 赋值给 x, 把 2 赋值给 y, 再计算 x+y 的值
[1] 3                  #给出计算结果
```

如果想每行写一条指令,但是多条指令一次性执行,可以用花括号“{ }”将批处理的指令标注起来,系统就会等你编辑完成之后一次性运行了。

```
> {x=1
+ y=2
+ x+y}
[1] 3
```

二、赋值命令的写法

在很多软件中,通常都是以等号“=”作为赋值符号,R 语言也认可等号赋值的写法。但是在 R 语言中,标准写法是使用带方向的箭头号“<-”或者“->”表示赋值。假如我们要完成如下赋值命令:

$$\begin{aligned}x &= 3 \\ y &= 2x + 1 \\ z &= y^2\end{aligned}$$

我们可以在 R 语言中编辑如下命令:

```
> x<-3
> 2*x+1->y
> z<-y^2
```

在完成上面三条赋值语句后,我们会发现 R 语言界面没有任何结果输出。这是 R 语言和其他统计专业软件的一个不同之处。很多统计专业软件(比如 SAS)采用的是最大输出结果模式。运行一个分析程序之后,系统会给出非常丰富的输出结果。但是 R 语言采用的是最小输出结果模式。R 语言会将丰富的分析与运算结果存入适当的对象中,所有对象存于软件后台,它不会主动在对话窗口输出赋值对象的结果,你需要调用对象名进行结果查询。

比如我们想知道上例中 x, y, z 的赋值结果,查询命令和输出结果如下:

```
> x
[1] 3
> y
[1] 7
> z
[1] 49
```


三、区分大小写

有些软件不区分大小写, 在编写程序时字母的大小写不影响程序的识别, SAS 软件就具有这种属性. 但 R 语言对大小写非常敏感, 所以在使用 R 语言时要非常注意字母的大小写录入准确.

四、命令的修改和再次调用

在编辑一条命令的过程中, 如果发现前面编辑的内容有误, 可以使用键盘上的向左键移动光标, 回到错误处进行修改. 修改之后, 再使用向右键回到命令的最后编辑位继续编辑.

如果想浏览或者重新调用之前已经运行过的一条或多条命令, 可以使用键盘上的向上键或向下键进行选择.

五、常用运算符和函数表达式

每种软件都有自己特定的运算符和函数表达式, R 语言常用的运算符和函数表达式如表 1—1 和表 1—2 所示.

表 1—1 R 语言常用运算符

运算符	描述	运算符	描述
+	加	>	大于
-	减	>=	大于等于
*	乘	==	严格等于
/	除	!=	不等于
^	幂次运算	!x	非 x
<	小于	x y	x 或 y
<=	小于等于	x&y	x 和 y

表 1—2 常用函数

函数	描述	函数	描述
abs(x)	绝对值	sum(x)	求和
sqrt(x)	平方根	range(x)	值域
sin(x)	正弦	diff(x)	差分运算
cos(x)	余弦	diff(x,n)	n 步差分
tan(x)	正切	lag(x)	延迟运算
log(x)	自然对数	min(x)	最小值
log(x,base=n)	以 n 为底求对数	max(x)	最大值
exp(x)	指数	set.seed(n)	产生以 n 为基数的随机数种子
mean(x)	均值	runif(n)	产生 n 个 (0,1) 区间的均匀分布随机数

续前表

函数	描述	函数	描述
median(x)	中位数	rnorm(n)	产生 n 个标准正态分布随机数
var(x)	方差	pnorm(x)	计算标准正态分布的分位点
sd(x)	标准差	qnorm(q)	标准正态分布分布函数为 q 的分位数
quantile(x,probs)	分位数		

1.4.4 生成时间序列数据

一、直接录入

如果数据不多,我们可以采用直接录入的方式生成时间序列数据.

1. 行输入

我们以表 1—3 中的数据为例,以行输入的方式读入 R 语言,形成一个时间序列数据对象.

表 1—3

时间	价格	时间	价格
2005 年 1 月	101	2005 年 4 月	35
2005 年 2 月	82	2005 年 5 月	31
2005 年 3 月	66	2005 年 6 月	7

```
> price<-c(101,82,66,35,31,7)
> price<-ts(price, start=c(2005,1), frequency=12)
> price
      Jan  Feb  Mar  Apr  May  Jun
2005 101   82   66   35   31   7
```

语句说明:

第一句命令是调用 c 函数,以行输入的方式将 6 个时序数据依次赋值给了 price 这个变量,price 以向量的方式存储了这 6 个数据.

第二句命令是指定 price 为时间序列变量,而且序列名仍然取为 price. start 选项指定序列的起始读入时间,本例中指定序列从 2005 年 1 月开始读入. frequency 选项指定序列每年读入的数据频率,本例中指定读入频率为每年 12 个,也就是说该序列为月度数据.以此类推,读入半年度数据,即 frequency=2;读入季度数据,即 frequency=4;读入星期数据,即 frequency=52;读入日数据,即 frequency=365.

第三句命令是查看时间序列 price 的情况.这条命令输入之后,在对话界面会立刻出现时间序列 price 的具体信息.其中行标识是年份,列标识是月份,对应的 6 个月度的序列值非常清晰地呈现出来.

2. 列输入

如果我们以列输入的方式读入一列时间序列数据, 相关命令和输出结果如下:

```
> price<-scan()
1: 101
2: 82
3: 66
4: 35
5: 31
6: 7
7:
Read 6 items
> price<-ts(price, start=c(2005,1), frequency=12)
> price
```

	Jan	Feb	Mar	Apr	May	Jun
2005	101	82	66	35	31	7

语句说明:

第一句命令是调用 scan 函数, 以列输入的方式将 6 个时序数据以列的方式读入. 如果是手工编辑, 每输入一个数据回车一次. 如果是从其他数据文件通过复制加粘贴的方式录入列数据, 则直接粘贴即可. 最后一个数据输完, 在空白数据行回车, 结束赋值过程.

二、外部数据文件转换

当序列数据量很大时, 采用直接输入或者粘贴的方法都不太方便, 最好是能从其他数据文件直接转换过来.

在 R 中, 最简单的数据转换是带分隔符的文本文件 (txt 格式或者 csv 格式) 转换. 其他各种数据文件比如 Excel 数据 (含 xls 后缀和xlsx 后缀)、XML 数据、SPSS 数据、SAS 数据等, R 也提供了不同的程序包处理它们的格式转换问题. 但是, 本着少而精的原则, 我们只介绍带分隔符的文本文件的转换方法. 因为理论上其他格式的数据在另存时都可以很方便地转换成带分隔符的文本格式, 所以掌握了这种数据转换方法, 基本上其他格式的数据都能以文本格式为桥梁, 非常便捷地转换为 R 语言数据文件. 这样既不需要下载多个数据转换功能包, 也不需要记很多命令格式.

在 R 中, 将多个变量构成的数据文件称为数据框 (date.frame), 所以其他格式的多变量数据文件转入 R 后都称为数据框文件.

【例 1—1】 1884—1939 年英格兰和威尔士小麦平均亩产序列存储在电脑中的路径为 E:/R/data/file1.csv, 现将该数据文件读入 R 并存入数据框 x. 读入命令和结果如下:

```
> x<-read.table("E:/R/data/file1.csv", sep=",", header=T)
> x
```

	year	yield
1	1884	15.2
2	1885	16.9
3	1886	15.3
4	1887	14.9
5	1888	15.7
6	1889	15.1
7	1890	16.7
8	1891	16.3
9	1892	16.5
10	1893	13.3
11	1894	16.5
12	1895	15.0
13	1896	15.9
14	1897	15.5
15	1898	16.9
16	1899	16.4
17	1900	14.9
18	1901	14.5
19	1902	16.6
20	1903	15.1
21	1904	14.6
22	1905	16.0
23	1906	16.8
24	1907	16.8
25	1908	15.5
26	1909	17.3
27	1910	15.5
28	1911	15.5
29	1912	14.2
30	1913	15.8
31	1914	15.7
32	1915	14.1
33	1916	14.8
34	1917	14.4
35	1918	15.6
36	1919	13.9
37	1920	14.7
38	1921	14.3
39	1922	14.0
40	1923	14.5
41	1924	15.4
42	1925	15.3

43	1926	16.0
44	1927	16.4
45	1928	17.2
46	1929	17.8
47	1930	14.4
48	1931	15.0
49	1932	16.0
50	1933	16.8
51	1934	16.9
52	1935	16.6
53	1936	16.2
54	1937	14.0
55	1938	18.1
56	1939	17.5

语句说明:

sep 指定制表分隔符. 如果是 txt 格式, sep="\t", 即文件以制表符分隔. 如果是 csv 格式, sep="," , 即以逗号分隔.

header 指定第一行是否包含变量名. header=TRUE (或简写为 "T") 表示要转换的文本文件第一行包含变量名, 这是系统指定的默认值. header=FALSE (或简写为 F) 表示该文件第一行不包含变量名.

原文件的两个变量 year 和 yield 在数据框中仍然命名为 year 和 yield. 假设我们想查看小麦产量序列的详情, 如果只输入 yield, 将会得到报错结果. 这是因为在 R 中, 数据框中的变量采用 "数据框 \$ 变量名" 的方式记录, 单独考察这两个变量的特征时, 需要采用 x\$year 和 x\$yield.

```
> yield
错误: 找不到对象 'yield'
> x$yield
[1] 15.2 16.9 15.3 14.9 15.7 15.1 16.7 16.3 16.5 13.3 16.5 15.0 15.9 15.5
[15] 16.9 16.4 14.9 14.5 16.6 15.1 14.6 16.0 16.8 16.8 15.5 17.3 15.5 15.5
[29] 14.2 15.8 15.7 14.1 14.8 14.4 15.6 13.9 14.7 14.3 14.0 14.5 15.4 15.3
[43] 16.0 16.4 17.2 17.8 14.4 15.0 16.0 16.8 16.9 16.6 16.2 14.0 18.1 17.5
```

1.4.5 时间序列数据的处理

一、变换

在时间序列分析中, 我们得到的是观察值序列, 但是需要分析的可能是这个观察值序列的某个函数变换, 比如对数变换序列. 我们可以通过简单的赋值命令实现这些变换.

```
> y<-log(x$yield)
> y
 [1] 2.721295 2.827314 2.727853 2.701361 2.753661 2.714695 2.815409 2.791165
 [9] 2.803360 2.587764 2.803360 2.708050 2.766319 2.740840 2.827314 2.797281
[17] 2.701361 2.674149 2.809403 2.714695 2.681022 2.772589 2.821379 2.821379
[25] 2.740840 2.850707 2.740840 2.740840 2.653242 2.760010 2.753661 2.646175
[33] 2.694627 2.667228 2.747271 2.631889 2.687847 2.660260 2.639057 2.674149
[41] 2.734368 2.727853 2.772589 2.797281 2.844909 2.879198 2.667228 2.708050
[49] 2.772589 2.821379 2.827314 2.809403 2.785011 2.639057 2.895912 2.862201
```

二、子集

使用 subset 函数可以得到原数据文件的子集. 采用下面的命令我们就能得到一个子集 z, 原数据文件中的变量 year 和 1925 年之前的数据将被剔除.

```
> z<-subset(x,year>1925,select=yield)
> z
  yield
43 16.0
44 16.4
45 17.2
46 17.8
47 14.4
48 15.0
49 16.0
50 16.8
51 16.9
52 16.6
53 16.2
54 14.0
55 18.1
56 17.5
```

三、缺失值插值

在时间序列分析中, 如果序列缺失一期或多期观察值, 那么该序列称为缺失值序列. 在 R 中, 缺失值用符号 “NA” 表示, 意思是 “not available”, 即无法获得的数据.

缺失值的存在是时间序列分析的大难题, 即使只有一个缺失值存在, R 也会罢工, 并提示你因为缺失值的存在, 相应的分析无法进行. 所以当序列有缺失值时, 我们通常会通过插值的方法对缺失值进行插补.

R 提供了非常丰富的插值方法, 不同的方法由不同的程序包提供. 在时间序列分析中, 最常用的插值法是简单的线性插值法和样条插值法, 它们都可以由 zoo 程序包

提供. 所以使用插值方法前, 需要先下载安装 zoo 程序包, 并用 library 函数加载.

```
> a<-c(1:7)      #将 1~7 赋值给变量 a
> a[4]<-NA        #定义序列 a 第四个序列值为缺失值
> a
[1] 1 2 3 NA 5 6 7
• 线性插值
> y1<-na.approx(a)
> y1
[1] 1 2 3 4 5 6 7
• 样条插值
> y2<- na.spline(a)
> y2
[1] 1 2 3 4 5 6 7
```

1.4.6 时间序列数据导出

对时间序列进行适当的处理和分析之后, 如果我们想把部分数据导出, 可以使用适当的函数完成这项任务. R 提供了很多种函数来完成数据导出任务, 我们在此介绍其中一种方法: 使用 write.table 函数来导出数据.

【例 1—1 续】 对 yield 序列进行对数变换, 并将对数序列和原序列值导出, 存入数据文件 E:/R/data/yield.csv.

```
> ln_yield<-log(x$yield)
> x_new<-data.frame(x,ln_yield)
> write.table(x_new,file="E:/R/data/yield.csv",sep=" ",row.names=F)
```

语句说明:

第一句命令是对数据框中的序列 yield 进行对数运算, 赋值给变量 ln_yield.

第二句命令是将原数据框和新变量 ln_yield 合并成一个新的数据框 x_new.

第三句命令是将数据框 x_new 以 csv 格式导出, 存入指定路径. 其中, row.names=F 是确定数据以列变量的方式存储, 没有行变量名.

1.5 习 题

1. 什么是时间序列? 请收集几个生活中的观察值序列.
2. 时域方法的特点是什么?
3. 时域方法的发展轨迹是怎样的?
4. 在附录 1 中选择几个感兴趣的序列, 创建 R 数据框.

拿到一个观察值序列之后,首先要对它的平稳性和纯随机性进行检验,这两个重要的检验称为序列的预处理.根据检验的结果可以将序列分为不同的类型,对不同类型的序列我们会采用不同的分析方法.

2.1 平稳时间序列

2.1.1 特征统计量

平稳性是某些时间序列具有的一种统计特征.要描述清楚这个特征,我们必须借助以下统计工具.

一、概率分布

数理统计的基础知识告诉我们分布函数或密度函数能够完整地描述一个随机变量的统计特征.同样,一个随机变量族 $\{X_t\}$ 的统计特性也完全由它们的联合分布函数或联合密度函数决定.

对于时间序列 $\{X_t, t \in T\}$, 这样来定义它的概率分布:

任取正整数 m , 任取 $t_1, t_2, \dots, t_m \in T$, 则 m 维随机向量 $(X_{t_1}, X_{t_2}, \dots, X_{t_m})'$ 的联合概率分布记为 $F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m)$, 由这些有限维分布函数构成的全体

$$\{F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m), \forall m \in \text{正整数}, \forall t_1, t_2, \dots, t_m \in T\}$$

就称为序列 $\{X_t\}$ 的概率分布族.

概率分布族是极其重要的统计特征描述工具,因为序列的所有统计性质理论上都可以通过概率分布推导出来,但是概率分布族的重要性也就停留在这样的理论意义上.在实际应用中,要得到序列的联合概率分布几乎是不可能的,而且联合概率分布通常涉及非常复杂的数学运算,这些原因使我们很少直接使用联合概率分布进行时间序列分析.

二、特征统计量

一种更简单、更实用的描述时间序列统计特征的方法是研究该序列的低阶矩,特别是均值、方差、自协方差和自相关系数,它们也称为特征统计量。

尽管这些特征统计量不能描述随机序列全部的统计性质,但由于它们概率意义明显,易于计算,而且往往能代表随机序列的主要概率特征,所以我们对时间序列进行分析,主要就是通过分析这些特征量的统计特性,推断出随机序列的性质。

1. 均值

对时间序列 $\{X_t, t \in T\}$ 而言,任意时刻的序列值 X_t 都是一个随机变量,都有它自己的概率分布,不妨记 X_t 的分布函数为 $F_t(x)$ 。只要满足条件

$$\int_{-\infty}^{\infty} x dF_t(x) < \infty$$

就一定存在某个常数 μ_t , 使得随机变量 X_t 总是围绕在常数值 μ_t 附近做随机波动。我们称 μ_t 为序列 $\{X_t\}$ 在 t 时刻的均值函数。

$$\mu_t = EX_t = \int_{-\infty}^{\infty} x dF_t(x)$$

当 t 取遍所有的观察时刻时,就得到一个均值函数序列 $\{\mu_t, t \in T\}$ 。它反映的是时间序列 $\{X_t, t \in T\}$ 每时每刻的平均水平。

2. 方差

当 $\int_{-\infty}^{\infty} x^2 dF_t(x) < \infty$ 时,可以定义时间序列的方差函数用以描述序列值围绕其均值做随机波动时的平均波动程度。

$$\sigma_t^2 = DX_t = E(X_t - \mu_t)^2 = \int_{-\infty}^{\infty} (x - \mu_t)^2 dF_t(x)$$

同样,当 t 取遍所有的观察时刻时,我们得到一个方差函数序列 $\{\sigma_t^2, t \in T\}$ 。

3. 自协方差函数和自相关系数

类似于协方差函数和相关系数的定义,在时间序列分析中我们定义自协方差函数 (autocovariance function) 和自相关系数 (autocorrelation coefficient) 的概念。

对于时间序列 $\{X_t, t \in T\}$, 任取 $t, s \in T$, 定义 $\gamma(t, s)$ 为序列 $\{X_t\}$ 的自协方差函数:

$$\gamma(t, s) = E[(X_t - \mu_t)(X_s - \mu_s)]$$

定义 $\rho(t, s)$ 为时间序列 $\{X_t\}$ 的自相关系数,简记为 ACF。

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{DX_t \cdot DX_s}}$$

之所以称它们为自协方差函数和自相关系数,是因为通常的协方差函数和相关系数度量的是两个不同事件彼此之间的相互影响程度,而自协方差函数和自相关系数度量的是同一事件在两个不同时期之间的相关程度,形象地讲就是度量自己过去的行为对自己现在的影响。

2.1.2 平稳时间序列的定义

平稳时间序列有两种定义, 根据限制条件的严格程度, 分为严平稳时间序列和宽平稳时间序列.

一、严平稳

所谓严平稳 (strictly stationary), 就是一种条件比较苛刻的平稳性定义, 它认为只有当序列所有的统计性质都不会随着时间的推移而发生变化时, 该序列才能被认为是平稳的. 而我们知道, 随机变量族的统计性质完全由它们的联合概率分布族决定, 所以严平稳时间序列的定义如下:

定义 2.1 设 $\{X_t\}$ 为一时间序列, 对任意正整数 m , 任取 $t_1, t_2, \dots, t_m \in T$, 对任意整数 τ , 有

$$F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m) = F_{t_1+\tau, t_2+\tau, \dots, t_m+\tau}(x_1, x_2, \dots, x_m)$$

则称时间序列 $\{X_t\}$ 为严平稳时间序列.

前面说过, 在实践中要获得随机序列的联合分布是非常困难的, 而且即使知道随机序列的联合分布, 计算和应用也非常不便. 所以严平稳时间序列通常只具有理论意义, 在实践中用得更多的是条件比较宽松的宽平稳时间序列.

二、宽平稳

宽平稳 (weak stationary) 是使用序列的特征统计量来定义的一种平稳性. 它认为序列的统计性质主要由它的低阶矩决定, 所以只要保证序列低阶矩 (二阶) 平稳, 就能保证序列的主要性质近似稳定.

定义 2.2 如果 $\{X_t\}$ 满足如下三个条件:

- (1) 任取 $t \in T$, 有 $EX_t^2 < \infty$;
- (2) 任取 $t \in T$, 有 $EX_t = \mu$, μ 为常数;
- (3) 任取 $t, s, k \in T$, 且 $k + s - t \in T$, 有 $\gamma(t, s) = \gamma(k, k + s - t)$,

则称 $\{X_t\}$ 为宽平稳时间序列. 宽平稳也称为弱平稳或二阶平稳 (second-order stationary).

显然, 严平稳比宽平稳的条件严格. 严平稳是对序列联合分布的要求, 以保证序列所有的统计特征都相同; 而宽平稳只要求序列二阶平稳, 对于高于二阶的矩没有任何要求. 所以通常情况下, 严平稳序列也满足宽平稳条件, 而宽平稳序列不能反推严平稳成立.

但这不是绝对的, 两种情况都有特例.

比如服从柯西分布的严平稳序列就不是宽平稳序列, 因为它不存在一、二阶矩, 所以无法验证它二阶平稳. 严格地讲, 只有存在二阶矩的严平稳序列才能保证它一定也

是宽平稳序列.

宽平稳一般推不出严平稳, 但当序列服从多元正态分布时, 二阶平稳可以推出严平稳.

定义 2.3 时间序列 $\{X_t\}$ 称为正态时间序列, 如果任取正整数 n , 任取 $t_1, t_2, \dots, t_n \in T$, 相对应的有限维随机变量 $X_1, X_2, \dots, X_n$ 服从 n 维正态分布, 密度函数为:

$$f_{t_1, t_2, \dots, t_n}(\widetilde{X}_n) = (2\pi)^{-\frac{n}{2}} |\Gamma_n|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\widetilde{X}_n - \widetilde{\mu}_n)' \Gamma_n^{-1} (\widetilde{X}_n - \widetilde{\mu}_n) \right]$$

式中, $\widetilde{X}_n = (X_1, X_2, \dots, X_n)'$; $\widetilde{\mu}_n = (EX_1, EX_2, \dots, EX_n)'$; Γ_n 为协方差阵.

$$\Gamma_n = \begin{pmatrix} \gamma(t_1, t_1) & \gamma(t_1, t_2) & \cdots & \gamma(t_1, t_n) \\ \gamma(t_2, t_1) & \gamma(t_2, t_2) & \cdots & \gamma(t_2, t_n) \\ \vdots & \vdots & & \vdots \\ \gamma(t_n, t_1) & \gamma(t_n, t_2) & \cdots & \gamma(t_n, t_n) \end{pmatrix}$$

从正态随机序列的密度函数可以看出, 它的 n 维分布仅由均值向量和协方差阵决定, 换言之, 对于正态随机序列而言, 只要二阶矩平稳, 就等于分布平稳了. 所以宽平稳正态时间序列一定是严平稳时间序列. 对于非正态过程, 就没有这个性质了.

在实际应用中, 研究最多的是宽平稳随机序列, 以后见到平稳随机序列, 如果不加特别注明, 指的都是宽平稳随机序列. 如果序列不满足平稳条件, 就称为非平稳序列.

2.1.3 平稳时间序列的统计性质

根据平稳时间序列的定义, 可以推断出它一定具有如下两个重要的统计性质.

一、常数均值

$$EX_t = \mu, \quad \forall t \in T$$

二、自协方差函数和自相关系数只依赖于时间的平移长度而与时间的起止点无关

$$\gamma(t, s) = \gamma(k, k + s - t), \quad \forall t, s, k \in T$$

根据这个性质, 可以将自协方差函数由二维函数 $\gamma(t, s)$ 简化为一维函数 $\gamma(s - t)$:

$$\gamma(s - t) \doteq \gamma(t, s), \quad \forall t, s \in T$$

由此引出延迟 k 自协方差函数的概念.

定义 2.4 对于平稳时间序列 $\{X_t, t \in T\}$, 任取 $t (t + k \in T)$, 定义 $\gamma(k)$ 为时间序列 $\{X_t\}$ 的延迟 k 自协方差函数:

$$\gamma(k) = \gamma(t, t + k)$$

根据平稳序列的这个性质, 容易推断出平稳随机序列一定具有常数方差:

$$DX_t = \gamma(t, t) = \gamma(0), \quad \forall t \in T$$

由延迟 k 自协方差函数的概念可以等价得到延迟 k 自相关系数的概念:

$$\rho_k = \frac{\gamma(t, t+k)}{\sqrt{DX_t \cdot DX_{t+k}}} = \frac{\gamma(k)}{\gamma(0)}$$

容易验证和相关系数一样, 自相关系数具有如下三个性质:

(1) 规范性.

$$\rho_0 = 1 \text{ 且 } |\rho_k| \leq 1, \quad \forall k$$

(2) 对称性.

$$\rho_k = \rho_{-k}$$

(3) 非负定性.

对任意正整数 m , 相关阵 Γ_m 为对称非负定阵.

$$\Gamma_m = \begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{m-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{m-2} \\ \vdots & \vdots & & \vdots \\ \rho_{m-1} & \rho_{m-2} & \cdots & \rho_0 \end{pmatrix}$$

值得注意的是, ρ_k 除了具有上述三个性质外, 还具有一个特别的性质: 对应模型的非唯一性.

一个平稳时间序列一定唯一决定了它的自相关函数, 但一个自相关函数未必唯一对应一个平稳时间序列. 我们在后面的章节中将证明这一点. 这个性质就给我们根据样本的自相关系数的特点来确定模型增加了一定的难度.

2.1.4 平稳时间序列的意义

时间序列分析方法作为数理统计学的一个分支, 遵循数理统计学的基本原理, 都是利用样本信息来推测总体信息.

传统的统计分析通常都拥有如下数据结构 (见表 2—1).

表 2—1

随机变量 样本	X_1	$\cdots$	X_m
1	x_{11}	$\cdots$	x_{m1}
2	x_{12}	$\cdots$	x_{m2}
$\vdots$	$\vdots$		$\vdots$
n	x_{1n}	$\cdots$	x_{mn}

根据数理统计学常识, 显然要分析的随机变量越少越好 (m 越小越好), 而每个变量获得的样本信息越多越好 (n 越大越好). 因为随机变量越少, 分析的过程就会越简单, 而样本容量越大, 分析的结果就会越可靠.

但是时间序列分析的数据结构有它的特殊性. 对随机序列 $\{\cdots, X_1, X_2, \cdots, X_t, \cdots\}$ 而言, 它在任意时刻 t 的序列值 X_t 都是一个随机变量, 而且由于时间的不可重复性, 该变量在任意一个时刻只能获得唯一的样本观察值. 因而时间序列分析的数据结构如下 (见表 2-2).

表 2-2

随机变量 \ 样本	$\cdots$	X_1	$\cdots$	X_t	$\cdots$
1	$\cdots$	x_1	$\cdots$	x_t	$\cdots$

由于样本信息太少, 如果没有其他的辅助信息, 通常这种数据结构是没有办法进行分析的. 而序列平稳性概念的提出可以有效地解决这个问题.

在平稳序列场合, 序列的均值等于常数意味着原本含有可列多个随机变量的均值序列

$$\{\mu_t, t \in T\}$$

变成了一个常数序列

$$\{\mu, t \in T\}$$

原本每个随机变量的均值 μ_t ($t \in T$) 只能依靠唯一的一个样本观察值 x_t 去估计

$$\hat{\mu} = x_t$$

现在由于 $\mu_t = \mu$ ($\forall t \in T$), 于是每一个样本观察值 x_t ($\forall t \in T$) 都变成了常数均值 μ 的样本观察值

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

这极大地减少了随机变量的个数, 并增加了待估参数的样本容量. 换句话说, 这大大降低了时序分析的难度, 同时也提高了对均值函数的估计精度.

同理, 根据平稳序列二阶矩平稳的性质, 可以得到基于全体观察样本计算出的延迟 k 自协方差函数的估计值

$$\hat{\gamma}(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{n - k}, \quad \forall 0 < k < n$$

并进一步推导出总体方差的估计值

$$\hat{\gamma}(0) = \frac{\sum_{t=1}^n (x_t - \bar{x})^2}{n - 1}$$

和延迟 k 自相关系数的估计值

$$\hat{\rho}_k = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad \forall 0 < k < n$$

当延迟阶数 k 远远小于样本容量 n 时,

$$\hat{\rho}_k \approx \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad \forall 0 < k < n$$

2.2 时序图与自相关图

2.2.1 时序图

一、默认格式输出

所谓时序图, 就是一个二维平面坐标图, 通常横轴表示时间, 纵轴表示序列取值. 时序图可以帮助我们直观地掌握时间序列的一些基本分布特征. 所以时序分析的第一步通常是绘制时序图.

R 具有非常便捷与强大的绘图功能. 利用 R 绘制时序图主要是调用 `plot` 函数. 如果对时序图的点线格式、坐标轴的比例、在图中添加文本等没有特殊要求, 则只需要使用最基本的 `plot` 命令就可以绘制出美观大方的时序图.

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 我们可以输入如下指令, 得到这 7 年的时序图 (见图 2—1).

```
> yield<-c(15.2, 16.9, 15.3, 14.9, 15.7, 15.1, 16.7)
> yield<-ts(yield, start=1884)
> plot(yield)
```

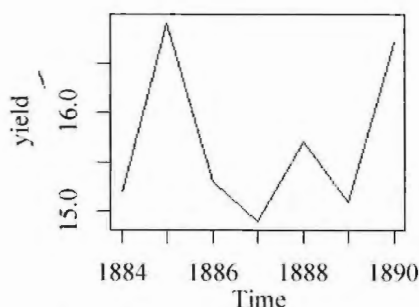


图 2—1 1884—1890 年英格兰和威尔士地区小麦平均亩产量时序图

语句说明:

- (1) 第一句是用行输入的方式将 7 个序列值赋值给向量 `yield`.
- (2) 第二句是指定 `yield` 为时序变量, 观察值起始时间为 1884 年, 数据频率为年度数据.
- (3) 第三句是绘制 `yield` 的时序图, 按 R 语言默认格式输出.

二、自定义图形参数

有时我们对图形输出有更高标准的个性化的要求, 为了满足用户的需求, R 语言提供了丰富的图形参数. 下面我们介绍常用绘图命令具体参数的意义.

1. 点线结构参数

在 `plot` 函数中, 使用参数 `type` 来控制点线输出结构. 具体的参数值及含义如表 2—3 所示.

表 2—3 参数 `type` 的取值及含义

参数取值	描述	参数取值	描述
<code>type="p"</code>	点	<code>type="o"</code>	线穿过点
<code>type="l"</code>	线	<code>type="h"</code>	悬垂线
<code>type="b"</code>	点连线	<code>type="s"</code>	阶梯线

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 考察 `type` 选项对时序图的影响.

• 散点图

> `plot(yield,type="p")` #结果如图 2—2 所示

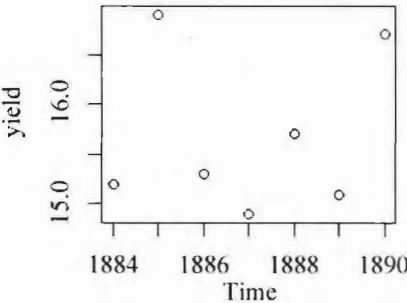


图 2—2 散点图

• 点线图

> `plot(yield,type="o")` #结果如图 2—3 所示

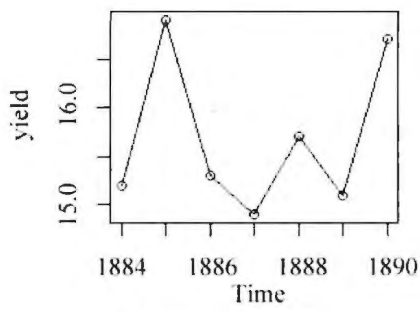


图 2—3 点线图

2. 符号参数

plot 函数中, 用 pch 选项设置观察点的符号. pch 一共有 25 个参数值, 它们对应的符号如图 2—4 所示.

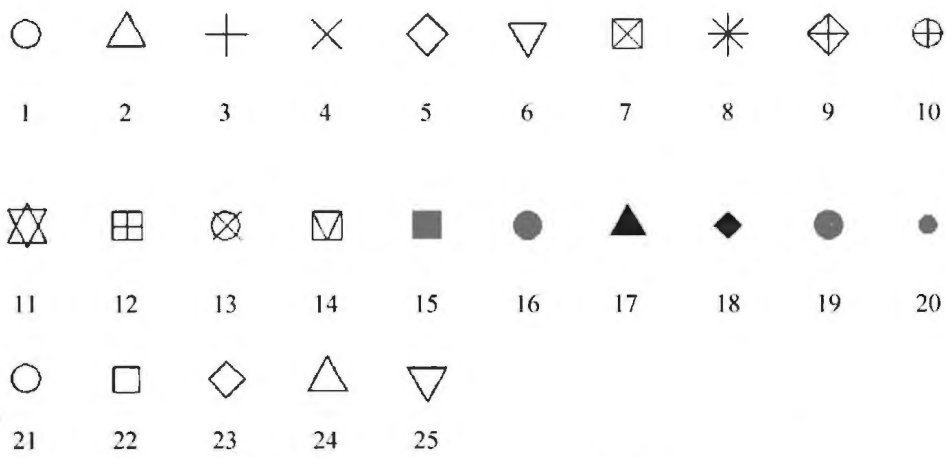


图 2—4 pch 选项参数值对应的绘图符号

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 考察 pch 选项对时序图的影响.

```
> plot(yield,type="o", pch=17) #结果如图 2—5 所示
```

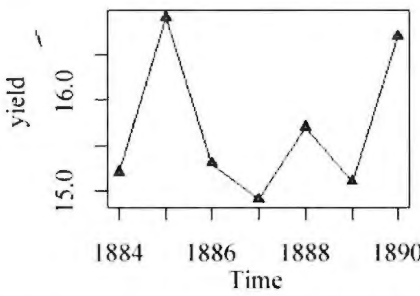


图 2—5 符号参数为实心三角的时序图

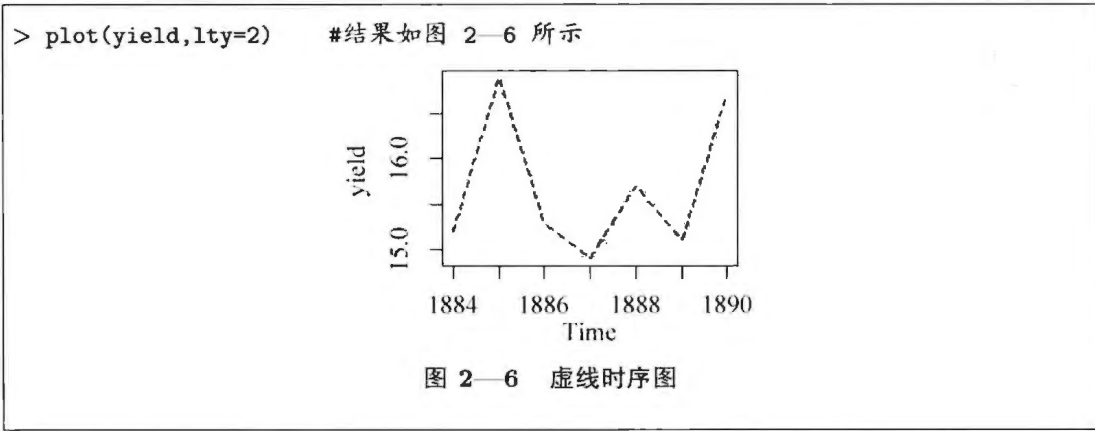
3. 连线类型参数

plot 函数中, 用 lty 选项设置线的类型. 具体的参数值与线型规定如表 2—4 所示.

表 2—4 参数 lty 的取值及含义

参数取值	描述
lty=1	实线
lty=2	虚线
lty=3	点线
lty=4	点 + 短虚线
lty=5	长虚线
lty=6	点 + 长虚线

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 考察 lty 选项对时序图的影响.



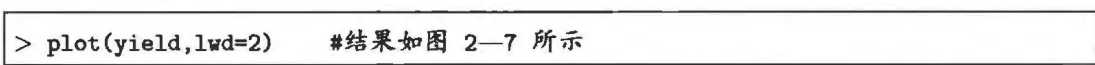
4. 线的宽度参数

plot 函数中, 用 lwd 选项设置线的宽度. 具体的参数值与线型规定如表 2—5 所示.

表 2—5 参数 lwd 的取值及含义

参数取值	描述
lwd=1	默认宽度
lwd=k	默认宽度的 k 倍
lwd= $\frac{1}{k}$	默认宽度的 $\frac{1}{k}$ 倍

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 考察 lwd 选项对时序图的影响.



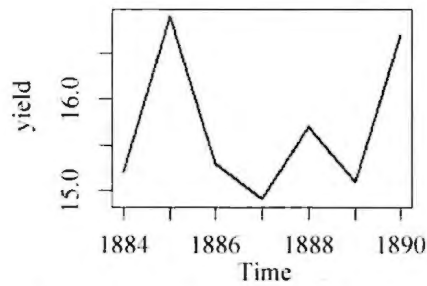


图 2—7 标准线宽两倍时序图

5. 颜色参数

plot 函数中, 用 col 选项设置点线颜色. 等号后面可以是具体的颜色, 例如 col="red", 也可以是某个数值, 例如 col=2. 表 2—6 给出了常用的四种颜色的对应表达, 而实际上 R 语言一共可以提供 657 种颜色.

表 2—6 参数 col 的取值及含义

参数取值	等价表达	颜色
col=1	col="black"	黑色
col=2	col="red"	红色
col=3	col="green"	绿色
col=4	col="blue"	蓝色

6. 添加文本

plot 函数中, 用 main 选项添加标题文本, 用 sub 选项添加副标题文本, 用 xlab 选项指定横坐标的名称, 用 ylab 选项指定纵坐标的名称.

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例添加标题, 并将横坐标显示为“年份”, 纵坐标显示为“亩产量”, 命令和显示结果如下.

```
> plot(yield,main="1884—1890 年英格兰和威尔士地区小麦平均亩产量", xlab="年份", ylab=" 亩产量") #结果如图 2—8 所示
```

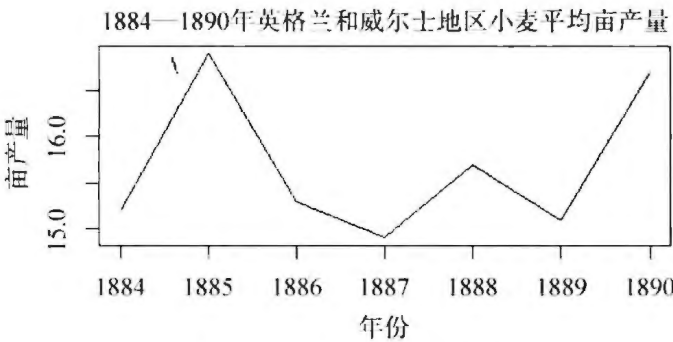


图 2—8 添加文本的时序图

7. 指定坐标轴范围

plot 函数中, 用 xlim 选项指定横坐标的范围, 用 ylim 选项指定纵坐标的范围.

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 我们可以通过限制坐标的范围控制时序图输出内容. 相关命令和显示结果如下.

- 指定输出横轴范围

```
> plot(yield,xlim=c(1886,1890)) #结果如图 2—9 所示
```

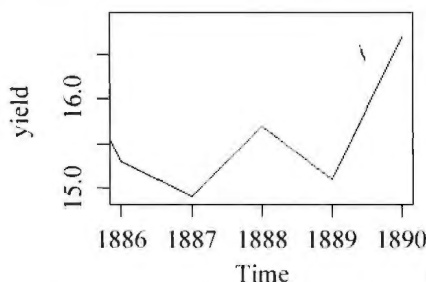


图 2—9 指定输出横轴范围

- 指定输出纵轴范围

```
> plot(yield,ylim=c(15,16)) #结果如图 2—10 所示
```

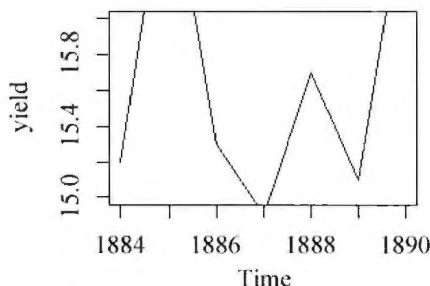


图 2—10 指定输出纵轴范围

8. 添加参照线

在绘图时, 可以使用 abline 函数为图形添加参照线. 参照线可以是垂线, 也可以是水平线, 还可以是线性回归线.

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 在时序图基础上添加参照线, 添加垂直参照线和水平参照线的相关命令和显示结果如下, 添加回归参照线的命令我们将在第 4 章介绍.

- 添加一条垂线

```
> plot(yield)
> abline(v=1887, lty=2) #结果如图 2—11 所示
```

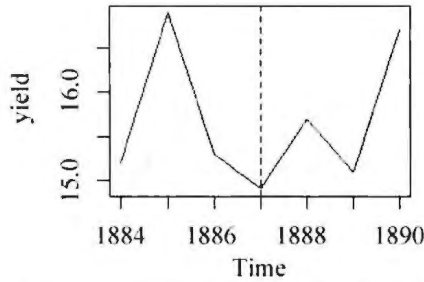


图 2—11 时序图添加一条垂直参照线

- 添加多条垂直参照线

```
> plot(yield)
> abline(v=c(1885,1889), lty=2)    #结果如图 2—12 所示
```

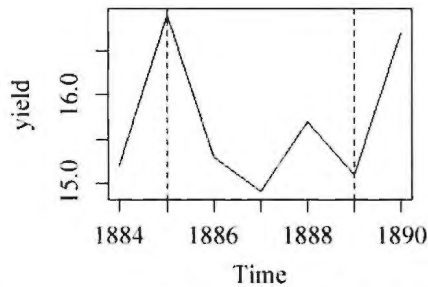


图 2—12 时序图添加多条垂直参照线

- 添加水平线

```
> plot(yield)
> abline(h=c(15.5,16.5), lty=2)    #结果如图 2—13 所示
```

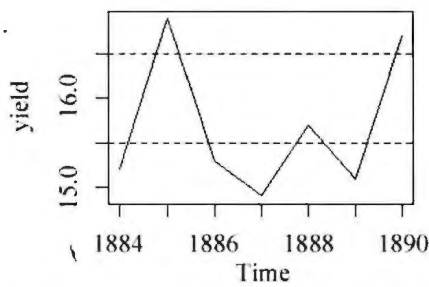


图 2—13 时序图添加水平参照线

2.2.2 绘制序列自相关图

自相关图是一个平面二维坐标悬垂线图. 横坐标表示延迟时期数, 纵坐标表示自相关系数. 悬垂线表示自相关系数的大小.

在 R 语言中使用 `acf` 函数绘制序列自相关图, 该函数的命令格式为:

```
acf(x, lag=)
```

式中:

-x: 变量名;

-lag: 延迟阶数, 若用户不特殊指定的话, 系统会根据序列长度自动指定延迟阶数.

以 1884—1890 年英格兰和威尔士地区小麦平均亩产量数据为例, 绘制序列的自相关图, 相关命令和显示结果如下.

```
> acf(yield) #结果如图 2—14 所示
```

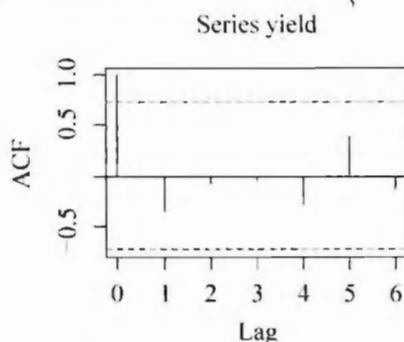


图 2—14 自相关图

说明: 虚线为自相关系数 2 倍标准差位置.

2.3 平稳性的检验

对序列的平稳性有两种检验方法, 一种是根据时序图和自相关图显示的特征做出判断的图检验方法; 一种是构造检验统计量进行假设检验的方法.

图检验方法是一种操作简便、运用广泛的平稳性判别方法, 它的缺点是判别结论带有很强的主观色彩, 所以最好能用统计检验方法加以辅助判断. 目前最常用的平稳性统计检验方法是单位根检验 (unit root test). 由于目前知识的局限性, 本章将主要介绍平稳性的图检验方法, 单位根检验将在第 6 章详细介绍.

2.3.1 时序图检验

根据平稳时间序列均值、方差为常数的性质, 平稳序列的时序图应该显示出该序列始终在一个常数值附近随机波动, 而且波动的范围有界的特点. 如果序列的时序图显示出该序列有明显的趋势性或周期性, 那么它通常不是平稳序列. 根据这个性质, 对于很多非平稳序列, 可以通过查看它的时序图将其识别出来.

【例 2—1】 绘制 1964—1999 年中国纱年产量序列时序图 (数据见表 A1—4).

```
> sha<-read.table("E:/R/data/file4.csv",sep=" ",header=T)
> output<-ts sha$output,start=1964)
> plot(output)    #结果如图 2—15 所示
```

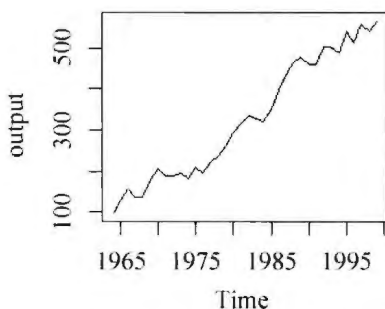


图 2—15 中国纱年产量序列时序图

时序图图 2—15 给我们提供的信息非常明确, 中国纱年产量序列有明显的递增趋势, 所以它一定不是平稳序列。

【例 2—2】 绘制 1962 年 1 月至 1975 年 12 月平均每头奶牛月产奶量序列时序图 (数据见表 A1—5)。

```
> a<-read.table("E:/R/data/file5.csv",sep=" ",header=T)
> milk<-ts(a$milk,start=c(1962,1),frequency=12)
> plot(milk)    #结果如图 2—16 所示
```

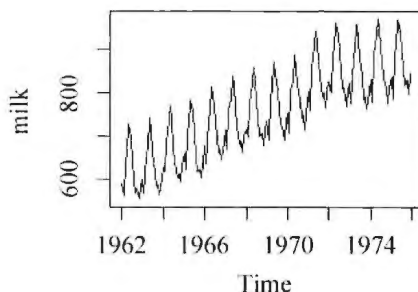


图 2—16 平均每头奶牛月产奶量序列时序图

时序图图 2—16 清晰地显示平均每头奶牛的月产奶量以年为周期呈现出规则的周期性, 除此之外, 还有明显的逐年递增趋势。显然该序列也一定不是平稳序列。

【例 2—3】 绘制 1949—1998 年北京市每年最高气温序列时序图 (数据见表 A1—6)。

```
> b<-read.table("E:/R/data/file6.csv",sep=" ",header=T)
> temp<-ts(b$temp,start=1949)
> plot(temp)    #结果如图 2—17 所示
```

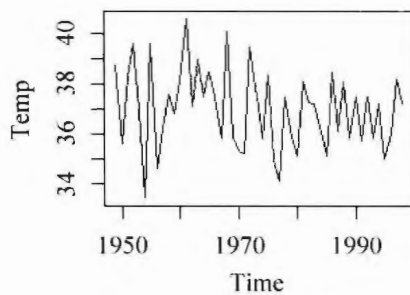


图 2—17 北京市每年最高气温序列时序图

时序图图 2—17 显示北京市每年的最高气温始终围绕在 37°C 附近随机波动, 没有明显趋势或周期, 基本可以视为平稳序列. 为了稳妥起见, 我们还需要利用自相关图进一步辅助识别.

2.3.2 自相关图检验

在后面的章节里我们会证明平稳序列通常具有短期相关性. 该性质用自相关系数来描述就是随着延迟期数 k 的增加, 平稳序列的自相关系数 $\hat{\rho}_k$ 会很快地衰减向零. 反之, 非平稳序列的自相关系数 $\hat{\rho}_k$ 衰减向零的速度通常比较慢, 这就是我们利用自相关图进行平稳性判断的标准.

【例 2—1 续】 绘制 1964—1999 年中国纱年产量序列自相关图.

> acf(output, lag=25) #结果如图 2—18 所示

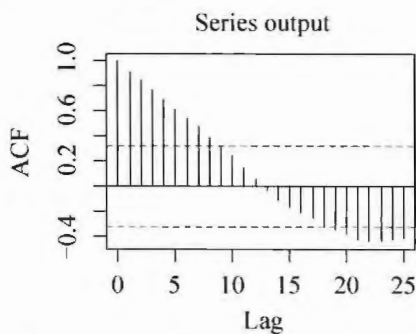


图 2—18 中国纱产量序列自相关图

从图 2—18 中我们发现序列的自相关系数递减到零的速度相当缓慢, 在很长的延迟时期里, 自相关系数一直为正, 而后又一直为负, 在自相关图上显示出明显的三角对称性, 这是具有单调趋势的非平稳序列的一种典型的自相关图形式. 这和该序列时序图图 2—15 显示的显著的单调递增性是一致的.

【例 2—2 续】 绘制 1962 年 1 月至 1975 年 12 月平均每头奶牛的月产奶量序列自相关图.

> acf(milk) #结果如图 2—19 所示

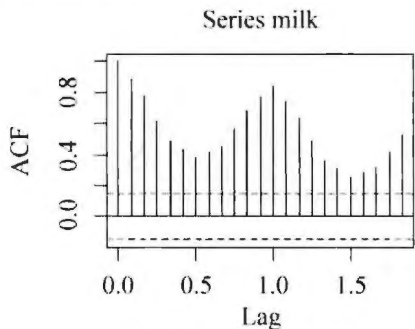


图 2—19 平均每头奶牛月产奶量序列自相关图

自相关图图 2—19 显示序列自相关系数长期位于零轴的一边, 这是具有单调趋势序列的典型特征, 同时自相关图呈现出明显的正弦波动规律, 这是具有周期变化规律的非平稳序列的典型特征. 自相关图显示出来的这两个性质和该序列时序图图 2—16 显示出的带长期递增趋势的周期性质是非常吻合的.

【例 2—3 续 (1)】 绘制 1949—1998 年北京市每年最高气温序列自相关图.

> acf(temp) #结果如图 2—20 所示

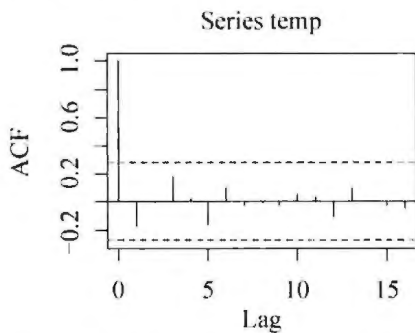


图 2—20 北京市每年最高气温序列的自相关图

自相关图图 2—20 显示该序列的自相关系数一直都比较小, 始终控制在 2 倍的标准差范围以内, 可以认为该序列自始至终都在零轴附近波动, 这是随机性非常强的平稳时间序列通常具有的自相关图特征.

2.4 纯随机性检验

拿到一个观察值序列之后, 首先是判断它的平稳性. 通过平稳性检验, 序列可以分为平稳序列和非平稳序列两大类.

对于非平稳序列, 由于它不具有二阶矩平稳的性质, 所以对它的统计分析要费一些周折, 通常要通过进一步的检验、变换或处理, 才能确定适当的拟合模型.

如果序列平稳, 情况就简单多了, 我们有一套非常成熟的平稳序列建模方法. 但是, 并不是所有的平稳序列都值得建模. 只有那些序列值之间具有密切的相关关系, 历史数据对未来的发展有一定影响的序列, 才值得我们花时间去挖掘历史数据中的有效信息, 用来预测序列未来的发展.

如果序列值彼此之间没有任何相关性, 那就意味着该序列是一个没有记忆的序列, 过去的行为对将来的发展没有丝毫影响, 这种序列称为纯随机序列. 从统计分析的角度而言, 纯随机序列是没有任何分析价值的序列.

为了确定平稳序列是否值得继续分析下去, 我们需要对平稳序列进行纯随机性检验.

2.4.1 纯随机序列的定义

定义 2.5 如果时间序列 $\{X_t\}$ 满足如下性质:

- (1) 任取 $t \in T$, 有 $EX_t = \mu$;
- (2) 任取 $t, s \in T$, 有

$$\gamma(t, s) = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

称序列 $\{X_t\}$ 为纯随机序列, 也称为白噪声 (white noise) 序列, 简记为 $X_t \sim WN(\mu, \sigma^2)$.

之所以称为白噪声序列, 是因为人们最初发现白光具有这种特性. 容易证明白噪声序列一定是平稳序列, 而且是最简单的平稳序列.

【例 2—4】 随机产生 1 000 个服从标准正态分布的白噪声序列观察值, 并绘制时序图.

R 提供了丰富的随机数发生器, 在时序模拟中, 我们最常用的是产生一批服从正态分布的随机数. 正态分布随机数生成函数是 `rnorm`. `rnorm` 函数的命令格式为:

```
rnorm(n=, mean=, sd=)
```

式中:

-n: 随机数个数;

-mean: 均值, 缺省值默认为 0;

-sd: 标准差, 缺省值默认为 1.

注: (1) `rnorm` 函数也可简写为 `rnorm(n, 均值, 标准差)`.

(2) 如果要产生 n 个服从标准正态分布的随机数, 可以简写为 `rnorm(n)`.

所以本例的命令与输出结果如下.

```
> white_noise<-rnorm(1000)
> white_noise<-ts(white_noise)
> plot(white_noise)    #结果如图 2—21 所示
```

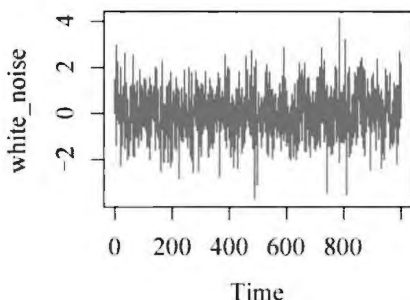


图 2—21 标准正态白噪声序列时序图

2.4.2 白噪声序列的性质

白噪声序列虽然很简单,但它在我們进行时间序列分析时所起的作用却非常大.它的两个重要性质在后面的分析过程中要经常用到.

一、纯随机性

由于白噪声序列具有如下性质:

$$\gamma(k) = 0, \quad \forall k \neq 0$$

这说明白噪声序列的各项之间没有任何相关关系,这种“没有记忆”的序列就是纯随机序列.

纯随机序列各项之间没有任何关联,序列在进行完全无序的随机波动.一旦某个随机事件呈现出纯随机波动的特征,就认为该随机事件没有包含任何值得提取的有用信息,我们就应该终止分析.

如果序列值之间呈现出某种显著的相关关系:

$$\gamma(k) \neq 0, \quad \exists k \neq 0$$

就说明该序列不是纯随机序列,该序列间隔 k 期的序列值之间存在一定程度的相互影响关系,这种相互影响关系统计上称为相关信息.我们分析的目的就是要想方设法把这种相关信息从观察值序列中提取出来.一旦观察值序列中蕴涵的相关信息完全提取出来了,那么剩下的残差序列就应该呈现出纯随机序列的性质.所以纯随机性还是判断相关信息提取是否充分的一个判别标准.

二、方差齐性

所谓方差齐性,就是指序列中每个变量的方差都相等,即

$$DX_t = \gamma(0) = \sigma^2$$

如果序列不满足方差齐性, 就称该序列具有异方差性质.

在时间序列分析中, 方差齐性是一个非常重要的限制条件. 因为根据马尔可夫定理, 只有方差齐性假定成立时, 我们用最小二乘法得到的未知参数估计值才是准确、有效的. 如果假定不成立, 最小二乘估计值就不是方差最小线性无偏估计, 拟合模型的精度会受到很大影响.

所以我们在进行模型拟合时, 检验内容之一就是要检验拟合模型的残差是否满足方差齐性假定. 如果不满足, 那就说明残差序列还不是白噪声序列, 即拟合模型没有充分提取随机序列中的相关信息, 这时拟合模型的精度是值得怀疑的. 在这种场合下, 我们通常需要使用适当的条件异方差模型来处理异方差信息.

2.4.3 纯随机性检验

纯随机性检验也称为白噪声检验, 是专门用来检验序列是否为纯随机序列的一种方法. 我们知道如果一个序列是纯随机序列, 那么它的序列值之间应该没有任何相关关系, 即满足

$$\gamma(k) = 0, \quad \forall k \neq 0$$

这是一种理论上才会出现的理想状况. 实际上, 由于观察值序列的有限性, 导致纯随机序列的样本自相关系数不会绝对为零.

【例 2—4 续 (1)】 绘制例 2—4 标准正态白噪声序列的样本自相关图.

> acf(white_noise) #结果如图 2—22 所示

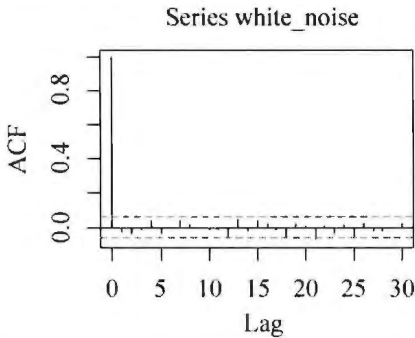


图 2—22 白噪声序列样本自相关图

样本自相关图显示这个纯随机序列没有一个样本自相关系数严格等于零. 但这些自相关系数确实都非常小, 都在零值附近以一个很小的幅度随机波动. 这就提醒我们应该考虑样本自相关系数的分布性质, 从统计意义上来判断序列的纯随机性质.

Barlett 证明, 如果一个时间序列是纯随机的, 得到一个观察期数为 n 的观察序列 $\{x_t, t = 1, 2, \dots, n\}$, 那么该序列的延迟非零期的样本自相关系数将近似服从均值为零、方差为序列观察期数倒数的正态分布, 即

$$\hat{\rho}_k \overset{\cdot}{\sim} N\left(0, \frac{1}{n}\right), \quad \forall k \neq 0$$

式中, n 为序列观察期数.

根据 Barlett 定理, 我们可以构造检验统计量来检验序列的纯随机性.

一、假设条件

由于序列值之间的变异性是绝对的, 而相关性是偶然的, 所以假设条件确定如下:

原假设: 延迟期数小于或等于 m 期的序列值之间相互独立.

备择假设: 延迟期数小于或等于 m 期的序列值之间有相关性.

该假设条件用数学语言描述即为:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0, \quad \forall m \geq 1$$

$$H_1: \text{至少存在某个 } \rho_k \neq 0, \quad \forall m \geq 1, k \leq m$$

二、检验统计量

1. Q 统计量

为了检验这个联合假设, Box 和 Pierce 推导出了 Q 统计量:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2$$

式中, n 为序列观测期数; m 为指定延迟期数.

根据正态分布和卡方分布之间的关系, 我们很容易推导出 Q 统计量近似服从自由度为 m 的卡方分布:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \sim \chi^2(m)$$

当 Q 统计量大于 $\chi_{1-\alpha}^2(m)$ 分位点, 或该统计量的 P 值小于 α 时, 则可以以 $1-\alpha$ 的置信水平拒绝原假设, 认为该序列为非白噪声序列; 否则, 接受原假设, 认为该序列为纯随机序列.

2. LB 统计量

在实际应用中人们发现 Q 统计量在大样本场合 (n 很大的场合) 检验效果很好, 但在小样本场合不太精确. 为了弥补这一缺陷, Box 和 Ljung 又推导出了 LB(Ljung-Box) 统计量:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right)$$

式中, n 为序列观测期数; m 为指定延迟期数.

Box 和 Ljung 证明 LB 统计量同样近似服从自由度为 m 的卡方分布.

实际上 LB 统计量就是 Box 和 Pierce 的 Q 统计量的修正, 所以人们习惯把它们统称为 Q 统计量, 分别记作 Q_{BP} 统计量 (Box 和 Pierce 的 Q 统计量) 和 Q_{LB} 统计量 (Box 和 Ljung 的 Q 统计量), 在各种检验场合普遍采用的 Q 统计量通常指的是 LB 统计量.

R 语言中使用 Box.test 函数进行纯随机性检验(白噪声检验). 该函数的命令格式为:

```
Box.test(x, type= , lag=)
```

式中:

-x: 变量名;

-type: 检验统计量类型.

(1) type='Box-Pierce', 输出白噪声检验的 Q 统计量. 该统计量为系统默认输出结果.

(2) type='Ljung-Box', 输出白噪声检验的 LB 统计量.

-lag: 延迟阶数. lag=n 表示输出滞后 n 阶的白噪声检验统计量. 忽略该选项时, 默认输出滞后 1 阶的检验统计量结果.

【例 2—4 续 (2)】 计算例 2—4 中白噪声序列延迟 6 期、延迟 12 期的 Q_{LB} 统计量的值, 并判断该序列的随机性 ($\alpha = 0.05$).

```
> Box.test(white_noise, lag=6)
```

```
Box-Pierce test
```

```
data: white_noise
```

```
X-squared=5.0327, df=6, p-value=0.5396
```

```
> Box.test(white_noise, lag=12)
```

```
Box-Pierce test
```

```
data: white_noise
```

```
X-squared=9.811, df=12, p-value=0.6325
```

R 是最小结果输出软件, 所以每次调用 Box.test 函数, 只能给出一个检验结果. 上面为了得到延迟 6 阶和延迟 12 阶两个 LB 统计量的结果, 我们调用了两次 Box.test 函数. 如果希望能够像其他统计软件一样, 一次得到多个白噪声检验结果, 可以编写一个循环命令. 最简单的循环函数是 for 函数, 它的命令格式如下:

```
for(x, in, n1:n2 ) state
```

式中:

-x: 循环变量名.

-n1:n2: 给出的循环取值区间.

-state: 需要循环执行的命令.

本例的等价命令和输出结果为:

```
> for(i in 1:2) print(Box.test(white_noise, lag=6*i))
```

```
Box-Pierce test
```

```
data: white_noise
X-squared=5.0327, df=6, p-value=0.5396

Box-Pierce test
.

data: white_noise
X-squared=9.811, df=12, p-value=0.6325
```

由于 P 值显著大于显著性水平 α , 所以该序列不能拒绝纯随机的原假设。换言之, 我们可以认为该序列的波动没有任何统计规律可循, 因而可以停止对该序列的统计分析。

还需要解释的一点是, 为什么在本例中只检验了前 6 期和前 12 期延迟的 Q 统计量和 LB 统计量就直接判断该序列是白噪声序列呢? 为什么不进行全部 999 期延迟检验呢?

这是因为平稳序列通常具有短期相关性, 如果序列值之间存在显著的相关关系, 通常只存在于延迟时期比较短的序列值之间。所以, 如果一个平稳序列短期延迟的序列值之间都不存在显著的相关关系, 通常长期延迟之间就更不会存在显著的相关关系了。

另一方面, 假如一个平稳序列显示出显著的短期相关性, 那么该序列就一定不是白噪声序列, 我们就可以对序列值之间存在的相关性进行分析。假如此时考虑的延迟时期数太长, 反而可能淹没了该序列的短期相关性。因为平稳序列只要延迟时期足够长, 自相关系数都会收敛于零。

【例 2—3 续 (2)】对 1949—1998 年北京市最高气温序列做白噪声检验 ($\alpha = 0.05$)。

```
> for(i in 1:2) print(Box.test(temp, lag=6*i))

Box-Pierce test

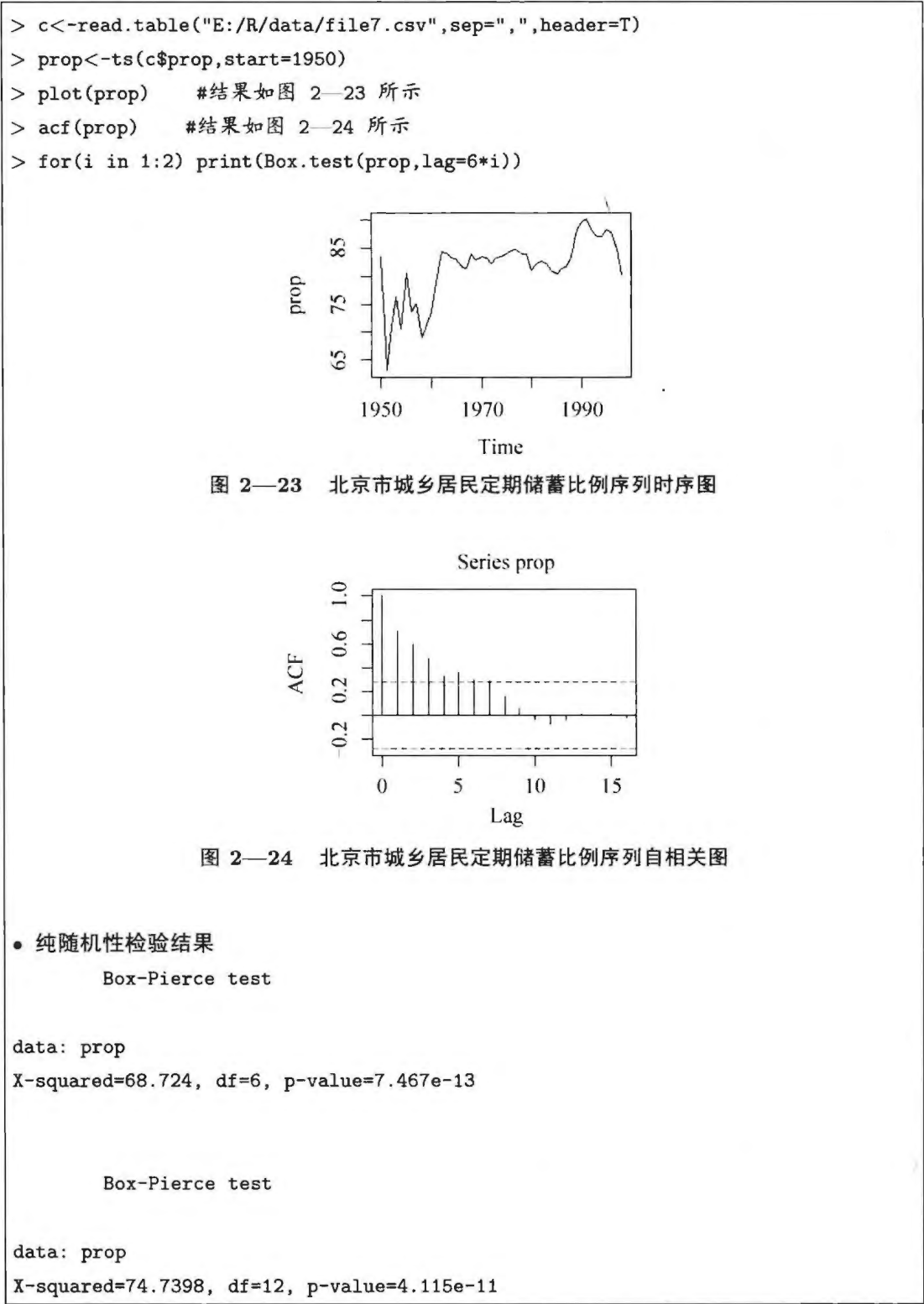
data: temp
X-squared=5.0166, df=6, p-value=0.5417

Box-Pierce test

data: temp
X-squared=5.8565, df=12, p-value=0.9231
```

根据这个检验结果, 不能拒绝序列纯随机的原假设。因而可以认为北京市最高气温的变动属于纯随机波动。这说明我们很难根据历史信息预测未来年份的最高气温。至此, 对该序列的分析也就结束了。

【例 2—5】 对 1950—1998 年北京市城乡居民定期储蓄所占比例序列的平稳性与纯随机性进行检验 (数据见表 A1—7).



(1) 时序图显示北京市城乡居民的定期储蓄始终占储蓄存款余额的 80%左右, 波动比较平稳.

(2) 样本自相关图显示延迟 7 阶之后, 自相关系数都落入 2 倍标准差范围以内, 而且自相关系数向零衰减的速度非常快, 延迟 8 阶之后自相关系数即在零值附近波动. 这是一个非常典型的短期相关的样本自相关图. 由时序图和样本自相关图的性质, 可以认为该序列平稳.

(3) 纯随机性检验结果显示, 在各阶延迟下 LB 检验统计量的 P 值都非常小 (< 0.05), 所以我们可以以很大的把握断定北京市城乡居民定期储蓄所占比例序列属于非白噪声序列.

结合前面的平稳性检验结果, 说明该序列不仅可以视为平稳的, 而且蕴涵着值得提取的相关信息. 这种平稳非白噪声序列是目前最容易分析的一种序列, 下一章我们将要详细介绍对这种平稳非白噪声序列的建模及预测方法.

2.5 习 题

- 1. 考虑系列 $\{1, 2, 3, 4, 5, \cdots, 20\}$:
 - (1) 判断该系列是否平稳.
 - (2) 计算该序列的样本自相关系数 $\hat{\rho}_k$ ($k = 1, 2, \cdots, 6$).
 - (3) 绘制该样本自相关图, 并解释该图形.
- 2. 1975—1980 年夏威夷岛莫那罗亚火山 (Mauna Loa) 每月释放的 CO_2 数据 (单位: ppm) 如表 2—7 所示 (行数据).

表 2—7

330.45	330.97	331.64	332.87	333.61	333.55
331.90	330.05	328.58	328.31	329.41	330.63
331.63	332.46	333.36	334.45	334.82	334.32
333.05	330.87	329.24	328.87	330.18	331.50
332.81	333.23	334.55	335.82	336.44	335.99
334.65	332.41	331.32	330.73	332.05	333.53
334.66	335.07	336.33	337.39	337.65	337.57
336.25	334.39	332.44	332.25	333.59	334.76
335.89	336.44	337.63	338.54	339.06	338.95
337.41	335.71	333.68	333.69	335.05	336.53
337.81	338.16	339.88	340.57	341.19	340.87
339.25	337.19	335.49	336.63	337.74	338.36

- (1) 绘制该序列时序图, 并判断该系列是否平稳.
- (2) 计算该序列的样本自相关系数 $\hat{\rho}_k$ ($k = 1, 2, \cdots, 24$).
- (3) 绘制该样本自相关图, 并解释该图形.
- 3. 1945—1950 年费城月度降雨量数据 (单位: mm) 如表 2—8 所示 (行数据).

表 2—8

69.3	80.0	40.9	74.9	84.6	101.1	225.0	95.3	100.6	48.3	144.5	128.3
38.4	52.3	68.6	37.1	148.6	218.7	131.6	112.8	81.8	31.0	47.5	70.1
96.8	61.5	55.6	171.7	220.5	119.4	63.2	181.6	73.9	64.8	166.9	48.0
137.7	80.5	105.2	89.9	174.8	124.0	86.4	136.9	31.5	35.3	112.3	143.0
160.8	97.0	80.5	62.5	158.2	7.6	165.9	106.7	92.2	63.2	26.2	77.0
52.3	105.4	144.3	49.5	116.1	54.1	148.6	159.3	85.3	67.3	112.8	59.4

- (1) 计算该序列的样本自相关系数 $\hat{\rho}_k$ ($k = 1, 2, \dots, 24$).
- (2) 判断该序列的平稳性.
- (3) 判断该序列的纯随机性.
4. 若序列长度为 100, 前 12 个样本自相关系数如下:

$$\begin{aligned} \rho_1 &= 0.02, \rho_2 = 0.05, \rho_3 = 0.10, \rho_4 = -0.02, \rho_5 = 0.05, \rho_6 = 0.01 \\ \rho_7 &= 0.12, \rho_8 = -0.06, \rho_9 = 0.08, \rho_{10} = -0.05, \rho_{11} = 0.02, \rho_{12} = -0.05 \end{aligned}$$

该序列能否视为纯随机序列 ($\alpha=0.05$)?

5. 表 2—9 是某公司在 2000—2003 年期间每月的销售量数据.

表 2—9

月份	2000 年	2001 年	2002 年	2003 年
1	153	134	145	117
2	187	175	203	178
3	234	243	189	149
4	212	227	214	178
5	300	298	295	248
6	221	256	220	202
7	201	237	231	162
8	175	165	174	135
9	123	124	119	120
10	104	106	85	96
11	85	87	67	90
12	78	74	75	63

- (1) 绘制该序列时序图及样本自相关图.
- (2) 判断该序列的平稳性.
- (3) 判断该序列的纯随机性.

6. 1969 年 1 月至 1973 年 9 月在芝加哥海德公园内每 28 天发生的抢包案件数如表 2—10 所示 (行数据).

表 2—10

10	15	10	10	12	10	7	7	10	14	8	17
14	18	3	9	11	10	6	12	14	10	25	29
33	33	12	19	16	19	19	12	34	15	36	29
26	21	17	19	13	20	24	12	6	14	6	12
9	11	17	12	8	14	14	12	5	8	10	3
16	8	8	7	12	6	10	8	10	5		

(1) 判断该序列 $\{x_t\}$ 的平稳性及纯随机性.

(2) 对该序列进行函数运算:

$$y_t = x_t - x_{t-1}$$

并判断序列 $\{y_t\}$ 的平稳性及纯随机性.

一个序列经过预处理被识别为平稳非白噪声序列, 那就说明该序列是一个蕴涵着相关信息的平稳序列. 在统计上, 我们通常是建立一个线性模型来拟合该序列的发展, 借此提取该序列中的有用信息. ARMA 模型是目前最常用的平稳序列拟合模型.

3.1 方法性工具

在时间序列分析中有一些方法性工具经常使用, 它们可以使我们的模型表达和序列分析更加简洁、方便, 所以在介绍具体的模型之前先简单介绍这些常用的方法性工具.

3.1.1 差分运算

一、 p 阶差分

相距一期的两个序列值之间的减法运算称为 1 阶差分运算. 记 ∇x_t 为 x_t 的 1 阶差分:

$$\nabla x_t = x_t - x_{t-1}$$

对 1 阶差分后序列再进行一次 1 阶差分运算称为 2 阶差分. 记 $\nabla^2 x_t$ 为 x_t 的 2 阶差分:

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$$

依此类推, 对 $p-1$ 阶差分后序列再进行一次 1 阶差分运算称为 p 阶差分. 记 $\nabla^p x_t$ 为 x_t 的 p 阶差分:

$$\nabla^p x_t = \nabla^{p-1} x_t - \nabla^{p-1} x_{t-1}$$

二、 k 步差分

相距 k 期的两个序列值之间的减法运算称为 k 步差分运算. 记 $\nabla_k x_t$ 为 x_t 的 k 步差分:

$$\nabla_k x_t = x_t - x_{t-k}$$

3.1.2 延迟算子

一、定义

延迟算子类似于一个时间指针, 当前序列值乘以一个延迟算子, 就相当于把当前序列值的时间向过去拨了一个时刻. 记 B 为延迟算子, 有

$$\begin{aligned} x_{t-1} &= Bx_t \\ x_{t-2} &= B^2x_t \\ &\vdots \\ x_{t-p} &= B^p x_t \end{aligned}$$

延迟算子有如下性质:

- (1) $B^0 = 1$
- (2) 若 c 为任一常数, 有 $B(c \cdot x_t) = c \cdot B(x_t) = c \cdot x_{t-1}$
- (3) 对任意两个序列 $\{x_t\}$ 和 $\{y_t\}$, 有 $B(x_t \pm y_t) = x_{t-1} \pm y_{t-1}$
- (4) $B^n x_t = x_{t-n}$
- (5) $(1 - B)^n = \sum_{i=0}^n (-1)^i C_n^i B^i$, 其中 $C_n^i = \frac{n!}{i!(n-i)!}$

二、用延迟算子表示差分运算

(1) p 阶差分

$$\nabla^p x_t = (1 - B)^p x_t = \sum_{i=0}^p (-1)^i C_p^i x_{t-i}$$

(2) k 步差分

$$\nabla_k x_t = x_t - x_{t-k} = (1 - B^k)x_t$$

3.1.3 线性差分方程

一、线性差分方程的定义

定义 3.1 称如下形式的方程为序列 $\{z_t, t = 0, \pm 1, \pm 2, \dots\}$ 的线性差分方程:

$$z_t + a_1 z_{t-1} + a_2 z_{t-2} + \dots + a_p z_{t-p} = h(t) \quad (3.1)$$

式中, $p \geq 1$; $a_1, a_2, \dots, a_p$ 为实数; $h(t)$ 为 t 的已知函数.

特别地, 若 $h(t) = 0$, 则差分方程

$$z_t + a_1 z_{t-1} + a_2 z_{t-2} + \dots + a_p z_{t-p} = 0 \quad (3.2)$$

称为齐次线性差分方程. 否则, 线性差分方程 (3.1) 称为非齐次线性差分方程.

二、齐次线性差分方程的解

齐次线性差分方程的求解要借助它的特征方程和特征根. 齐次线性差分方程 (3.2) 的特征方程为:

$$\lambda^p + a_1\lambda^{p-1} + a_2\lambda^{p-2} + \cdots + a_p = 0 \quad (3.3)$$

这是一个一元 p 次线性方程, 它应该有 p 个非零根, 称这 p 个非零根为特征方程的特征根, 不妨记作

$$\lambda_1, \lambda_2, \cdots, \lambda_p$$

特征根的取值不同, 齐次线性差分方程 (3.2) 的解会有不同的表达式. 下面分情况讨论.

(1) $\lambda_1, \lambda_2, \cdots, \lambda_p$ 为 p 个不同的实数.

这时齐次线性差分方程 (3.2) 的解为:

$$z_t = c_1\lambda_1^t + c_2\lambda_2^t + \cdots + c_p\lambda_p^t$$

式中, $c_1, c_2, \cdots, c_p$ 为任意实数.

(2) $\lambda_1, \lambda_2, \cdots, \lambda_p$ 中有相同实根.

不妨假定 $\lambda_1 = \lambda_2 = \cdots = \lambda_d$ 为 d 个相同实根, 而 $\lambda_{d+1}, \lambda_{d+2}, \cdots, \lambda_p$ 为互不相同实根. 这时齐次线性差分方程 (3.2) 的解为:

$$z_t = (c_1 + c_2t + \cdots + c_dt^{d-1})\lambda_1^t + c_{d+1}\lambda_{d+1}^t + \cdots + c_p\lambda_p^t$$

式中, $c_1, c_2, \cdots, c_p$ 为任意实数.

(3) $\lambda_1, \lambda_2, \cdots, \lambda_p$ 中有复根.

由于差分方程的系数 $a_1, a_2, \cdots, a_p$ 为实数, 所以其复根必呈共轭出现. 不妨假定

$$\lambda_1 = a + ib = re^{i\omega}, \lambda_2 = a - ib = re^{-i\omega}$$

为一对共轭复根, 其中 $r = \sqrt{a^2 + b^2}$, $\omega = \arccos \frac{a}{r}$, 而 $\lambda_3, \lambda_4, \cdots, \lambda_p$ 为互不相同实根. 这时齐次线性差分方程 (3.2) 的解为:

$$\begin{aligned} z_t &= c_1\lambda_1^t + c_2\lambda_2^t + \cdots + c_p\lambda_p^t \\ &= r^t(c_1e^{it\omega} + c_2e^{-it\omega}) + c_3\lambda_3^t + \cdots + c_p\lambda_p^t \end{aligned}$$

式中, $c_1, c_2, \cdots, c_p$ 为任意实数.

三、非齐次线性差分方程的解

求非齐次线性差分方程 (3.1) 的解通常需要进行两步运算. 首先求出齐次线性差分方程 (3.2) 的通解 z_t' ; 然后求出该非齐次线性差分方程 (3.1) 的一个特解 z_t'' .

所谓特解, 就是任意一个使得非齐次线性差分方程 (3.1) 成立的值 z_t'' , 即

$$z_t'' + a_1 z_{t-1}'' + a_2 z_{t-2}'' + \cdots + a_p z_{t-p}'' = h(t)$$

而非齐次线性差分方程 (3.1) 的通解即为齐次线性差分方程 (3.2) 的通解 z_t' 和非齐次线性差分方程 (3.1) 的特解 z_t'' 之和, 即

$$z_t = z_t' + z_t''$$

四、时间序列模型与线性差分方程

线性差分方程在时间序列分析中有重要的应用. 常用的时间序列模型和某些模型的自协方差函数和自相关函数都可以视为线性差分方程, 而线性差分方程对应的特征根的性质对判断模型的平稳性有非常重要的意义.

3.2 ARMA 模型的性质

ARMA 模型的全称是自回归移动平均模型, 它是目前最常用的拟合平稳序列的模型. 它又可以细分为 AR 模型、MA 模型和 ARMA 模型三大类.

3.2.1 AR 模型

一、定义

定义 3.2 具有如下结构的模型称为 p 阶自回归模型, 简记为 $AR(p)$:

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \\ \phi_p \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (3.4)$$

$AR(p)$ 模型有三个限制条件:

条件一: $\phi_p \neq 0$. 这个限制条件保证了模型的最高阶数为 p .

条件二: $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t$. 这个限制条件实际上是要求随机干扰序列 $\{\varepsilon_t\}$ 为零均值白噪声序列.

条件三: $E(x_s \varepsilon_t) = 0, \forall s < t$. 这个限制条件说明当期的随机干扰与过去的序列值无关.

通常会缺省默认式 (3.4) 的限制条件, 把 $AR(p)$ 模型简记为:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \quad (3.5)$$

当 $\phi_0 = 0$ 时, 自回归模型 (3.4) 又称为中心化 $AR(p)$ 模型. 非中心化 $AR(p)$ 序列都可以通过下面的变换转化为中心化 $AR(p)$ 序列.

令

$$\mu = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}, \quad y_t = x_t - \mu$$

则 $\{y_t\}$ 为 $\{x_t\}$ 的中心化序列. 中心化变换实际上就是非中心化序列整个平移了一个常数位移. 这种整体移动对序列值之间的相关关系没有任何影响, 所以今后在分析 AR 模型的相关关系时, 都简化为对它的中心化模型进行分析.

引进延迟算子, 中心化 AR(p) 模型又可以简记为:

$$\Phi(B)x_t = \varepsilon_t \quad (3.6)$$

式中, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, 称为 p 阶自回归系数多项式.

二、AR 模型平稳性判别

要拟合一个平稳序列的发展, 用来拟合的模型显然也应该是平稳的. AR 模型是常用的平稳序列的拟合模型之一, 但并非所有的 AR 模型都是平稳的.

R 提供了多种序列拟合函数, 每种函数各有利弊. 我们介绍两种最常用的序列拟合方法.

1. arima.sim 函数拟合

arima.sim 函数是一个便捷的序列拟合函数, 它可以拟合平稳 AR 序列、MA 序列、平稳 ARMA 序列和第 5 章要介绍的 ARIMA 序列. 它的命令格式为:

```
arima.sim(n, list(ar=, ma=, order=), sd=)
```

式中:

-n: 拟合序列长度.

-list: 指定具体的模型参数. 其中:

(1) 拟合平稳 AR(p) 模型, 要给出自回归系数. 如果指定拟合的 AR 模型为非平稳模型, 系统会报错.

(2) 拟合 MA(q) 模型, 要给出移动平均系数.

(3) 拟合平稳 ARMA(p, q) 模型, 需要同时给出自回归系数和移动平均系数. 如果指定拟合的 ARMA 模型为非平稳模型, 系统会报错.

(4) 拟合 ARIMA(p, d, q) 模型(第 5 章介绍), 除了需要给出自回归系数和移动平均系数, 还需要增加 order 选项. order=c(p, d, q), 其中, p 为自回归阶数; d 为差分阶数; q 为移动平均阶数.

-sd: 指定序列的标准差, 不特殊指定, 系统默认 sd=1.

2. filter 函数拟合

filter 函数可以直接拟合 AR 序列(无论是否平稳)和 MA 序列. filter 函数的命令格式为:

```
filter(e, filter=, method=, circular =)
```

式中:

-e: 随机波动序列的变量名.

-filter: 指定模型系数. 其中:

(1) AR(p) 模型为 $\text{filter} = c(\phi_1, \phi_2, \dots, \phi_p)$;

(2) MA(q) 模型为 $\text{filter} = c(1, -\theta_1, -\theta_2, \dots, -\theta_q)$.

-method: 指定拟合的是 AR 模型还是 MA 模型.

(1) method="recursive" 为 AR 模型;

(2) method="convolution" 为 MA 模型.

-circular: 拟合 MA 模型时专用的一个选项, circular = T 可以避免 NA 数据出现.

【例 3—1】 考察如下四个 AR 模型的平稳性:

$$(1) x_t = 0.8x_{t-1} + \varepsilon_t$$

$$(2) x_t = -1.1x_{t-1} + \varepsilon_t$$

$$(3) x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

$$(4) x_t = x_{t-1} + 0.5x_{t-2} + \varepsilon_t$$

拟合这四个序列的序列值, 并绘制时序图.

```
> x1<-arima.sim(n=100,list(ar=0.8))
> x3<-arima.sim(n=100,list(ar=c(1,-0.5)))
> e<-rnorm(100)
> x2<-filter(e,filter=-1.1,method="recursive")
> x4<-filter(e,filter=c(1,0.5),method="recursive")
> ts.plot(x1)
> ts.plot(x2)
> ts.plot(x3)
> ts.plot(x4)
```

#结果如图 3—1 所示

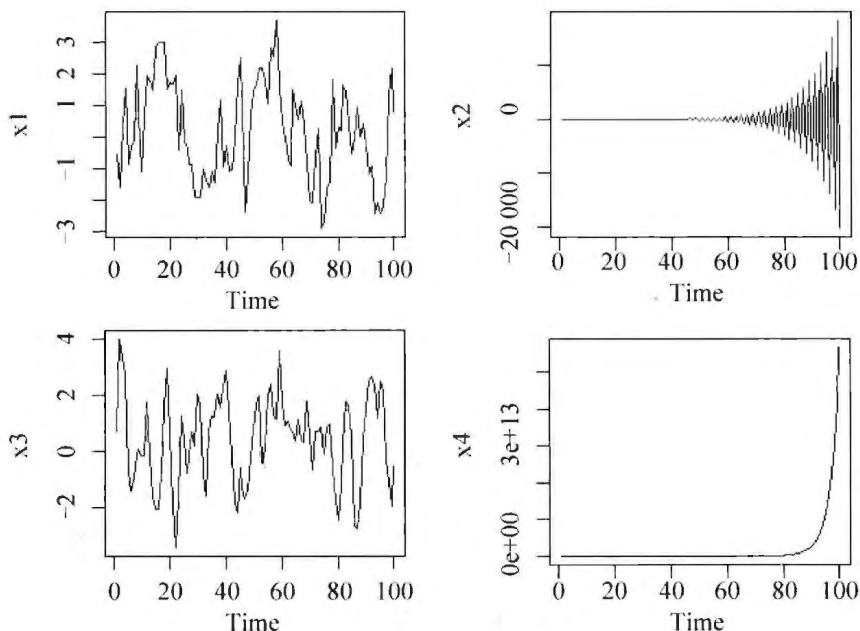


图 3—1 四个 AR 序列时序图

根据图 3—1 可以直观判断出 (1)、(3) 模型平稳, (2)、(4) 模型非平稳. 图示法只是一种粗糙的直观判别方法, 我们有两种准确的平稳性判别方法: 特征根判别和平稳域判别.

1. 特征根判别

任一中心化 $AR(p)$ 模型 $\Phi(B)x_t = \varepsilon_t$ 都可以被视为一个非齐次线性差分方程:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p} = \varepsilon_t \quad (3.7)$$

方程 (3.7) 的通解为:

$$x_t = x'_t + x''_t$$

式中, x'_t 为齐次线性差分方程 $\Phi(B)x_t = 0$ 的通解; x''_t 为方程 (3.7) 的一个特解.

(1) 求齐次线性差分方程 $\Phi(B)x_t = 0$ 的一个通解 x'_t . 假定 $\lambda_1, \lambda_2, \cdots, \lambda_p$ 是该特征方程的 p 个特征根. 为了有代表性, 不妨假设这 p 个特征根取值如下:

$\lambda_1 = \lambda_2 = \cdots = \lambda_d$ 为 d 个相等实根.

$\lambda_{d+1}, \lambda_{d+2}, \cdots, \lambda_{p-2m}$ 为 $p-d-2m$ 个互不相等的实根.

$\lambda_{j1} = r_j e^{i\omega_j}, \lambda_{j2} = r_j e^{-i\omega_j} (j = 1, \cdots, m)$ 为 m 对共轭复根.

那么齐次线性差分方程 $\Phi(B)x_t = 0$ 的通解为:

$$x'_t = \sum_{j=1}^d c_j t^{j-1} \lambda_1^t + \sum_{j=d+1}^{p-2m} c_j \lambda_j^t + \sum_{j=1}^m r_j^t (c_{1j} e^{it\omega_j} + c_{2j} e^{-it\omega_j})$$

式中, $c_1, \cdots, c_{p-2m}, c_{1j}, c_{2j} (j = 1, \cdots, m)$ 为任意实数.

(2) 求非齐次线性差分方程 $\Phi(B)x_t = \varepsilon_t$ 的一个特解 x''_t . 首先, 可以证明 $AR(p)$ 模型的自回归系数多项式方程 $\Phi(u) = 0$ 的根是齐次线性差分方程 $\Phi(B)x_t = 0$ 的特征根的倒数.

证明: 设 $\lambda_1, \lambda_2, \cdots, \lambda_p$ 为齐次线性差分方程 $\Phi(B)x_t = 0$ 的 p 个特征根, 任取 $\lambda_i (i \in (1, 2, \cdots, p))$ 代入特征方程, 有

$$\lambda_i^p - \phi_1 \lambda_i^{p-1} - \phi_2 \lambda_i^{p-2} - \cdots - \phi_p = 0$$

把 $u_i = \frac{1}{\lambda_i}$ 代入 $AR(p)$ 模型的自回归系数多项式, 有

$$\begin{aligned} \Phi(u_i) &= 1 - \phi_1 \frac{1}{\lambda_i} - \cdots - \phi_p \frac{1}{\lambda_i^p} \\ &= \frac{1}{\lambda_i^p} [\lambda_i^p - \phi_1 \lambda_i^{p-1} - \cdots - \phi_p] = 0 \end{aligned}$$

证毕.

根据这个性质, $\Phi(B)$ 可以因子分解成

$$\Phi(B) = \prod_{i=1}^p (1 - \lambda_i B)$$

由此可以得到非齐次线性差分方程 (3.7) 的一个特解为:

$$x''_t = \frac{\varepsilon_t}{\Phi(B)} = \frac{\varepsilon_t}{\prod_{i=1}^p (1 - \lambda_i B)} = \sum_{i=1}^p \frac{k_i}{1 - \lambda_i B} \varepsilon_t$$

式中, $k_i (i = 1, 2, \cdots, p)$ 为常数.

(3) 求非齐次线性差分方程 (3.7) 的通解 x_t .

$$\begin{aligned} x_t &= x'_t + x''_t \\ &= \sum_{j=1}^d c_j t^{j-1} \lambda_1^t + \sum_{j=d+1}^{p-2m} c_j \lambda_j^t + \sum_{j=1}^m r_j^t (c_{1j} e^{itw_j} + c_{2j} e^{-itw_j}) + \sum_{i=1}^p \frac{k_i}{1 - \lambda_i B} \varepsilon_t \end{aligned}$$

要使得中心化 $AR(p)$ 模型平稳, 即要求对任意实数 $c_1, \dots, c_{p-2m}, c_{1j}, c_{2j} (j = 1, 2, \dots, m)$, 有

$$\lim_{t \rightarrow \infty} x_t = 0 \quad (3.8)$$

式 (3.8) 成立的充要条件是

$$\begin{aligned} |\lambda_i| &< 1, i = 1, 2, \dots, p - 2m \\ |r_i| &< 1, i = 1, 2, \dots, m \end{aligned} \quad (3.9)$$

式 (3.9) 实际上就是要求 $AR(p)$ 模型的 p 个特征根都在单位圆内. 所以 $AR(p)$ 模型平稳的充要条件是它的 p 个特征根都在单位圆内.

根据特征根和自回归系数多项式的根成倒数的性质, AR 模型平稳的等价判别条件是: 该 AR 模型的自回归系数多项式的根, 即 $\Phi(u) = 0$ 的根, 都在单位圆外.

2. 平稳域判别

对于一个 $AR(p)$ 模型而言, 如果没有平稳性的要求, 实际上也就意味着对参数向量 $(\phi_1, \phi_2, \dots, \phi_p)'$ 没有任何限制, 它们可以取遍 p 维欧氏空间的任意一点, 但是如果加上了平稳性限制, 参数向量 $(\phi_1, \phi_2, \dots, \phi_p)'$ 就只能取 p 维欧氏空间的一个子集, 使得特征根都在单位圆内的系数集合

$$\{(\phi_1, \phi_2, \dots, \phi_p) \mid \text{特征根都在单位圆内}\}$$

称为 $AR(p)$ 模型的平稳域.

对于低阶 AR 模型用平稳域的方法判别模型的平稳性通常更为简便.

(1) $AR(1)$ 模型的平稳域.

$AR(1)$ 模型为: $x_t = \phi_1 x_{t-1} + \varepsilon_t$, 其特征方程为: $\lambda - \phi_1 = 0$, 特征根为: $\lambda = \phi_1$. 根据 AR 模型平稳的充要条件, 容易推出 $AR(1)$ 模型平稳的充要条件是

$$|\phi_1| < 1 \quad (3.10)$$

所以, $AR(1)$ 模型的平稳域就是 $\{|\phi_1| < 1\}$.

(2) $AR(2)$ 模型的平稳域.

$AR(2)$ 模型为: $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$, 其特征方程为: $\lambda^2 - \phi_1 \lambda - \phi_2 = 0$, 特征根为: $\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$. 根据 AR 模型平稳的充要条件, $AR(2)$ 模型平稳的充要条件是 $|\lambda_1| < 1$ 且 $|\lambda_2| < 1$.

根据一元二次方程的性质和 $AR(2)$ 模型的平稳条件, 有

$$\begin{cases} \lambda_1 + \lambda_2 = \phi_1, \text{ 且 } |\lambda_1| < 1, |\lambda_2| < 1 \\ \lambda_1 \lambda_2 = -\phi_2 \end{cases} \quad (3.11)$$

可以导出:

1) $|\phi_2| = |\lambda_1 \lambda_2| < 1$

2) $\phi_2 + \phi_1 = -\lambda_1 \lambda_2 + \lambda_1 + \lambda_2 = 1 - (1 - \lambda_1)(1 - \lambda_2) < 1$

3) $\phi_2 - \phi_1 = -\lambda_1 \lambda_2 - \lambda_1 - \lambda_2 = 1 - (1 + \lambda_1)(1 + \lambda_2) < 1$

这三个限制条件意味着 AR(2) 模型的平稳域是一个三角形区域, 如图 3—2 所示.

$\{\phi_1, \phi_2 \mid |\phi_2| < 1, \text{ 且 } \phi_2 \pm \phi_1 < 1\}$

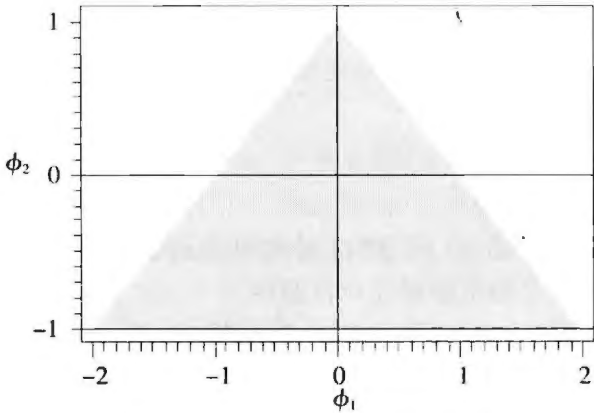


图 3—2 AR(2) 模型的平稳域

【例 3—1 续】 分别用特征根判别法和平稳域判别法检验例 3.1 中四个 AR 模型的平稳性:

(1) $x_t = 0.8x_{t-1} + \varepsilon_t$

(2) $x_t = -1.1x_{t-1} + \varepsilon_t$

(3) $x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$

(4) $x_t = x_{t-1} + 0.5x_{t-2} + \varepsilon_t$

其中 $\{\varepsilon_t\}$ 均为服从标准正态分布的白噪声序列.

结论如表 3—1 所示.

表 3—1

模型	特征根判别	平稳域判别	结论
(1)	$\lambda_1 = 0.8$	$\phi_1 = 0.8$	平稳
(2)	$\lambda_2 = -1.1$	$\phi_1 = -1.1$	非平稳
(3)	$\lambda_1 = \frac{1+i}{2}, \lambda_2 = \frac{1-i}{2}$	$ \phi_2 = 0.5, \phi_2 + \phi_1 = 0.5, \phi_2 - \phi_1 = -1.5$	平稳
(4)	$\lambda_1 = \frac{1+\sqrt{3}}{2}, \lambda_2 = \frac{1-\sqrt{3}}{2}$	$ \phi_2 = 0.5, \phi_2 + \phi_1 = 1.5, \phi_2 - \phi_1 = -0.5$	非平稳

理论判别得到的结论支持例 3.1 根据时序图图 3—1 所得出的直观判断.

三、平稳 AR 模型的统计性质

1. 均值

假如 $AR(p)$ 模型 (3.5) 满足了平稳性条件, 在等式两边取期望, 得

$$Ex_t = E(\phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t) \quad (3.12)$$

根据平稳序列均值为常数的性质, 有 $Ex_t = \mu$ ($\forall t \in T$), 且因为 $\{\varepsilon_t\}$ 为白噪声序列, 有 $E\varepsilon_t = 0$, 所以式 (3.12) 等价于

$$\begin{aligned} (1 - \phi_1 - \cdots - \phi_p)\mu &= \phi_0 \\ \Rightarrow \mu &= \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p} \end{aligned}$$

特别地, 对于中心化 AR(p) 模型, 有 $Ex_t = 0$.

2. 方差

(1) Green 函数. 要得到平稳 AR(p) 模型的方差, 需要借助 Green 函数的帮助, 下面给出 Green 函数的定义.

设 $\lambda_1, \lambda_2, \cdots, \lambda_p$ 为平稳 AR(p) 模型的特征根, 则平稳 AR(p) 模型可以写成如下形式:

$$\begin{aligned} x_t &= \frac{\varepsilon_t}{\Phi(B)} \\ &= \sum_{i=1}^p \frac{k_i}{1 - \lambda_i B} \varepsilon_t \\ &= \sum_{i=1}^p \sum_{j=0}^{\infty} k_i (\lambda_i B)^j \varepsilon_t \\ &= \sum_{j=0}^{\infty} \sum_{i=1}^p k_i \lambda_i^j \varepsilon_{t-j} \\ &\triangleq \sum_{j=0}^{\infty} G_j \varepsilon_{t-j} \end{aligned} \quad (3.13)$$

式中, $G_0 = 1, G_j = \sum_{i=1}^p k_i \lambda_i^j$ ($j = 1, 2, \cdots$).

式 (3.13) 称为 AR 模型的传递形式, 系数 G_j ($j = 0, 1, 2, \cdots$) 称为 Green 函数. 因为 $\lambda_1, \lambda_2, \cdots, \lambda_p$ 都在单位圆内, 所以 Green 函数应该呈指数下降, 且 $\lim_{j \rightarrow \infty} |G_j| = 0$. 通过待定系数法, 可以确定 Green 函数的递推公式.

记 $G(B) = \sum_{j=0}^{\infty} G_j B^j$, 则中心化 AR(p) 模型 (3.13) 可以简记为:

$$x_t = G(B)\varepsilon_t \quad (3.14)$$

把式 (3.14) 代入 AR(p) 模型 $\Phi(B)x_t = \varepsilon_t$, 得到

$$\Phi(B)G(B)\varepsilon_t = \varepsilon_t$$

展开上式, 得

$$\left(1 - \sum_{k=1}^p \phi_k B^k\right) \left(\sum_{j=0}^{\infty} G_j B^j\right) \varepsilon_t = \varepsilon_t$$

整理得

$$\left[1 + \sum_{j=1}^{\infty} \left(G_j - \sum_{k=1}^j \phi'_k G_{j-k} \right) B^j \right] \varepsilon_t = \varepsilon_t$$

根据待定系数法, 有

$$G_j - \sum_{k=1}^j \phi'_k G_{j-k} = 0, \quad j = 1, 2, \dots$$

由此可以得到 Green 函数递推公式:

$$\begin{cases} G_0 = 1 \\ G_j = \sum_{k=1}^j \phi'_k G_{j-k}, \quad j = 1, 2, \dots \end{cases} \quad (3.15)$$

式中

$$\phi'_k = \begin{cases} \phi_k, & k \leq p \\ 0, & k > p \end{cases}$$

(2) 平稳 AR 模型的方差. 对平稳 AR 模型 (3.14) 等式两边求方差, 有

$$\text{Var}(x_t) = \sum_{j=0}^{\infty} G_j^2 \text{Var}(\varepsilon_t)$$

式中, $\{\varepsilon_t\}$ 为白噪声序列; $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. 所以

$$\text{Var}(x_t) = \sum_{j=0}^{\infty} G_j^2 \sigma_\varepsilon^2 \quad (3.16)$$

因为 G_j 呈指数下降, 所以 $\sum_{j=0}^{\infty} G_j^2 < \infty$, 这说明平稳序列 $\{x_t\}$ 方差有界, 等于常数

$$\sum_{j=0}^{\infty} G_j^2 \sigma_\varepsilon^2.$$

【例 3—2】 求平稳 AR(1) 模型的方差.

把平稳 AR(1) 模型转化为传递形式:

$$(1 - \phi_1 B)x_t = \varepsilon_t$$

Green 函数为:

$$G_j = \phi_1^j, \quad j = 0, 1, \dots$$

所以平稳 AR(1) 模型的方差为:

$$\text{Var}(x_t) = \sum_{j=0}^{\infty} G_j^2 \text{Var}(\varepsilon_t) = \sum_{j=0}^{\infty} \phi_1^{2j} \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

3. 协方差函数

在平稳模型 $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t$ 等号两边同乘 $x_{t-k} (\forall k \geq 1)$, 再求期望, 得

$$E(x_t x_{t-k}) = \phi_1 E(x_{t-1} x_{t-k}) + \cdots + \phi_p E(x_{t-p} x_{t-k}) + E(\varepsilon_t x_{t-k}), \quad \forall k \geq 1$$

根据式 (3.4)AR(p) 模型的条件三, 有

$$E(\varepsilon_t x_{t-k}) = 0, \quad \forall k \geq 1$$

于是可以得到如下自协方差函数的递推公式:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \quad (3.17)$$

【例 3—3】求平稳 AR(1) 模型的自协方差函数.

平稳 AR(1) 模型的自协方差函数递推公式为:

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1} \\ &= \phi_1^k \gamma_0 \end{aligned}$$

根据例 3—2 已知

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

所以平稳 AR(1) 模型的自协方差函数递推公式如下:

$$\gamma_k = \phi_1^k \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}, \quad \forall k \geq 1$$

【例 3—4】求平稳 AR(2) 模型的自协方差函数.

平稳 AR(2) 模型的递推公式为:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad \forall k \geq 1$$

特别地, 当 $k = 1$ 时, 有

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

即

$$\gamma_1 = \frac{\phi_1 \gamma_0}{1 - \phi_2}$$

利用 Green 函数可以推导出 AR(2) 模型的方差为:

$$\gamma_0 = \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)} \sigma_\varepsilon^2$$

所以平稳 AR(2) 模型的自协方差函数的递推公式如下:

$$\begin{cases} \gamma_0 = \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)} \sigma_\varepsilon^2 \\ \gamma_1 = \frac{\phi_1 \gamma_0}{1 - \phi_2} \\ \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, k \geq 2 \end{cases}$$

4. 自相关系数

(1) 平稳 AR 模型自相关系数递推公式. 由于 $\rho_k = \frac{\gamma_k}{\gamma_0}$, 在自协方差函数的递推公式 (3.17) 等号两边同除以方差函数 γ_0 , 就得到自相关系数的递推公式:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \quad (3.18)$$

容易验证平稳 AR(1) 模型的自相关系数递推公式为:

$$\rho_k = \phi_1^k, \quad k \geq 0$$

平稳 AR(2) 模型的自相关系数递推公式为:

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\phi_1}{1 - \phi_2}, & k = 1 \\ \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, & k \geq 2 \end{cases}$$

(2) 自相关系数的性质. 平稳 AR(p) 模型的自相关系数有两个显著的性质: 一是拖尾性; 二是呈指数衰减. 这两个性质都可以从自相关系数的通解推出.

根据式 (3.18), 容易看出 AR(p) 模型的自相关系数的表达式实际上是一个 p 阶齐次差分方程. 那么滞后任意 k 阶的自相关系数的通解为:

$$\rho_k = \sum_{i=1}^p c_i \lambda_i^k \quad (3.19)$$

式中, $|\lambda_i| < 1$ ($i = 1, 2, \dots, p$) 为该差分方程的特征根; $c_1, c_2, \dots, c_p$ 为任意常数. 如果已知任意 p 个自相关系数的值就可以确定 $c_1, c_2, \dots, c_p$ 的解, 显然 $c_1, c_2, \dots, c_p$ 不能全为零.

通过这个通解形式, 容易推出 ρ_k 始终有非零取值, 不会在 k 大于某个常数之后就恒等于零, 这个性质就是拖尾性.

可以直观地解释 AR(p) 模型自相关系数拖尾的原因. 对于一个平稳 AR(p) 模型:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

虽然它的表达式显示 x_t 只受当期随机误差 ε_t 和最近 p 期的序列值 $x_{t-1}, \dots, x_{t-p}$ 的影响, 但是由于 x_{t-1} 的值又依赖于 x_{t-1-p} , 所以实际上 x_{t-1-p} 对 x_t 也有影响, 依此类推, x_t 之前的每一个序列值 $x_{t-1}, \dots, x_{t-k}, \dots$ 都会对 x_t 构成影响. 自回归模型的这种特性体现在自相关系数上就是自相关系数的拖尾性.

同时随着时间的推移, ρ_k 会迅速衰减, 因为 $|\lambda_i| < 1$ ($i = 1, 2, \dots, p$), 所以 $k \rightarrow \infty$ 时, $\lambda_i^k \rightarrow 0$ ($i = 1, 2, \dots, p$), 继而导致 $\rho_k = \sum_{i=1}^p c_i \lambda_i^k \rightarrow 0$, 而且这种影响是以指数 λ^k 的速度在衰减.

这种自相关系数以指数衰减的性质就是在第 2 章利用自相关图判断平稳序列时所说的“短期相关性”. 它是平稳序列的一个重要特征. 这个特征表明对平稳序列而言, 通常只有近期的序列值对现时值的影响比较明显, 间隔越远的过去值对现时值的影响越小.

【例 3—5】考察如下四个平稳 AR 模型的自相关图.

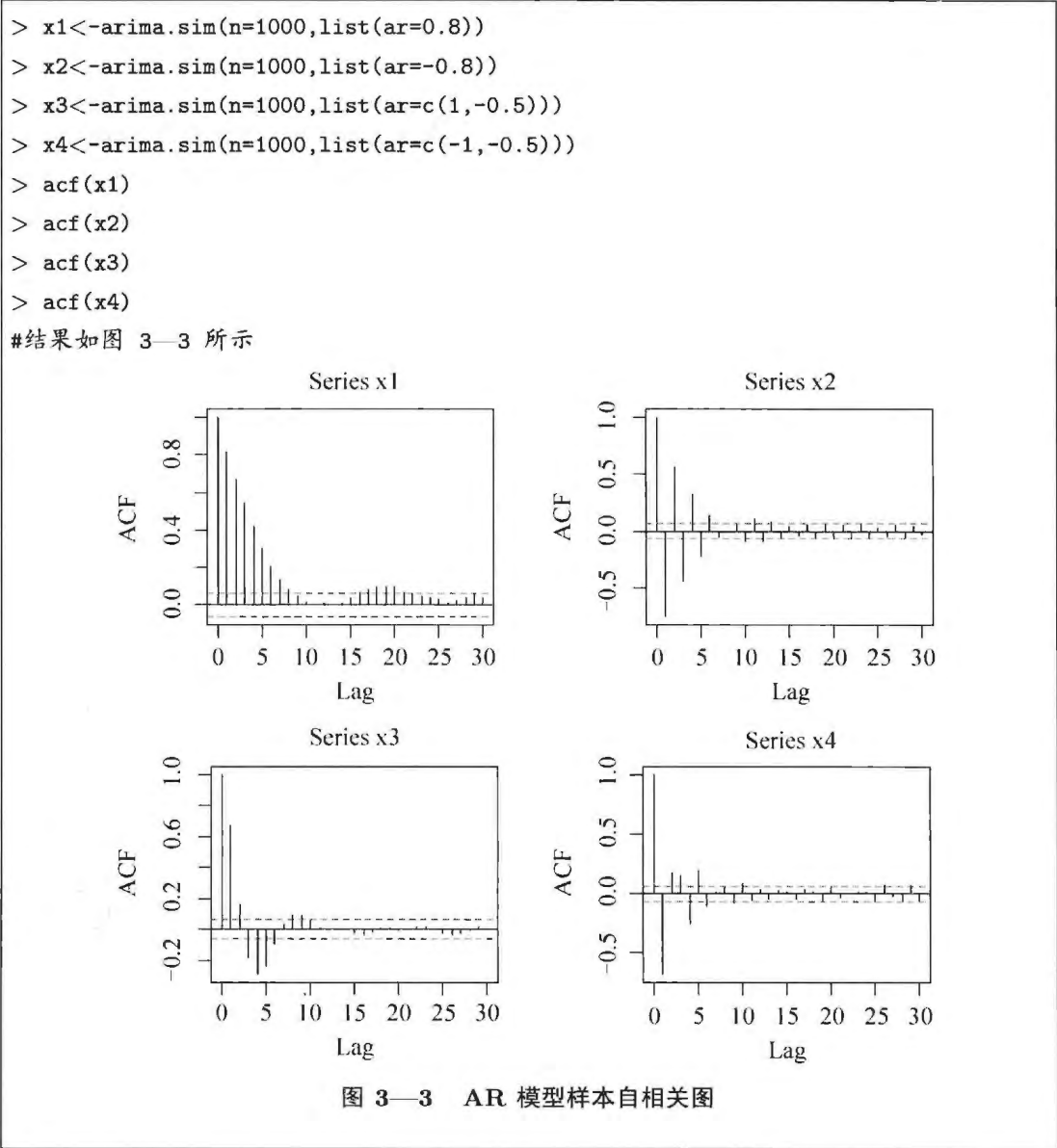
$$(1) x_t = 0.8x_{t-1} + \varepsilon_t$$

$$(2) x_t = -0.8x_{t-1} + \varepsilon_t$$

$$(3) x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

$$(4) x_t = -x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

假定 $\{\varepsilon_t\}$ 为标准正态白噪声序列, 拟合这四个 AR 模型, 相关命令和样本自相关图如下.



从图 3—3 中看到, 这四个平稳 AR 模型, 不论它们是 AR(1) 模型还是 AR(2) 模型, 不论它们的特征根是实根还是复根, 是正根还是负根, 它们的自相关系数都呈现出拖尾性和呈指数衰减到零值附近的性质.

但由于特征根的不同, 它们的自相关系数衰减的方式也不一样. 有的自相关系数是按负指数单调收敛到零 (如模型 (1)); 有的是呈现正负相间地衰减 (如模型 (2)); 还有些自回归系数会呈现出类似于周期性的余弦衰减, 即具有“伪周期”特征 (如模型 (3)), 这些都是平稳模型自相关系数常见的特征.

5. 偏自相关系数

(1) 偏自相关系数的定义. 对于一个平稳 $AR(p)$ 模型, 求出滞后 k 自相关系数 ρ_k 时, 实际上得到的并不是 x_t 与 x_{t-k} 之间单纯的相关关系. 因为 x_t 同时还会受到中间

$k-1$ 个随机变量 $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ 的影响, 而这 $k-1$ 个随机变量又都和 x_{t-k} 具有相关关系, 所以自相关系数 ρ_k 里实际上掺杂了其他变量对 x_t 与 x_{t-k} 的相关影响. 为了能单纯测度 x_{t-k} 对 x_t 的影响, 引进偏自相关系数的概念.

定义 3.3 对于平稳序列 $\{x_t\}$, 所谓滞后 k 偏自相关系数, 就是指在给定中间 $k-1$ 个随机变量 $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ 的条件下, 或者说, 在剔除了中间 $k-1$ 个随机变量 $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ 的干扰之后, x_{t-k} 对 x_t 相关影响的度量. 用数学语言描述就是:

$$\rho_{x_t, x_{t-k} | x_{t-1}, \dots, x_{t-k+1}} = \frac{E[(x_t - \hat{E}x_t)(x_{t-k} - \hat{E}x_{t-k})]}{E[(x_{t-k} - \hat{E}x_{t-k})^2]} \quad (3.20)$$

式中, $\hat{E}x_t = E[x_t | x_{t-1}, \dots, x_{t-k+1}]$, $\hat{E}x_{t-k} = E[x_{t-k} | x_{t-1}, \dots, x_{t-k+1}]$. 这就是滞后 k 偏自相关系数的定义.

(2) 偏自相关系数的计算. 偏自相关系数的定义和回归分析中偏相关系数的定义非常相似. 这启发我们可以从线性回归的角度得到偏自相关系数的另一层含义.

假定 $\{x_t\}$ 为中心化平稳序列, 用过去的 k 期序列值 $x_{t-1}, x_{t-2}, \dots, x_{t-k}$ 对 x_t 作 k 阶自回归拟合, 即

$$x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \dots + \phi_{kk}x_{t-k} + \varepsilon_t \quad (3.21)$$

式中, $E(\varepsilon_t) = 0$, $E(\varepsilon_t x_s) = 0$ ($\forall s < t$). 对 $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ 取条件, 记

$$\hat{E}x_t = E[x_t | x_{t-1}, \dots, x_{t-k+1}], \hat{E}x_{t-k} = E[x_{t-k} | x_{t-1}, \dots, x_{t-k+1}]$$

则

$$\begin{aligned} \hat{E}x_t &= \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \dots + \phi_{k(k-1)}x_{t-k+1} + \phi_{kk}\hat{E}(x_{t-k}) \\ &\quad + E(\varepsilon_t | x_{t-1}, \dots, x_{t-k+1}) \end{aligned} \quad (3.22)$$

已知 $E(\varepsilon_t) = 0$, $E(\varepsilon_t x_s) = 0$ ($\forall s < t$), 所以

$$E(\varepsilon_t | x_{t-1}, \dots, x_{t-k+1}) = E(\varepsilon_t) = 0$$

式 (3.22) 等价于

$$\hat{E}x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \dots + \phi_{k(k-1)}x_{t-k+1} + \phi_{kk}\hat{E}(x_{t-k})$$

则

$$x_t - \hat{E}x_t = \phi_{kk}(x_{t-k} - \hat{E}x_{t-k}) + \varepsilon_t \quad (3.23)$$

在式 (3.23) 等号两边同时乘以 $x_{t-k} - \hat{E}x_{t-k}$, 并求期望:

$$\begin{aligned} E[(x_t - \hat{E}x_t)(x_{t-k} - \hat{E}x_{t-k})] &= \phi_{kk}E[(x_{t-k} - \hat{E}x_{t-k})^2] \\ &\quad + E[\varepsilon_t(x_{t-k} - \hat{E}x_{t-k})] \end{aligned} \quad (3.24)$$

因为 $E(\varepsilon_t x_s) = 0$ ($\forall s < t$), 所以

$$E[\varepsilon_t(x_{t-k} - \hat{E}x_{t-k})] = 0$$

式 (3.24) 等价于

$$E[(x_t - \hat{E}x_t)(x_{t-k} - \hat{E}x_{t-k})] = \phi_{kk}E[(x_{t-k} - \hat{E}x_{t-k})^2]$$

由此得出

$$\phi_{kk} = \frac{E[(x_t - \hat{E}x_t)(x_{t-k} - \hat{E}x_{t-k})]}{E[(x_{t-k} - \hat{E}x_{t-k})^2]} \quad (3.25)$$

式 (3.25) 等号右边的结果正好等于式 (3.20) 所定义的滞后 k 偏自相关系数.

这说明滞后 k 偏自相关系数实际上就等于 k 阶自回归模型第 k 个回归系数 ϕ_{kk} 的值. 根据这个性质容易计算偏自相关系数的值.

在式 (3.21) 等号两边同乘以 x_{t-l} , 并求期望, 得

$$\rho_l = \phi_{k1}\rho_{l-1} + \phi_{k2}\rho_{l-2} + \cdots + \phi_{kk}\rho_{l-k}, \quad \forall l \geq 1$$

取前 k 个方程构成的方程组:

$$\begin{cases} \rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ \vdots \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0 \end{cases} \quad (3.26)$$

该方程组称为 Yule-Walker 方程. 通过解该方程组, 可以得到参数 $(\phi_{k1}, \phi_{k2}, \cdots, \phi_{kk})'$ 的解, 参数向量中最后一个参数的解, 即为滞后 k 偏自相关系数 ϕ_{kk} 的值.

用矩阵形式表示为:

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (3.27)$$

根据线性方程组求解的 Cramer 法则, 有

$$\phi_{kk} = \frac{D_k}{D} \quad (3.28)$$

式中

$$D = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}, \quad D_k = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{vmatrix}$$

D 为式 (3.27) 中系数矩阵的行列式, D_k 是把 D 中的第 k 个列向量换成式 (3.27) 等号右边的自相关系数向量后构成的行列式.

(3) 偏自相关系数的截尾性. 可以证明: 平稳 $AR(p)$ 模型的偏自相关系数具有 p 步截尾性. 所谓 p 步截尾是指, $\phi_{kk} = 0$ ($\forall k > p$). 要证明这一点实际上只要能证明当 $k > p$ 时, $D_k = 0$ 即可.

证明: 对任一 $AR(p)$ 模型

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \quad \forall k > p$$

有如下 Yule-Walker 方程成立:

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_{k-p} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (3.29)$$

记 ξ_i ($i = 1, 2, \dots, p$) 为式 (3.29) 系数矩阵中 p 个列向量, η 为式 (3.29) 中等号右边的自相关系数向量, 即

$$\xi_i = \begin{pmatrix} \rho_{i-1} \\ \rho_{i-2} \\ \vdots \\ \rho_{k-i} \end{pmatrix}, i = 1, 2, \dots, p; \eta = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}$$

则有

$$\eta = \phi_1 \xi_1 + \phi_2 \xi_2 + \cdots + \phi_p \xi_p \quad (3.30)$$

因为 AR(p) 模型限制条件之一是 $\phi_p \neq 0$, 所以向量 η 一定可以表示成向量 ξ_i ($i = 1, 2, \dots, p$) 的非零线性组合.

当 $k > p$ 时

$$D_k = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{p-2} & \cdots & \rho_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_{k-p} & \cdots & \rho_k \end{vmatrix} \quad (3.31)$$

显然 D_k 的前 p 个列向量正好就是 ξ_i ($i = 1, 2, \dots, p$), 而最后一个列向量正好就是向量 η . 根据式 (3.30), 说明在行列式 (3.31) 中最后一个列向量可以用另外 p 个列向量的线性组合表示. 根据行列式的性质, 具有这种线性相关关系的行列式的值一定为零, 即 $D_k = 0$. 由 $D_k = 0$, 等价推出 $\phi_{kk} = 0$.

证毕.

由此证明了 AR(p) 模型偏自相关系数的 p 步截尾性. 这个性质连同前面的自相关系数拖尾性是 AR(p) 模型识别的重要依据.

【例 3—5 续】 考察例 3—5 中四个平稳 AR 模型的偏自相关系数截尾性.

$$(1) x_t = 0.8x_{t-1} + \varepsilon_t$$

$$(2) x_t = -0.8x_{t-1} + \varepsilon_t$$

$$(3) x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

$$(4) x_t = -x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

```
> pacf(x1)
```

```
> pacf(x2)
```

```
> pacf(x3)
```

```
> pacf(x4)
```

```
#结果如图 3—4 所示
```

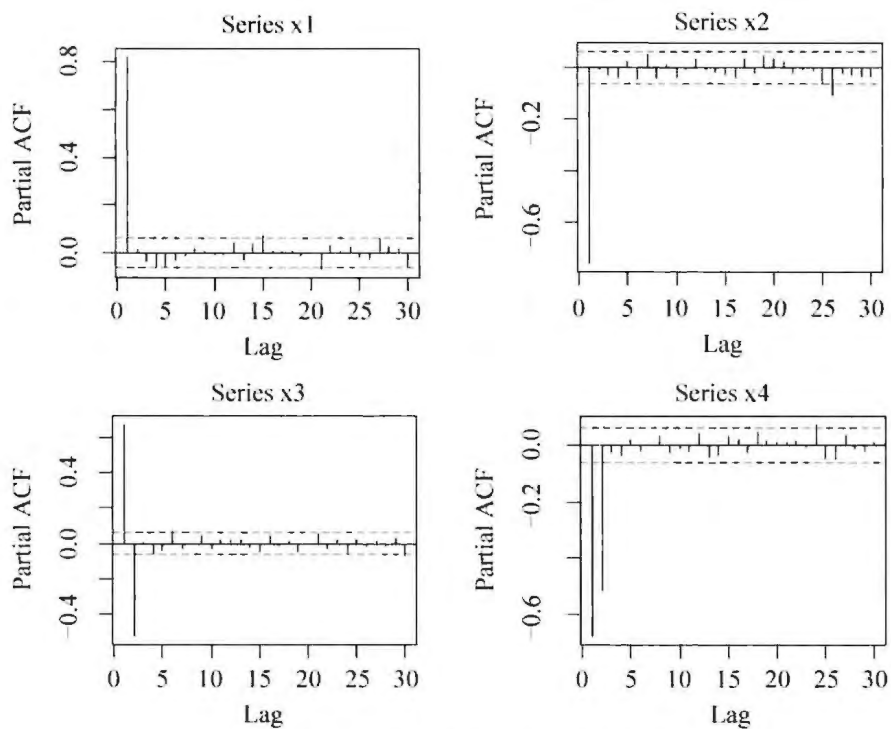


图 3—4 AR 模型样本偏自相关图

根据式 (3.28) 容易算出, AR(1) 模型的偏自相关系数为:

$$\phi_{kk} = \begin{cases} \phi_1, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

AR(2) 模型的偏自相关系数为:

$$\phi_{kk} = \begin{cases} \frac{\phi_1}{1 - \phi_2}, & k = 1 \\ \phi_2, & k = 2 \\ 0, & k \geq 3 \end{cases}$$

所以, 这四个 AR 模型的理论偏自相关系数如表 3—2 所示.

表 3—2

(1) $x_t = 0.8x_{t-1} + \varepsilon_t$	$\phi_{kk} = \begin{cases} 0.8, & k = 1 \\ 0, & k \geq 2 \end{cases}$
(2) $x_t = -0.8x_{t-1} + \varepsilon_t$	$\phi_{kk} = \begin{cases} -0.8, & k = 1 \\ 0, & k \geq 2 \end{cases}$

(3) $x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$	$\phi_{kk} = \begin{cases} \frac{2}{3}, & k = 1 \\ -0.5, & k = 2 \\ 0, & k \geq 3 \end{cases}$
(4) $x_t = -x_{t-1} - 0.5x_{t-2} + \varepsilon_t$	$\phi_{kk} = \begin{cases} -\frac{2}{3}, & k = 1 \\ -0.5, & k = 2 \\ 0, & k \geq 3 \end{cases}$

由于样本的随机性, 样本偏自相关系数不会和理论偏自相关系数一样严格截尾, 但可以看出两个 AR(1) 模型的样本偏自相关系数 1 阶显著不为零, 1 阶之后都近似为零; 而两个 AR(2) 模型的样本偏自相关系数 2 阶显著不为零, 2 阶之后都近似为零. 样本偏自相关图可以直观地验证 AR 模型偏自相关系数截尾性.

3.2.2 MA 模型

一、定义

定义 3.4 具有如下结构的模型称为 q 阶移动平均 (moving average) 模型, 简记为 $MA(q)$:

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \end{cases} \quad (3.32)$$

使用 $MA(q)$ 模型需要满足两个限制条件:

条件一: $\theta_q \neq 0$, 这个限制条件保证了模型的最高阶数为 q .

条件二: $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0 (s \neq t)$. 这个条件保证了随机干扰序列 $\{\varepsilon_t\}$ 为零均值白噪声序列.

通常缺省默认式 (3.32) 的限制条件, 把模型简记为:

$$x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (3.33)$$

当 $\mu = 0$ 时, 模型 (3.32) 称为中心化 $MA(q)$ 模型. 对非中心化 $MA(q)$ 模型只要做一个简单的位移 $y_t = x_t - \mu$, 就可以转化为中心化 $MA(q)$ 模型. 这种中心化运算不会影响序列值之间的相关关系, 所以今后在分析 MA 模型的相关关系时, 常常简化为对它的中心化模型进行分析.

使用延迟算子, 中心化 $MA(q)$ 模型又可以简记为:

$$x_t = \Theta(B)\varepsilon_t$$

式中, $\Theta(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q$, 称为 q 阶移动平均系数多项式.

二、MA 模型的统计性质

1. 常数均值

当 $q < \infty$ 时, $MA(q)$ 模型具有常数均值

$$Ex_t = E(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}) = \mu$$

特别地, 如果该模型为中心化 $MA(q)$ 模型, 该模型均值为零.

2. 常数方差

$$\text{Var}(x_t) = \text{Var}(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}) = (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma_\varepsilon^2$$

3. 自协方差函数只与滞后阶数相关, 且 q 阶截尾

$$\begin{aligned} \gamma_k &= E(x_t x_{t-k}) \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1} - \cdots - \theta_q \varepsilon_{t-k-q})] \\ &= \begin{cases} (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma_\varepsilon^2, & k = 0 \\ \left(-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i} \right) \sigma_\varepsilon^2, & 1 \leq k \leq q \\ 0, & k > q \end{cases} \end{aligned}$$

4. 自相关系数 q 阶截尾

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & k = 0 \\ \frac{-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i}}{1 + \theta_1^2 + \cdots + \theta_q^2}, & 1 \leq k \leq q \\ 0, & k > q \end{cases}$$

容易验证, $MA(1)$ 模型的自相关系数为:

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1}{1 + \theta_1^2}, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

$MA(2)$ 模型的自相关系数为:

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2 \\ 0, & k \geq 3 \end{cases}$$

三、MA 模型的可逆性

【例 3—6】 绘制下列 MA 模型的样本自相关图, 直观考察 MA 模型自相关系数截尾的特性.

(1) $x_t = \varepsilon_t - 2\varepsilon_{t-1}$

(2) $x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$

(3) $x_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$

(4) $x_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$

假定 $\{\varepsilon_t\}$ 为标准正态白噪声序列.

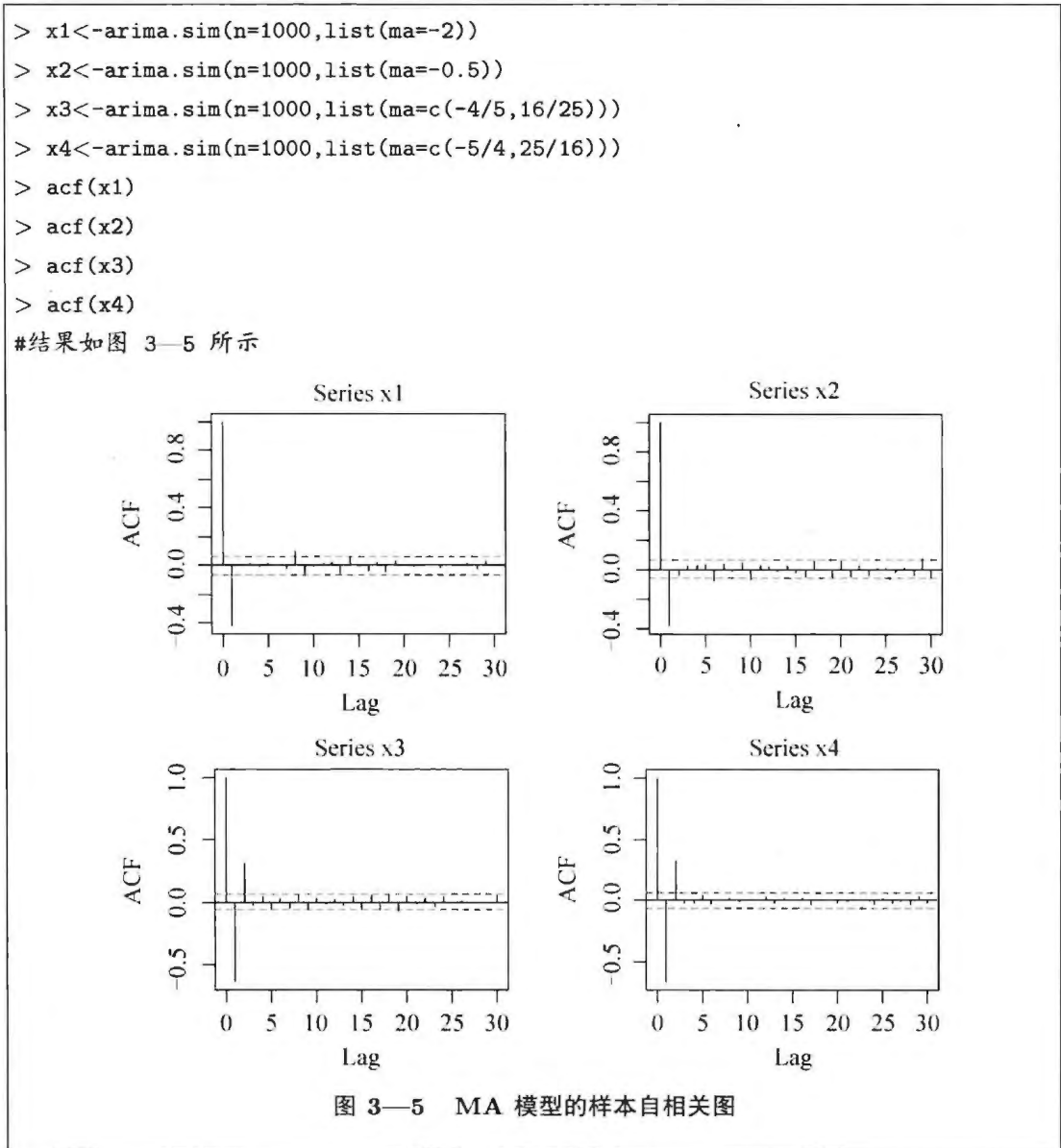


图 3—5 MA 模型的样本自相关图

排除样本随机性的影响, 样本自相关图清晰显示出 MA(1) 模型自相关系数一阶截尾, MA(2) 模型自相关系数二阶截尾的特征.

再次观察例 3—6 中四个 MA 模型的自相关图 (见图 3—5), 可以发现两个不同的 MA(1) 模型:

$$(1) x_t = \varepsilon_t - 2\varepsilon_{t-1}$$

$$(2) x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$$

具有完全相同的样本自相关图. 容易验证它们的理论自相关系数也相等.

$$\rho_k = \begin{cases} -0.4, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

而另两个 MA(2) 模型:

$$(3) x_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$$

$$(4) x_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$$

也出现了同样的情况, 不同的模型却拥有完全相同的自相关系数:

$$\rho_k = \begin{cases} -0.64012, & k = 1 \\ 0.321256, & k = 2 \\ 0, & k \geq 3 \end{cases}$$

产生这种现象的原因就是我们在第 2 章中提到的: 一个自相关系数未必唯一对应一个平稳时间序列模型.

这种自相关系数对应模型的不唯一性会给我们以后的工作带来麻烦. 因为我们将根据样本自相关系数显示出来的特征选择合适的模型拟合序列的发展, 如果自相关系数和模型之间不是一一对应的关系, 就会导致拟合模型和随机序列之间不是一一对应关系.

为了保证一个给定的自相关系数能够对应唯一的 MA 模型, 我们就要给模型增加约束条件. 这个约束条件称为 MA 模型的可逆性 (invertibility) 条件.

1. 可逆的定义

容易验证当两个 MA(1) 模型具有如下结构时, 它们的自相关系数正好相等:

$$\text{模型 1: } x_t = \varepsilon_t - \theta\varepsilon_{t-1} \quad \text{模型 2: } x_t = \varepsilon_t - \frac{1}{\theta}\varepsilon_{t-1}$$

把这两个 MA(1) 模型表示成两个自相关模型形式:

$$\text{模型 1: } \frac{x_t}{1 - \theta B} = \varepsilon_t \quad \text{模型 2: } \frac{x_t}{1 - \frac{1}{\theta}B} = \varepsilon_t$$

显然, 如果 $|\theta| < 1$, 模型 1 收敛, 而模型 2 不收敛; 如果 $|\theta| > 1$, 则模型 2 收敛, 而模型 1 不收敛. 若一个 MA 模型能够表示成收敛的 AR 模型形式, 那么该 MA 模型称为可逆模型. 一个自相关系数唯一对应一个可逆 MA 模型.

2. MA(q) 模型的可逆性条件

与分析 AR(p) 模型的平稳性条件类似, MA(q) 模型可以表示为:

$$\varepsilon_t = \frac{x_t}{\Theta(B)} \tag{3.34}$$

式中, $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ 为移动平均系数多项式. 假定 $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_q}$ 是该系数多项式的 q 个根, 则 $\Theta(B)$ 可以分解成:

$$\Theta(B) = \prod_{k=1}^q (1 - \lambda_k B) \quad (3.35)$$

把式 (3.35) 代入式 (3.34), 得

$$\varepsilon_t = \frac{x_t}{(1 - \lambda_1 B) \cdots (1 - \lambda_q B)} \quad (3.36)$$

式 (3.36) 收敛的充要条件是 $|\lambda_i| < 1$, 等价于 $MA(q)$ 模型的系数多项式的根都在单位圆外, $\left| \frac{1}{\lambda_i} \right| > 1$. 这个条件也称为 $MA(q)$ 模型的可逆性条件.

显然, $MA(q)$ 模型的可逆概念和 $AR(p)$ 模型的平稳概念是完全对偶的. 容易验证, $MA(1)$ 模型可逆的条件是: $-1 < \theta_1 < 1$; $MA(2)$ 模型可逆的条件是: $|\theta_2| < 1$, 且 $\theta_2 \pm \theta_1 < 1$.

3. 逆函数的递推公式

如果一个 $MA(q)$ 模型满足可逆性条件, 它就可以写成如下两种等价形式:

$$\begin{cases} \Theta(B)\varepsilon_t = x_t \\ \varepsilon_t = I(B)x_t \end{cases} \quad \begin{matrix} (a) \\ (b) \end{matrix}$$

把式 (b) 代入式 (a), 得

$$\Theta(B)I(B)x_t = x_t$$

展开上式, 得

$$\left(1 - \sum_{k=1}^q \theta_k B^k\right) \left(1 + \sum_{j=0}^{\infty} I_j B^j\right) x_t = x_t$$

和 Green 函数的递推公式完全类似, 由待定系数法容易得到逆函数的递推公式为:

$$\begin{cases} I_0 = 1 \\ I_j = \sum_{k=1}^j \theta'_k I_{j-k}, \quad j \geq 1 \end{cases} \quad (3.37)$$

式中

$$\theta'_k = \begin{cases} \theta_k, & k \leq q \\ 0, & k > q \end{cases}$$

【例 3—6 续 (1)】 考察例 3—6 中四个 MA 模型的可逆性, 并写出可逆 MA 模型的逆转形式.

$$(1) x_t = \varepsilon_t - 2\varepsilon_{t-1}$$

$$(2) x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$$

$$(3) x_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$$

$$(4) x_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$$

MA 模型是否可逆如表 3—3 所示.

表 3—3

模型	条件	结论
(1) $x_t = \varepsilon_t - \varepsilon_{t-1}$	$ \theta_1 = 2 > 1$	不可逆
(2) $x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$	$ \theta_1 = 0.5 < 1$	可逆
(3) $x_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$	$ \theta_2 = \frac{16}{25} < 1$ $\theta_2 + \theta_1 = -\frac{16}{25} + \frac{4}{5} = \frac{4}{25} < 1$ $\theta_2 - \theta_1 = -\frac{16}{25} - \frac{4}{5} = -\frac{36}{25} < 1$	可逆
(4) $x_t = \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2}$	$ \theta_2 = \frac{25}{16} > 1$	不可逆

模型 (2): $x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ 的逆函数为:

$$\begin{cases} I_0 = 1 \\ I_j = 0.5^j, \quad j \geq 1 \end{cases}$$

则该模型的逆转形式为:

$$\varepsilon_t = \sum_{j=0}^{\infty} 0.5^j x_{t-j}$$

模型 (3): $x_t = \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2}$, 根据逆函数的递推公式, 并根据该模型特有的 $\theta_2 = -\theta_1^2$ 的性质, 整理之后该模型的逆函数为:

$$I_k = \begin{cases} (-1)^n \theta_1^k, & k = 3n \text{ 或 } 3n + 1 \quad (n = 0, 1, \dots) \\ 0, & k = 3n + 2 \end{cases}$$

则该模型的逆转形式为:

$$\varepsilon_t = \sum_{n=0}^{\infty} (-1)^n 0.8^{3n} x_{t-3n} + \sum_{n=0}^{\infty} (-1)^n 0.8^{3n+1} x_{t-3n-1}$$

四、偏自相关系数拖尾

对于一个可逆 $MA(q)$ 模型, 可以等价写成 $AR(\infty)$ 模型形式:

$$I(B)x_t = \varepsilon_t$$

式中

$$\begin{cases} I_0 = 1 \\ I_j = \sum_{k=1}^j \theta'_k I_{j-k}, \quad j \geq 1 \end{cases}$$

$AR(p)$ 模型偏自相关系数 p 阶截尾, 所以可逆 $MA(q)$ 模型偏自相关系数 ∞ 截尾, 即具有偏自相关系数拖尾性.

一个可逆 $MA(q)$ 模型一定对应一个与它具有相同自相关系数和偏自相关系数的不可逆 $MA(q)$ 模型, 这个不可逆 $MA(q)$ 模型也同样具有偏自相关系数拖尾性.

【例 3—7】求 $MA(1)$ 模型偏自相关系数的表达式.

假设 $MA(1)$ 模型的表达式为 $x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$. 根据偏自相关系数的定义, 我们知道延迟 k 阶偏自相关系数 ϕ_{kk} 是方程组

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \cdots, k$$

的最后一个系数, 则对 $j = 1, 2, \cdots, k$ 依次求解方程, 得

$$\begin{aligned} \phi_{11} &= \rho_1 = \frac{-\theta_1}{1 + \theta_1^2} \\ \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{-\rho_1^2}{1 - \rho_1^2} = \frac{-\theta_1^2}{1 + \theta_1^2 + \theta_1^4} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{vmatrix}} \\ &= \frac{\rho_1^3}{1 - 2\rho_1^2} = \frac{-\theta_1^3}{1 + \theta_1^2 + \theta_1^4 + \theta_1^6} \end{aligned}$$

依此类推, 可以得到 $MA(1)$ 模型任意 k 阶偏自相关系数 ϕ_{kk} 的通解为:

$$\phi_{kk} = \frac{-\theta_1^k}{\sum_{j=0}^k \theta_1^{2j}}, \quad k \geq 1$$

$MA(1)$ 模型任意 k 阶偏自相关系数 ϕ_{kk} 的通解形式也说明了 $MA(1)$ 模型偏自相关系数拖尾.

【例 3—6 续 (2)】绘制下列 MA 模型的偏自相关系数图, 直观考察 MA 模型偏自相关系数的拖尾性.

$$\begin{aligned} (1) \quad x_t &= \varepsilon_t - 2\varepsilon_{t-1} & (2) \quad x_t &= \varepsilon_t - 0.5\varepsilon_{t-1} \\ (3) \quad x_t &= \varepsilon_t - \frac{4}{5}\varepsilon_{t-1} + \frac{16}{25}\varepsilon_{t-2} & (4) \quad x_t &= \varepsilon_t - \frac{5}{4}\varepsilon_{t-1} + \frac{25}{16}\varepsilon_{t-2} \end{aligned}$$

假定 $\{\varepsilon_t\}$ 为标准正态白噪声序列.

```
> pacf(x1)
> pacf(x2)
> pacf(x3)
> pacf(x4)
#结果如图 3—6 所示
```

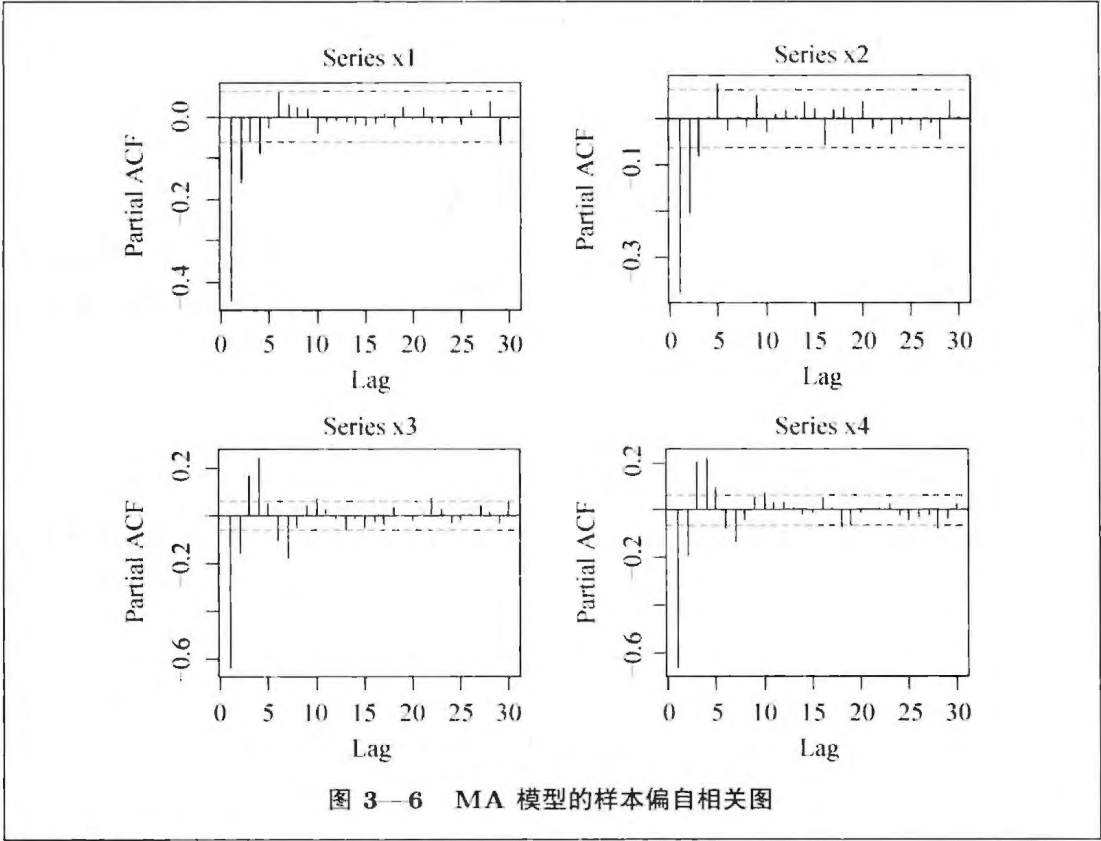


图 3—6 MA 模型的样本偏自相关图

3.2.3 ARMA 模型

一、定义

定义 3.5 把具有如下结构的模型称为自回归移动平均模型, 简记为 ARMA(p, q):

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (3.38)$$

若 $\phi_0 = 0$, 该模型称为中心化 ARMA(p, q) 模型. 缺省默认条件, 中心化 ARMA(p, q) 模型可以简写为:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (3.39)$$

默认条件与 AR 模型、MA 模型相同.

引进延迟算子, ARMA(p, q) 模型简记为:

$$\Phi(B)x_t = \Theta(B)\varepsilon_t$$

式中

$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, 为 p 阶自回归系数多项式.

$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, 为 q 阶移动平均系数多项式.

显然

当 $q = 0$ 时, ARMA(p, q) 模型就退化成了 AR(p) 模型;

当 $p = 0$ 时, ARMA(p, q) 模型就退化成了 MA(q) 模型.

所以, AR(p) 模型和 MA(q) 模型实际上是 ARMA(p, q) 模型的特例, 它们都统称为 ARMA 模型. 而 ARMA(p, q) 模型的统计性质也正是 AR(p) 模型和 MA(q) 模型统计性质的有机组合.

二、平稳条件与可逆条件

对于一个 ARMA(p, q) 模型, 令 $z_t = \Theta(B)\varepsilon_t$, 显然 $\{z_t\}$ 是一个均值为零、方差为 $(1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2$ 的平稳序列. 于是 ARMA(p, q) 模型可以改写为如下形式:

$$\Phi(B)x_t = z_t$$

类似于 AR(p) 模型平稳性的分析, 容易推导出 ARMA(p, q) 模型的平稳条件是: $\Phi(B) = 0$ 的根都在单位圆外. 也就是说, ARMA(p, q) 模型的平稳性完全由其自回归部分的平稳性所决定.

同理, 可以推导出 ARMA(p, q) 模型的可逆条件和 MA(q) 模型的可逆条件完全相同: 当 $\Theta(B) = 0$ 的根都在单位圆外时, ARMA(p, q) 模型可逆.

当 $\Phi(B) = 0, \Theta(B) = 0$ 的根都在单位圆外时, ARMA(p, q) 模型称为平稳可逆模型, 这是一个由它的自相关系数唯一识别的模型.

三、传递形式与逆转形式

对于一个平稳可逆 ARMA(p, q) 模型, 它的传递形式为:

$$x_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j}$$

式中, $\{G_0, G_1, G_2, \cdots\}$ 为 Green 函数.

通过待定系数法, 容易得到 ARMA(p, q) 模型场合下 Green 函数的递推公式为:

$$\begin{cases} G_0 = 1 \\ G_k = \sum_{j=1}^k \phi'_j G_{k-j} - \theta'_k, & k \geq 1 \end{cases} \quad (3.40)$$

式中

$$\phi'_j = \begin{cases} \phi_j, & 1 \leq j \leq p \\ 0, & j > p \end{cases}, \quad \theta'_k = \begin{cases} \theta_k, & 1 \leq k \leq q \\ 0, & k > q \end{cases}$$

同理, 可以得到 ARMA(p, q) 模型的逆转形式为:

$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)x_t = \sum_{j=0}^{\infty} I_j x_{t-j}$$

式中, $\{I_1, I_2, \dots\}$ 为逆函数.

通过待定系数法容易得到逆函数的递推公式为:

$$\begin{cases} I_0 = 1 \\ I_k = \sum_{j=1}^k \theta'_j I_{k-j} - \phi'_k, k \geq 1 \end{cases} \quad (3.41)$$

式中, θ'_j 和 ϕ'_k 的定义同上.

四、ARMA(p, q) 模型的统计性质

1. 均值

对于一个非中心化平稳可逆的 ARMA(p, q) 模型

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

两边同求均值, 有

$$Ex_t = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}$$

2. 自协方差函数

$$\begin{aligned} \gamma(k) &= E(x_t x_{t+k}) \\ &= E \left[\left(\sum_{i=0}^{\infty} G_i \varepsilon_{t-i} \right) \left(\sum_{j=0}^{\infty} G_j \varepsilon_{t+k-j} \right) \right] \\ &= E \left[\sum_{i=0}^{\infty} G_i \sum_{j=0}^{\infty} G_j \varepsilon_{t-i} \varepsilon_{t+k-j} \right] \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} G_i G_{i+k} \end{aligned}$$

3. 自相关系数

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2}$$

根据自相关系数的表达式很容易判断 ARMA(p, q) 模型的自相关系数不截尾. 这和 ARMA(p, q) 模型可以转化为无穷阶移动平均模型的性质是一致的. 同理, 根据 ARMA(p, q) 模型可以转化为无穷阶自回归模型, 可以判断它的偏自相关系数也不截尾.

【例 3—8】 拟合 ARMA(1,1) 模型: $x_t - 0.5x_{t-1} = \varepsilon_t - 0.8\varepsilon_{t-1}$, 并直观地考察该模型自相关系数和偏自相关系数的拖尾性. 假定 $\{\varepsilon_t\}$ 为标准正态白噪声序列.

```
> x<-arima.sim(n=1000,list(ar=0.5,ma=-0.8))
> acf(x)
> pacf(x)
```

#结果如图 3—7 所示

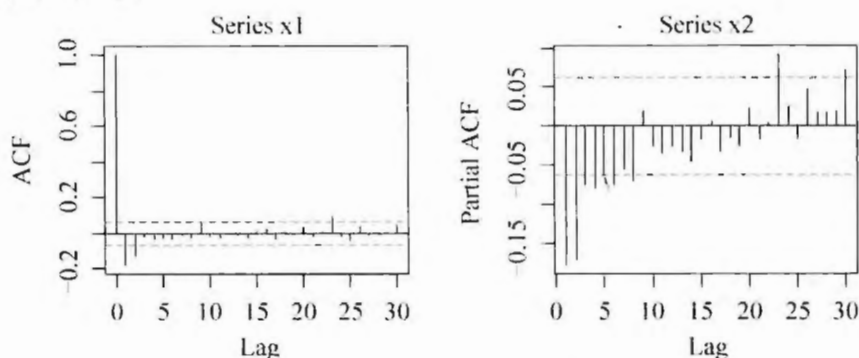


图 3—7 ARMA 模型的样本相关图。

综合考察 $AR(p)$ 模型、 $MA(q)$ 模型和 $ARMA(p, q)$ 模型自相关系数和偏自相关系数的性质, 我们可以总结出如下规律 (表 3—4)。

表 3—4

模型	自相关系数	偏自相关系数
$AR(p)$	拖尾	p 阶截尾
$MA(q)$	q 阶截尾	拖尾
$ARMA(p, q)$	拖尾	拖尾

3.3 平稳序列建模

3.3.1 建模步骤

假如某个观察值序列通过序列预处理可以判定为平稳非白噪声序列, 就可以利用 ARMA 模型对该序列建模. 建模的基本步骤如图 3—8 所示.

(1) 求出该观察值序列的样本自相关系数 (ACF) 和样本偏自相关系数 (PACF) 的值.

(2) 根据样本自相关系数和偏自相关系数的性质, 选择阶数适当的 $ARMA(p, q)$ 模型进行拟合.

(3) 估计模型中未知参数的值.

(4) 检验模型的有效性. 如果拟合模型未通过检验, 回到步骤 (2), 重新选择模型拟合.

(5) 模型优化. 如果拟合模型通过检验, 仍然回到步骤 (2), 充分考虑各种可能, 建立多个拟合模型, 从所有通过检验的拟合模型中选择最优模型.

(6) 利用拟合模型, 预测序列将来的走势.

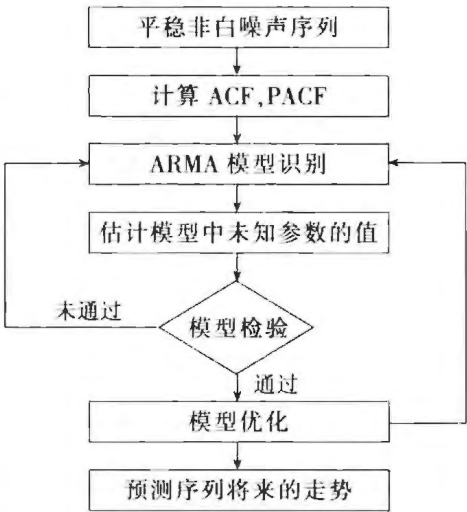


图 3—8 建模步骤

3.3.2 样本自相关系数与偏自相关系数

我们是通过考察平稳序列样本自相关系数和偏自相关系数的性质选择适合的模型拟合观察值序列的, 所以模型拟合的第一步是要根据观察值序列的取值求出该序列的样本自相关系数 $\{\hat{\rho}_k, 0 < k < n\}$ 和样本偏自相关系数 $\{\hat{\phi}_{kk}, 0 < k < n\}$ 的值.

样本自相关系数可以根据以下公式求得

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad \forall 0 < k < n$$

样本偏自相关系数可以利用样本自相关系数的值, 根据以下公式求得

$$\hat{\phi}_{kk} = \frac{\hat{D}_k}{\hat{D}}, \quad \forall 0 < k < n$$

式中

$$\hat{D} = \begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2} \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & 1 \end{vmatrix}, \quad \hat{D}_k = \begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_2 \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_k \end{vmatrix}$$

3.3.3 模型识别

计算出样本自相关系数和偏自相关系数的值之后, 就要根据它们表现出来的性质,

选择适当的 ARMA 模型拟合观察值序列. 这个过程实际上就是要根据样本自相关系数和偏自相关系数的性质估计自相关阶数 $\hat{p}$ 和移动平均阶数 $\hat{q}$, 因此, 模型识别过程也称为模型定阶过程.

ARMA 模型定阶的基本原则如表 3—5 所示.

表 3—5

$\hat{\rho}_k$	$\hat{\phi}_{kk}$	模型定阶
拖尾	p 阶截尾	AR(p)模型
q 阶截尾	拖尾	MA(q) 模型
拖尾	拖尾	ARMA(p, q) 模型

但是在实践中, 这个定阶原则在操作上具有一定的困难. 由于样本的随机性, 样本的相关系数不会呈现出理论截尾的完美情况, 本应截尾的样本自相关系数或偏自相关系数仍会呈现出小值振荡. 同时, 由于平稳时间序列通常都具有短期相关性, 随着延迟阶数 $k \rightarrow \infty$, $\hat{\rho}_k$ 与 $\hat{\phi}_{kk}$ 都会衰减至零值附近作小值波动. 这种现象促使我们必须思考, 当样本自相关系数或偏自相关系数在延迟若干阶之后衰减为小值波动时, 什么情况下该看做相关系数截尾, 什么情况下该看做相关系数在延迟若干阶之后正常衰减到零值附近作拖尾波动呢?

这实际上没有绝对的标准, 很大程度上依靠分析人员的主观经验. 但样本自相关系数和偏自相关系数的近似分布可以帮助缺乏经验的分析人员做出尽量合理的判断.

Jankins 和 Watts 于 1968 年证明

$$E(\hat{\rho}_k) = \left(1 - \frac{k}{n}\right) \rho_k$$

也就是说, 该样本自相关系数是总体自相关系数的有偏估计值. 当 k 足够大时, 根据平稳序列自相关系数呈负指数衰减, 有 $\rho_k \rightarrow 0$.

根据 Bartlett 公式计算样本自相关系数的方差

$$\text{Var}(\hat{\rho}_k) \approx \frac{1}{n} \sum_{m=-j}^j \hat{\rho}_m^2 = \frac{1}{n} \left(1 + 2 \sum_{m=1}^j \hat{\rho}_m^2\right), \quad k > j$$

当样本容量 n 充分大时, 样本自相关系数近似服从正态分布:

$$\hat{\rho}_k \overset{\circ}{\sim} N\left(0, \frac{1}{n}\right)$$

Quenouille 证明, 样本偏自相关系数也同样近似服从这个正态分布:

$$\hat{\phi}_{kk} \overset{\circ}{\sim} N\left(0, \frac{1}{n}\right)$$

根据正态分布的性质, 有

$$\begin{aligned} \Pr\left(-\frac{2}{\sqrt{n}} \leq \hat{\rho}_k \leq \frac{2}{\sqrt{n}}\right) &\geq 0.95 \\ \Pr\left(-\frac{2}{\sqrt{n}} \leq \hat{\phi}_{kk} \leq \frac{2}{\sqrt{n}}\right) &\geq 0.95 \end{aligned}$$

所以可以利用 2 倍标准差范围辅助判断。

如果样本自相关系数或偏自相关系数在最初的 d 阶明显超过 2 倍标准差范围, 而后几乎 95% 的自相关系数都落在 2 倍标准差的范围以内, 而且由非零自相关系数衰减为小值波动的过程非常突然, 这时, 通常视为自相关系数截尾, 截尾阶数为 d 。

如果有超过 5% 的样本自相关系数落入 2 倍标准差范围之外, 或者由显著非零的自相关系数衰减为小值波动的过程比较缓慢或者非常连续, 这时, 通常视为自相关系数不截尾。

【例 3—9】选择合适的模型拟合 1950—2008 年我国邮路及农村投递线路每年新增里程数序列 (数据见表 A1—8)。

• 读入数据, 并绘制时序图

```
> a<-read.table("E:/R/data/file8.csv",sep=",",header=T)
> x<-ts(a$kilometer,start=1950)
> plot(x)    #结果如图 3—9 所示
```

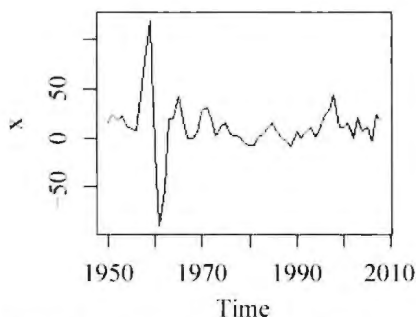


图 3—9 我国邮路及农村投递线路每年新增里程数序列时序图

• 白噪声检验

```
> for(i in 1: 2) print(Box.test(x,type="Ljung-Box",lag=6*i))
```

Box-Ljung test

data: x

X-squared=37.7538, df=6, p-value=1.255e-06

Box-Ljung test

data: x

X-squared=44.6196, df=12, p-value=1.197e-05

• 绘制自相关图和偏自相关图

```
> acf(x)    #结果如图 3—10 所示
```

```
> pacf(x)   #结果如图 3—11 所示
```

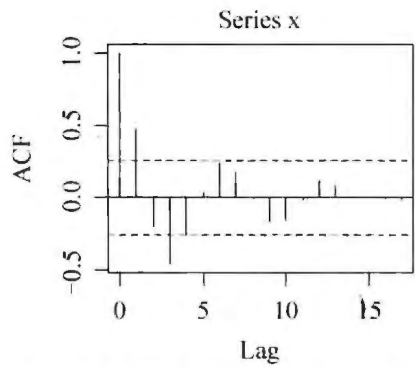


图 3—10 我国邮路及农村投递线路每年新增里程数序列自相关图

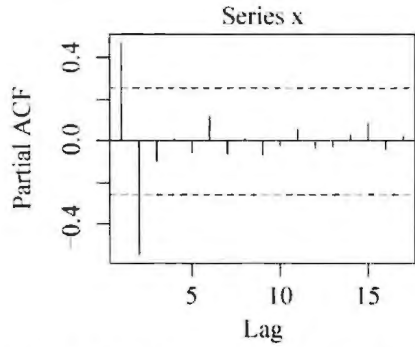


图 3—11 我国邮路及农村投递线路每年新增里程数序列偏自相关图

时序图图 3—9 显示, 序列没有显著非平稳特征. 白噪声检验显示序列值彼此之间蕴涵着相关关系, 为非白噪声序列.

图 3—10 显示, 除了延迟 1~3 阶的自相关系数在 2 倍标准差范围之外, 其他阶数的自相关系数都在 2 倍标准差范围内波动. 根据自相关系数的这个特点可以判断该序列具有短期相关性, 进一步确定序列平稳.

再进一步考察自相关系数衰减到零的过程, 可以看到有明显的正弦波动轨迹, 这说明自相关系数衰减到零不是一个突然的过程, 而是一个连续渐变的过程, 这是自相关系数拖尾的典型特征, 我们可以把拖尾特征形象地描述为“坐着滑梯落水”.

最后考察偏自相关系数衰减到零的过程, 除了 1~2 阶偏自相关系数在 2 倍标准差范围之外, 其他阶数的偏自相关系数都在 2 倍标准差范围内, 这是一个偏自相关系数 2 阶截尾的典型特征, 我们可以把这种截尾特征形象地描述为“2 阶之后高台跳水, 溅起水花点点”.

本例中, 根据自相关系数拖尾, 偏自相关系数 2 阶截尾属性, 我们可以初步确定拟合模型为 AR(2) 模型.

【例 3—10】 选择合适的 ARMA 模型拟合美国科罗拉多州某一加油站连续 57 天的 OVERSHORT 序列 (数据见表 A1—9).

- 读入数据, 并绘制时序图

```
> overshoot<-read.table("E:/R/data/file9.csv",sep=",",header=T)
> overshoot<-ts(overshoot)
> plot(overshoot)    #结果如图 3—12 所示
```

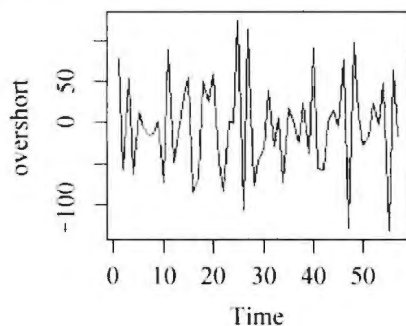


图 3—12 OVERSHORT 序列时序图

- 白噪声检验

```
> for(i in 1: 2) print(Box.test(overshoot,type="Ljung-Box",lag=6*i))
```

Box-Ljung test

data: overshoot

X-squared=20.2386, df=6, p-value=0.002511

Box-Ljung test

data: overshoot

X-squared=31.3659, df=12, p-value=0.001732

- 绘制自相关图和偏自相关图

```
> acf(overshoot)    #结果如图 3—13 所示
> pacf(overshoot)   #结果如图 3—14 所示
```

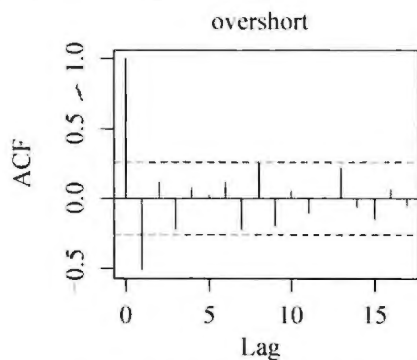


图 3—13 OVERSHORT 序列自相关图

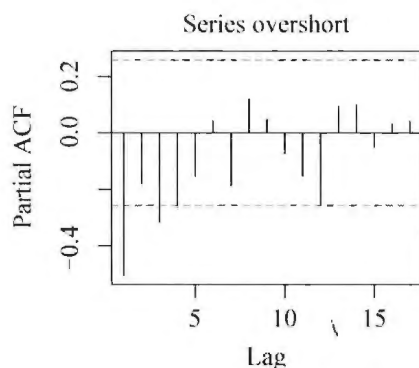


图 3—14 OVERSHORT 序列偏自相关图

时序图显示序列没有显著的非平稳特征。

自相关图显示除了延迟 1 阶的自相关系数在 2 倍标准差范围之外, 其他阶数的自相关系数都在 2 倍标准差范围内波动。根据自相关系数的这个特点可以判断该序列具有短期相关性, 进一步确定序列平稳。同时, 可以认为该序列自相关系数 1 阶截尾。

偏自相关系数显示出非截尾的性质。综合该序列自相关系数和偏自相关系数的性质, 为拟合模型定阶为 MA(1)。

【例 3—11】选择合适的 ARMA 模型拟合 1880—1985 年全球气表平均温度改变值差分序列 (全球气表平均温度改变值序列见表 A1—10)。

• 读入数据, 并绘制时序图

```
> b<-read.table("E:/R/data/file10.csv",sep=";",header=T)
> dif_x<-ts(diff(b$change_temp),start=1880)
> plot(dif_x) #结果如图 3—15 所示
```

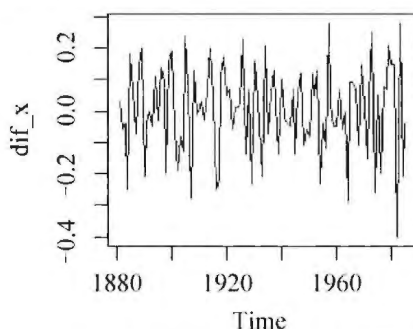


图 3—15 全球气表平均温度改变值差分序列时序图

• 白噪声检验

```
> for(i in 1: 2) print(Box.test(dif_x,type="Ljung-Box",lag=6*i))
```

Box-Ljung test

data: dif_x

X-squared=17.4362, df=6, p-value=0.007807

Box-Ljung test

data: dif_x

X-squared=26.2659, df=12, p-value=0.009841

• 绘制自相关图和偏自相关图

> acf(dif_x) #结果如图 3—16 所示

> pacf(dif_x) #结果如图 3—17 所示

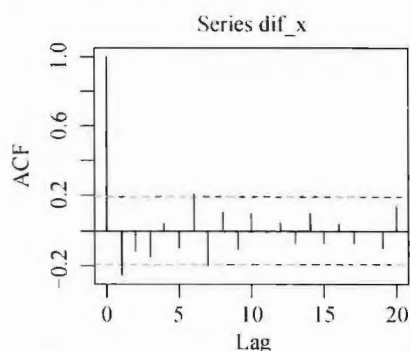


图 3—16 全球气表平均温度改变值差分序列自相关图

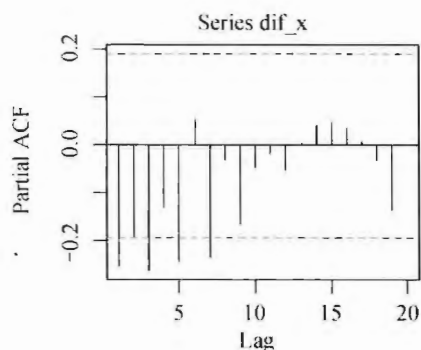


图 3—17 全球气表平均温度改变值差分序列偏自相关图

自相关系数和偏自相关系数均显示出不截尾的性质, 因此可以尝试使用 ARMA(1, 1) 拟合该序列。

为了尽量避免因个人经验不足而导致的模型识别不准确的问题, R 提供了 auto.arima 函数。该函数基于信息量最小原则自动识别模型阶数, 并给出该模型的参数估计值。要使用 auto.arima 函数, 需要先下载 zoo 和 forecast 程序包, 并用 library 调用这两个程序包。auto.arima 函数的命令格式如下:

```
auto.arima(x, max.p=5, max.q=, ic=)
```

式中:

-x: 需要定阶的序列名。

-max.p: 自相关系数最高阶数, 不特殊指定的话, 系统默认值为 5.
 -max.q: 移动平均系数最高阶数, 不特殊指定的话, 系统默认值为 5.
 -ic: 指定信息量准则.ic 有"aicc","aic" 和"bic"三个选项. 系统默认 AIC 准则.

有时选择不同的信息量准则会导致模型的识别阶数不同. 有时系统指定的模型阶数会高于真实阶数.

将例 3—9 至例 3—11 用 auto.arima 函数定阶, 可以发现这三个例子系统自动定阶结果与我们根据自相关图和偏自相关图定阶的结果一致.

```
> library(zoo)
> library(forecast)
• 例 3—9 系统自动定阶
> auto.arima(x)
Series:x
ARIMA(2, 0, 0) with non-zero mean

Coefficients:
          ar1          ar2  intercept
          0.7185        -0.5294         11.0226
s.e.        0.1083         0.1067         3.0906

sigma^2 estimated as 365.2: log likelihood=-258.23
AIC=524.46   AICc=525.2   BIC=532.77
• 例 3—10 系统自动定阶
> auto.arima(overshort)
Series: overshort
ARIMA(0,0,1) with non-zero mean

Coefficients:
          ma1          intercept
          -0.8477         -4.7945
s.e.        0.1206         1.0252

sigma^2 estimated as 2020: log likelihood=-298.42
AIC=602.84  AICc=603.29  BIC=608.97
• 例 3—11 系统自动定阶
> auto.arima(dif_x)
Series:dif_x
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1          ma1          intercept
          0.3926        -0.8867         0.0053
```

s.e.	0.1180	0.0604	0.0024
sigma^2 estimated as 0.01541: log likelihood=69.66			
AIC=-131.32	AICc=-130.92	BIC=-120.71	

3.3.4 参数估计

选择好拟合模型之后, 下一步就是要利用序列的观察值确定该模型的口径, 即估计模型中未知参数的值.

对于一个非中心化 ARMA(p, q) 模型, 有

$$x_t = \mu \frac{\Theta_q(B)}{\Phi_p(B)} \varepsilon_t$$

式中 $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

$$\Theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

$$\Phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

该模型共含有 $p + q + 2$ 个未知参数: $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \mu, \sigma_\varepsilon^2$.

参数 μ 是序列均值, 通常采用矩估计方法, 用样本均值估计总体均值即可得到它的估计值:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

对原序列中心化, 有

$$y_t = x_t - \bar{x}$$

原 $p + q + 2$ 个待估参数减少为 $p + q + 1$ 个: $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_\varepsilon^2$. 对这 $p + q + 1$ 个未知参数的估计方法有三种: 矩估计、极大似然估计和最小二乘估计.

一、矩估计

运用 $p + q$ 个样本自相关系数估计总体自相关系数

$$\begin{cases} \rho_1(\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q) = \hat{\rho}_1 \\ \vdots \\ \rho_{p+q}(\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q) = \hat{\rho}_{p+q} \end{cases}$$

从中解出的参数值 $\hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\theta}_1, \cdots, \hat{\theta}_q$ 就是 $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q$ 的矩估计.

用序列样本方差估计序列总体方差

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

在 ARMA(p, q) 模型两边同时求方差, 整理得到 σ_ε^2 关于 σ_x^2 的函数形式

$$\sigma_\varepsilon^2 = \frac{1 + \phi_1^2 + \cdots + \phi_p^2}{1 + \theta_1^2 + \cdots + \theta_q^2} \sigma_x^2 \quad (3.42)$$

把 $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ 及 $\hat{\sigma}_x^2$ 代入式 (3.42), 即得到白噪声序列方差的矩估计

$$\hat{\sigma}_\varepsilon^2 = \frac{1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_p^2}{1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2} \hat{\sigma}_x^2$$

【例 3—12】求 AR(2) 模型 $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$ 中未知参数 ϕ_1, ϕ_2 的矩估计.

根据 Yule-Walker 方程, 有

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 \\ \rho_2 = \phi_1 \rho_1 + \phi_2 \end{cases}$$

则

$$\hat{\phi}_1 = \frac{1 - \hat{\rho}_2}{1 - \hat{\rho}_1^2} \hat{\rho}_1, \quad \hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}$$

【例 3—13】求 MA(1) 模型 $x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ 中未知参数 θ_1 的矩估计.

根据 MA(1) 模型自协方差函数的性质, 有

$$\begin{cases} \gamma_0 = (1 + \theta_1^2) \sigma_\varepsilon^2 \\ \gamma_1 = -\theta_1 \sigma_\varepsilon^2 \end{cases} \Rightarrow \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta_1}{1 + \theta_1^2}$$

解一元二次方程

$$\theta_1^2 \rho_1 + \theta_1 + \rho_1 = 0$$

得

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

考虑 MA(1) 模型的可逆性条件 $|\theta_1| < 1$, 可得到未知参数的唯一解:

$$\hat{\theta}_1 = \frac{-1 + \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

【例 3—14】求 ARMA(1,1) 模型 $x_t = \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ 中未知参数 ϕ_1, θ_1 的矩估计.

根据 ARMA 模型 Green 函数的递推公式, 可以确定该 ARMA(1,1) 模型的 Green 函数为:

$$G_0 = 1$$

$$G_k = (\phi_1 - \theta_1) \phi_1^{k-1}, \quad k = 1, 2, \dots$$

推导出

$$\begin{cases} \gamma_0 = \sum_{k=0}^{\infty} G_k^2 \sigma_{\varepsilon}^2 = \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi_1^2} \sigma_{\varepsilon}^2 \\ \gamma_1 = \sum_{k=0}^{\infty} G_k G_{k+1} \sigma_{\varepsilon}^2 = \frac{(\phi_1 - \theta_1)(1 - \theta_1\phi_1)}{1 - \phi_1^2} \sigma_{\varepsilon}^2 \\ \gamma_2 = \sum_{k=0}^{\infty} G_k G_{k+2} \sigma_{\varepsilon}^2 = \phi_1 \gamma_1 \end{cases}$$

则

$$\begin{cases} \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\phi_1 - \theta_1)(1 - \theta_1\phi_1)}{1 + \theta_1^2 - 2\theta_1\phi_1} \\ \rho_2 = \phi_1 \rho_1 \end{cases} \quad (3.43)$$

整理方程组 (3.43), 得

$$\begin{cases} \theta_1^2 - \frac{1 + \phi_1^2 - 2\rho_2}{\phi_1 - \rho_1} \theta_1 + 1 = 0 \\ \phi_1 = \frac{\rho_2}{\rho_1} \end{cases}$$

考虑可逆条件: $|\theta_1| < 1$, 得到未知参数矩估计的唯一解

$$\begin{cases} \hat{\phi}_1 = \frac{\hat{\rho}_2}{\hat{\rho}_1} \\ \hat{\theta}_1 = \begin{cases} \frac{c + \sqrt{c^2 - 4}}{2}, & c \leq -2 \\ \frac{c - \sqrt{c^2 - 4}}{2}, & c \geq 2 \end{cases}, \quad c = \frac{1 + \hat{\phi}_1^2 - 2\hat{\rho}_2}{\hat{\phi}_1 - \hat{\rho}_1} \end{cases}$$

矩估计方法, 尤其是低阶 ARMA 模型场合下的矩估计方法具有计算量小、估计思想简单直观, 且不需要假设总体分布的优点. 但是在这种估计方法中只用到了 $p+q$ 个样本自相关系数, 即样本二阶矩的信息, 观察值序列中的其他信息都被忽略了. 这导致矩估计方法是一种比较粗糙的估计方法, 它的估计精度一般不高, 因此它常用作确定极大似然估计和最小二乘估计迭代计算的初始值.

二、极大似然估计

在极大似然准则下, 认为样本来自使该样本出现概率最大的总体. 因此未知参数的极大似然估计就是使得似然函数 (即联合密度函数) 达到最大的参数值.

$$L(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k; x_1, x_2, \dots, x_n) = \max\{p(x_1, x_2, \dots, x_n); \beta_1, \beta_2, \dots, \beta_k\} \quad (3.44)$$

使用极大似然估计必须已知总体的分布函数, 而在时间序列分析中, 序列总体的分布通常是未知的. 为便于分析和计算, 通常假设序列服从多元正态分布.

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

记

$$\tilde{x} = (x_1, \dots, x_n)$$

$$\begin{aligned}\tilde{\beta} &= (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)' \\ \Sigma_n &= E(\tilde{x}'\tilde{x}) = \Omega\sigma_\varepsilon^2\end{aligned}$$

式中

$$\Omega = \begin{pmatrix} \sum_{i=0}^{\infty} G_i^2 & \cdots & \sum_{i=0}^{\infty} G_i G_{i+n-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{\infty} G_i G_{i+n-1} & \cdots & \sum_{i=0}^{\infty} G_i^2 \end{pmatrix}$$

$\tilde{x}$ 的似然函数为:

$$\begin{aligned}L(\tilde{\beta}; \tilde{x}) &= p(x_1, x_2, \dots, x_n; \tilde{\beta}) \\ &= (2\pi)^{-\frac{n}{2}} |\Sigma_n|^{\frac{1}{2}} \exp \left\{ -\frac{\tilde{x}' \Sigma_n^{-1} \tilde{x}}{2} \right\} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma_\varepsilon^2)^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{\tilde{x}' \Omega^{-1} \tilde{x}}{2\sigma_\varepsilon^2} \right\}\end{aligned}$$

对数似然函数为:

$$l(\tilde{\beta}; \tilde{x}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma_\varepsilon^2} [\tilde{x}' \Omega^{-1} \tilde{x}]$$

对对数似然函数中的未知参数求偏导数, 得到似然方程组

$$\begin{cases} \frac{\partial}{\partial \sigma_\varepsilon^2} l(\tilde{\beta}; \tilde{x}) = -\frac{n}{2\sigma_\varepsilon^2} + \frac{S(\tilde{\beta})}{2\sigma_\varepsilon^4} = 0 \\ \frac{\partial}{\partial \tilde{\beta}} l(\tilde{\beta}, \tilde{x}) = -\frac{1}{2} \frac{\partial \ln |\Omega|}{\partial \tilde{\beta}} - \frac{1}{\sigma_\varepsilon^2} \frac{\partial S(\tilde{\beta})}{2\partial \tilde{\beta}} = 0 \end{cases} \quad (3.45)$$

式中, $S(\tilde{\beta}) = \tilde{x}' \Omega^{-1} \tilde{x}$.

理论上, 求解方程组 (3.45) 即得到未知参数的极大似然估计值. 但是, 由于 $S(\tilde{\beta})$ 和 $\ln |\Omega|$ 都不是 $\tilde{\beta}$ 的显式表达式的, 因此似然方程组 (3.45) 实际上是由 $p+q+1$ 个超越方程构成的, 通常需要经过复杂的迭代算法才能求出未知参数的极大似然估计值.

幸运的是, 目前计算机技术比较发达, 有许多统计软件可以辅助分析, 使得求 ARMA 模型的极大似然估计值成为一件很容易实现的事情.

极大似然估计充分应用了每一个观察值所提供的信息, 因而它的估计精度高, 同时还具有估计的一致性、渐近正态性和渐近有效性等许多优良的统计性质, 是一种非常优良的参数估计方法.

三、最小二乘估计

在 ARMA(p, q) 模型场合, 记

$$\begin{aligned}\tilde{\beta} &= (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)' \\ F_t(\tilde{\beta}) &= \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}\end{aligned}$$

残差项为:

$$\varepsilon_t = x_t - F_t(\tilde{\beta})$$

残差平方和为:

$$\begin{aligned} Q(\tilde{\beta}) &= \sum_{t=1}^n \varepsilon_t^2 \\ &= \sum_{t=1}^n (x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})^2 \end{aligned}$$

使残差平方和达到最小的那组参数值即为 $\tilde{\beta}$ 的最小二乘估计值.

由于随机扰动 $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ 不可观测, 所以 $Q(\tilde{\beta})$ 也不是 $\tilde{\beta}$ 的显性函数, 未知参数的最小二乘估计值通常也得借助迭代法求出. 由于充分利用了序列观察值的信息, 因此最小二乘估计的精度很高.

在实际运用中, 最常用的是条件最小二乘估计方法. 它假定过去未观测到的序列值等于零, 即

$$x_t = 0, t \leq 0$$

根据这个假定可以得到残差序列的有限项表达式:

$$\varepsilon_t = \frac{\Phi(B)}{\Theta(B)} x_t = x_t - \sum_{i=1}^t \pi_i x_{t-i}$$

于是残差平方和为:

$$\begin{aligned} Q(\tilde{\beta}) &= \sum_{t=1}^n \varepsilon_t^2 \\ &= \sum_{t=1}^n \left[x_t - \sum_{i=1}^t \pi_i x_{t-i} \right]^2 \end{aligned} \quad (3.46)$$

通过迭代法, 使式 (3.46) 达到最小值的估计值即为参数 β 的条件最小二乘估计.

R 语言中, 参数估计通过调用 `arima` 函数完成, 该函数的命令格式为:

```
arima(x, order=, include.mean=, method=)
```

式中:

-x: 要进行模型拟合的序列名.

-order: 指定模型阶数. `order = c(p,d,q)`.

(1) p 为自回归阶数.

(2) d 为差分阶数, 本章不涉及差分问题, 所以 `d=0`.

(3) q 为移动平均阶数.

-include.mean: 要不要包含常数项.

(1) `include.mean=T`, 需要拟合常数项, 这也是系统默认设置.

(2) 如果不需要拟合常数项, 需要特别指定 `include.mean = F`.

-method: 指定参数估计方法.

(1) `method="CSS-ML"`, 默认的是条件最小二乘与极大似然估计混合方法.

(2) `method="ML"`, 极大似然估计.

(3) `method="CSS"`, 条件最小二乘估计.

【例 3—9 续 (1)】 确定 1950—2008 年我国邮路及农村投递线路每年新增里程数序列拟合模型的口径。

```
> a<-read.table(file="E:/R/data/file8.csv",sep=" ",header=T)
> x<-ts(a$kilometer,start=1950)
> x.fit<-arima(x,order=c(2,0,0),method="ML")
> x.fit
Series: x
ARIMA(2,0,0) with non-zero mean

Coefficients:
          ar1      ar2  intercept
      0.7185  -0.5294    11.0223
s.e.      0.1083    0.1067     3.0906

sigma^2 estimated as 365.2: log likelihood=-258.23
AIC=524.46   AICc=525.2   BIC=532.77
```

根据该序列自相关图和偏自相关图, 我们将该序列定阶为 AR(2) 模型, 使用极大似然估计确定该模型口径为:

$$x_t - 11.0223 = \frac{\varepsilon_t}{1 - 0.7185B + 0.5294B^2}, \text{Var}(\varepsilon_t) = 365.2$$

或者等价写为:

$$x_t = 8.9383 + 0.7185x_{t-1} - 0.5294x_{t-2} + \varepsilon_t, \text{Var}(\varepsilon_t) = 365.2$$

【例 3—10 续 (1)】 确定美国科罗拉多州某一加油站连续 57 天的 OVERSHORT 序列拟合模型的口径。

```
> overshoot<-read.table("E:/R/data/file9.csv",sep=" ",header=T)
> overshoot<-ts(overshoot)
> overshoot.fit<-arima(overshoot,order=c(0,0,1))
> overshoot.fit
Series: overshoot
ARIMA(0,0,1) with non-zero mean

Coefficients:
      ma1  intercept
    -0.8477    -4.7945
s.e.    0.1206     1.0252

sigma^2 estimated as 2020: log likelihood=-298.42
AIC=602.84   AICc=603.29   BIC=608.97
```

在例 3—10 中, 我们将该序列的拟合模型定阶为 MA(1) 模型, 使用条件最小二乘估计, 容易确定该模型口径为:

$$x_t = -4.794\ 5 + (1 - 0.847\ 7B)\varepsilon_t, \text{Var}(\varepsilon_t) = 2\ 020$$

【例 3—11 续 (1)】 确定 1880—1985 年全球气表平均温度改变值差分序列拟合模型的口径。

```
> b<-read.table("E:/R/data/file10.csv",sep=",",header=T)
> dif_x<-ts(diff(b$change_temp),start=1880)
> dif_x.fit<-arima(dif_x,order=c(1,0,1))
> dif_x.fit
Series:dif_x
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1          ma1      intercept
      0.3926      -0.8867         0.0053
s.e.      0.1180       0.0604         0.0024

sigma^2 estimated as 0.01541: log likelihood=69.66
AIC=-131.32    AICc=-130.92    BIC=-120.71
```

对序列尝试拟合 ARMA(1,1) 模型, 使用条件最小二乘估计, 得到该模型的口径为:

$$x_t = 0.005\ 3 + 0.392\ 6x_{t-1} + \varepsilon_t - 0.886\ 7\varepsilon_{t-1}, \text{Var}(\varepsilon_t) = 0.015\ 41$$

3.3.5 模型检验

确定拟合模型的口径之后, 我们还要对该拟合模型进行必要的检验。

一、模型的显著性检验

模型的显著性检验主要是检验模型的有效性。一个模型是否显著有效主要看它提取的信息是否充分。一个好的拟合模型应该能够提取观察值序列中几乎所有的样本相关信息, 换言之, 拟合残差项中将不再蕴涵任何相关信息, 即残差序列应该为白噪声序列。这样的模型称为显著有效模型。

反之, 如果残差序列为非白噪声序列, 那就意味着残差序列中还残留着相关信息未被提取, 这就说明拟合模型不够有效, 通常需要选择其他模型, 重新拟合。

所以模型的显著性检验即为残差序列的白噪声检验。原假设和备择假设分别为:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0, \quad \forall m \geq 1$$

$$H_1: \text{至少存在某个 } \rho_k \neq 0, \quad \forall m \geq 1, k \leq m$$

检验统计量为 LB(Ljung-Box) 检验统计量:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2(m), \quad \forall m > 0$$

如果拒绝原假设,就说明残差序列中还残留着相关信息,拟合模型不显著.如果不能拒绝原假设,就认为拟合模型显著有效.

【例 3—9 续 (2)】 检验 1950—2008 年我国邮路及农村投递线路每年新增里程数序列拟合模型的显著性 ($\alpha=0.05$).

```
> a<-read.table(file="E:/R/data/file8.csv",sep=" ",header=T)
> x<-ts(a$kilometer,start=1950)
> x.fit<-arima(x,order=c(2,0,0),method="ML")
> for(i in 1: 2) print(Box.test(x.fit$residual,lag=6*i))
#对残差序列 x.fit$residual 进行白噪声检验
```

Box-Pierce test

data:x.fit\$residual

X-squared=2.0949, df=6, p-value=0.9108

Box-Pierce test

data:x.fit\$residual

X-squared=2.8341, df=12, p-value=0.9966

由于各阶延迟下 LB 统计量的 P 值都显著大于 0.05, 可以认为这个拟合模型的残差序列属于白噪声序列, 即该拟合模型显著有效.

【例 3—10 续 (2)】 对 OVERSHORT 序列的拟合模型进行检验.

```
> overshoot<-read.table("E:/R/data/file9.csv",sep=" ",header=T)
> overshoot<-ts(overshoot)
> overshoot.fit<-arima(overshoot,order=c(0,0,1))
> for(i in 1: 2) print(Box.test(overshoot.fit$residual,lag=6*i))
```

Box-Pierce test

data:overshoot.fit\$residual

X-squared=2.984, df=6, p-value=0.8108

Box-Pierce test

data:overshoot.fit\$residual

X-squared=8.4545, df=12, p-value=0.7487

残差检验结果显示, 残差序列可视为白噪声序列, 这说明拟合模型 MA(1) 显著有效.

【例 3—11 续 (2)】 对 1880—1985 年全球气表平均温度改变值差分序列拟合模型进行检验.

```
> b<-read.table("E:/R/data/file10.csv",sep=",",header=T)
> dif_x<-ts(diff(b$change_temp),start=1880)
> dif_x.fit<-arima(dif_x,order=c(1,0,1),method="CSS")
> for(i in 1: 2) print(Box.test(dif_x.fit$residual,lag=6*i))
```

Box-Pierce test

data:dif_x.fit\$residual

X-squared=4.593, df=6, p-value=0.597

Box-Pierce test

data:dif_x.fit\$residual

X-squared=9.1007, df=12, p-value=0.6943

残差检验结果显示, 残差序列可视为白噪声序列, 这说明拟合模型 ARMA(1, 1) 显著有效.

二、参数的显著性检验

参数的显著性检验就是要检验每一个未知参数是否显著非零. 这个检验的目的是使模型最精简.

如果某个参数不显著, 即表示该参数所对应的那个自变量对因变量的影响不明显, 该自变量可以从拟合模型中剔除. 最终模型将由一系列参数显著非零的自变量表示.

检验假设:

$$H_0: \beta_j = 0 \leftrightarrow H_1: \beta_j \neq 0, \quad \forall 1 \leq j \leq m$$

$$E(\hat{\beta}) = E[(X'X)^{-1}X'\tilde{y}] = (X'X)^{-1}X'X\tilde{\beta} = \tilde{\beta}$$

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}[(X'X)^{-1}X'\tilde{y}] = (X'X)^{-1}X'X(X'X)^{-1}\sigma_\varepsilon^2 \\ &= (X'X)^{-1}\sigma_\varepsilon^2 \end{aligned}$$

对于线性拟合模型, 记 $\hat{\beta}$ 为 $\tilde{\beta}$ 的最小二乘估计, 有

$$\Omega = (X'X)^{-1} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}$$

在正态分布假定下, 第 j 个未知参数的最小二乘估计值 $\hat{\beta}_j$ 服从正态分布:

$$\hat{\beta}_j \sim N(0, a_{jj}\sigma_\varepsilon^2), \quad 1 \leq j \leq m \quad (3.47)$$

由于 σ_ε^2 不可观测, 用最小残差平方和估计 σ_ε^2 :

$$\hat{\sigma}_\varepsilon^2 = \frac{Q(\tilde{\beta})}{n-m}$$

根据正态分布的性质, 有

$$\frac{Q(\tilde{\beta})}{\sigma_\varepsilon^2} \sim \chi^2(n-m) \quad (3.48)$$

由式 (3.47) 和式 (3.48) 可以构造出用于检验未知参数显著性的 t 检验统计量

$$t = \sqrt{n-m} \frac{\hat{\beta}_j}{\sqrt{a_{jj}Q(\tilde{\beta})}} \sim t(n-m)$$

当该检验统计量的绝对值大于自由度为 $n-m$ 的 t 分布的 $1-\alpha$ 分位点, 即

$$|t| \geq t_{1-\alpha}(n-m)$$

或者该检验统计量的 P 值小于 α 时, 拒绝原假设, 认为该参数显著; 否则, 认为该参数不显著. 这时, 应该剔除不显著参数所对应的自变量重新拟合模型, 构造出新的、结构更精练的拟合模型.

R 不提供参数的显著性检验结果, 一般默认输出参数均显著非零. 如果用户想获得参数检验统计量的 P 值, 需要自己计算参数的 t 统计量的值及统计量的 P 值.

根据 t 统计量的定义, 参数估计值除以参数标准差即为该参数的 t 统计量

$$t = \frac{\hat{\phi}_i}{\sigma}$$

再调用 t 分布 P 值函数 `pt` 即可获得该统计量的 P 值. `pt` 函数的命令格式为:

```
pt(t, df=, lower.tail=)
```

式中:

-`t`: t 统计量的值.

-`df`: 自由度.

-`lower.tail`: 确定计算概率的方向.

(1) `lower.tail=T`, 计算 $\Pr(X \leq x)$. 对于参数显著性检验, 如果参数估计值为负, 选择 `lower.tail=T`.

(2) `lower.tail=F`, 计算 $\Pr(X > x)$. 对于参数显著性检验, 如果参数估计值为正, 选择 `lower.tail=F`.

【例 3—9 续 (3)】 检验 1950—2008 年我国邮路及农村投递线路每年新增里程数序列拟合模型参数的显著性 ($\alpha=0.05$).

```
> a<-read.table(file="E:/R/data/file8.csv",sep="," ,header=T)
> x<-ts(a$kilometer,start=1950)
> x.fit<-arima(x,order=c(2,0,0),method="ML")
> x.fit
Series:x
ARIMA(2,0,0) with non-zero mean

Coefficients:
```

```

          ar1      ar2      intercept
      0.7185   -0.5294      11.0223
s.e.    0.1083    0.1067      3.0906

sigma^2 estimated as 365.2: log likelihood=-258.23
AIC=524.46   AICc=525.2   BIC=532.77
• ar1 系数显著性检验
> t1<-0.7185/0.1083
> pt(t1,df=56,lower.tail=F)
[1] 6.94276e-09
• ar2 系数显著性检验
> t2<-0.5294/0.1067
> pt(t2,df=56,lower.tail=T)
[1] 3.43633e-06
• 常数的显著性检验
> t0=11.0223/3.0906
> pt(t0,df=56,lower.tail=F)
[1] 0.0003748601

```

检验结果显示, 三个系数均显著非零.

3.3.6 模型优化

一、问题的提出

若一个拟合模型通过了检验, 说明在一定的置信水平下, 该模型能够有效地拟合观察值序列的波动, 但这种有效模型并不是唯一的.

【例 3—15】 等时间间隔, 连续读取某次化学反应的 70 个过程数据, 构成一时间序列 (数据见表 A1—11). 试对该序列进行拟合 ($\alpha = 0.05$).

```

• 读入数据, 绘制时序图
> x<-read.table(file="E:/R/data/file11.csv",sep=",",header=T)
> x<-ts(x)
> plot(x)      #结果如图 3—18 所示

```

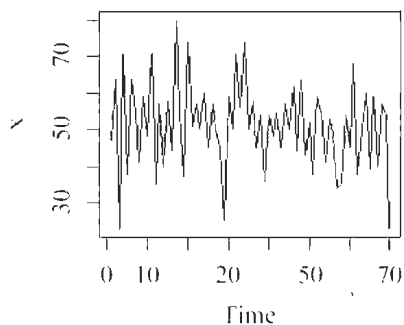


图 3-18 化学反应过程时序图

• 序列白噪声检验

```
> for(i in 1: 2)    print(Box.test(x,lag=6*i))
```

```
Box-Pierce test
```

```
data:x
```

```
X-squared=20.2091, df=6, p-value=0.002542
```

```
Box-Pierce test
```

```
data: x
```

```
X-squared=21.6219, df=12, p-value=0.04198
```

• 绘制自相关图与偏自相关图

```
> acf(x)    #结果如图 3-19 所示
```

```
> pacf(x)    #结果如图 3-20 所示
```

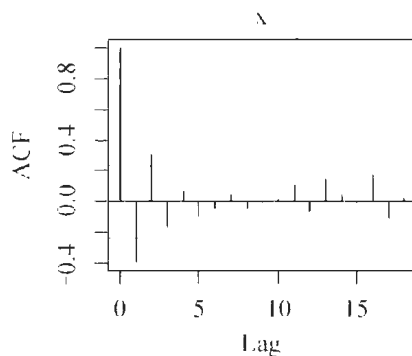


图 3—19 化学反应序列自相关图

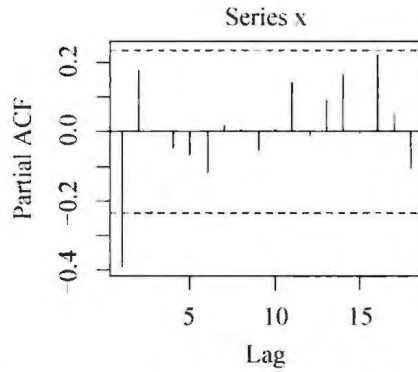


图 3—20 化学反应序列偏自相关图

• 拟合 MA(2) 模型

```
> x.fit1<-arima(x,order=c(0,0,2))
```

```
> x.fit1
```

Series: x

ARIMA(0,0,2) with non-zero mean

Coefficients:

	ma1	ma2	intercept
	-0.3194	0.3019	51.1695
s.e.	0.1160	0.1233	1.2516

sigma^2 estimated as 114.4: log likelihood=-265.35

AIC=538.71 AICc=539.32 BIC=547.7

• MA(2) 模型显著性检验

```
> for(i in 1: 2) print(Box.test(x.fit1$residual,lag=6*i))
```

Box-Pierce test

data:x.fit1\$residual

X-squared=2.1105, df=6, p-value=0.9093

Box-Pierce test

data: x.fit1\$residual

X-squared=3.9217, df=12, p-value=0.9848

• 拟合 AR(1) 模型

```
> x.fit2<-arima(x,order=c(1,0,0))
```

```
> x.fit2
```

Series: x

ARIMA(1,0,0) with non-zero mean

```
Coefficients:
      ar1      intercept
    -0.4191      51.2658
s.e.      0.1129      0.9137

sigma^2 estimated as 116.6: log likelihood=-265.98
AIC=537.96      AICc=538.32      BIC=544.7
• AR(1) 模型显著性检验
> for(i in 1: 2) print(Box.test(x.fit2$residual,lag=6*i))

      Box-Pierce test

data: x.fit2$residual
X-squared=4.1678, df=6, p-value=0.654

      Box-Pierce test

data:x.fit2$residual
X-squared=6.1411, df=12, p-value=0.9088
```

同一个序列可以构造两个拟合模型, 两个模型都显著有效, 那么到底该选择哪个模型用于统计推断呢?

为了解决这个问题, 引进 AIC 和 SBC (BIC) 信息准则的概念进行模型优化.

二、AIC 准则

AIC 准则是由日本统计学家 Akaike 于 1973 年提出的, 它的全称是最小信息量准则 (Akaike information criterion).

该准则的指导思想是一个拟合模型的优劣可以从两方面去考察: 一方面是大家非常熟悉的常用来衡量拟合程度的似然函数值; 另一方面是模型中未知参数的个数.

通常似然函数值越大, 说明模型拟合的效果越好. 模型中未知参数个数越多, 说明模型中包含的自变量越多, 自变量越多, 模型变化越灵活, 模型拟合的准确度就会越高. 模型拟合程度高是我们所希望的, 但是我们又不能单纯地以拟合精度来衡量模型的优劣, 因为这样势必会导致未知参数的个数越多越好.

未知参数越多, 说明模型中自变量越多, 未知的风险越多. 而且参数越多, 参数估计的难度就越大, 估计的精度也就越差. 所以一个好的拟合模型应该是拟合精度和未知参数个数的综合最优配置.

AIC 准则就是在这种考虑下提出的, 它是拟合精度和参数个数的加权函数:

$$AIC = -2\ln(\text{模型的极大似然函数值}) + 2(\text{模型中未知参数个数})$$

使 AIC 函数达到最小的模型被认为是最优模型.

在 ARMA(p, q) 模型场合, 对数似然函数为:

$$l(\tilde{\beta}; x_1, \dots, x_n) = - \left[\frac{n}{2} \ln(\sigma_\varepsilon^2) + \frac{1}{2} \ln |\Omega| + \frac{1}{2\sigma_\varepsilon^2} S(\tilde{\beta}) \right]$$

因为 $\frac{1}{2} \ln |\Omega|$ 有界, $\frac{1}{2\sigma_\varepsilon^2} S(\tilde{\beta}) \rightarrow \frac{n}{2}$, 所以对数似然函数与 $-\frac{n}{2} \ln(\sigma_\varepsilon^2)$ 成正比.

$$l(\tilde{\beta}; x_1, \dots, x_n) \propto -\frac{n}{2} \ln(\sigma_\varepsilon^2)$$

中心化 ARMA(p, q) 模型的未知参数个数为 $p + q + 1$, 非中心化 ARMA(p, q) 模型的未知参数个数为 $p + q + 2$.

所以, 中心化 ARMA(p, q) 模型的 AIC 函数为:

$$\text{AIC} = n \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q + 1)$$

非中心化 ARMA(p, q) 模型的 AIC 函数为:

$$\text{AIC} = n \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q + 2)$$

三、SBC (BIC) 准则

AIC 准则为选择最优模型提供了有效的规则, 但其也有不足之处. 对于一个观察值序列而言, 序列越长, 相关信息就越分散, 要很充分地提取其中的有用信息, 或者说要使拟合精度比较高, 通常需要包含多个自变量的复杂模型. 在 AIC 准则中拟合误差提供的信息要受到样本容量的放大, 它等于 $n \ln(\hat{\sigma}_\varepsilon^2)$, 但参数个数的惩罚因子却和样本容量没关系, 它的权重始终是常数 2. 因此在样本容量趋于无穷大时, 由 AIC 准则选择的模型不收敛于真实模型, 它通常比真实模型所含的未知参数个数要多.

为了弥补 AIC 准则的不足, Akaike 于 1976 年提出了 BIC 准则. 而 Schwartz 在 1978 年根据 Bayes 理论也得出了同样的判别准则, 称为 SBC 准则. SBC 准则定义为:

$$\text{SBC} = -2 \ln(\text{模型的极大似然函数值}) + \ln(n) \quad (\text{模型中未知参数个数})$$

SBC 对 AIC 的改进就是将未知参数个数的惩罚权重由常数 2 变成了样本容量的对数函数 $\ln(n)$. 理论上已证明, SBC 准则确定的最优模型是真实阶数的相合估计.

容易得到, 中心化 ARMA(p, q) 模型的 SBC 函数为:

$$\text{SBC} = n \ln(\hat{\sigma}_\varepsilon^2) + \ln(n)(p + q + 1)$$

非中心化 ARMA(p, q) 模型的 SBC 函数为:

$$\text{SBC} = n \ln(\hat{\sigma}_\varepsilon^2) + \ln(n)(p + q + 2)$$

在所有通过检验的模型中使得 AIC 或 SBC 函数达到最小的模型为相对最优模型. 之所以称为相对最优模型而不是绝对最优模型, 是因为我们不可能比较所有模型的 AIC 和 SBC 函数值, 我们总是在尽可能全面的范围里考察有限多个模型的 AIC 和

SBC 函数值, 再选择其中 AIC 和 SBC 函数值达到最小的那个模型作为最终的拟合模型, 因而这样得到的最优模型就是一个相对最优模型.

【例 3—15 续】 用 AIC 准则和 SBC 准则评判例 3—15 中两个拟合模型的相对优劣.

检验结果如表 3—6 所示.

表 3—6

模型	AIC	SBC
MA(2)	538.71	547.7
AR(1)	537.96	544.7

最小信息量检验显示无论是使用 AIC 准则还是使用 SBC 准则, AR(1) 模型都要优于 MA(2) 模型, 所以本例中 AR(1) 模型是相对最优模型.

AIC 准则和 SBC 准则的提出, 可以有效弥补根据自相关图和偏自相关图定阶的主观性, 在有限的阶数范围内帮助我们寻找相对最优拟合模型.

3.4 序列预测

到目前为止, 我们对观察值序列做了许多工作, 包括平稳性判别、白噪声判别、模型选择、参数估计及模型检验. 这些工作的最终目的常常就是要利用这个拟合模型对随机序列的未来发展进行预测.

所谓预测, 就是要利用序列已观测到的样本值对序列在未来某个时刻的取值进行估计. 目前对平稳序列最常用的预测方法是线性最小方差预测. 线性是指预测值为观察值序列的线性函数, 最小方差是指预测方差达到最小.

3.4.1 线性预测函数

根据 ARMA(p, q) 模型的平稳性和可逆性, 可以用传递形式和逆转形式等价描述该序列:

$$x_t = \sum_{i=0}^{\infty} G_i \varepsilon_{t-i} \tag{3.49}$$

$$\varepsilon_t = \sum_{j=0}^{\infty} I_j x_{t-j} \tag{3.50}$$

式中, $\{G_i\}$ 为 Green 函数值; $\{I_i\}$ 为逆转函数值.

把式 (3.50) 代入式 (3.49), 有

$$x_t = \sum_{i=0}^{\infty} G_i \left(\sum_{j=0}^{\infty} I_j x_{t-i-j} \right)$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_i I_j x_{t-i-j}$$

显然 x_t 是历史数据 $x_{t-1}, x_{t-2}, \dots$ 的线性函数. 不妨简记为:

$$x_t = \sum_{i=0}^{\infty} C_i x_{t-1-i}$$

对任意一个未来时刻 $t+l$ ($\forall l \geq 1$) 而言, 该时刻的序列值 x_{t+l} 也可以表示成它的历史数据 $x_{t+l-1}, \dots, x_{t+1}, x_t, x_{t-1}, \dots$ 的线性函数:

$$x_{t+l} = \sum_{i=0}^{\infty} C_i x_{t+l-1-i}$$

但问题是其中只有部分历史信息 $x_t, x_{t-1}, \dots$ 是已知的, 还有部分历史信息 $x_{t+l-1}, \dots, x_{t+1}$ 是未知的.

一个有趣的现象是, 根据线性函数的可加性, 所有的未知历史信息 $x_{t+l-1}, \dots, x_{t+1}$ 都可以用已知历史信息 $x_t, x_{t-1}, \dots$ 的线性函数表示出来. 以 x_{t+2} 为例, 已知

$$x_{t+2} = \sum_{i=0}^{\infty} C_i x_{t+1-i} = C_0 x_{t+1} + \sum_{i=0}^{\infty} C_{i+1} x_{t-i} \quad (3.51)$$

式中, x_{t+1} 是未知信息.

把 $x_{t+1} = \sum_{i=0}^{\infty} C_i x_{t-i}$ 代入式 (3.51), 得

$$\begin{aligned} x_{t+2} &= C_0 \sum_{i=0}^{\infty} C_i x_{t-i} + \sum_{i=0}^{\infty} C_{i+1} x_{t-i} \\ &= \sum_{i=0}^{\infty} (C_0 C_i + C_{i+1}) x_{t-i} \end{aligned}$$

由此, x_{t+2} 最终表示为已知历史信息 $x_t, x_{t-1}, \dots$ 的线性函数.

同理, 对于未来任意 l 时刻的序列值 x_{t+l} ($\forall l \geq 1$), 最终都可以表示成已知历史信息 $x_t, x_{t-1}, \dots$ 的线性函数, 并用该函数形式估计 x_{t+l} 的值:

$$\hat{x}_t(l) = \sum_{i=0}^{\infty} \hat{D}_i x_{t-i} \quad (3.52)$$

$\hat{x}_t(l)$ 也称为序列 $\{x_t\}$ 的第 l 步预测值.

3.4.2 预测方差最小原则

用 $e_t(l)$ 衡量预测误差:

$$e_t(l) = x_{t+l} - \hat{x}_t(l)$$

显然, 预测误差越小, 预测精度就越高. 因此, 目前最常用的预测原则是预测方差最小原则, 即

$$\text{Var}_{\hat{x}_t(l)}[e_t(l)] = \min\{\text{Var}[e_t(l)]\}$$

因为 $\hat{x}_t(l)$ 为 $x_t, x_{t-1}, \dots$ 的线性函数, 所以该原则也称为线性预测方差最小原则.

为便于分析, 使用传递形式来描述序列值, 根据 $\text{ARMA}(p, q)$ 平稳模型的性质和线性函数的可加性, 显然有

$$\begin{cases} x_{t+l} = \sum_{i=0}^{\infty} G_i \varepsilon_{t+l-i} \\ \hat{x}_t(l) = \sum_{i=0}^{\infty} \hat{D}_i x_{t-i} = \sum_{i=0}^{\infty} \hat{D}_i \left(\sum_{j=0}^{\infty} G_j \varepsilon_{t-i-j} \right) = \sum_{i=0}^{\infty} W_i \varepsilon_{t-i} \end{cases}$$

则 $e_t(l) = x_{t+l} - \hat{x}_t(l)$

$$= \sum_{i=0}^{\infty} G_i \varepsilon_{t+l-i} - \sum_{i=0}^{\infty} W_i \varepsilon_{t-i} = \sum_{i=0}^{l-1} G_i \varepsilon_{t+l-i} + \sum_{i=0}^{\infty} (G_{l+i} - W_i) \varepsilon_{t-i}$$

预测方差为:

$$\text{Var}[e_t(l)] = \left[\sum_{i=0}^{l-1} G_i^2 + \sum_{i=0}^{\infty} (G_{l+i} - W_i)^2 \right] \sigma_{\varepsilon}^2 \geq \sum_{i=0}^{l-1} G_i^2 \sigma_{\varepsilon}^2$$

显然, 要使得预测方差达到最小, 必须有

$$W_i = G_{l+i}, \quad i = 0, 1, 2, \dots$$

这时, x_{t+l} 的预测值为:

$$\hat{x}_t(l) = \sum_{i=0}^{\infty} G_{l+i} \varepsilon_{t-i}, \quad \forall l \geq 1$$

预测误差为:

$$e_t(l) = \sum_{i=0}^{l-1} G_i \varepsilon_{t+l-i}$$

由于 $\{\varepsilon_t\}$ 为白噪声序列, 所以

$$\begin{aligned} E[e_t(l)] &= 0 \\ \text{Var}[e_t(l)] &= \sum_{i=0}^{l-1} G_i^2 \sigma_{\varepsilon}^2, \quad \forall l \geq 1 \end{aligned}$$

3.4.3 线性最小方差预测的性质

一、条件无偏最小方差估计值

序列值 x_{t+l} 可以如下分解:

$$\begin{aligned} x_{t+l} &= (\varepsilon_{t+l} + G_1 \varepsilon_{t+l-1} + \dots + G_{l-1} \varepsilon_{t+1}) + (G_l \varepsilon_t + G_{l+1} \varepsilon_{t-1} + \dots) \\ &= e_t(l) + \hat{x}_t(l) \end{aligned}$$

由式 (3.52)

$$\hat{x}_t(l) = \sum_{i=0}^{\infty} \hat{D}_i x_{t-i}$$

即在 $x_t, x_{t-1}, \dots$ 已知的条件下, $\hat{x}_t(l)$ 为常数, 有

$$E(\hat{x}_t(l)|x_t, x_{t-1}, \dots) = \hat{x}_t(l), \text{Var}(\hat{x}_t(l)|x_t, x_{t-1}, \dots) = 0$$

推导出

$$\begin{aligned} E(x_{t+l}|x_t, x_{t-1}, \dots) &= \hat{E}[e_t(l)|x_t, x_{t-1}, \dots] + E[\hat{x}_t(l)|x_t, x_{t-1}, \dots] = \hat{x}_t(l) \\ \text{Var}(x_{t+l}|x_t, x_{t-1}, \dots) &= \text{Var}[e_t(l)|x_t, x_{t-1}, \dots] + \text{Var}[\hat{x}_t(l)|x_t, x_{t-1}, \dots] \\ &= \text{Var}[e_t(l)] \end{aligned}$$

这说明在预测方差最小原则下得到的估计值 $\hat{x}_t(l)$ 是序列值 x_{t+l} 在 $x_t, x_{t-1}, \dots$ 已知的情况下得到的条件无偏最小方差估计值, 且预测方差只与预测步长 l 有关, 而与预测起始点 t 无关. 但预测步长 l 越大, 预测值的方差也越大, 因而为了保证预测的精度, 时间序列数据通常只适合做短期预测.

在正态假定下, 有

$$x_{t+l}|x_t, x_{t-1}, \dots \sim N(\hat{x}_t(l), \text{Var}[e_t(l)])$$

式中, $x_{t+l}|x_t, x_{t-1}, \dots$ 的置信水平为 $1 - \alpha$ 的置信区间为:

$$(\hat{x}_t(l) \pm z_{1-\frac{\alpha}{2}} \cdot (1 + G_1^2 + \dots + G_{l-1}^2)^{\frac{1}{2}} \cdot \sigma_\varepsilon)$$

二、AR(p) 序列预测

在 AR(p) 序列场合:

$$\begin{aligned} \hat{x}_t(l) &= E(x_{t+l}|x_t, x_{t-1}, \dots) \\ &= E(\phi_1 x_{t+l-1} + \dots + \phi_p x_{t+l-p} + \varepsilon_{t+l}|x_t, x_{t-1}, \dots) \\ &= \phi_1 \hat{x}_t(l-1) + \dots + \phi_p \hat{x}_t(l-p) \end{aligned}$$

式中

$$\hat{x}_t(k) = \begin{cases} \hat{x}_t(k), & k \geq 1 \\ x_{t+k}, & k \leq 0 \end{cases}$$

预测方差为:

$$\text{Var}[e_t(l)] = (1 + G_1^2 + \dots + G_{l-1}^2) \sigma_\varepsilon^2$$

【例 3—16】 已知某超市月销售额近似服从 AR(2) 模型 (单位: 万元/每月):

$$x_t = 10 + 0.6x_{t-1} + 0.3x_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 36)$$

某年第一季度该超市月销售额分别为: 101 万元、96 万元、97.2 万元. 请确定该超市第二季度每月销售额的 95% 的置信区间.

(1) 预测值的计算.

4 月: $\hat{x}_3(1) = 10 + 0.6x_3 + 0.3x_2 = 97.12$
5 月: $\hat{x}_3(2) = 10 + 0.6\hat{x}_3(1) + 0.3x_3 = 97.432$
6 月: $\hat{x}_3(3) = 10 + 0.6\hat{x}_3(2) + 0.3\hat{x}_3(1) = 97.595\ 2$

(2) 预测方差的计算.

首先, 根据 Green 函数的递推公式, 计算得

$$\begin{aligned} G_0 &= 1 \\ G_1 &= \phi_1 G_0 = 0.6 \\ G_2 &= \phi_1 G_1 + \phi_2 G_0 = 0.36 + 0.3 = 0.66 \end{aligned}$$

则 $\text{Var}[e_3(1)] = G_0^2 \sigma_\varepsilon^2 = 36$
 $\text{Var}[e_3(2)] = (G_0^2 + G_1^2) \sigma_\varepsilon^2 = 48.96$
 $\text{Var}[e_3(3)] = (G_0^2 + G_1^2 + G_2^2) \sigma_\varepsilon^2 = 64.641\ 6$

(3) l 步预测销售额的 95% 置信区间为:

$$(\hat{x}_3(l) - 1.96\sqrt{\text{Var}[e_3(l)]}, \hat{x}_3(l) + 1.96\sqrt{\text{Var}[e_3(l)]})$$

计算结果如表 3—7 所示.

表 3—7

预测时期	95% 置信区间
4 月	(85.36,108.88)
5 月	(83.72,111.15)
6 月	(81.84,113.35)

下载并调用 forecast 程序包后, 我们可以调用 forecast 函数完成预测工作. forecast 函数的命令格式为:

```
forecast(object, h=, level=)
```

式中:

- object: 拟合信息文件名.
- h: 预测期数.
- level: 置信区间的置信水平. 不特殊指定的话, 系统会自动给出置信水平分别为 80% 和 95% 的双层置信区间.

【例 3—9 续 (4)】 根据 1950—2008 年的观察值序列预测 2009—2013 年我国邮路及农村投递线路每年新增里程数.

```
> a<-read.table(file="E:/R/data/file8.csv",sep=" ",header=T)
> x<-ts(a$kilometer,start=1950)
> x.fit<-arima(x,order=c(2,0,0))
> x.fore<-forecast(x.fit,h=5)
> x.fore
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2009	9.465515	-15.02434	33.95537	-27.98848	46.91951
2010	6.215194	-23.94089	36.37128	-39.90456	52.33495
2011	8.392736	-21.76507	38.55054	-37.72964	54.51512
2012	11.678108	-19.95476	43.31098	-36.70019	60.05640
2013	12.885903	-19.44653	45.21833	-36.56228	62.33409

● 系统默认输出预测图

> plot(x.fore) #结果如图 3—21 所示

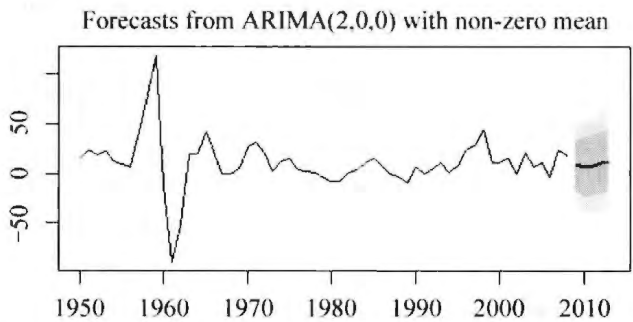


图 3—21 我国邮路及农村投递线路每年新增里程数序列预测图

● 个性化输出预测图

```
> L1<-x.fore$fitted-1.96*sqrt(x.fit$sigma2)
> U1<-x.fore$fitted+1.96*sqrt(x.fit$sigma2)
> L2<-ts(x.fore$lower[,2],start=2009)
> U2<-ts(x.fore$upper[,2],start=2009)
> c1<-min(x,L1,L2)
> c2<-max(x,L2,U2)
> plot(x,type="p",pch=8,xlim=c(1950,2013),ylim=c(c1,c2))
> lines(x.fore$fitted,col=2,lwd=2)
> lines(x.fore$mean,col=2,lwd=2)
> lines(L1,col=4,lty=2)
> lines(U1,col=4,lty=2)
> lines(L1,col=4,lty=2)
> lines(L2,col=4,lty=2)
> lines(U2,col=4,lty=2)
```

#结果如图 3—22 所示

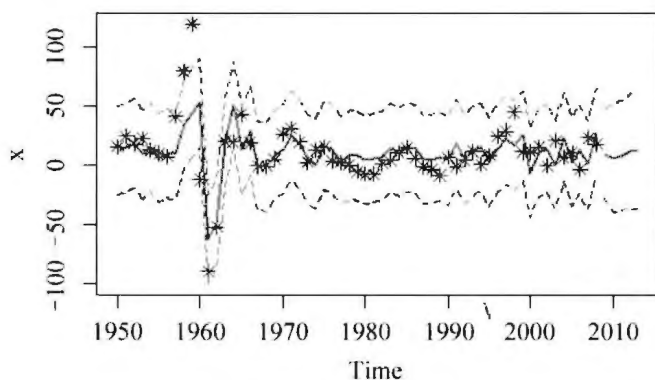


图 3—22 我国邮路及农村投递线路每年新增里程数序列拟合与预测图

说明: 星号为观察值序列, 实线为拟合值, 虚线为置信水平为 95% 的置信线。

三、MA(q) 序列预测

对一个 MA(q) 序列 $x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$ 而言, 有

$$x_{t+l} = \mu + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \cdots - \theta_q \varepsilon_{t+l-q}$$

在 $x_t, x_{t-1}, \cdots$ 已知的条件下求 x_{t+l} 的估计值, 就等价于在 $\varepsilon_t, \varepsilon_{t-1}, \cdots$ 已知的条件下求 x_{t+l} 的估计值, 而未来时刻的随机扰动 $\varepsilon_{t+1}, \varepsilon_{t+2}, \cdots$ 是不可观测的, 属于预测误差。所以:

当预测步长小于等于 MA 模型的阶数 ($l \leq q$) 时, x_{t+l} 可以分解为:

$$\begin{aligned} x_{t+l} &= \mu + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \cdots - \theta_q \varepsilon_{t+l-q} \\ &= (\varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \cdots - \theta_{l-1} \varepsilon_{t+1}) + (\mu - \theta_l \varepsilon_t - \cdots - \theta_q \varepsilon_{t+l-q}) \\ &= e_t(l) + \hat{x}_t(l) \end{aligned}$$

当预测步长大于 MA 模型的阶数 ($l > q$) 时, x_{t+l} 可以分解为:

$$\begin{aligned} x_{t+l} &= \mu + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \cdots - \theta_q \varepsilon_{t+l-q} \\ &= (\varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \cdots - \theta_q \varepsilon_{t+l-q}) + \mu \\ &= e_t(l) + \hat{x}_t(l) \end{aligned}$$

即 MA(q) 序列 l 步的预测值为:

$$\hat{x}_t(l) = \begin{cases} \mu - \sum_{i=l}^q \theta_i \varepsilon_{t+l-i}, & l \leq q \\ \mu, & l > q \end{cases}$$

这说明 MA(q) 序列理论上只能预测 q 步之内的序列走势, 超过 q 步预测值恒等于序列均值。这是由 MA(q) 序列自相关 q 步截尾的性质决定的。

MA(q) 序列预测方差为:

$$\text{Var}[e_t(l)] = \begin{cases} (1 + \theta_1^2 + \cdots + \theta_{l-1}^2)\sigma_\varepsilon^2, & l \leq q \\ (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2, & l > q \end{cases}$$

【例 3—17】 已知某地区每年常住人口数量近似服从 MA(3) 模型 (单位: 万人):

$$x_t = 100 + \varepsilon_t - 0.8\varepsilon_{t-1} + 0.6\varepsilon_{t-2} - 0.2\varepsilon_{t-3}, \quad \sigma_\varepsilon^2 = 25$$

最近 3 年的常住人口数量及一步预测数量如表 3—8 所示.

表 3—8

年份	统计人数	预测人数
2002	104	110
2003	108	100
2004	105	109

预测未来 5 年该地区常住人口的 95% 置信区间.

(1) 随机扰动项的计算:

$$\varepsilon_{t-2} = x_{2002} - \hat{x}_{2001}(1) = 104 - 110 = -6$$

$$\varepsilon_{t-1} = x_{2003} - \hat{x}_{2002}(1) = 108 - 100 = 8$$

$$\varepsilon_t = x_{2004} - \hat{x}_{2003}(1) = 105 - 109 = -4$$

(2) 未来常住人口预测值的计算:

$$\hat{x}_t(1) = 100 - 0.8\varepsilon_t + 0.6\varepsilon_{t-1} - 0.2\varepsilon_{t-2} = 109.2$$

$$\hat{x}_t(2) = 100 + 0.6\varepsilon_t - 0.2\varepsilon_{t-1} = 96$$

$$\hat{x}_t(3) = 100 - 0.2\varepsilon_t = 100.8$$

$$\hat{x}_t(4) = 100$$

$$\hat{x}_t(5) = 100$$

(3) 预测方差的计算:

$$\text{Var}[e_t(1)] = \sigma_\varepsilon^2 = 25$$

$$\text{Var}[e_t(2)] = (1 + \theta_1^2)\sigma_\varepsilon^2 = 41$$

$$\text{Var}[e_t(3)] = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2 = 50$$

$$\text{Var}[e_t(4)] = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2)\sigma_\varepsilon^2 = 51$$

$$\text{Var}[e_t(5)] = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2)\sigma_\varepsilon^2 = 51$$

(4) 95% 置信区间的计算:

$$(\hat{x}_t(l) - 1.96\sqrt{\text{Var}[e_t(l)]}, \hat{x}_t(l) + 1.96\sqrt{\text{Var}[e_t(l)]})$$

容易计算未来 5 年该地区常住人口 95% 的置信区间如表 3—9 所示.

表 3—9

预测年份	95% 置信区间
2005	(99,119)
2006	(83,109)
2007	(87,115)
2008	(86,114)
2009	(86,114)

四、ARMA(p,q) 序列预测

在 ARMA(p,q) 模型场合:

$$\begin{aligned} x_t(l) &= E(\phi_1 x_{t+l-1} + \cdots + \phi_p x_{t+l-p} + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} \\ &\quad - \cdots - \theta_q \varepsilon_{t+l-q} | x_t, x_{t-1}, \dots) \\ &= \begin{cases} \phi_1 \hat{x}_t(l-1) + \cdots + \phi_p \hat{x}_t(l-p) - \sum_{i=l}^q \theta_i \varepsilon_{t+l-i}, & l \leq q \\ \phi_1 \hat{x}_t(l-1) + \cdots + \phi_p \hat{x}_t(l-p), & l > q \end{cases} \end{aligned}$$

式中 $\hat{x}_t(k) = \begin{cases} \hat{x}_t(k), & k \geq 1 \\ x_{t+k}, & k \leq 0 \end{cases}$

预测方差为:

$$\text{Var}[e_t(l)] = (G_0^2 + G_1^2 + \cdots + G_{l-1}^2) \sigma_\varepsilon^2$$

【例 3—18】 已知 ARMA(1,1) 模型为:

$$x_t = 0.8x_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}, \sigma_\varepsilon^2 = 0.0025$$

且 $x_{100} = 0.3, \varepsilon_{100} = 0.01$, 预测未来 3 期序列值的 95% 的置信区间.

(1) 预测值的计算:

$$\begin{aligned} \hat{x}_{100}(1) &= 0.8x_{100} - 0.6\varepsilon_{100} = 0.234 \\ \hat{x}_{100}(2) &= 0.8\hat{x}_{100}(1) = 0.1872 \\ \hat{x}_{100}(3) &= 0.8\hat{x}_{100}(2) = 0.14976 \end{aligned}$$

(2) 预测方差的计算:

首先, 根据 Green 函数的递推公式, 计算得

$$\begin{aligned} G_0 &= 1 \\ G_1 &= \phi_1 G_0 - \theta_1 = 0.2 \\ G_2 &= \phi_1 G_1 = 0.16 \end{aligned}$$

则
$$\begin{aligned} \text{Var}[e_{100}(1)] &= G_0^2 \sigma_\varepsilon^2 = 0.0025 \\ \text{Var}[e_{100}(2)] &= (G_0^2 + G_1^2) \sigma_\varepsilon^2 = 0.0026 \\ \text{Var}[e_{100}(3)] &= (G_0^2 + G_1^2 + G_2^2) \sigma_\varepsilon^2 = 0.002664 \end{aligned}$$

(3) 95% 置信区间的计算:

$$(\hat{x}_{100}(l) - 1.96\sqrt{\text{Var}[e_{100}(l)]}, \hat{x}_{100}(l) + 1.96\sqrt{\text{Var}[e_{100}(l)]})$$

计算结果如表 3—10 所示.

表 3—10

时期	95% 置信区间
101	(0.136, 0.332)
102	(0.087, 0.287)
103	(0.049, 0.251)

3.5 习 题

1. 已知 AR(1) 模型为: $x_t = 0.7x_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, 1)$. 求 $E(x_t), \text{Var}(x_t), \rho_2$ 和 ϕ_{22} .
2. 已知某 AR(2) 模型为: $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, 且 $\rho_1 = 0.5, \rho_2 = 0.3$, 求 ϕ_1, ϕ_2 的值.
3. 已知某 AR(2) 模型为: $(1 - 0.5B)(1 - 0.3B)x_t = \varepsilon_t, \varepsilon_t \sim WN(0, 1)$, 求 $E(x_t), \text{Var}(x_t), \rho_k, \phi_{kk}$, 其中 $k = 1, 2, 3$.
4. 已知 AR(2) 序列为 $x_t = x_{t-1} + cx_{t-2} + \varepsilon_t$, 其中 $\{\varepsilon_t\}$ 为白噪声序列. 确定 c 的取值范围, 以保证 $\{x_t\}$ 为平稳序列, 并给出该序列 ρ_k 的表达式.
5. 证明对任意常数 c , 如下定义的 AR(3) 序列一定是非平稳序列:

$$x_t = x_{t-1} + cx_{t-2} - cx_{t-3} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

6. 对于 AR(1) 模型: $x_t = \phi_1 x_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, 判断如下命题是否正确:

- (1) $\gamma_0 = (1 + \phi_1^2)\sigma_\varepsilon^2$;
- (2) $E[(x_t - \mu)(x_{t-1} - \mu)] = -\phi_1$;
- (3) $\hat{x}_T(l) = \phi_1^l x_T$;
- (4) $e_T(l) = \varepsilon_{T+l} + \phi_1^2 \varepsilon_{T+l-1} + \phi_1^4 \varepsilon_{T+l-2} + \cdots + \phi_1^{2l} \varepsilon_T$;
- (5) $\lim_{l \rightarrow \infty} \text{Var}[x_{T+l} - \hat{x}_T(l)] = \frac{\sigma_\varepsilon^2}{\phi_1^2}$.

7. 已知某中心化 MA(1) 模型 1 阶自相关系数 $\rho_1 = 0.4$, 求该模型的表达式.

8. 确定常数 C 的值, 以保证如下表达式为 MA(2) 模型:

$$x_t = 10 + 0.5x_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-2} + C\varepsilon_{t-3}$$

9. 已知 MA(2) 模型为: $x_t = \varepsilon_t - 0.7\varepsilon_{t-1} + 0.4\varepsilon_{t-2}, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. 求 $E(x_t), \text{Var}(x_t)$, 及 $\rho_k (k \geq 1)$.

10. 证明:

- (1) 对任意常数 C , 如下定义的无穷阶 MA 序列一定是非平稳序列:

$$x_t = \varepsilon_t + C(\varepsilon_{t-1} + \varepsilon_{t-2} + \cdots), \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

(2) $\{x_t\}$ 的 1 阶差分序列一定是平稳序列, 并求 $\{y_t\}$ 自相关系数表达式:

$$y_t = x_t - x_{t-1}$$

11. 检验下列模型的平稳性与可逆性, 其中 $\{\varepsilon_t\}$ 为白噪声序列:

$$(1) x_t = 0.5x_{t-1} + 1.2x_{t-2} + \varepsilon_t \quad (2) x_t = 1.1x_{t-1} - 0.3x_{t-2} + \varepsilon_t$$

$$(3) x_t = \varepsilon_t - 0.9\varepsilon_{t-1} + 0.3\varepsilon_{t-2} \quad (4) x_t = \varepsilon_t + 1.3\varepsilon_{t-1} - 0.4\varepsilon_{t-2}$$

$$(5) x_t = 0.7x_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1} \quad (6) x_t = -0.8x_{t-1} + 0.5x_{t-2} + \varepsilon_t - 1.1\varepsilon_{t-1}$$

12. 已知 ARMA(1,1) 模型为: $x_t = 0.6x_{t-1} + \varepsilon_t - 0.3\varepsilon_{t-1}$, 确定该模型的 Green 函数, 使该模型可以等价表示为无穷 MA 阶模型形式.

13. 某 ARMA(2,2) 模型为: $\Phi(B)x_t = 3 + \Theta(B)\varepsilon_t$, 求 $E(x_t)$. 其中, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, $\Phi(B) = (1 - 0.5B)^2$.

14. 证明 ARMA(1,1) 序列 $x_t = 0.5x_{t-1} + \varepsilon_t - 0.25\varepsilon_{t-1}$, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ 的自相关系数为:

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0.27, & k = 1 \\ 0.5\rho_{k-1}, & k \geq 2 \end{cases}$$

15. 对于平稳时间序列, 以下等式哪些一定成立?

$$(1) \sigma_\varepsilon^2 = E(\varepsilon_1^2);$$

$$(2) \text{Cov}(y_t, y_{t+k}) = \text{Cov}(y_t, y_{t-k});$$

$$(3) \rho_k = \rho_{-k};$$

$$(4) \hat{y}_t(k+1) = \hat{y}_{t+1}(k).$$

16. 对于 AR(1) 模型: $x_t - \mu = \phi_1(x_{t-1} - \mu) + \varepsilon_t$, 根据 t 个历史观察值数据: $\cdots, 10.1, 9.6$, 已求出 $\hat{\mu} = 10, \hat{\phi}_1 = 0.3, \hat{\sigma}_\varepsilon^2 = 9$, 求 x_{t+3} 的 95% 的置信区间.

17. 某城市过去 63 年中每年降雪量数据 (单位: mm) 如表 3—11 所示 (行数据).

表 3—11

126.4	82.4	78.1	51.1	90.9	76.2	104.5	87.4
110.5	25	69.3	53.5	39.8	63.6	46.7	72.9
79.6	83.6	80.7	60.3	79	74.4	49.6	54.7
71.8	49.1	103.9	51.6	82.4	83.6	77.8	79.3
89.6	85.5	58	120.7	110.5	65.4	39.9	40.1
88.7	71.4	83	55.9	89.9	84.8	105.2	113.7
124.7	114.5	115.6	102.4	101.4	89.8	71.5	70.9
98.3	55.5	66.1	78.4	120.5	97	110	

(1) 判断该序列的平稳性与纯随机性.

(2) 如果序列平稳且非白噪声, 选择适当模型拟合该序列的发展.

(3) 利用拟合模型, 预测该城市未来 5 年的降雪量.

18. 某地区连续 74 年的谷物产量 (单位: 千吨) 如表 3—12 所示 (行数据).

表 3—12

0.97	0.45	1.61	1.26	1.37	1.43	1.32	1.23	0.84	0.89	1.18
1.33	1.21	0.98	0.91	0.61	1.23	0.97	1.10	0.74	0.80	0.81
0.80	0.60	0.59	0.63	0.87	0.36	0.81	0.91	0.77	0.96	0.93
0.95	0.65	0.98	0.70	0.86	1.32	0.88	0.68	0.78	1.25	0.79
1.19	0.69	0.92	0.86	0.86	0.85	0.90	0.54	0.32	1.40	1.14
0.69	0.91	0.68	0.57	0.94	0.35	0.39	0.45	0.99	0.84	0.62
0.85	0.73	0.66	0.76	0.63	0.32	0.17	0.46			

(1) 判断该序列的平稳性与纯随机性.

(2) 选择适当模型拟合该序列的发展.

(3) 利用拟合模型, 预测该地区未来 5 年的谷物产量.

19. 现有 201 个连续的生产记录, 如表 3—13 所示 (行数据).

表 3—13

81.9	89.4	79.0	81.4	84.8	85.9	88.0	80.3	82.6
83.5	80.2	85.2	87.2	83.5	84.3	82.9	84.7	82.9
81.5	83.4	87.7	81.8	79.6	85.8	77.9	89.7	85.4
86.3	80.7	83.8	90.5	84.5	82.4	86.7	83.0	81.8
89.3	79.3	82.7	88.0	79.6	87.8	83.6	79.5	83.3
88.4	86.6	84.6	79.7	86.0	84.2	83.0	84.8	83.6
81.8	85.9	88.2	83.5	87.2	83.7	87.3	83.0	90.5
80.7	83.1	86.5	90.0	77.5	84.7	84.6	87.2	80.5
86.1	82.6	85.4	84.7	82.8	81.9	83.6	86.8	84.0
84.2	82.8	83.0	82.0	84.7	84.4	88.9	82.4	83.0
85.0	82.2	81.6	86.2	85.4	82.1	81.4	85.0	85.8
84.2	83.5	86.5	85.0	80.4	85.7	86.7	86.7	82.3
86.4	82.5	82.0	79.5	86.7	80.5	91.7	81.6	83.9
85.6	84.8	78.4	89.9	85.0	86.2	83.0	85.4	84.4
84.5	86.2	85.6	83.2	85.7	83.5	80.1	82.2	88.6
82.0	85.0	85.2	85.3	84.3	82.3	89.7	84.8	83.1
80.6	87.4	86.8	83.5	86.2	84.1	82.3	84.8	86.6
83.5	78.1	88.8	81.9	83.3	80.0	87.2	83.3	86.6
79.5	84.1	82.2	90.8	86.5	79.7	81.0	87.2	81.6
84.4	84.4	82.2	88.9	80.9	85.1	87.1	84.0	76.5
82.7	85.1	83.3	90.4	81.0	80.3	79.8	89.0	83.7
80.9	87.3	81.1	85.6	86.6	80.0	86.6	83.3	83.1
82.3	86.7	80.2						

- (1) 判断该序列的平稳性与纯随机性.
- (2) 如果序列平稳且非白噪声, 选择适当模型拟合该序列的发展.
- (3) 利用拟合模型, 预测该序列下一时刻 95% 的置信区间.

20. 1971 年 9 月—1993 年 6 月澳大利亚季度常住人口变动 (单位: 千人) 情况如表 3—14 所示 (行数据).

表 3—14

63.2	67.9	55.8	49.5	50.2	55.4
49.9	45.3	48.1	61.7	55.2	53.1
49.5	59.9	30.6	30.4	33.8	42.1
35.8	28.4	32.9	44.1	45.5	36.6
39.5	49.8	48.8	29.0	37.3	34.2
47.6	37.3	39.2	47.6	43.9	49.0
51.2	60.8	67.0	48.9	65.4	65.4
67.6	62.5	55.1	49.6	57.3	47.3
45.5	44.5	48.0	47.9	49.1	48.8
59.4	51.6	51.4	60.9	60.9	56.8
58.6	62.1	64.0	60.3	64.6	71.0
79.4	59.9	83.4	75.4	80.2	55.9
58.5	65.2	69.5	59.1	21.5	62.5
170.0	47.4	62.2	60.0	33.1	35.3
43.4	42.7	58.4	34.4		

- (1) 判断该序列的平稳性与纯随机性.
- (2) 选择适当模型拟合该序列的发展.
- (3) 绘制该序列拟合及未来 5 年预测序列图.

前面介绍了对平稳时间序列的分析方法. 实际上, 在自然界中绝大部分序列都是非平稳的, 因而对非平稳序列的分析更普遍、更重要, 人们创造的分析方法也更多.

对非平稳时间序列的分析方法可以分为确定性时序分析和随机时序分析两大类, 本章主要介绍一些常用的确定性时序分析方法.

4.1 时间序列的分解

4.1.1 Wold 分解定理

1938 年, H. Wold 在他的博士论文 “A Study in the Analysis of Stationary Time Series” 中提出了著名的 Wold 分解定理.

Wold 分解定理 对于任何一个离散平稳过程 $\{x_t\}$, 它都可以分解为两个不相关的平稳序列之和, 其中一个为确定性的, 另一个为随机性的, 不妨记作

$$x_t = V_t + \xi_t$$

式中, $\{V_t\}$ 为确定性序列; $\{\xi_t\}$ 为随机序列, 且 $\xi_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}$. 它们需要满足如下条件:

$$(1) \varphi_0 = 1, \sum_{j=0}^{\infty} \varphi_j^2 < \infty$$

$$(2) \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

$$(3) \text{Cov}(V_t, \varepsilon_s) = E(V_t \cdot \varepsilon_s), \quad \forall t \neq s$$

这里, 确定性序列和随机序列的概念如下定义:

对任意序列 $\{y_t\}$ 而言, 令 y_t 关于 q 期之前的序列值 $\{y_{t-q}, y_{t-q-1}, \dots\}$ 作线性回归

$$y_t = \alpha_0 + \alpha_1 y_{t-q} + \alpha_2 y_{t-q-1} + \dots + v_t$$

式中, $\{v_t\}$ 为回归残差序列; $\text{Var}(v_t) = \tau_q^2$.

显然, $\tau_q^2 \leq \text{Var}(y_t)$, 且随着 q 的增大而增大, 也就是说 τ_q^2 是非减的有界序列, 它的大小可以衡量历史信息对现时值的预测精度. τ_q^2 越小, 说明预测得越准确; τ_q^2 越大,

说明预测得越差.

如果 $\lim_{q \rightarrow \infty} \tau_q^2 = 0$, 说明序列的发展有很强的规律性 (或称确定性), 历史数据可以很好地预测将来, 这时称 $\{y_t\}$ 是确定性序列.

如果 $\lim_{q \rightarrow \infty} \tau_q^2 = \text{Var}(y_t)$, 说明序列随着时间的发展随机性很强, 历史信息对于现时值的估计或者说是预测效果很差, 这时称 $\{y_t\}$ 是随机序列.

显然, 针对之前我们对平稳序列进行 ARMA(p, q) 模型拟合时得到的拟合模型

$$x_t = \mu + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

其中, $\{\mu_t = \mu, t \in T\}$ 即为确定性平稳序列; $\left\{ \frac{\Theta(B)}{\Phi(B)} \varepsilon_t, t \in T \right\}$ 即为随机平稳序列. 而对 ARMA(p, q) 模型的分析实际上主要是对随机平稳序列进行的分析.

4.1.2 Cramer 分解定理

Wold 分解定理是现代时间序列分析理论的灵魂. 尽管 Wold 提出这个分解定理只是为了分析平稳序列的构成, 但 Cramer 已于 1961 年证明这种分解思路同样可以用于非平稳序列.

Cramer 分解定理 任何一个时间序列 $\{x_t\}$ 都可以分解为两部分的叠加, 其中一部分是由多项式决定的确定性趋势成分, 另一部分是平稳的零均值误差成分, 即

$$x_t = \mu_t + \varepsilon_t = \sum_{j=0}^d \beta_j t^j + \Psi(B)a_t$$

式中, $d < \infty$; $\beta_1, \dots, \beta_d$ 为常数系数; $\{a_t\}$ 为一个零均值白噪声序列; B 为延迟算子.

因为

$$E(\varepsilon_t) = \Psi(B)E(a_t) = 0$$

所以有

$$E(x_t) = E(\mu_t) = \sum_{j=0}^d \beta_j t^j$$

即均值序列 $\left\{ \sum_{j=0}^d \beta_j t^j \right\}$ 反映了 $\{x_t\}$ 受到的确定性影响, 而 $\{\varepsilon_t; \varepsilon_t = \Psi(B)a_t\}$ 反映了 $\{x_t\}$ 受到的随机影响.

Cramer 分解定理说明任何一个序列的波动都可以视为同时受到了确定性影响和随机性影响的作用. 平稳序列要求这两方面的影响都是稳定的, 而非平稳序列产生的机理就在于它所受到的这两方面的影响至少有一方面是不稳定的.

4.2 确定性因素分解

在自然界中,由确定性因素导致的非平稳通常显示出非常明显的规律性,比如有显著的趋势或者有固定的变化周期,这种规律性信息通常比较容易提取,而由随机因素导致的波动则非常难以确定、分析.根据这种性质,传统的时序分析方法通常都把分析的重点放在确定性信息的提取上,忽视对随机信息的提取,将序列简单地假定为:

$$x_t = \mu_t + \varepsilon_t$$

式中, $\{\varepsilon_t\}$ 为零均值白噪声序列. 这种分析方法就称为确定性分析方法.

最常用的确定性分析方法是确定性因素分解方法. 该方法产生于长期的观察实践. 人们发现尽管不同的序列波动特征千差万别,但是序列的各种变化都可以归纳成四大类因素的综合影响:

(1) 长期趋势 (trend). 该因素的影响会导致序列呈现出明显的长期趋势 (递增、递减等).

(2) 循环波动 (circle). 该因素会导致序列呈现出从低到高再由高至低的反复循环波动.

(3) 季节性变化 (season). 该因素会导致序列呈现出和季节变化相关的稳定的周期波动.

(4) 随机波动 (immediate). 除了长期趋势、循环波动和季节性变化之外,序列还会受到各种其他因素的综合影响,而这些影响导致序列呈现出一定的随机波动.

在进行确定性时序分析时,我们假定序列会受到这四种因素中的全部或部分的影响,呈现出不同的波动特征. 在实务中,常假定这四种因素主要有两种相互作用模式:加法模型和乘法模型.

(1) 加法模型:

$$x_t = T_t + C_t + S_t + I_t$$

(2) 乘法模型:

$$x_t = T_t \cdot C_t \cdot S_t \cdot I_t$$

Durbin 在 1975 年还介绍过加法与乘法的混合模型,但是一般的统计软件中只提供加法模型和乘法模型的分析程序,混合模型的分析程序只能自编.

确定性因素分解方法在宏观经济领域中有广泛的应用,但我们从大量的使用经验中也发现了一些问题. 一是如果观察时期不是足够长,那么循环因素和趋势因素的影响很难严格分解开. 二是有些经济现象显示出交易日是一个很显著的影响因素,但是在传统因素分解模型中,它却没有被纳入研究. 比如研究股票交易序列,成交量、开盘价、收盘价明显会受到交易日的影响,同一只股票每周一和每周五的波动情况可能有显著的不同. 超市销售情况更是明显受到交易日的影响,平时和周末的销售特征相差很大.

近年来, 针对这两个问题, 人们对确定性因素分解模型做了改进, 如果观察时期不是足够长, 人们将循环 (circle) 因素改为了交易日 (day) 因素, 新的四因素模型为: 趋势 (T)、季节 (S)、交易日 (D) 和随机波动 (I). 假定这四个因素的相互作用模式也是两种: 加法模型和乘法模型.

(1) 加法模型:

$$x_t = T_t + S_t + D_t + I_t$$

(2) 乘法模型:

$$x_t = T_t \cdot S_t \cdot D_t \cdot I_t$$

四大因素的综合影响会导致序列呈现出各种变化情况. 而我们进行确定性时序分析的目的不外乎以下两种:

一是克服其他因素的影响, 单纯测度出某一个确定性因素 (长期趋势波动或季节效应) 对序列的影响.

二是推断出各种确定性因素彼此之间的相互作用关系以及它们对序列的综合影响.

4.3 趋势分析

有些时间序列具有非常显著的趋势, 有时我们分析的目的就是要找到序列中的这种趋势, 并利用这种趋势对序列未来的发展作出合理的预测.

4.3.1 趋势拟合法

趋势拟合法就是把时间作为自变量, 相应的序列观察值作为因变量, 建立序列值随时间变化的回归模型的方法. 根据序列所表现出的线性或非线性特征, 拟合方法又可以具体分为线性拟合和曲线拟合.

一、线性拟合

如果长期趋势呈现出线性特征, 那么我们可以用线性模型来拟合, 模型可以具体写为:

$$\begin{cases} x_t = a + bt + I_t \\ E(I_t) = 0, \text{Var}(I_t) = \sigma^2 \end{cases}$$

式中, $\{I_t\}$ 为随机波动; $T_t = a + bt$ 就是消除随机波动的影响之后该序列的长期趋势.

R 语言中使用 `lm` 函数拟合线性趋势. `lm` 函数的命令格式为:

```
lm(Y ~ a + X1 + X2 + ... + Xn, data=)
```

式中:

-Y: 响应变量。
-a: 指定是否需要常数项。
(1) a=1, 模型有非零常数项, 这是默认设置。
(2) a=0, 模型不需要常数项。
-X1, ..., Xn: 自变量。
-data: 数据框名。如果自变量和响应变量不是独立输入变量而是共同存在于某个数据框中, 则需要指定数据框名。

【例 4—1】 澳大利亚政府 1981—1990 年每季度的消费支出数据 (单位: 百万澳元) 如表 4—1 所示。

表 4—1

8 444	9 215	8 879	8 990	8 115	9 457	8 590	9 294	8 997	9 574
9 051	9 724	9 120	10 143	9 746	10 074	9 578	10 817	10 116	10 779
9 901	11 266	10 686	10 961	10 121	11 333	10 677	11 325	10 698	11 624
11 052	11 393	10 609	12 077	11 376	11 777	11 225	12 231	11 884	12 109

该序列时序图显示该序列有显著的线性递增趋势。于是考虑使用线性模型

$$\begin{cases} x_t = a + bt + I_t, t = 1, 2, \cdots, 40 \\ E(I_t) = 0, \text{Var}(I_t) = \sigma^2 \end{cases}$$

拟合该序列的发展。

```
• 读入数据
> x<-c(8444,9215,8879,8990,8115,9457,8590,9294,8997,9574,9051,9724,9120,
+ 10143,9746,10074,9578,10817,10116,10779,9901,11266,10686,10961,10121,
+ 11333,10677,11325,10698,11624,11052,11393,10609,12077,11376,11777,
+ 11225,12231,11884,12109)
• 构造时间自变量
> t<-c(1:40)
• 拟合回归模型
> x.fit<-lm(x~t)
• 查看拟合信息
> summary(x.fit)

Call:
lm(formula=x~t)

Residuals:
    Min       1Q   Median       3Q      Max
-830.73  -324.15   57.66  334.57  806.62

Coefficients:
      Estimate      Std. Error    t value    Pr(> |t|)
```



```
(Intercept) 8498.688 137.917 61.62 <2e-16 ***
t            89.123   5.862 15.20 <2e-16 ***
—
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 428 on 38 degrees of freedom
Multiple R-squared: 0.8588, Adjusted R-squared: 0.8551
F-statistic: 231.1 on 1 and 38 DF, p-value: < 2.2e-16
```

• 绘制拟合效果图

```
> x<-ts(x)
> plot(x)
> abline(lm(x~t),col=2)
#结果如图 4—1 所示
```

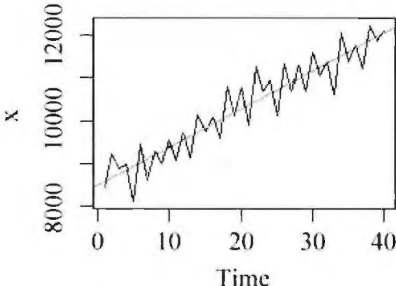


图 4—1 澳大利亚政府季度消费支出序列线性拟合图

根据模型输出结果, 可知澳大利亚政府季度消费支出序列线性拟合模型为:

$$x_t = 8\,498.688 + 89.123t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 428^2)$$

二、曲线拟合

如果长期趋势呈现出非线性特征, 那么我们可以用曲线模型来拟合它.

对曲线模型进行参数估计时, 指导思想是: 能转换成线性模型的都转换成线性模型, 用线性最小二乘法进行参数估计; 实在不能转换成线性模型的, 就用迭代法进行参数估计.

常用的曲线模型和对应的参数估计方法如表 4—2 所示.

表 4—2

模型	变换	参数估计方法
二次型: $T_t = a + bt + ct^2$	令 $t_2 = t^2$, 原模型变换为: $T_t = a + bt + ct_2$	线性最小二乘法
指数型: $T_t = ab^t$	对原模型求对数, 再令 $T'_t = \ln T_t, a' = \ln a, b' = \ln b$, 原模型变换为: $T'_t = a' + b't$	用线性最小二乘法求出 a', b' , 再做变换: $a = e^{a'}$, $b = e^{b'}$

续前表

模型	变换	参数估计方法
修正指数型: $T_t = a + bc^t$	不能转换成线性模型	迭代法
Gompertz 型: $T_t = e^{a+bc^t}$	不能转换成线性模型	迭代法
Logistic 型: $T_t = \frac{1}{a + bc^t}$	不能转换成线性模型	迭代法

R 语言针对非线性趋势的拟合也分为两类: 一类可以写成关于时间 t 的多项式, 这时仍然可以用 `lm` 函数拟合. 另一类无法通过适当的变换变成线性回归模型, 只能通过非线性回归解决, 这时需要调用 `nls` 函数.

`nls` 函数的命令格式如下:

```
nls(Y~ f(X1,... ,Xn ), data=, start=)
```

式中:

- Y: 响应变量.
- X1, ..., Xn: 自变量.
- f: 非线性函数.
- data: 数据框名. 如果自变量和响应变量不是独立输入变量而是存在于某个数据框中, 则需要指定数据框名.
- start: 如果需要利用迭代法计算未知参数, 可以指定迭代初始值.

【例 4—2】 对我国 1949—2008 年化肥产量序列进行曲线拟合 (数据见表 A1—12).

时序图显示该序列有显著的曲线递增趋势, 我们尝试使用二次型模型

$$x_t = a_0 + a_1t + a_2t^2, \quad t = 1, 2, \dots, 60$$

拟合该序列的趋势.

```
• 读入数据
> a<-read.table("E:/R/data/file12.csv",sep=" ",header=T)
> x<-ts(a$output,start=1949)

• lm 函数拟合
> t1<-c(1:60)
> t2<-t1^2
> x.fit1<-lm(x~t1+t2)
> summary(x.fit1)

Call:
lm(formula=x ~ t1+t2)

Residuals:
    Min       1Q   Median       3Q      Max
-532.12  -164.92   24.68  105.51  716.37
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	319.0255	105.3371	3.029	0.00369 **
t1	-57.7690	7.9679	-7.250	1.22e-09 ***
t2	2.3551	0.1266	18.601	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 263 on 57 degrees of freedom

Multiple R-squared: 0.9755, Adjusted R-squared: 0.9746

F-statistic: 1133 on 2 and 57 DF, p-value: < 2.2e-16

• nls 函数拟合

> x.fit2<-nls(x~a+b*t1+c*t1^2,start=list(a=1,b=1,c=1))

> summary(x.fit2)

Formula: x ~ a + b * t1 + c * t1^2

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
a	319.0255	105.3371	3.029	0.00369 **
b	-57.7690	7.9679	-7.250	1.22e-09 ***
c	2.3551	0.1266	18.601	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 263 on 57 degrees of freedom

Number of iterations to convergence: 1

Achieved convergence tolerance: 3.059e-08

根据输出结果可以知道, 这两个函数得到的拟合结果完全一致, 1949—2008 年我国化肥产量序列的拟合模型为:

$$x_t = 319.0255 - 57.769t + 2.3551t^2 + \varepsilon_t, \quad \varepsilon_t \sim N(0, 263^2)$$

绘图命令及结果如下:

```
> y<-predict(x.fit2) #把 nls 函数得到的拟合值赋值给 y
> y<-ts(y,start=1949)
> plot(x,type="p")
> lines(y,col=2,lwd=2)
#结果如图 4—2 所示
```

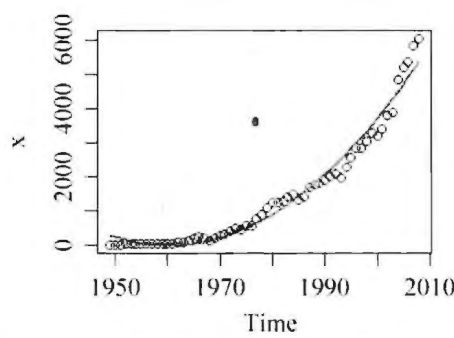


图 4—2 中国化肥产量序列非线性拟合图

说明: 圆点为观察值, 曲线为二次型拟合值.

4.3.2 平滑法

平滑法是进行趋势分析和预测时常用的一种方法. 它是利用修匀技术, 削弱短期随机波动对序列的影响, 使序列平滑化, 从而显示出变化的规律. 它具有调节灵活、计算简便的特征, 广泛应用于计量经济、人口研究等诸多领域. 根据所用的平滑技术的不同, 平滑法又可以具体分为移动平均法和指数平滑法.

一、移动平均法

移动平均法的基本思想是对于一个时间序列 $\{X_t\}$, 我们可以假定在一个比较短的时间间隔里, 序列的取值是比较稳定的, 它们之间的差异主要是由随机波动造成的. 根据这种假定, 我们可以用一定时间间隔内的平均值作为下一期的估计值.

n 期移动平均是一种常用的平稳序列预测方法, $\tilde{x}_t = \frac{1}{n}(x_t + x_{t-1} + \cdots + x_{t-n+1})$. 假定最后一期的观察值为 x_t , 那么使用 n 期移动平均方法, 向前预测 l 期, 各期的预测值为:

$$\begin{aligned}\hat{x}_t(1) &= \frac{x_t + x_{t-1} + \cdots + x_{t-n+1}}{n} \\ \hat{x}_t(2) &= \frac{\hat{x}_t(1) + x_t + \cdots + x_{t-n+2}}{n} \\ &\vdots \\ \hat{x}_t(l) &= \frac{\hat{x}_t(l-1) + \cdots + \hat{x}_t(1) + x_t + \cdots + x_{t-n+l}}{n}\end{aligned}$$

移动平均的期数对原序列的修匀效果影响很大, 要确定移动平均的期数, 一般会从如下三个方面加以考虑:

(1) 事件的发展有无周期性. 如果事件的发展具有一定的周期性, 一般以周期长度作为移动平均的间隔长度. 比如研究每月的平均气温变化趋势, 就应该做 12 期移动平均, 通过周期平滑消除季节效应的影响.

(2) 对趋势平滑性的要求. 一般移动平均的期数越多, 修匀曲线越平滑, 表现出的长期趋势就越清晰.

(3) 对趋势反映近期变化敏感程度的要求. 用移动平均方法确定事件的发展趋势都具有一定的滞后性. 移动平均的期数越多, 滞后性越强; 移动平均的期数越少, 所得的趋势图对近期变化的反应就越敏感.

综合以上三方面的考虑, 如果想得到长期趋势, 就应该做期数较多的移动平均, 如果想密切关注序列的短期趋势, 就应该做期数较少的移动平均.

【例 4—3】 某一观察值序列 $\{x_t\}$ 最后 4 期的观察值 $x_{T-3}, x_{T-2}, x_{T-1}, x_T$ 为:

5, 5.4, 5.8, 6.2

(1) 使用 4 期移动平均法预测 $\hat{x}_T(2)$.

(2) 在 2 期预测值 $\hat{x}_T(2)$ 中 x_T 前面的系数等于多少?

解: (1) $\hat{x}_T(1) = \frac{1}{4}(x_T + x_{T-1} + x_{T-2} + x_{T-3}) = \frac{5 + 5.4 + 5.8 + 6.2}{4} = 5.6$

$\hat{x}_T(2) = \frac{1}{4}(\hat{x}_T(1) + x_T + x_{T-1} + x_{T-2}) = \frac{5.6 + 6.2 + 5.8 + 5.4}{4} = 5.75$

(2) 因为

$$\begin{aligned}\hat{x}_T(2) &= \frac{1}{4}(\hat{x}_T(1) + x_T + x_{T-1} + x_{T-2}) \\ &= \frac{1}{4} \left[\frac{1}{4}(x_T + x_{T-1} + x_{T-2} + x_{T-3}) + x_T + x_{T-1} + x_{T-2} \right] \\ &= \frac{5}{16}(x_T + x_{T-1} + x_{T-2}) + \frac{1}{16}x_{T-3}\end{aligned}$$

所以使用 4 期移动平均法, 在 2 期预测值 $\hat{x}_T(2)$ 中 x_T 前面的系数等于 $\frac{5}{16}$.

R 语言中, TTR 程序包中的 SMA 函数是专门作简单移动平均趋势拟合的函数. 在调用 SMA 函数之前, 需要先下载 TTR 程序包并用 library 函数加载. SMA 函数的命令格式为:

SMA(x, n=)

式中:

-x: 需要做简单移动平均的序列名.

-n: 移动平均期数.

【例 4—4】 对 1949—1998 年北京市每年最高气温序列 (数据见附表 A1—6) 进行 5 期移动平均拟合.

```
> library(TTR)
> a<-read.table("E:/R/data/file6.csv",",",header=T)
> x<-ts(a$temp,start=1949)
> x.ma<-SMA(x,n=5)
> plot(x,type="o")
> lines(x.ma,col=2,lwd=2)
#结果如图 4—3 所示
```

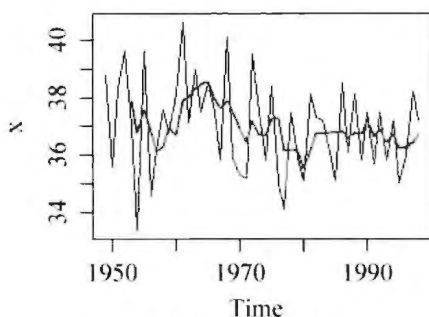


图 4—3 北京市最高气温 5 期移动平均拟合图

说明: 无滞后期的是观察值序列, 有滞后期的是 5 期移动平均拟合值序列。

二、指数平滑法

移动平均法实际上就是用简单的加权平均数作为某一期趋势的估计值。以 n 期移动平均为例, $\tilde{x}_t = \frac{1}{n}(x_t + x_{t-1} + \cdots + x_{t-n+1})$, 相当于用近 n 期的加权平均数作为最后一期趋势的估计值, 它们的权重都取为 $\frac{1}{n}$, 实际上也就是假定无论时间的远近, 这 n 期的观察值 $x_t, x_{t-1}, \cdots, x_{t-n+1}$ 对最后一期的影响都是一样的。但在实际生活中, 我们会发现对大多数随机事件而言, 一般都是近期的结果对现在的影响会大些, 远期的结果对现在的影响会小些。为了更好地反映这种影响作用, 我们将考虑时间间隔对事件发展的影响, 各期权重随时间间隔的增大而呈指数衰减。这就是指数平滑法的基本思想。不考虑季节 (周期) 因素的影响, 常用的进行趋势拟合的指数平滑公式有如下两种。

1. 简单指数平滑

$$\tilde{x}_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \cdots$$

式中, α 为平滑系数, 它满足 $0 < \alpha < 1$ 。

因为

$$\tilde{x}_{t-1} = \alpha x_{t-1} + \alpha(1 - \alpha)x_{t-2} + \alpha(1 - \alpha)^2 x_{t-3} + \cdots$$

所以 $\tilde{x}_t$ 又等价于

$$\tilde{x}_t = \alpha x_t + (1 - \alpha)\tilde{x}_{t-1}$$

简单指数平滑面临一个确定 $\tilde{x}_0$ 初始值的问题。我们有许多方法可以确定 $\tilde{x}_0$ 的初始值, 最简单的方法是指定 $\tilde{x}_0 = x_1$ 。

平滑系数 α 的值由研究人员根据经验给出。一般对于变化缓慢的序列, 常取较小的 α 值; 相反, 对于变化迅速的序列, 常取较大的 α 值。经验表明 α 的值介于 $0.05 \sim 0.3$ 之间, 修匀效果比较好。

指数平滑也是一种常用的平稳序列预测方法. 假定最后一期的观察值为 x_t , 那么使用指数平滑法, 向前预测 1 期的预测值为:

$$\hat{x}_t(1) = \tilde{x}_t = \alpha x_t + (1 - \alpha)\tilde{x}_{t-1} = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \cdots$$

从理论上我们可以证明使用指数平滑法预测任意 l 期的预测值都为常数.

证明: 已知 $\hat{x}_t(1) = \tilde{x}_t = \alpha x_t + (1 - \alpha)\tilde{x}_{t-1}$, 则

$$\hat{x}_t(2) = \alpha \hat{x}_t(1) + (1 - \alpha)\tilde{x}_t = \alpha \tilde{x}_t + (1 - \alpha)\tilde{x}_t = \tilde{x}_t$$

任取 l 期, 有

$$\hat{x}_t(l) = \alpha \hat{x}_t(l-1) + (1 - \alpha)\tilde{x}_t(l-1) = \tilde{x}_t, \quad l \geq 3$$

所以使用指数平滑最好只做 1 期预测.

2. Holt 两参数指数平滑

Holt 两参数指数平滑适用于对含有线性趋势的序列进行修匀. 它的基本思想是假定序列有一个比较固定的线性趋势——每期都递增 r 或递减 r , 那么第 t 期的估计值就应该等于第 $t-1$ 期的观察值加上每期固定的趋势变动值, 即

$$\hat{x}_t = x_{t-1} + r$$

但是由于随机因素的影响, 每期的递增或递减值不会恒定为 r , 它会随时间变化上下波动, 所以趋势序列实际上是一个随机序列 $\{r_t\}$, 因而

$$\hat{x}_t = x_{t-1} + r_{t-1}$$

考虑用第 t 期的观察值和第 t 期的估计值的加权平均数作为第 t 期的修匀值:

$$\begin{aligned} \tilde{x}_t &= \alpha x_t + (1 - \alpha)\hat{x}_t \\ \Rightarrow \tilde{x}_t &= \alpha x_t + (1 - \alpha)(x_{t-1} + r_{t-1}), \quad 0 < \alpha < 1 \end{aligned} \quad (4.1)$$

因为趋势序列 $\{r_t\}$ 也是一个随机序列, 为了使修匀序列 $\{\tilde{x}_t\}$ 更平滑, 我们对 $\{r_t\}$ 也进行一次修匀处理:

$$r_t = \beta(\tilde{x}_t - \tilde{x}_{t-1}) + (1 - \beta)r_{t-1}, \quad 0 < \beta < 1 \quad (4.2)$$

把式 (4.2) 代入式 (4.1), 就能得到比较光滑的修匀序列 $\{\tilde{x}_t\}$. 这就是 Holt 两参数指数平滑法的构造思想, 它的平滑公式为:

$$\begin{cases} \tilde{x}_t = \alpha x_t + (1 - \alpha)(\tilde{x}_{t-1} + r_{t-1}) \\ r_t = \beta(\tilde{x}_t - \tilde{x}_{t-1}) + (1 - \beta)r_{t-1} \end{cases}$$

式中, α, β 为两个平滑系数, 也称平滑参数, 它们满足 $0 < \alpha, \beta < 1$. 平滑系数的选择原则和简单指数平滑的选择原则一样.

和简单指数平滑一样, 我们也面临确定初始值的问题, 在此需要确定两个序列的初始值:

(1) 平滑序列的初始值 $\tilde{x}_0$. 最简单的方法是指定 $\tilde{x}_0 = x_1$.

(2) 趋势序列的初始值 r_0 . 和确定 $\tilde{x}_0$ 一样, 有许多方法确定它, 最简单的方法是: 任意指定一个区间长度 n , 用这段区间的平均趋势作为趋势初始值:

$$r_0 = \frac{x_{n+1} - x_1}{n}$$

假定最后一期的修匀值为 $\tilde{x}_T$, 那么使用 Holt 两参数指数平滑方法, 向前 l 期的预测值为:

$$\hat{x}_T(l) = \tilde{x}_T + l \cdot r_T$$

3. Holt-Winters 三参数指数平滑

假设要进行指数平滑的序列为 $\{x_t\}$, $\{x_t\}$ 序列既含有趋势又含有季节. 不妨记 a_t 为该序列的水平部分, b_t 为该序列的趋势部分, s_t 为该序列的季节因子 (假设一个季节周期长度为 π), 这个季节因子可以随着每年的具体情况波动. 因为引入了季节因子, 所以 Winters 1960 年在 Holt 两参数指数平滑的基础上构造了 Holt-Winters 三参数指数平滑模型.

如果季节和趋势是加法模型, 则 Holt-Winters 三参数指数平滑模型的构造如下:

$$a_t = \alpha(x_t - s_{t-\pi}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-\pi}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-\pi}$$

如果季节和趋势是乘法模型, 则 Holt-Winters 三参数指数平滑模型的构造如下:

$$a_t = \alpha(x_t/s_{t-\pi}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-\pi}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(x_t/a_t) + (1 - \gamma)s_{t-\pi}$$

在 R 语言中, HoltWinters 函数可以完成上述三种平滑趋势拟合. HoltWinters 函数的命令格式如下:

```
HoltWinters(x, alpha =, beta =, gamma =, seasonal = )
```

式中:

-x: 要进行指数平滑的序列名.

-alpha: 随机波动部分的参数.

-beta: 趋势部分的参数.

-gamma: 季节部分的参数.

这三个指数联合起来, 确定要拟合的指数平滑模型类型:

(1) 当 alpha 不指定, beta=F, gamma=F 时, 表示拟合简单指数平滑模型.

(2) 当 alpha 和 beta 不指定, gamma=F 时, 表示拟合 Holt 两参数指数平滑模型.

(3) 当三个参数都不指定时, 表示拟合 Holt-Winters 三参数指数平滑模型.

-seasonal: 当既含有季节又含有趋势时, 指定季节与趋势的关系.

(1) seasonal="additive" 表示加法关系, 这是系统默认选项.

(2) seasonal="multiplicative" 表示乘法关系.

【例 4—5】对 1964—1999 年中国纱年产量序列 (数据见表 A1—4) 进行 Holt 两参数指数平滑, 并预测未来 10 年的中国纱产量.

• 读入序列

```
> a<-read.table("E:/R/data/file4.csv",sep=" ",header=T)
> x<-ts(a$output,start=1964)
```

• 进行 Holt 两参数平滑

```
> x.fit<-HoltWinters(x,gamma=F)
> x.fit
```

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

HoltWinters(x = x, gamma = F)

Smoothing parameters:

alpha: 0.855644

beta : 0.158537

gamma: FALSE

Coefficients:

[,1]

a 565.55301

b 12.29066

• 绘制 Holt 两参数指数平滑拟合效果图

```
> plot(x.fit) #结果如图 4-4 所示
```

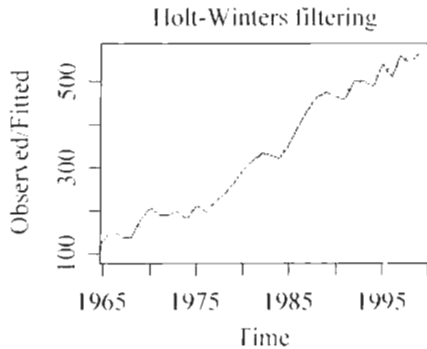


图 4—4 Holt 两参数指数平滑拟合效果图

说明：无滞后序列为观察值序列，滞后序列为 Holt 两参数指数平滑序列。

• 预测序列并绘制预测效果图

```
> x.fore<-forecast(x.fit,h=10)
> x.fore
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2000	577.8437	545.8643	609.8230	528.9355	626.7519
2001	590.1343	545.1051	635.1636	521.2681	659.0006
2002	602.4250	544.7497	660.1003	514.2182	690.6318
2003	614.7157	544.3115	685.1198	507.0417	722.3896

```

2004 627.0063 543.6025 710.4101 499.4512 754.5614
2005 639.2970 542.5375 736.0565 491.3161 787.2779
2006 651.5876 541.0751 762.1002 482.5733 820.6020
2007 663.8783 539.1960 788.5606 473.1931 854.5635
2008 676.1690 536.8922 815.4458 463.1635 889.1744
2009 688.4596 534.1622 842.7570 452.4821 924.4371

```

```
> plot(x.fore)    #结果如图 4—5 所示
```

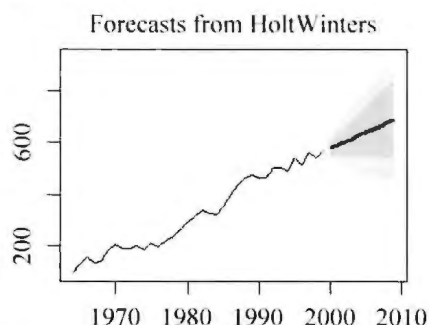


图 4—5 中国纱产量序列 Holt 两参数指数平滑序列预测图

【例 4—6】 对 1962 年 1 月—1975 年 12 月平均每头奶牛月产奶量序列 (数据见附表 A1—5) 进行 Holt-Winters 三参数指数平滑, 并预测未来两年平均每头奶牛的月度产奶量。

• 读入序列

```

> b<-read.table("E:/R/data/file5.csv",sep=",",header=T)
> x<-ts(b$milk,start=c(1962,1),frequency=12)

```

• 进行 Holt-winters 三参数指数平滑

```

> x.fit<-HoltWinters(x)
> x.fit

```

Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:

HoltWinters(x = x)

Smoothing parameters: \

alpha: 0.68933

beta : 0

gamma: 0.8362592

Coefficients:

[,1]

a 885.775547

b 1.278118

s1 -16.743296

```
s2 -59.730034
s3  47.492731
s4  56.203890
s5 115.537545
s6  84.554817
s7  39.580306
s8  -4.702033
s9 -54.554684
s10 -51.582594
s11 -85.953466
s12 -42.907363
```

• 绘制 Holt-Winters 三参数指数平滑拟合效果图

> plot(x.fit) #结果如图 4—6 所示

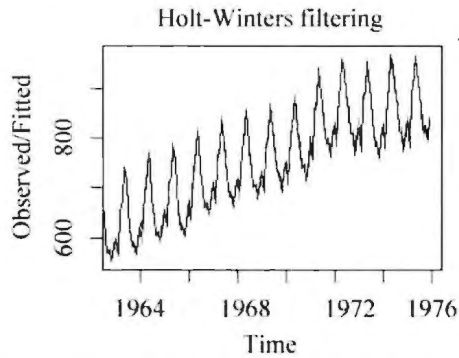


图 4—6 Holt-Winters 三参数指数平滑拟合效果图

说明：无滞后序列为观察值序列，滞后序列为 Holt 两参数指数平滑序列。

• 预测序列并绘制预测效果图

```
> x.fore<-forecast(x.fit,h=24)
```

> plot(x.fore) #结果如图 4—7 所示

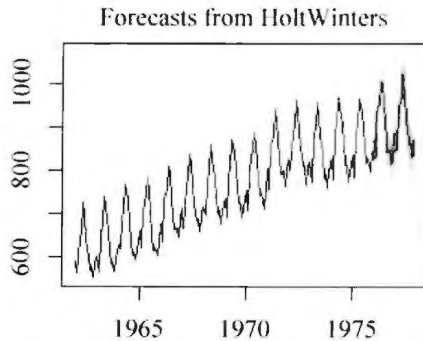


图 4—7 奶牛月度产奶量序列 Holt-Winters 三参数指数平滑序列预测图

4.4 季节效应分析

在日常生活中,我们可以见到许多有季节效应的时间序列,比如四季的气温、每个月的商品零售额、某自然景点每季度的旅游人数,等等,它们都会呈现出明显的季节变动规律.

我们还可以把“季节”广义化,凡是呈现出固定的周期性变化的事件,都称它具有“季节”效应.现在“季节”效应已经变成周期效应的代名词.而“季”也变成了周期内每一期的代名词.

【例 4—7】以北京市 1995—2000 年月平均气温序列为例(数据见表 A1—13),介绍季节效应分析的基本思想和具体操作步骤.

首先绘制该序列的时序图,如图 4—8 所示.

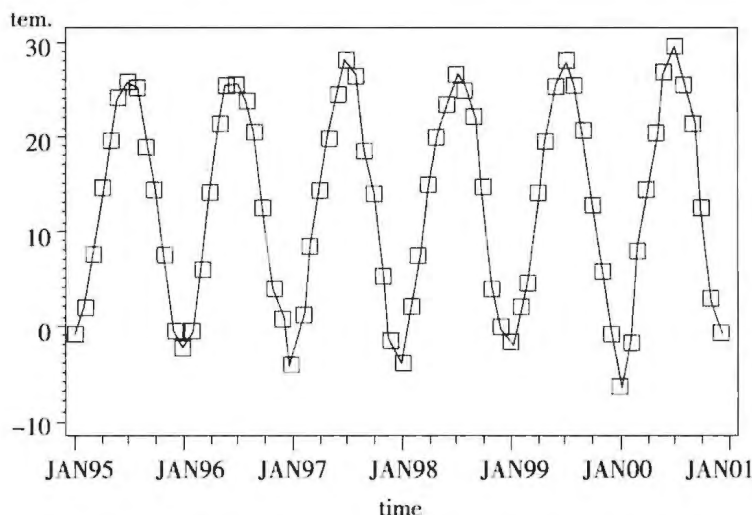


图 4—8 北京市 1995—2000 年月平均气温序列时序图

通过时序图,我们发现北京市 1995—2000 年每月的平均气温随着季节的变动有非常规律的变化.气温的波动主要受到两个因素的影响:一个是季节效应;一个是随机波动.

假如没有季节效应的影响,北京市的气温应该始终在某个均值附近随机波动,季节效应的存在使得气温会在不同年份的相同月份呈现出相似的性质.为便于得到数量化的季节信息,我们构造出季节指数的概念.所谓季节指数,就是用简单平均法计算的周期内各时期季节性影响的相对数.

以本例的数据结构为例,我们需要求出每个月的季节指数 $S_1, S_2, \dots, S_{12}$, 那么第 i 年第 j 个月的平均气温可以表示为:

$$x_{ij} = \bar{x} \cdot S_j + I_{ij}, \quad j = 1, 2, \dots, 12$$

式中, $\bar{x}$ 为各月总平均气温; S_j 为第 j 个月的季节指数; I_{ij} 为第 i 年第 j 个月气温的随机波动.

季节指数的计算分为三步.

第一步: 计算周期内各期平均数, 得到长期以来该时期的平均水平. 假定序列的数据结构为 m 期为一周期, 共有 n 个周期, 则

$$\bar{x}_k = \frac{\sum_{i=1}^n x_{ik}}{n}, \quad k = 1, 2, \dots, m$$

第二步: 计算总平均数.

$$\bar{x} = \frac{\sum_{i=1}^n \sum_{k=1}^m x_{ik}}{nm}$$

第三步: 用时期平均数除以总平均数就可以得到各时期的季节指数 $S_k (k = 1, 2, \dots, m)$, 即

$$S_k = \frac{\bar{x}_k}{\bar{x}}, \quad k = 1, 2, \dots, m$$

这就是季节指数的构造方法. 季节指数反映了该季度与总平均值之间的一种比较稳定的关系. 如果这个比值大于 1, 就说明该季度的值常常会高于总平均值; 如果这个比值小于 1, 就说明该季度的值常常低于总平均值; 如果序列的季节指数都近似等于 1, 那就说明该序列没有明显的季节效应.

下面我们来具体计算本例的季节指数. 计算结果如表 4—3 所示.

表 4—3

年 月	平均气温 (°C)						月平均	季节指数
	1995	1996	1997	1998	1999	2000	$\bar{x}_j$	S_j
1	-0.7	-2.2	-3.8	-3.9	-1.6	-6.4	-3.10	-0.238
2	2.1	-0.4	1.3	2.4	2.2	-1.5	1.02	0.078
3	7.7	6.2	8.7	7.6	4.8	8.1	7.18	0.551
4	14.7	14.3	14.5	15.0	14.4	14.6	14.58	1.119
5	19.8	21.6	20.0	19.9	19.5	20.4	20.20	1.550
6	24.3	25.4	24.6	23.6	25.4	26.7	25.00	1.919
7	25.9	25.5	28.2	26.5	28.1	29.6	27.30	2.095
8	25.4	23.9	26.6	25.1	25.6	25.7	25.38	1.948
9	19.0	20.7	18.6	22.2	20.9	21.8	20.53	1.576
10	14.5	12.8	14.0	14.8	13.0	12.6	13.62	1.045
11	7.7	4.2	5.4	4.0	5.9	3.0	5.03	0.386
12	-0.4	0.9	-1.5	0.1	-0.6	-0.6	-0.35	-0.027
总平均 $\bar{x}=13.03$								

把季节指数绘制成图 (见图 4—9), 可以帮助我们更清楚地看到月度变迁对气温的影响.

本例中, 7 月的季节指数最大, 说明 7 月是北京最热的月份; 1 月的季节指数最小, 说明 1 月是北京最冷的月份; 最接近于年平均气温的是 10 月, 这个月的平均气温是

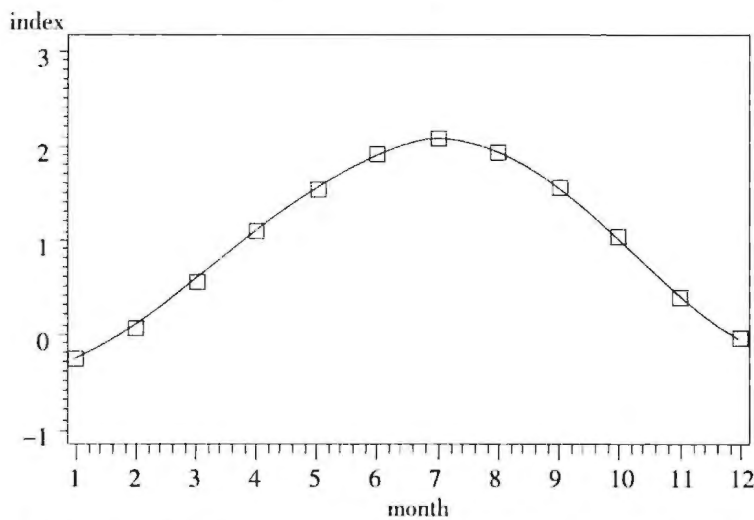


图 4—9 北京市 1995—2000 年月平均气温序列季节指数图

13.62°C, 和年平均气温的差异只有 4.5%.

由任意两个季节指数, 我们可以得到单纯的“季节”变动对事件影响的大小. 比如 3 月的季节指数为 0.551, 而 6 月的季节指数为 1.919, 这说明 6 月的平均气温一般是 3 月的 3.48 倍, 假如明年 3 月北京市的平均气温为 7°C, 那么根据季节效应我们有理由预测在大气环境不发生大的改变的前提下, 北京市 6 月的平均气温应该在 21°C 左右.

4.5 综合分析

在上面两节里我们分别介绍了单纯的趋势分析方法和单纯的季节效应分析方法. 在这一节里主要介绍既有趋势起伏变动又有季节效应的复杂序列的分析方法.

要对趋势起伏变动和季节效应同时进行分析, 首先必须了解它们之间的相互作用关系. 但是, 通常我们并不知道它们到底是怎样综合作用而影响序列变化的, 这时只能根据序列的表现, 选择一些经验模型来估计各因素之间的相互作用关系. 常用的模型有:

(1) 加法模型

$$x_t = T_t + S_t + I_t \tag{4.3}$$

(2) 乘法模型

$$x_t = T_t \cdot S_t \cdot I_t \tag{4.4}$$

式中, T_t 代表序列的长期趋势波动; S_t 代表序列的季节性 (周期性) 变化; I_t 代表随机波动.

在确定性影响很强劲时, 选择合适的确定性模型通常会得到非常不错的分析预测效果.

R 语言中进行确定性因素分解的函数是 `decompose`, 该函数的命令格式如下:

```
decompose(x, type = )
```

式中:

-x: 序列名.

-type: 指定是加法模型还是乘法模型.

(1) 加法模型, type="additive", 这是系统默认设置.

(2) 乘法模型, type="multiplicative".

【例 4—8】 对 1993—2000 年中国社会消费品零售总额序列 (数据见表 A1—14) 进行确定性因素分解.

• 读入序列, 并绘制时序图

```
> c<-read.table("E:/R/data/file14.csv",sep=",",header=T)
```

```
> x<-ts(c$sales,start=c(1993,1),frequency=12)
```

```
> plot(x) #结果如图 4—10 所示
```

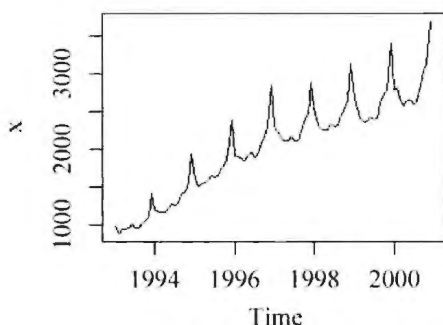


图 4—10 中国社会消费品零售总额序列时序图

• 确定性因素分解

```
> x.fit<-decompose(x,type="mult")
```

• 查看季节指数, 并绘制季节指数图

```
> x.fit$figure
```

```
[1] 1.0439030 0.9939439 0.9592626 0.9397644 0.9438897 0.9588798 0.9286603
```

```
[8] 0.9260807 0.9814290 1.0074970 1.0472403 1.2694493
```

```
> plot(x.fit$figure,type="o") #结果如图 4—11 所示
```

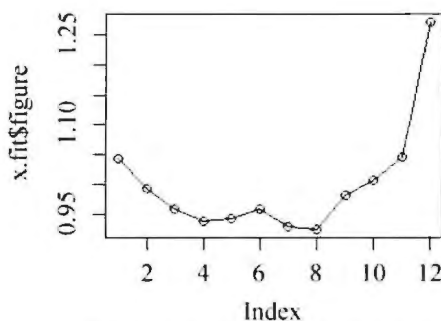


图 4—11 中国社会消费品零售总额序列季节指数图

• 查看趋势拟合值, 并绘制趋势拟合图

```
> x.fit$trend
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
1993	NA	NA	NA	NA	NA	NA	1028.696	1048.900
1994	1153.913	1179.492	1208.621	1240.533	1275.733	1316.087	1354.700	1385.483
1995	1537.554	1567.687	1597.308	1627.879	1659.350	1694.296	1726.138	1756.412
1996	1890.954	1913.637	1938.921	1966.329	1994.879	2028.796	2063.733	2092.137
1997	2190.733	2207.475	2221.812	2236.637	2251.825	2260.071	2272.325	2287.071
1998	2350.200	2366.608	2383.846	2400.158	2416.213	2434.837	2449.933	2464.292
1999	2513.642	2521.579	2532.308	2547.687	2561.737	2578.554	2594.675	2610.475
2000	2707.971	2726.379	2746.162	2768.446	2793.929	2818.963	NA	NA
2001	NA							

	Sep	Oct	Nov	Dec
1993	1069.542	1088.471	1108.496	1130.450
1994	1414.425	1445.429	1476.679	1507.108
1995	1787.517	1813.887	1839.679	1866.312
1996	2116.017	2137.537	2156.521	2173.546
1997	2297.142	2309.683	2322.567	2335.829
1998	2479.100	2488.579	2497.033	2505.433
1999	2630.912	2649.208	2669.550	2689.933
2000	NA	NA	NA	NA
2001				

```
> plot(x.fit$trend) #结果如图 4—12 所示
```

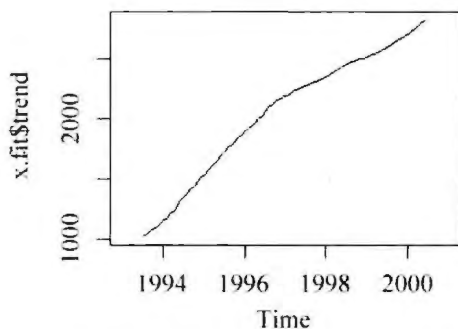


图 4—12 中国社会消费品零售总额序列趋势拟合图

• 查看随机波动 (残差) 值, 并绘制残差图

```
> x.fit$random
```

	Jan	Feb	Mar	Apr	May	Jun	Jul
1993	NA	NA	NA	NA	NA	NA	1.0088882
1994	0.9897286	0.9917699	1.0069996	1.0039379	1.0079296	1.0151591	0.9947887
1995	0.9982198	0.9571982	1.0006930	1.0123394	1.0122310	1.0092784	1.0128538
1996	0.9671360	1.0048115	1.0000891	1.0037414	1.0081543	1.0106040	0.9854906
1997	1.0006939	1.0088390	0.9998115	0.9993280	0.9918727	0.9988758	0.9963425
1998	1.0391783	0.9804972	0.9969239	0.9987216	0.9932308	0.9962666	1.0048103


```
1999 1.0145206 1.0128044 0.9892766 0.9843682 0.9776685 0.9823163 0.9878519
2000 0.9815488 1.0351058 0.9972322 0.9885893 0.9999390 0.9785257 NA
2001 NA
      Aug      Sep      Oct      Nov      Dec
1993 0.9880930 0.9748693 0.9584809 0.9492950 0.9863775
1994 1.0022841 1.0057935 0.9916461 1.0047607 1.0099323
1995 1.0064677 1.0009573 0.9948091 1.0044711 1.0085729
1996 0.9891157 1.0032647 0.9975563 1.0140385 1.0323984
1997 0.9935731 0.9933992 1.0090249 1.0092977 0.9718346
1998 1.0142266 1.0041262 1.0114724 1.0142280 0.9845558
1999 0.9972656 1.0086157 1.0280362 0.9949349 0.9973544
2000 NA      NA      NA      NA      NA
2001
```

> plot(x.fit\$random) #结果如图 4—13 所示

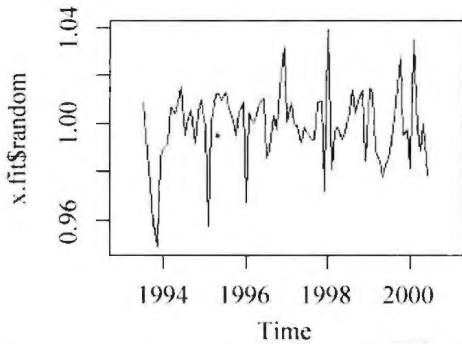


图 4—13 中国社会消费品零售总额序列因素分解残差图

我们在前面详细介绍了因素分解模型每一因素的获取和显示. 如果研究人员并不需要对某个因素进行深入考察或单独使用, 可以直接让 R 提供信息集成显示和图形集成输出, 相关命令如下:

```
> x.fit
$x
      Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct
1993  977.5   892.5   942.3   941.3   962.2  1005.7   963.8   959.8  1023.3  1051.1
1994  1192.2  1162.7  1167.5  1170.4  1213.7  1281.1  1251.5  1286.0  1396.2  1444.1
1995  1602.2  1491.5  1533.3  1548.7  1585.4  1639.7  1623.6  1637.1  1756.0  1818.0
1996  1909.1  1911.2  1860.1  1854.8  1898.3  1966.0  1888.7  1916.4  2083.5  2148.3
1997  2288.5  2213.5  2130.9  2100.5  2108.2  2164.7  2102.5  2104.4  2239.6  2348.0
1998  2549.5  2306.4  2279.7  2252.7  2265.2  2326.0  2286.1  2314.6  2443.1  2536.0
1999  2662.1  2538.4  2403.1  2356.8  2364.0  2428.8  2380.3  2410.9  2604.3  2743.9
2000  2774.7  2805.0  2627.0  2572.0  2637.0  2645.0  2597.0  2636.0  2854.0  3029.0
2001 NA
      Nov      Dec
1993 1102.0  1415.5
```

1994	1553.8	1932.2						
1995	1935.2	2389.5						
1996	2290.1	2848.6						
1997	2454.9	2881.7						
1998	2652.2	3131.4						
1999	2781.5	3405.7						
2000	3108.0	3680.0						
2001								
\$seasonal								
	Jan	Feb	Mar	Apr	May	Jun	Jul	
1993	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1994	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1995	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1996	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1997	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1998	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
1999	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
2000	1.0439030	0.9939439	0.9592626	0.9397644	0.9438897	0.9588798	0.9286603	
2001	1.0439030							
	Aug	Sep	Oct	Nov	Dec			
1993	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1994	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1995	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1996	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1997	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1998	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
1999	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
2000	0.9260807	0.9814290	1.0074970	1.0472403	1.2694493			
2001								
\$trend								
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
1993	NA	NA	NA	NA	NA	NA	1028.696	1048.900
1994	1153.913	1179.492	1208.621	1240.533	1275.733	1316.087	1354.700	1385.483
1995	1537.554	1567.687	1597.308	1627.879	1659.350	1694.296	1726.138	1756.412
1996	1890.954	1913.637	1938.921	1966.329	1994.879	2028.796	2063.733	2092.137
1997	2190.733	2207.475	2221.812	2236.637	2251.825	2260.071	2272.325	2287.071
1998	2350.200	2366.608	2383.846	2400.158	2416.213	2434.837	2449.933	2464.292
1999	2513.642	2521.579	2532.308	2547.687	2561.737	2578.554	2594.675	2610.475
2000	2707.971	2726.379	2746.162	2768.446	2793.929	2818.963	NA	NA
2001	NA							
	Sep	Oct	Nov	Dec				

```

1993 1069.542 1088.471 1108.496 1130.450
1994 1414.425 1445.429 1476.679 1507.108
1995 1787.517 1813.887 1839.679 1866.312
1996 2116.017 2137.537 2156.521 2173.546
1997 2297.142 2309.683 2322.567 2335.829
1998 2479.100 2488.579 2497.033 2505.433
1999 2630.912 2649.208 2669.550 2689.933
2000      NA      NA      NA      NA
2001

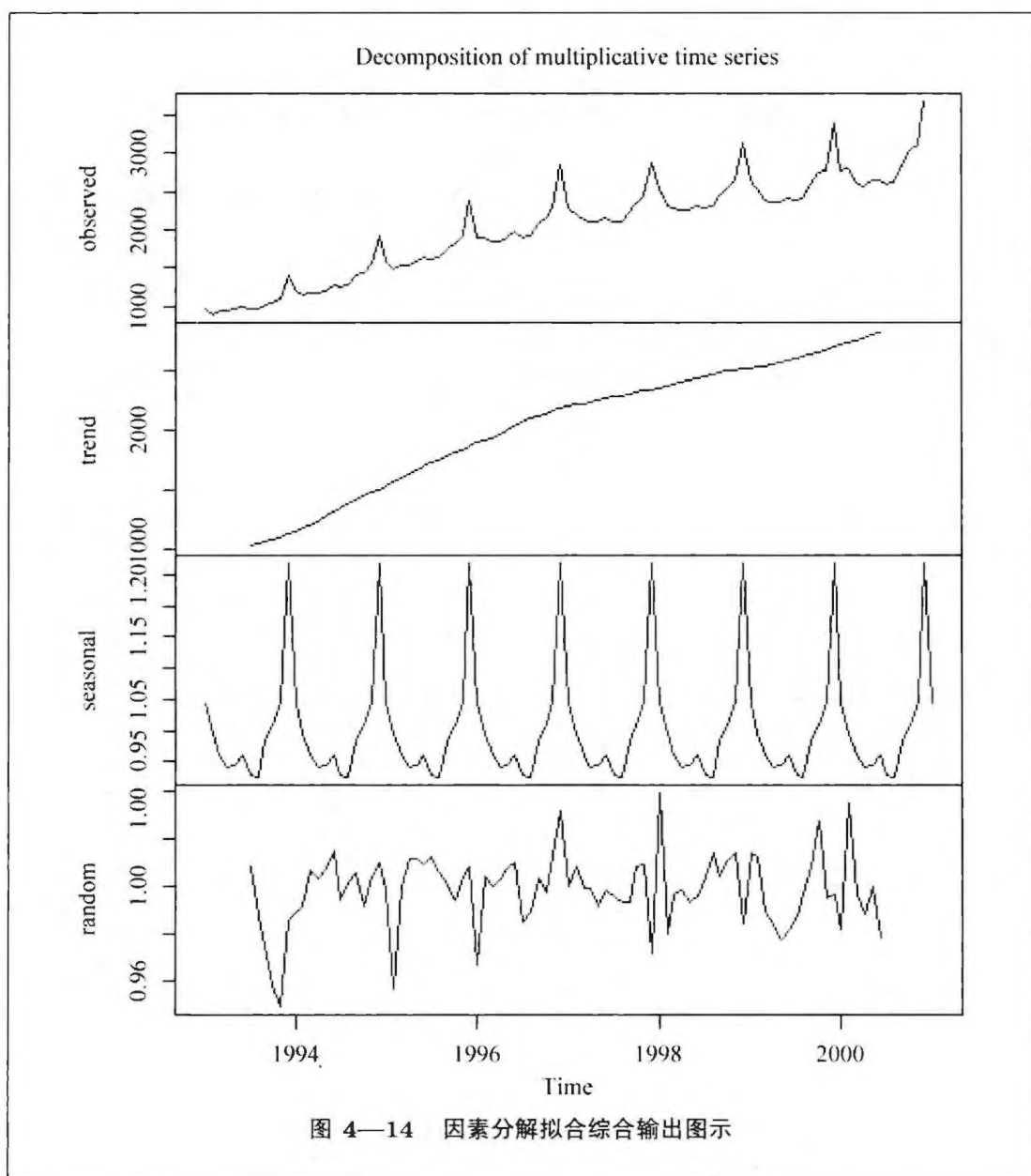
$random
      Jan      Feb      Mar      Apr      May      Jun      Jul
1993      NA      NA      NA      NA      NA      NA 1.0088882
1994 0.9897286 0.9917699 1.0069996 1.0039379 1.0079296 1.0151591 0.9947887
1995 0.9982198 0.9571982 1.0006930 1.0123394 1.0122310 1.0092784 1.0128538
1996 0.9671360 1.0048115 1.0000891 1.0037414 1.0081543 1.0106040 0.9854906
1997 1.0006939 1.0088390 0.9998115 0.9993280 0.9918727 0.9988758 0.9963425
1998 1.0391783 0.9804972 0.9969239 0.9987216 0.9932308 0.9962666 1.0048103
1999 1.0145206 1.0128044 0.9892766 0.9843682 0.9776685 0.9823163 0.9878519
2000 0.9815488 1.0351058 0.9972322 0.9885893 0.9999390 0.9785257      NA
2001      NA
      Aug      Sep      Oct      Nov      Dec
1993 0.9880930 0.9748693 0.9584809 0.9492950 0.9863775
1994 1.0022841 1.0057935 0.9916461 1.0047607 1.0099323
1995 1.0064677 1.0009573 0.9948091 1.0044711 1.0085729
1996 0.9891157 1.0032647 0.9975563 1.0140385 1.0323984
1997 0.9935731 0.9933992 1.0090249 1.0092977 0.9718346
1998 1.0142266 1.0041262 1.0114724 1.0142280 0.9845558
1999 0.9972656 1.0086157 1.0280362 0.9949349 0.9973544
2000      NA      NA      NA      NA      NA
2001

$figure
[1] 1.0439030 0.9939439 0.9592626 0.939764 4 0.9438897 0.9588798 0.9286603
[8] 0.9260807 0.9814290 1.0074970 1.0472403 1.2694493

$type
[1] "multiplicative"

attr("class")
[1] "decomposed.ts"
> plot(x.fit)      #结果如图 4 14 所示

```



4.6 习 题

1. 使用 4 期移动平均作预测, 求在 2 期预测值 $\hat{x}_T(2)$ 中 x_{T-3} 与 x_{T-1} 前面的系数分别等于多少.
2. 使用指数平滑法得到 $\tilde{x}_{t-1} = 5, \tilde{x}_{t+1} = 5.26$, 已知序列观察值 $x_t = 5.25, x_{t+1} = 5.5$, 求指数平滑系数 α .
3. 有一 20 期的观察值序列 $\{x_t\}$ 为:

10 11 12 10 11 14 12 13 11 15 12 14 13 12 14 12 10 10 11 13

- (1) 使用 5 期移动平均法预测 $\hat{x}_{22}$.
- (2) 使用指数平滑法确定 $\hat{x}_{22}$, 其中平滑系数为 $\alpha = 0.4$.
- (3) 假设 a 为 5 期移动平均法预测 $\hat{x}_{22}$ 中 x_{20} 前的系数, b 为在 $\hat{x}_{22}$ 的平滑系数为 $\alpha = 0.4$ 的指数平滑法预测中 x_{20} 前的系数, 求 $b - a$.
- 4. 现有序列 $\{x_i = t, t = 1, 2, \cdots\}$, 使用平滑系数为 α 的指数平滑法修匀该序列. 假定 $\tilde{x}_0 = 0$, 求 $\lim_{t \rightarrow \infty} \frac{\tilde{x}_t}{t}$.
- 5. 我国 1949—2008 年年末人口总数 (单位: 万人) 序列如表 4—4 所示 (行数据).

表 4—4

54 167	55 196	56 300	57 482	58 796	60 266	61 465	62 828
64 653	65 994	67 207	66 207	65 859	67 295	69 172	70 499
72 538	74 542	76 368	78 534	80 671	82 992	85 229	87 177
89 211	90 859	92 420	93 717	94 974	96 259	97 542	98 705
100 072	101 654	103 008	104 357	105 851	107 507	109 300	111 026
112 704	114 333	115 823	117 171	118 517	119 850	121 121	122 389
123 626	124 761	125 786	126 743	127 627	128 453	129 227	129 988
130 756	131 448	132 129	132 802				

选择适当模型拟合该序列的长期趋势, 并作 5 期预测.

- 6. 艾奥瓦州 1948—1979 年非农产品季度收入数据如表 4—5 所示 (行数据).

表 4—5

601	604	620	626	641	642	645	655	682	678	692	707
736	753	763	775	775	783	794	813	823	826	829	831
830	838	854	872	882	903	919	937	927	962	975	995
1 001	1 013	1 021	1 028	1 027	1 048	1 070	1 095	1 113	1 143	1 154	1 173
1 178	1 183	1 205	1 208	1 209	1 223	1 238	1 245	1 258	1 278	1 294	1 314
1 323	1 336	1 355	1 377	1 416	1 430	1 455	1 480	1 514	1 545	1 589	1 634
1 669	1 715	1 760	1 812	1 809	1 828	1 871	1 892	1 946	1 983	2 013	2 045
2 048	2 097	2 140	2 171	2 208	2 272	2 311	2 349	2 362	2 442	2 479	2 528
2 571	2 634	2 684	2 790	2 890	2 964	3 085	3 159	3 237	3 358	3 489	3 588
3 624	3 719	3 821	3 934	4 028	4 129	4 205	4 349	4 463	4 598	4 725	4 827
4 939	5 067	5 231	5 408	5 492	5 653	5 828	5 965				

选择适当模型拟合该序列长期趋势.

- 7. 某地区 1962—1970 年平均每头奶牛的月度产奶量数据 (单位: 磅) 如表 4—6 所示 (行数据).

表 4—6

589	561	640	656	727	697	640	599	568	577	553	582
600	566	653	673	742	716	660	617	583	587	565	598
628	618	688	705	770	736	678	639	604	611	594	634
658	622	709	722	782	756	702	653	615	621	602	635
677	635	736	755	811	798	735	697	661	667	645	688
713	667	762	784	837	817	767	722	681	687	660	698
717	696	775	796	858	826	783	740	701	706	677	711
734	690	785	805	871	845	801	764	725	723	690	734
750	707	807	824	886	859	819	783	740	747	711	751

(1) 绘制该序列时序图, 直观考察该序列的特点.

(2) 使用因素分解方法拟合该序列的发展, 并预测下一年该地区奶牛的月度产奶量.

8. 某城市 1980 年 1 月至 1995 年 8 月每月屠宰生猪数量 (单位: 头) 如表 4—7 所示 (行数据).

表 4—7

76 378	71 947	33 873	96 428	105 084	95 741	110 647	100 331	94 133	103 055
90 595	101 457	76 889	81 291	91 643	96 228	102 736	100 264	103 491	97 027
95 240	91 680	101 259	109 564	76 892	85 773	95 210	93 771	98 202	97 906
100 306	94 089	102 680	77 919	93 561	117 062	81 225	88 357	106 175	91 922
104 114	109 959	97 880	105 386	96 479	97 580	109 490	110 191	90 974	98 981
107 188	94 177	115 097	113 696	114 532	120 110	93 607	110 925	103 312	120 184
103 069	103 351	111 331	106 161	111 590	99 447	101 987	85 333	86 970	100 561
89 543	89 265	82 719	79 498	74 846	73 819	77 029	78 446	86 978	75 878
69 571	75 722	64 182	77 357	63 292	59 380	78 332	72 381	55 971	69 750
85 472	70 133	79 125	85 805	81 778	86 852	69 069	79 556	88 174	66 698
72 258	73 445	76 131	86 082	75 443	73 969	78 139	78 646	66 269	73 776
80 034	70 694	81 823	75 640	75 540	82 229	75 345	77 034	78 589	79 769
75 982	78 074	77 588	84 100	97 966	89 051	93 503	84 747	74 531	91900
81 635	89 797	81 022	78 265	77 271	85 043	95 418	79 568	103 283	95 770
91 297	101 244	114 525	101 139	93 866	95 171	100 183	103 926	102 643	108 387
97 077	90 901	90 336	88 732	83 759	99 267	73 292	78 943	94 399	92 937
90 130	91 055	106 062	103 560	104 075	101 783	93 791	102 313	82 413	83 534
109 011	96 499	102 430	103 002	91 815	99 067	110 067	101 599	97 646	104 930
88 905	89 936	106 723	84 307	114 896	106 749	87 892	100 506		

选择适当模型拟合该序列的发展, 并预测 1995 年 9 月至 1997 年 9 月该城市生猪屠宰数量.

第 4 章介绍了非平稳序列的确定性因素分解方法. 这种方法具有原理简单、操作简便、易于解释等优点, 在宏观经济管理与预测领域有广泛的应用.

但是随着研究方法的深入和研究领域的拓宽, 人们发现确定性因素分解方法还存在一些问题, 主要有:

(1) 确定性因素分解方法只能提取强劲的确定性信息, 对随机性信息浪费严重.

(2) 确定性因素分解方法把所有序列的变化都归结为四因素的综合影响, 却始终无法提供明确、有效的方法判断各大因素之间确切的作用关系.

这些问题导致确定性因素分解方法不能充分提取观察值序列中的有效信息, 导致模型拟合精度通常不够理想. 随机时序分析方法的发展就是为了弥补确定性因素分解方法的不足, 为人们提供更加丰富、更加精确的时序分析工具.

5.1 差分运算

5.1.1 差分运算的实质

拿到观察值序列之后, 无论是采用确定性时序分析方法还是随机时序分析方法, 分析的第一步都是要通过有效的手段提取序列中所蕴涵的确定性信息.

确定性信息的提取方法非常多, 前面介绍的构造季节指数、拟合长期趋势模型、移动平均、指数平滑等诸多方法都是确定性信息提取方法, 但是它们对确定性信息的提取都不够充分.

Box 和 Jenkins 在 *Time Series Analysis: Forecasting and Control* 一书中特别强调差分方法的使用, 他们使用大量的案例分析证明差分方法是一种非常简便、有效的确定性信息提取方法. 而 Cramer 分解定理则在理论上保证了适当阶数的差分一定可以充分提取确定性信息.

根据 Cramer 分解定理, 方差齐性非平稳序列都可以分解为如下形式:

$$x_t = \sum_{j=0}^d \beta_j \cdot t^j + \Psi(B)a_t$$

式中, $\{a_t\}$ 为零均值白噪声序列.

离散序列的 d 阶差分就相当于连续变量的 d 阶求导, 显然, 在 Cramer 分解定理的保证下, d 阶差分就可以将 $\{x_t\}$ 中蕴涵的确定性信息充分提取.

$$\nabla^d \sum_{j=0}^d \beta_j t^j = c, \quad c \text{ 为某一常数}$$

展开 1 阶差分, 有

$$\nabla x_t = x_t - x_{t-1}$$

等价于

$$x_t = x_{t-1} + \nabla x_t$$

这意味着 1 阶差分实质上就是一个自回归过程, 它是用延迟一期的历史数据 $\{x_{t-1}\}$ 作为自变量来解释当期序列值 $\{x_t\}$ 的变动状况, 差分序列 $\{\nabla x_t\}$ 度量的是 $\{x_t\}$ 1 阶自回归过程中产生的随机误差的大小.

展开任意一个 d 阶差分, 有

$$\nabla^d x_t = (1 - B)^d x_t = \sum_{i=0}^d (-1)^i C_d^i x_{t-i}$$

它实质上就是一个 d 阶自回归过程:

$$x_t = \sum_{i=1}^d (-1)^{i+1} C_d^i x_{t-i} + \nabla^d x_t$$

这意味着差分运算的实质是使用自回归的方式提取确定性信息.

R 语言中, diff 函数可以完成各种差分运算. diff 函数的命令格式如下:

```
diff( x, lag= , differences= )
```

式中:

-x: 变量名.

-lag: 差分的步长. 不特意指定, 系统默认 lag=1.

-differences: 差分次数. 不特意指定, 系统默认 differences=1

根据 diff 函数的参数定义, 如果差分命令写成 diff(x, d, k), 意思是进行 k 次 d 步差分. 常用的差分运算相关命令为:

1 阶差分: diff(x);

2 阶差分: diff(x, 1, 2);

k 阶差分: diff(x, 1, k);

d 步差分: diff(x, d, 1), 或者简写为 diff(x, d);

一阶差分后再进行 d 步差分: diff(diff(x), d).

5.1.2 差分方式的选择

实践中, 我们会根据序列的不同特点选择合适的差分方式, 常见情况有以下三种.

一、序列蕴涵显著的线性趋势, 1 阶差分就可以实现趋势平稳

【例 5—1】 在例 2—1 的分析中, 我们知道 1964—1999 年中国纱年产量序列蕴涵一个近似线性的递增趋势. 现对该序列进行 1 阶差分运算, 考察差分运算对该序列线性趋势信息的提取作用.

```
> a<-read.table("E:/R/data/file4.csv",",",header=T)
> x<-ts(a$output,start=1964)
> x.dif<-diff(x)
> plot(x.dif)    #绘制 1 阶差分序列时序图, 结果如图 5—1 所示
```

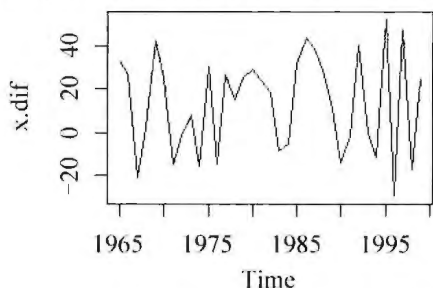


图 5—1 1964—1999 年中国纱年产量 1 阶差分序列时序图

时序图清晰地显示, 1 阶差分运算已经非常成功地从原序列中提取出了线性趋势, 差分后序列呈现出非常平稳的随机波动.

二、序列蕴涵曲线趋势, 通常低阶 (2 阶或 3 阶) 差分就可以提取出曲线趋势的影响

【例 5—2】 尝试提取 1950—1999 年北京市民用车辆拥有量序列 (数据见表 A1—15) 的确定性信息.

• 读入数据, 并绘制时序图

```
> b<-read.table("E:/R/data/file15.csv",",",header=T)
> x<-ts(b$vehicle,start=1950)
> plot(x)    #结果如图 5—2 所示
```

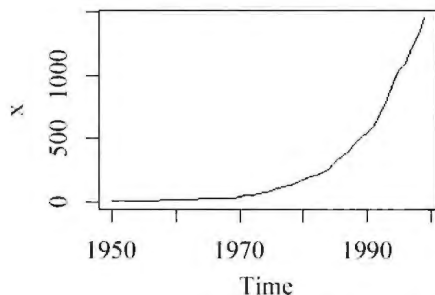


图 5—2 1950—1999 年北京市民用车辆拥有量序列时序图

• 1 阶差分, 并绘制差分后序列时序图

```
> x.dif<-diff(x)
```

```
> plot(x.dif)    #结果如图 5—3 所示
```

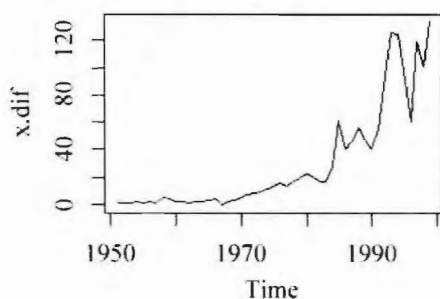


图 5—3 1950—1999 年北京市民用车辆拥有量序列 1 阶差分时序图

• 2 阶差分, 并绘制差分后序列时序图

```
> x.dif2<-diff(x,1,2)
```

```
> plot(x.dif2)    #结果如图 5—4 所示
```

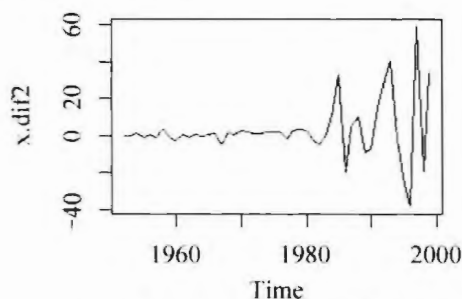


图 5—4 1950—1999 年北京市民用车辆拥有量序列 2 阶差分时序图

从时序图图 5—2 中我们可以清楚地看到该序列蕴涵曲线递增的长期趋势. 对该序列进行 1 阶差分运算:

$$\nabla x_t = x_t - x_{t-1}$$

1 阶差分序列时序图显示, 1 阶差分提取了原序列中的部分长期趋势, 但长期趋势信息提取不充分, 1 阶差分后序列中仍蕴涵长期递增的趋势, 于是对 1 阶差分后序列再做一次差分运算:

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$$

2 阶差分后序列 $\{\nabla^2 x_t\}$ 的时序图 (见图 5—4) 显示, 2 阶差分比较充分地提取了原序列中蕴涵的长期趋势, 使得差分后序列不再呈现确定性趋势.

三、蕴涵固定周期的序列

对于蕴涵固定周期的序列进行步长为周期长度的差分运算, 通常可以较好地提取周期信息.

【例 5—3】 利用差分运算提取 2001 年 1 月至 2014 年 11 月中国月度出口额序列中的确定性信息 (数据见表 A1—16).

• 读入数据, 并绘制时序图

```
> c<-read.table("E:/R/data/file16.csv",",",header=T)
> x<-ts(c$export,start=c(2001,1),frequency=12)
> plot(x) #结果如图 5—5 所示
```

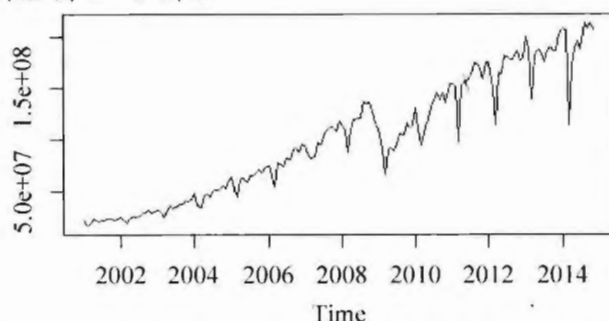


图 5—5 2001 年 1 月至 2014 年 11 月中国月度出口额序列时序图

• 1 阶差分, 并绘制差分后序列时序图

```
> x.dif<-diff(x)
> plot(x.dif) #结果如图 5—6 所示
```

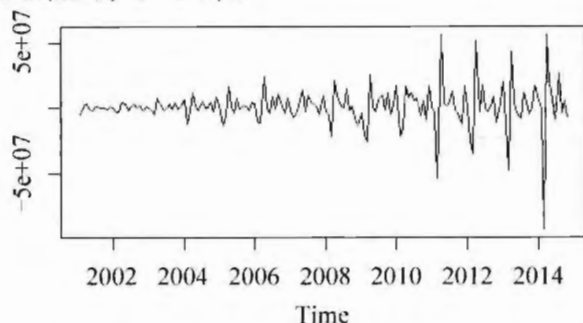


图 5—6 2001 年 1 月至 2014 年 11 月中国月度出口额序列 1 阶差分时序图

• 1 阶差分加 12 步差分, 并绘制差分后时序图

```
> x.dif1.12<-diff(diff(x),12)
> plot(x.dif1.12) #结果如图 5—7 所示
```

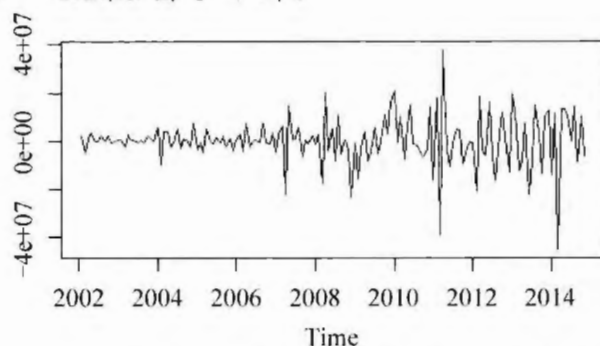


图 5—7 2001 年 1 月至 2014 年 11 月中国月度出口额序列 1 阶 12 步差分后序列时序图

时序图显示该序列具有一个线性递增的长期趋势和一个周期长度为一年的稳定的季节变动. 所以对原序列先做 1 阶差分, 提取线性递增趋势:

$$\nabla x_t = x_t - x_{t-1}$$

图 5—6 显示, 1 阶差分后线性递增信息被提取, 1 阶差分后序列具有稳定的季节波动和随机波动. 对 1 阶差分后序列再进行 12 步的周期差分, 提取季节波动信息:

$$\nabla_{12} \nabla x_t = \nabla x_t - \nabla x_{t-12}$$

图 5—7 显示, 周期差分可以非常好地提取周期信息. 至此, 差分运算已经比较充分地提取了原序列中蕴涵的季节效应和长期趋势效应等确定性信息.

5.1.3 过差分

理论上, 足够多次的差分运算可以充分地提取原序列中的非平稳确定性信息. 但应当注意的是, 差分运算的阶数并不是越多越好. 因为差分运算是一种对信息的提取、加工过程, 每次差分都会有信息的损失, 所以在实际应用中差分运算的阶数要适当, 避免出现过度差分 (简称过差分) 的现象.

【例 5—4】 假定线性非平稳序列 $\{x_t\}$ 形如

$$x_t = \beta_0 + \beta_1 t + a_t$$

式中, $E(a_t) = 0$, $\text{Var}(a_t) = \sigma^2$, $\text{Cov}(a_t, a_{t-1}) = 0$, $\forall t \geq 1$.

对 x_t 作 1 阶差分:

$$\begin{aligned} \nabla x_t &= x_t - x_{t-1} \\ &= \beta_0 + \beta_1 t + a_t - [\beta_0 + \beta_1(t-1) + a_{t-1}] \\ &= \beta_1 + a_t - a_{t-1} \end{aligned}$$

显然差分后序列 $\{\nabla x_t\}$ 为平稳序列. 这说明 1 阶差分运算有效地提取了 $\{x_t\}$ 中的非平稳确定性信息.

对 1 阶差分后序列 ∇x_t 再作一次差分:

$$\begin{aligned} \nabla^2 x_t &= \nabla x_t - \nabla x_{t-1} \\ &= \beta_1 + a_t - a_{t-1} - [\beta_1 + a_{t-1} - a_{t-2}] \\ &= a_t - 2a_{t-1} + a_{t-2} \end{aligned}$$

显然 2 阶差分后序列 $\{\nabla^2 x_t\}$ 也是平稳序列, 它也将原序列中的非平稳趋势提取充分了.

考察它们的方差状况:

$$\begin{aligned} \text{Var}(\nabla x_t) &= \text{Var}(a_t - a_{t-1}) = 2\sigma^2 \\ \text{Var}(\nabla^2 x_t) &= \text{Var}(a_t - 2a_{t-1} + a_{t-2}) = 6\sigma^2 \end{aligned}$$

显然 2 阶差分后序列 $\{\nabla^2 x_t\}$ 的方差大于 1 阶差分后序列 $\{\nabla x_t\}$ 的方差, 在这种

场合下, 2 阶差分就属于过差分. 过差分实际上是因为过多次数的差分导致有效信息的无谓浪费而降低了估计的精度.

5.2 ARIMA 模型

差分运算具有强大的确定性信息提取能力, 许多非平稳序列差分后会显示出平稳序列的性质, 这时我们称这个非平稳序列为差分平稳序列. 对差分平稳序列可以使用 ARIMA 模型进行拟合.

5.2.1 ARIMA 模型的结构

具有如下结构的模型称为求和自回归移动平均模型, 简记为 ARIMA(p, d, q) 模型:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \\ E(x_s \varepsilon_t) = 0, \quad \forall s < t \end{cases} \quad (5.1)$$

式中, $\nabla^d = (1 - B)^d$; $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, 为平稳可逆 ARMA(p, q) 模型的自回归系数多项式; $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, 为平稳可逆 ARMA(p, q) 模型的移动平滑系数多项式.

求和自回归移动平均模型这个名字的由来是因为 d 阶差分后序列可以表示为:

$$\nabla^d x_t = \sum_{i=0}^d (-1)^i C_d^i x_{t-i}$$

式中, $C_d^i = \frac{d!}{i!(d-i)!}$, 即差分后序列等于原序列的若干序列值的加权和, 而对它又可以拟合自回归移动平均 (ARMA) 模型, 所以称它为求和自回归移动平均模型.

式 (5.1) 可以简记为:

$$\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \quad (5.2)$$

式中, $\{\varepsilon_t\}$ 为零均值白噪声序列.

由式 (5.2) 容易看出, ARIMA 模型的实质就是差分运算与 ARMA 模型的组合. 这一关系意义重大. 这说明任何非平稳序列如果能够通过适当阶数的差分实现差分后平稳, 就可以对差分后序列进行 ARMA 模型拟合了. 而 ARMA 模型的分析方法非常成熟, 这意味着对差分平稳序列的分析也将是非常简单、非常可靠的.

特别地

当 $d = 0$ 时, ARIMA(p, d, q) 模型实际上就是 ARMA(p, q) 模型.

当 $p = 0$ 时, ARIMA($0, d, q$) 模型可以简记为 IMA(d, q) 模型.

当 $q = 0$ 时, ARIMA($p, d, 0$) 模型可以简记为 ARI(p, d) 模型.

当 $d = 1, p = q = 0$ 时, ARIMA($0, 1, 0$) 模型为:

$$\begin{cases} x_t = x_{t-1} + \varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, \quad s \neq t \\ E(x_s \varepsilon_t) = 0, \quad \forall s < t \end{cases} \quad (5.3)$$

该模型又称为随机游走 (random walk) 模型, 或醉汉模型.

随机游走模型的产生有一个有趣的典故. 它最早于 1905 年 7 月由 Karl Pearson 在《自然》杂志上作为一个问题提出: 假如有个醉汉醉得非常严重, 完全丧失方向感, 把他放在荒郊野外, 一段时间之后再去找他, 在什么地方找到他的概率最大呢?

考虑到他完全丧失方向感, 那么他第 t 步的位置将是他第 $t-1$ 步的位置再加一个完全随机的位移. 用数学模型来描述任意时刻这个醉汉可能的位置, 即为一个随机游走模型.

1905 年 8 月, Lord Rayleigh 对 Karl Pearson 的这个问题作出了解答. 他算出这个醉汉离初始点的距离为 $r \sim r + \delta r$ 的概率为:

$$\frac{2}{nl^2} e^{-r^2/nl^2} r \delta r$$

且当 n 很大时, 该醉汉离初始点的距离服从零均值正态分布. 这意味着, 假如有人想去寻找该醉汉的话, 最好是去初始点附近找他, 该地点是醉汉未来位置的无偏估计值.

作为一个最简单的 ARIMA 模型, 随机游走模型目前广泛应用于计量经济学领域. 传统的经济学家普遍认为投机价格的走势类似于随机游走模型, 随机游走模型也是有效市场理论 (efficient market theory) 的核心.

5.2.2 ARIMA 模型的性质

一、平稳性

假如 $\{x_t\}$ 服从 ARIMA(p, d, q) 模型:

$$\Phi(B) \nabla^d x_t = \Theta(B) \varepsilon_t$$

式中 $\nabla^d = (1 - B)^d$

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

记 $\varphi(B) = \Phi(B) \nabla^d$, $\varphi(B)$ 称为广义自回归系数多项式. 显然 ARIMA 模型的平稳性完全由 $\varphi(B) = 0$ 的根的性质决定.

因为 $\{x_t\}$ d 阶差分后平稳, 服从 ARMA(p, q) 模型, 所以不妨设

$$\Phi(B) = \prod_{i=1}^p (1 - \lambda_i B), \quad |\lambda_i| < 1; i = 1, 2, \cdots, p$$

$$\text{则} \quad \varphi(B) = \Phi(B) \nabla^d = \left[\prod_{i=1}^p (1 - \lambda_i B) \right] (1 - B)^d \quad (5.4)$$

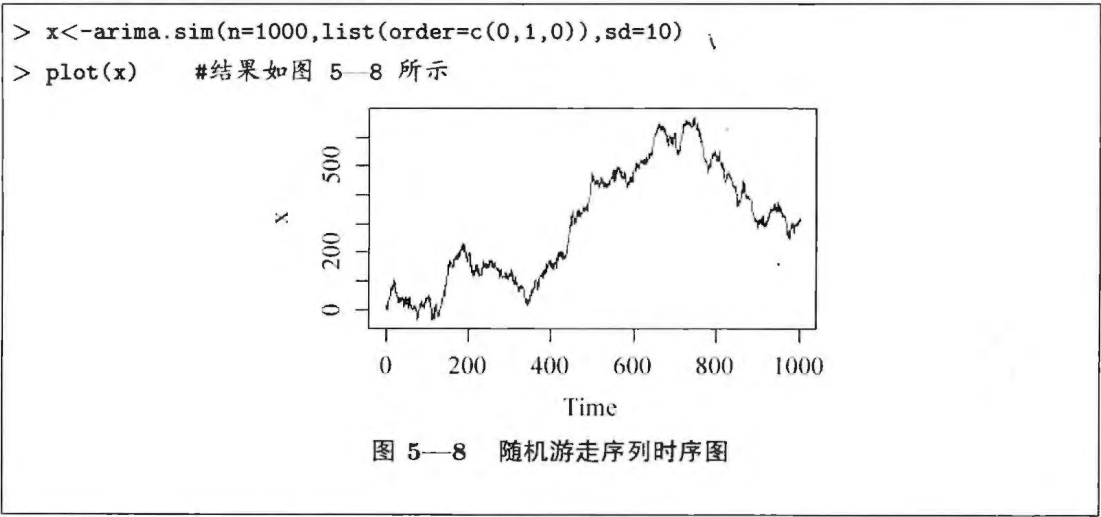
由式 (5.4) 容易判断, ARIMA(p, d, q) 模型的广义自回归系数多项式共有 $p + d$ 个根,

其中 p 个根 $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_p}$ 在单位圆外, d 个根在单位圆上.

自回归系数多项式的根即为特征根的倒数, 所以 $ARIMA(p, d, q)$ 模型共有 $p + d$ 个特征根, 其中 p 个在单位圆内, d 个在单位圆上.

因为有 d 个特征根在单位圆上而非单位圆内, 所以当 $d \neq 0$ 时, $ARIMA(p, d, q)$ 模型不平稳.

【例 5—5】 拟合随机游走序列: $x_t = x_{t-1} + \varepsilon_t, \varepsilon_t \sim NID(0, 100)$.



时序图清晰显示, 该序列非平稳.

二、方差齐性

对于 $ARIMA(p, d, q)$ 模型, 当 $d \neq 0$ 时, 不仅均值非齐性, 序列方差也非齐性. 以最简单的随机游走模型 $ARIMA(0,1,0)$ 为例:

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t \\ &= x_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &\vdots \\ &= x_0 + \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1 \end{aligned}$$

则 $Var(x_t) = Var(x_0 + \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1) = t\sigma_\varepsilon^2$
这是一个时间 t 的递增函数, 随着时间趋向无穷, 序列 $\{x_t\}$ 的方差也趋向无穷.
但 1 阶差分之后

$$\nabla x_t = \varepsilon_t$$

差分后序列方差齐性

$$Var(\nabla x_t) = \sigma_\varepsilon^2$$

5.2.3 ARIMA 模型建模

在掌握 ARMA 模型的建模方法之后, 尝试使用 ARIMA 模型对观察序列建模是一件比较简单的事情. 它遵循如下操作流程 (见图 5—9).

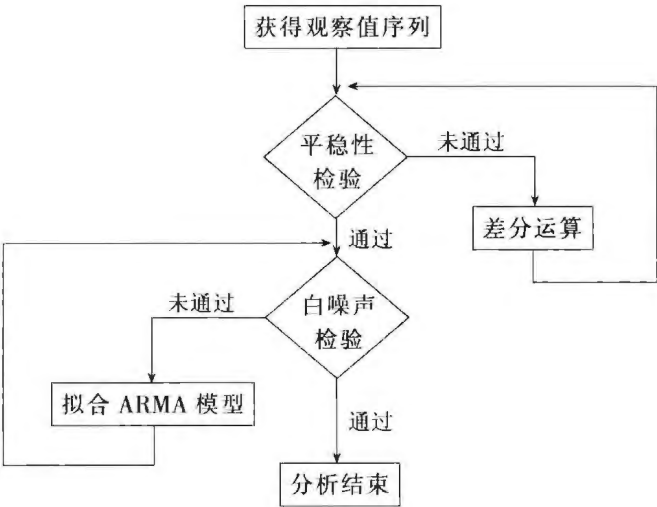


图 5—9 建模流程

根据这种建模流程, 对一个真实序列建模.

【例 5—6】对 1952—1988 年中国农业实际国民收入指数序列 (数据见表 A1—17) 建模.

- 读入数据, 并绘制时序图

```
> d<-read.table("E:/R/data/file17.csv",",",header=T)
> x<-ts(d$index,start=1952)
> plot(x)    #结果如图 5—10 所示
```

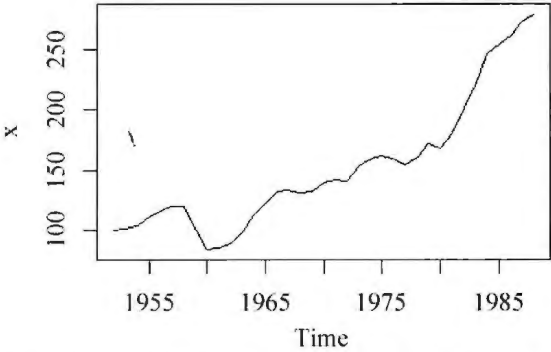


图 5—10 1952—1988 年中国农业实际国民收入指数时序图

- 1 阶差分, 并绘制差分后序列时序图

```
> x.dif<-diff(x)
> plot(x.dif)    #结果如图 5—11 所示
```

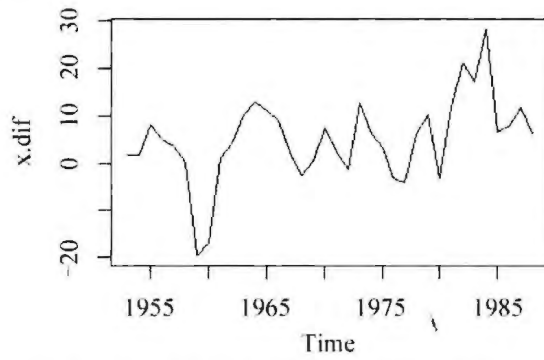



图 5—11 1952—1988 年中国农业实际国民收入指数 1 阶差分时序图

• 绘制差分后序列自相关图和偏自相关图

```
> acf(x.dif)    #结果如图 5—12 所示  
> pacf(x.dif)   #结果如图 5—13 所示
```

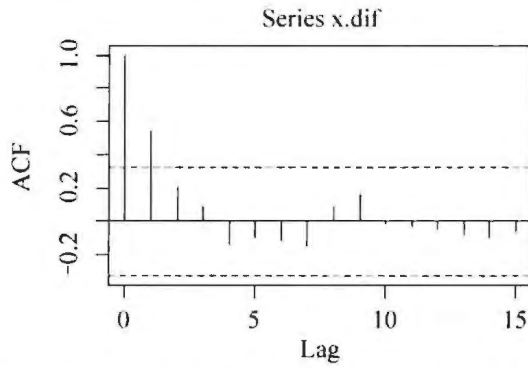


图 5—12 中国农业实际国民收入指数 1 阶差分后序列自相关图

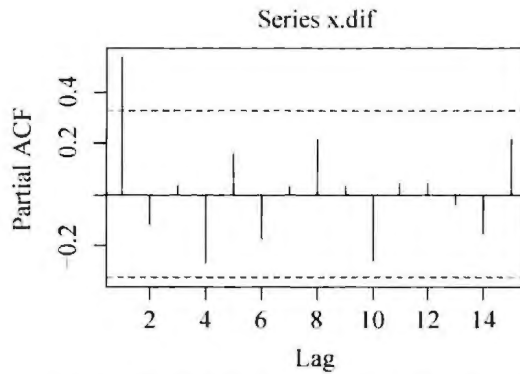


图 5—13 中国农业实际国民收入指数 1 阶差分后序列偏自相关图

• 拟合 ARIMA(0,1,1) 模型

```
> x.fit<-arima(x,order=c(0,1,1))
```

```

> x.fit

Call:
arima(x = x, order = c(0, 1, 1))

Coefficients:
          ma1
          0.7355
s.e.      0.1545

sigma^2 estimated as 61.95: log likelihood=-125.74, aic=255.49
• 残差白噪声检验
> for(i in 1:2) print(Box.test(x.fit$residual,lag=6*i))

          Box-Pierce test

data: x.fit$residual
X-squared=3.3169, df=6, p-value=0.7681

          Box-Pierce test

data: x.fit$residual
X-squared=6.0284, df=12, p-value=0.9146

```

时序图图 5—10 显示序列有线性趋势, 所以选择 1 阶差分. 1 阶差分后序列时序图图 5—11 显示序列在均值附近比较稳定地波动. 借助差分后序列自相关图图 5—12 进一步考察差分后序列的平稳性. 图 5—12 显示出除了延迟 1 阶的自相关系数显著非零, 其他阶数的自相关系数均在 2 倍标准差范围之内, 显示出很强的短期相关性, 所以可以认为 1 阶差分后序列平稳. 综合考察自相关图和偏自相关图的属性, 可以认为自相关系数 1 阶截尾, 偏自相关系数拖尾, 所以对原序列拟合 ARIMA(0,1,1)模型, 得到的拟合模型为:

$$x_t = x_{t-1} + \varepsilon_t + 0.7355\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, 61.95)$$

残差序列白噪声检验说明该模型显著成立. 这说明 ARIMA(0,1,1) 模型对该序列拟合成功.

5.2.4 ARIMA 模型预测

在最小均方误差预测原理下, ARIMA 模型和 ARMA 模型的预测方法非常相似. ARIMA(p, d, q) 模型的一般表示方法为:

$$\Phi(B)(1-B)^d x_t = \Theta(B)\varepsilon_t$$

和 ARMA 模型一样, 也可以用随机扰动项的线性函数表示:

$$\begin{aligned}x_t &= \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots \\&= \Psi(B) \varepsilon_t\end{aligned}$$

式中, $\Psi_1, \Psi_2, \dots$ 的值由如下等式确定:

$$\Phi(B)(1-B)^d \Psi(B) = \Theta(B)$$

如果把 $\Phi^*(B)$ 记为广义自相关函数, 有

$$\Phi^*(B) = \Phi(B)(1-B)^d = 1 - \tilde{\phi}_1 B - \tilde{\phi}_2 B^2 - \cdots$$

容易验证 $\Psi_1, \Psi_2, \dots$ 的值满足如下递推公式:

$$\begin{cases} \Psi_1 = \tilde{\phi}_1 - \theta_1 \\ \Psi_2 = \tilde{\phi}_1 \Psi_1 + \tilde{\phi}_2 - \theta_2 \\ \vdots \\ \Psi_j = \tilde{\phi}_1 \Psi_{j-1} + \cdots + \tilde{\phi}_{p+d} \Psi_{j-p-d} - \theta_j \end{cases} \quad (5.5)$$

式中, $\Psi_j = \begin{cases} 0, & j < 0 \\ 1, & j = 0 \end{cases}; \theta_j = 0 \ (j > q)$.

那么, x_{t+l} 的真实值为:

$$x_{t+l} = (\varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \cdots + \Psi_{l-1} \varepsilon_{t+1}) + (\Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \cdots)$$

由于 $\varepsilon_{t+l}, \varepsilon_{t+l-1}, \dots, \varepsilon_{t+1}$ 的不可获得性, 所以 x_{t+l} 的估计值只能为:

$$\hat{x}_t(l) = \Psi_0^* \varepsilon_t + \Psi_1^* \varepsilon_{t-1} + \Psi_2^* \varepsilon_{t-2} + \cdots$$

真实值与预测值之间的均方误差为:

$$E[x_{t+l} - \hat{x}_t(l)]^2 = (1 + \Psi_1^2 + \cdots + \Psi_{l-1}^2) \sigma_\varepsilon^2 + \sum_{j=0}^{\infty} (\Psi_{l+j} - \Psi_j^*)^2 \sigma_\varepsilon^2$$

要使均方误差最小, 当且仅当

$$\Psi_j^* = \Psi_{l+j}$$

所以在均方误差最小原则下, l 期预测值为:

$$\hat{x}_t(l) = \Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \Psi_{l+2} \varepsilon_{t-2} + \cdots$$

l 期预测误差为:

$$e_t(l) = \varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \cdots + \Psi_{l-1} \varepsilon_{t+1}$$

真实值等于预测值加上预测误差:

$$\begin{aligned}x_{t+l} &= (\varepsilon_{t+l} + \Psi_1 \varepsilon_{t+l-1} + \cdots + \Psi_{l-1} \varepsilon_{t+1}) + (\Psi_l \varepsilon_t + \Psi_{l+1} \varepsilon_{t-1} + \cdots) \\&= e_t(l) + \hat{x}_t(l)\end{aligned}$$

l 期预测误差的方差为:

$$\text{Var}[e_t(l)] = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_\varepsilon^2$$

【例5—7】已知 ARIMA(1, 1, 1) 模型为 $(1 - 0.8B)(1 - B)x_t = (1 - 0.6B)\varepsilon_t$, 且 $x_{t-1} = 4.5, x_t = 5.3, \varepsilon_t = 0.8, \sigma_\varepsilon^2 = 1$. 求 x_{t+3} 的 95% 的置信区间.

展开原模型, 等价形式为:

$$(1 - 1.8B + 0.8B^2)x_t = (1 - 0.6B)\varepsilon_t$$

$$x_t = 1.8x_{t-1} - 0.8x_{t-2} + \varepsilon_t - 0.6\varepsilon_{t-1}$$

则预测值的递推公式为:

$$\hat{x}_t(1) = 1.8x_t - 0.8x_{t-1} - 0.6\varepsilon_t = 5.46$$

$$\hat{x}_t(2) = 1.8\hat{x}_t(1) - 0.8x_t = 5.59$$

$$\hat{x}_t(3) = 1.8\hat{x}_t(2) - 0.8\hat{x}_t(1) = 5.69$$

3 期预测误差的方差为:

$$\text{Var}[e(3)] = (1 + \psi_1^2 + \psi_2^2)\sigma_\varepsilon^2$$

广义自相关函数为:

$$\begin{aligned}\Psi^*(B) &= \Phi(B)(1 - B)^d \\ &= (1 - 0.8B)(1 - B) \\ &= 1 - 1.8B + 0.8B^2\end{aligned}$$

则 $\tilde{\phi}_1 = 1.8, \tilde{\phi}_2 = -0.8$, 根据递推公式 (5.5) 可以得到

$$\begin{cases} \psi_1 = 1.8 - 0.6 = 1.2 \\ \psi_2 = 1.8\psi_1 - 0.8 = 1.36 \end{cases}$$

则

$$\text{Var}[e(3)] = (1 + \psi_1^2 + \psi_2^2)\sigma_\varepsilon^2 = 4.2896$$

x_{t+3} 的 95% 的置信区间为 $(\hat{x}_t(3) - 1.96\sqrt{\text{Var}(e(3))}, \hat{x}_t(3) + 1.96\sqrt{\text{Var}(e(3))})$, 即 (1.63, 9.75).

【例5—6 续 (1)】对 1952—1988 年中国农业实际国民收入指数序列作为期 10 年的预测.

<pre>> library(zoo) > library(forecast) > d<-read.table("E:/R/data/file17.csv", ",", header=T) > x<-ts(d\$index, start=1952) > x.fit<-arima(x, order=c(0,1,1)) > x.fore<-forecast(x.fit, h=10) > x.fore</pre>					
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1989	283.0082	272.9213	293.0950	267.5816	298.4347

1990	283.0082	262.8045	303.2118	252.1094	313.9070
1991	283.0082	256.2756	309.7407	242.1243	323.8921
1992	283.0082	251.0540	314.9624	234.1384	331.8779
1993	283.0082	246.5731	319.4432	227.2856	338.7308
1994	283.0082	242.5860	323.4304	221.1878	344.8286
1995	283.0082	238.9582	327.0581	215.6396	350.3767
1996	283.0082	235.6073	330.4090	210.5149	355.5015
1997	283.0082	232.4782	333.5382	205.7292	360.2871
1998	283.0082	229.5318	336.4845	201.2231	364.7932

> plot(x.fore) #结果如图 5—14 所示

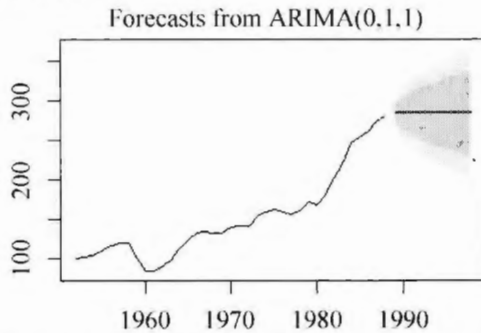


图 5—14 中国农业实际国民收入指数序列预测图

5.2.5 疏系数模型

ARIMA(p, d, q) 模型是指 d 阶差分后自相关最高阶数为 p , 移动平均最高阶数为 q 的模型, 通常它包含 $p+q$ 个独立的未知系数: $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$.

如果该模型中有部分自相关系数 $\phi_j (1 \leq j < p)$ 或部分移动平滑系数 $\theta_k (1 \leq k < q)$ 为零, 即原 ARIMA(p, d, q) 模型中有部分系数省缺了, 那么该模型称为疏系数模型.

如果只是自相关部分有省缺系数, 那么该疏系数模型可以简记为:

$$\text{ARIMA}((p_1, \dots, p_m), d, q)$$

式中, $p_1, \dots, p_m$ 为非零自相关系数的阶数.

如果只是移动平滑部分有省缺系数, 那么该疏系数模型可以简记为:

$$\text{ARIMA}(p, d, (q_1, \dots, q_n))$$

式中, $q_1, \dots, q_n$ 为非零移动平滑系数的阶数.

如果自相关和移动平滑部分都有省缺, 可以简记为:

$$\text{ARIMA}((p_1, \dots, p_m), d, (q_1, \dots, q_n))$$

在实际操作中, 疏系数模型时有应用.

R 语言用 `arima` 函数拟合 ARIMA 疏系数模型. `arima` 函数对疏系数模型的拟合命令如下:

```
arima(x, order=, include.mean=, method=, transform.pars=, fixed=)
```

式中:

-x: 要进行模型拟合的序列名.

-order: 指定模型阶数. $order = c(p, d, q)$, p 为自回归阶数, d 为差分阶数, q 为移动平均阶数.

-include.mean: $include.mean = T$, 需要拟合常数项; $include.mean = F$, 不需要拟合常数项.

-method: 指定参数估计方法.

-transform.pars: 指定参数估计是否由系统自动完成.

(1) $transform.pars = T$: 系统默认设置是系统根据 $order$ 选项设置的模型阶数自动完成参数估计.

(2) $transform.pars = F$: 需要拟合疏系数模型, 不能让系统根据模型的最高阶数自动完成所有参数的估计, 我们需要进行人为干预.

-fixed: 对疏系数模型指定疏系数的位置.

【例 5—8】对 1917—1975 年美国 23 岁妇女每万人生育率序列建模 (数据见表 A1—18).

• 读入数据, 并绘制时序图

```
> e<-read.table("E:/R/data/file18.csv",sep=";",header=T)
```

```
> x<-ts(e$fertility,start=1917)
```

```
> plot(x) #结果如图 5—15 所示
```

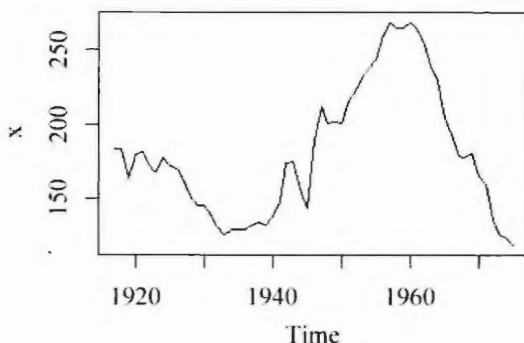


图 5—15 美国 23 岁妇女每万人生育率序列时序图

• 1 阶差分, 并绘制差分后序列时序图

```
> x.dif<-diff(x)
```

```
> plot(x.dif) #结果如图 5—16 所示
```

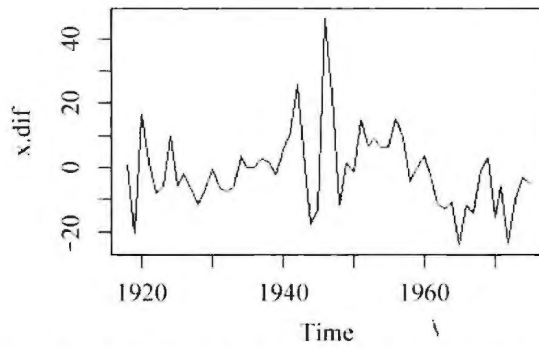


图 5—16 美国 23 岁妇女每万人生育率 1 阶差分后序列时序图

• 绘制差分后序列自相关图和偏自相关图

```
> acf(x.dif)    #结果如图 5—17 所示  
> pacf(x.dif)   #结果如图 5—18 所示
```

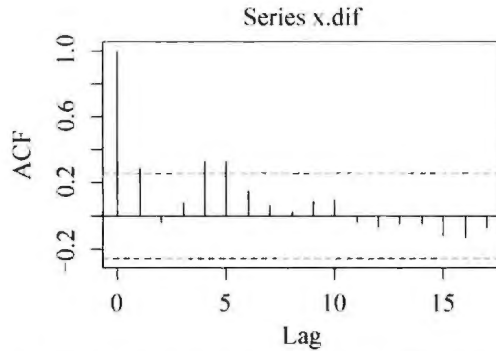


图 5—17 美国 23 岁妇女每万人生育率 1 阶差分后序列自相关图

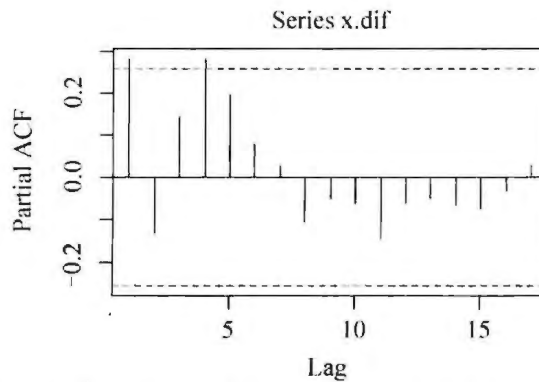


图 5—18 美国 23 岁妇女每万人生育率 1 阶差分后序列偏自相关图

• 拟合疏系数模型 $ARIMA((1,4),1,0)$

```
> x.fit<-arima(x,order=c(4,1,0),transform.pars=F,fixed=c(NA,0,0,NA))  
> x.fit  
Series: x  
ARIMA(4,1,0)
```

```
Coefficients:
      ar1  ar2  ar3    ar4
      0.2583    0    0 0.3408
s.e.  0.1159    0    0 0.1225

sigma^2 estimated as 118.2: log likelihood=-221
AIC=448.01  AICc=448.45  BIC=454.19
• 残差序列白噪声检验
> for(i in 1:2) print(Box.test(x.fit$residual,lag=6*i))

Box-Pierce test

data: x.fit$residual
X-squared=4.0909, df=6, p-value=0.6644

Box-Pierce test

data: x.fit$residual
X-squared=5.3742, df=12, p-value=0.9443
• 做5期预测, 并绘制预测图
> x.fore<-forecast(x.fit,h=5)
> x.fore
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
1976      109.18385  95.25187 123.1158  87.87673 130.4910
1977      103.43765  81.04480 125.8305  69.19075 137.6846
1978      100.93090  71.90947 129.9523  56.54645 145.3154
1979       98.64765  64.12267 133.1726  45.84624 151.4491
1980       94.88318  53.11326 136.6531  31.00160 158.7648
> plot(x.fore)      #结果如图 5—19 所示
```

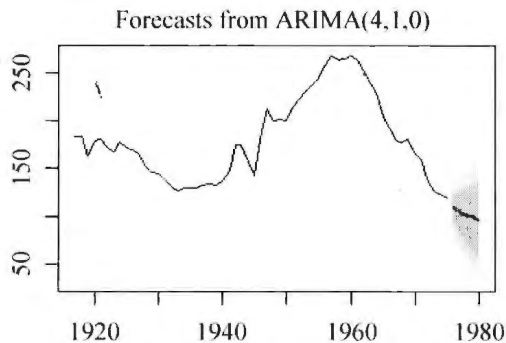


图 5—19 美国 23 岁妇女每万人生育率序列预测图

由时序图图 5—15 可知, 序列非平稳. 1 阶差分后序列 (见图 5—16) 没有明显趋

势或周期特征. 自相关图图 5—17 5 阶之后都在 2 倍标准差范围之内, 满足短期自相关性. 所以可以认为 1 阶差分后序列平稳. 自相关图和偏自相关图都呈现出截尾的特征, 可以尝试拟合多个模型. 在此根据偏自相关系数 1 阶和 4 阶显著非零, 4 阶后截尾, `airma` 函数指定:

(1) `order=c(4,1,0)`: 构造 1 阶差分后 AR(4) 模型.

(2) `transform.pars=F`: 指定系数要进行人为干预.

(3) `fixed=c(NA,0,0,NA)` 意味着 AR 部分四个参数只有 1,4 两个参数非零, 2,3 两个参数指定恒等于零.

这三个选项综合提供的信息是构造 ARIMA((1,4),1,0) 疏系数模型. 然后对模型拟合效果进行白噪声检验. 白噪声检验结果显示, 该疏系数模型显著成立. 利用该拟合模型进行 5 期预测, 并绘制预测图.

5.2.6 季节模型

ARIMA 模型可以对具有季节效应的序列建模. 根据季节效应提取的难易程度, 可以分为简单季节模型和乘积季节模型.

一、简单季节模型

简单季节模型是指序列中的季节效应和其他效应之间是加法关系, 即

$$x_t = S_t + T_t + I_t$$

这时, 各种效应信息的提取都非常容易. 通常简单的周期步长差分即可将序列中的季节信息提取充分, 简单的低阶差分即可将趋势信息提取充分, 提取完季节信息和趋势信息之后的残差序列就是一个平稳序列, 可以用 ARMA 模型拟合.

所以简单季节模型实际上就是通过趋势差分、季节差分将序列转化为平稳序列, 再对其进行拟合. 它的模型结构通常如下:

$$\nabla_D \nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

式中

(1) D 为周期步长, d 为提取趋势信息所用的差分阶数.

(2) $\{\varepsilon_t\}$ 为白噪声序列, 且 $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

(3) $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, 为 q 阶移动平均系数多项式.

(4) $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, 为 p 阶自回归系数多项式.

R 语言用 `arima` 函数中的 `seasonal` 选项拟合季节模型, 相关命令如下:

```
arima(x, order=, include.mean=, method=, transform.pars=, fixed=, seasonal=)
```

式中:

-x: 要进行模型拟合的序列名.

-order: 指定模型阶数.

-include.mean: 指定是否需要拟合常数项.

-method: 指定参数估计方法.

-transform.pars: 指定是否需要人为干预参数估计.

-fixed: 对疏系数模型指定疏系数的位置.

-seasonal: 指定季节模型的阶数与季节周期. 该选项的命令格式为:

seasonal=list(order=c(P,D,Q), period = π)

(1) 加法模型: $P=0, Q=0$;

(2) 乘法模型: P, Q 不全为零.

【例 5—9】 拟合 1962—1991 年德国工人季度失业率序列 (数据见表 A1—19).

• 读入数据, 并绘制时序图

```
> f<-read.table("E:/R/data/file19.csv",sep=" ",header=T)
```

```
> x<-ts(f$unemployment_rate,start=c(1962,1),frequency=4)
```

```
> plot(x) #结果如图 5—20 所示
```

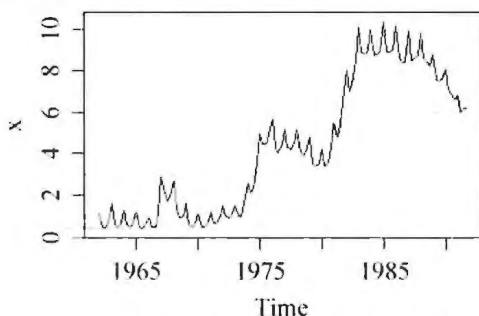


图 5—20 德国工人季度失业率序列时序图

• 1 阶 4 步差分, 并绘制差分后序列时序图

```
> x.dif<-diff(diff(x),4)
```

```
> plot(x.dif) #结果如图 5—21 所示
```

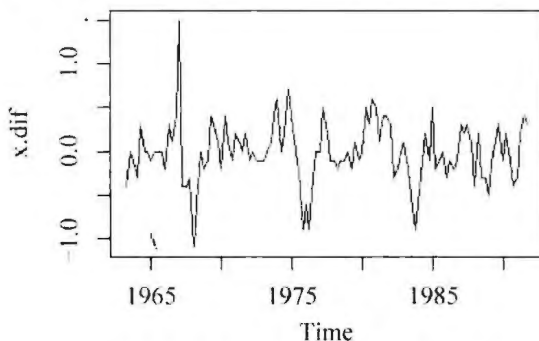


图 5—21 德国工人季度失业率 1 阶 4 步差分后序列时序图

• 差分后序列自相关图和偏自相关图

```
> acf(x.dif) #结果如图 5—22 所示
```

```
> pacf(x.dif) #结果如图 5—23 所示
```

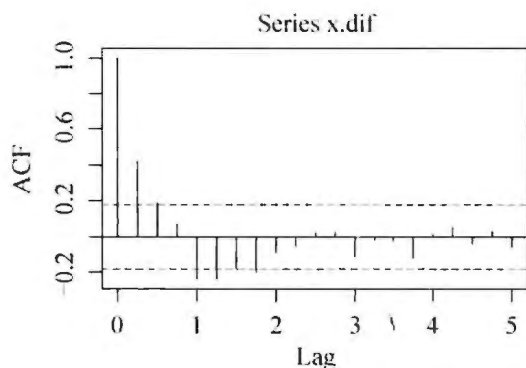


图 5—22 德国工人季度失业率 1 阶 4 步差分后序列自相关图

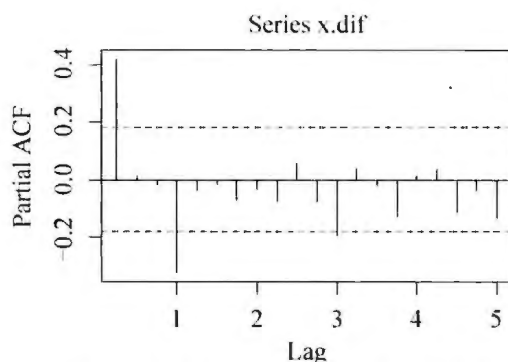


图 5—23 德国工人季度失业率 1 阶 4 步差分后序列偏自相关图

• 拟合加法季节模型 $ARIMA((1,4), (1,4), 0)$

```
> x.fit<-arima(x,order=c(4,1,0),seasonal=list(order=c(0,1,0),period=4),
transform.par=F,fixed=c(NA,0,0,NA))
```

```
> x.fit
```

Series: x

$ARIMA(4,1,0)(0,1,0)[4]$

Coefficients:

	ar1	ar2	ar3	ar4
	0.4449	0	0	-0.2720
s.e.	0.0807	0	0	0.0804

σ^2 estimated as 0.09266: log likelihood=-26.7

AIC=59.39 AICc=59.61 BIC=67.63

• 残差序列白噪声检验

```
> for(i in 1:2) print(Box.test(x.fit$residual,lag=6*i))
```

Box-Pierce test

```
data: x.fit$residual
X-squared=2.1452, df=6, p-value=0.9059
```

Box-Pierce test

```
data: x.fit$residual
X-squared=10.546, df=12, p-value=0.5682
```

• 做 3 年期预测, 并绘制预测图

```
> x.fore<-forecast(x.fit,h=12)
> x.fore
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1992 Q1	6.615077	6.224975	7.005178	6.0184679	7.211686
1992 Q2	5.856360	5.170875	6.541845	4.8080015	6.904719
1992 Q3	5.965918	5.027511	6.904325	4.5307483	7.401088
1992 Q4	5.844074	4.687986	7.000161	4.0759905	7.612157
1993 Q1	6.146436	4.634425	7.658448	3.8340140	8.458859
1993 Q2	5.326343	3.479538	7.173148	2.5018984	8.150787
1993 Q3	5.433197	3.294966	7.571428	2.1630545	8.703339
1993 Q4	5.343294	2.949603	7.736985	1.6824590	9.004128
1994 Q1	5.690528	2.925569	8.455487	1.4618876	9.919168
1994 Q2	4.907094	1.762628	8.051559	0.0980490	9.716138
1994 Q3	5.030993	1.528164	8.533821	-0.3261208	10.388106
1994 Q4	4.939984	1.103738	8.776231	-0.9270484	10.807017

```
> plot(x.fore) #结果如图 5—24 所示
```

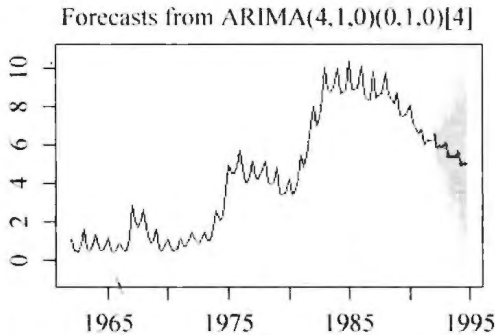


图 5—24 德国工人季度失业率序列预测图

时序图图 5—20 显示德国工人季度失业率序列具有趋势和周期. 进行 1 阶差分提取趋势信息, 进行 4 步差分提取周期信息, 1 阶 4 步差分后序列时序图 (见图 5—21) 显示差分后序列平稳. 考察差分后序列自相关图 (见图 5—22) 和偏自相关图 (见图 5—23) 的特征, 自相关系数拖尾, 而偏自相关系数 1 阶, 4 阶显著非零, 4 阶之后截尾. 所以综合差分和自相关信息, 拟合加法模型 $ARIMA((1, 4), (1, 4), 0)$.

arima 函数的命令语句中, `seasonal=list(order=c(0,1,0), period=4)` 是指定进行周期为 4 的 1 次差分, 也就是进行一次 4 步差分. `order=c(4,1,0)`, 指定了 1 阶差分和 AR(4) 模型. `transform.pars=F`, `fixed=c(NA,0,0,NA)`, 指定了拟合 AR(1,4) 疏系数模型.

根据模型拟合输出结果, 我们得到的拟合模型为:

$$(1-B)(1-B^4)x_t = \frac{1}{1-0.4449B+0.272B^4}\varepsilon_t, \quad \varepsilon_t \sim N(0, 0.09266)$$

残差序列白噪声检验显示该模型拟合效果良好, 对序列相关信息提取充分, 最后利用该模型进行 3 年期预测, 并得到预测图.

二、乘积季节模型

例 5—9 中的序列是一个既含有季节效应又含有长期趋势效应的简单序列, 说它简单是因为这种序列季节效应、趋势效应和随机波动彼此之间很容易分开, 这时简单的 ARIMA 模型即可拟合该序列的发展.

但更为常见的情况是, 序列的季节效应、长期趋势效应和随机波动之间存在复杂的交互影响关系, 简单的 ARIMA 模型并不足以提取其中的相关关系, 这时通常需要采用乘积季节模型.

【例 5—10】 拟合 1948—1981 年美国女性 (20 岁以上) 月度失业率序列 (数据见表 A1—20).

• 读入数据, 并绘制时序图

```
> g<-read.table("E:/data/file20.csv",sep=",",header=T)
> x<-ts(g$unemployment_rate,start=c(1948,1),frequency=12)
> plot(x)    #结果如图 5—25 所示
```

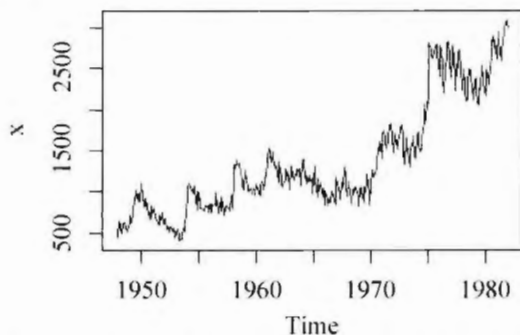


图 5—25 美国女性月度失业率序列时序图

• 作 1 阶 12 步差分, 并绘制差分后序列时序图

```
> x.dif<-diff(diff(x),12)
> plot(x.dif)    #结果如图 5—26 所示
```

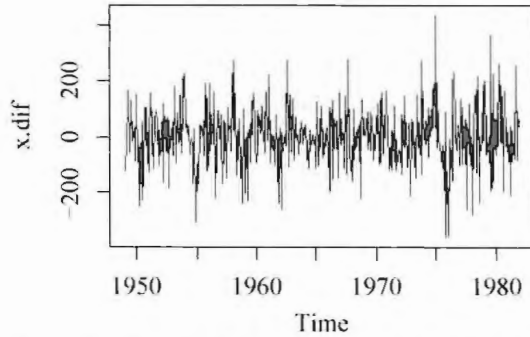


图 5—26 美国女性月度失业率差分后序列时序图

• 绘制差分后序列自相关图和偏自相关图

> acf(x.dif) #结果如图 5—27 所示

> pacf(x.dif) #结果如图 5—28 所示

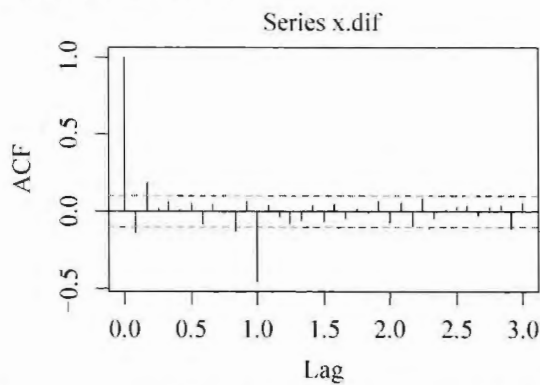


图 5—27 美国女性月度失业率差分后序列自相关图

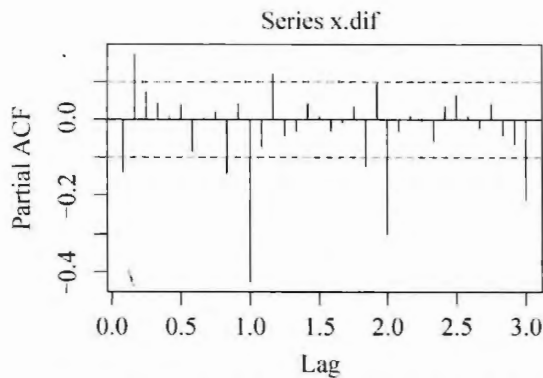


图 5—28 美国女性月度失业率差分后序列偏自相关图

• 拟合 ARIMA(1, (1, 12), 1)模型

```
> x.fit<-arima(x,order=c(1,1,1),seasonal=list(order=c(0,1,0),period=12))
```

```
> for(i in 1:2) print(Box.test(x.fit$residual,lag=6*i))
```

```

Box-Pierce test

data: x.fit$residual
X-squared=11.2036, df=6, p-value=0.08228

Box-Pierce test

data: x.fit$residual
X-squared=105.7755, df=12, p-value < 2.2e-16

```

时序图图 5—25 显示序列既有趋势也有周期特征. 1 阶 12 步差分之后序列呈现出平稳特征 (见图 5—26). 差分后序列的自相关图 (见图 5—27) 和偏自相关图 (见图 5—28) 都显示出拖尾属性, 所以首先尝试拟合加法季节模型 $ARIMA(1, (1, 12), 1)$ 模型. 残差序列的白噪声检验显示, 残差序列非白噪声, 该模型拟合效果不理想. 这说明加法季节模型并不适合拟合这个序列. 这时, 需要考虑乘积季节模型.

乘积模型的构造原理如下:

当序列具有短期相关性时, 通常可以使用低阶 $ARMA(p, q)$ 模型提取.

当序列具有季节效应, 季节效应本身还具有相关性时, 季节相关性可以使用以周期步长为单位的 $ARMA(P, Q)$ 模型提取.

由于短期相关性和季节效应之间具有乘积关系, 所以拟合模型实质为 $ARMA(p, q)$ 和 $ARMA(P, Q)$ 的乘积. 综合前面的 d 阶趋势差分 and D 阶以周期 S 为步长的季节差分运算, 对原观察值序列拟合的乘积模型完整的结构如下:

$$\nabla^d \nabla_S^D x_t = \frac{\Theta(B) \Theta_S(B)}{\Phi(B) \Phi_S(B)} \varepsilon_t$$

式中

$$\begin{aligned} \Theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ \Phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \Theta_S(B) &= 1 - \theta_1 B^S - \dots - \theta_Q B^{QS} \\ \Phi_S(B) &= 1 - \phi_1 B^S - \dots - \phi_P B^{PS} \end{aligned}$$

该乘积模型简记为 $ARIMA(p, d, q) \times (P, D, Q)_S$.

回到例 5—10 的模型定阶阶段, 首先考虑 1 阶 12 步差分之后, 序列 12 阶以内的自相关系数和偏自相关系数的特征, 以确定短期相关模型. 自相关图 (见图 5—27) 和偏自相关图 (见图 5—28) 显示 12 阶以内的自相关系数和偏自相关系数均不截尾, 所以尝试使用 $ARMA(1, 1)$ 模型提取差分后序列的短期自相关信息.

再考虑季节自相关特征, 这时考察延迟 12 阶、24 阶等以周期长度为单位的自相关系数和偏自相关系数的特征. 自相关图 (图 5—27) 显示延迟 12 阶自相关系数显著非零, 但是延迟 24 阶自相关系数落入 2 倍标准差范围. 而偏自相关图 (图 5—28) 显示延迟 12 阶和延迟 24 阶的偏自相关系数都显著非零. 所以可以认为季节自相关特征是自相关系数截尾, 偏自相关系数拖尾, 这时以 12 步为周期的 $ARMA(0, 1)_{12}$ 模型提取差分后序列的季节自相关信息.

综合前面的差分信息, 我们要拟合的乘积模型为 $ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$.

$$\nabla \nabla_{12} x_t = \frac{1 - \theta_1 B}{1 - \phi_1 B} (1 - \theta_{12} B^{12}) \varepsilon_t$$

• 拟合 ARIMA(1,1,1)×ARIMA(0,1,1)12 模型

```
> x.fit<-arima(x,order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12))
```

```
> x.fit
```

```
Series: x
```

```
ARIMA(1,1,1)(0,1,1)[12]
```

```
Coefficients:
```

```
          ar1      ma1      sma1
      -0.7290  0.6059  -0.7918
s.e.   0.1497  0.1728  0.0337
```

```
sigma^2 estimated as 7444: log likelihood=-2327.14
```

```
AIC=4662.28 AICc=4662.38 BIC=4678.2
```

• 残差序列白噪声检验

```
> for(i in 1:2) print(Box.test(x.fit$residual,lag=6*i))
```

```
Box-Pierce test
```

```
data: x.fit$residual
```

```
X-squared=4.5564, df=6, p-value=0.6018
```

```
Box-Pierce test
```

```
data: x.fit$residual
```

```
X-squared=9.6288, df=12, p-value=0.6485
```

根据拟合输出结果, 我们的拟合模型为:

$$\nabla \nabla_{12} x_t = \frac{1 + 0.6059B}{1 + 0.729B} (1 - 0.7918B^{12}) \varepsilon_t, \quad \varepsilon_t \sim N(0, 7444)$$

残差序列白噪声检验显示, 该拟合模型显著成立。

5.3 残差自回归模型

ARIMA 模型的提出使人们对非平稳序列的拟合精度大大提高, 自 1970 年 Box 和 Jenkins 提出这个模型以来, 它已经成为最经典的时间序列拟合模型。

但和传统的确定性因素分解方法比较, ARIMA 模型仍然有一些缺憾。它使用差分方法提取确定性信息, 差分方法的优点是对确定性信息的提取比较充分, 缺点是很难对模型进行直观解释。所以当序列具有非常显著的确定性趋势或季节效应时, 人们会

怀念确定性因素分解方法对各种确定性效应的解释,但又因为它对残差信息的浪费而不敢轻易使用.

为了解决这个问题,人们构造了残差自回归 (auto-regressive) 模型.

5.3.1 模型结构

残差自回归模型的构造思想是首先通过确定性因素分解方法提取序列中主要的确定性信息:

$$x_t = T_t + S_t + \varepsilon_t \quad (5.6)$$

式中, T_t 为趋势效应拟合; S_t 为季节效应拟合.

考虑到因素分解方法对确定性信息的提取可能不够充分,因而需要进一步检验残差序列 $\{\varepsilon_t\}$ 的自相关性.

如果检验结果显示残差序列的自相关性不显著,说明确定性回归模型 (5.6) 对信息的提取比较充分,可以停止分析.

如果检验结果显示残差序列的自相关性显著,说明确定性回归模型 (5.6) 对信息的提取不充分,这时可以考虑对残差序列拟合自回归模型,进一步提取相关信息:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t$$

这样构造的模型

$$x_t = T_t + S_t + \varepsilon_t$$

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t$$

$$E(a_t) = 0, \text{Var}(a_t) = \sigma^2, \text{Cov}(a_t, a_{t-i}) = 0, \quad \forall i \geq 1$$

称为 (残差) 自回归模型 (auto-regressive model).

实践中,对趋势效应的拟合常用如下两种方式:

(1) 自变量为时间 t 的幂函数.

$$T_t = \beta_0 + \beta_1 \cdot t + \cdots + \beta_k \cdot t^k + \varepsilon_t$$

(2) 自变量为历史观察值 $\{x_{t-1}, x_{t-2}, \cdots, x_{t-k}\}$.

$$T_t = \beta_0 + \beta_1 \cdot x_{t-1} + \cdots + \beta_k \cdot x_{t-k} + \varepsilon_t$$

第二种方式和差分方式的原理相同.

对季节效应的拟合也有两种常用的方式:

(1) 给定季节指数.

$$S_t = S'_t$$

式中, S'_t 是某个已知的季节指数.

(2) 建立季节自回归模型.

$$S_t = \alpha_0 + \alpha_1 \cdot x_{t-m} + \cdots + \alpha_l \cdot x_{t-lm}$$

假定固定周期为 m .

【例 5—6 续 (2)】 使用残差自回归模型分析 1952—1988 年中国农业实际国民收入指数序列.

图 5—10 显示该序列有显著的线性递增趋势, 但没有季节效应, 所以考虑建立如下结构的残差自回归模型:

$$\begin{cases} x_t = T_t + \varepsilon_t, & t = 1, 2, 3, \dots \\ \varepsilon_t = \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + a_t \\ E(a_t) = 0, \text{Var}(a_t) = \sigma^2, \text{Cov}(a_t, a_{t-i}) = 0, & \forall i \geq 1 \end{cases}$$

对 T_t 分别尝试构造如下两个确定性趋势模型:

(1) 变量为时间 t 的幂函数.

$$T_t = \beta_0 + \beta_1 \cdot t, \quad t = 1, 2, 3, \dots$$

(2) 变量为 1 阶延迟序列值 x_{t-1} .

$$T_t = \beta_0 + \beta_1 \cdot x_{t-1}$$

• 拟合关于时间 t 的线性回归模型

```
> d<-read.table("E:/R/data/file17.csv",",",header=T)
> x<-ts(d$index,start=1952)
> t<-c(1:37)
> x.fit1<-lm(x~t)
> summary(x.fit1)
```

Call:

```
lm(formula=x ~ t)
```

Residuals:

Min	1Q	Median	3Q	Max
-28.71	-20.48	-10.81	26.42	46.17

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	66.1491	8.1197	8.147	1.35e-09 ***
t	4.5158	0.3726	12.121	4.40e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.2 on 35 degrees of freedom

Multiple R-squared: 0.8076, Adjusted R-squared: 0.8021

F-statistic: 146.9 on 1 and 35 DF, p-value: 4.404e-14

• 拟合关于延迟变量的自回归模型

```
> xlag<-x[2:37]
```

```

> x2<-x[1:36]
> x.fit2<-lm(x2 ~ xlag)
> summary(x.fit2)

Call:
lm(formula=x2 ~ xlag)

Residuals:
    Min       1Q   Median       3Q      Max
-15.764  -5.066  -0.703   5.539  20.424

Coefficients:
            Estimate Std. Error t value Pr(> |t|)
(Intercept)  7.39226    4.00444   1.846  0.0736.
xlag         0.91932    0.02464  37.309 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.936 on 34 degrees of freedom
Multiple R-squared:  0.9762, Adjusted R-squared:  0.9755
F-statistic: 1392 on 1 and 34 DF, p-value: < 2.2e-16

```

● 两个趋势拟合模型的拟合效果图

```

> fit1<-ts(x.fit1$fitted.value,start=1952)
> fit2<-ts(x.fit2$fitted.value,start=1952)
> plot(x,type="p",pch=8)
> lines(fit1,col=2)
> lines(fit2,col=4)

```

#结果如图 5—29 所示

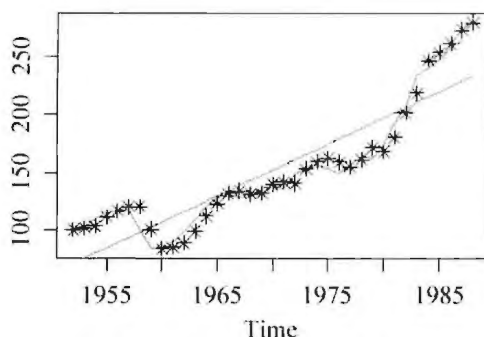


图 5—29 趋势拟合效果图

说明：图中，星号为序列观察值，直线为关于时间 t 的线性拟合模型估计值，曲线为关于延迟变量的自回归拟合模型估计值。

根据系统输出结果，我们可以得到如下两个确定性趋势拟合模型：

$$(1) x_t = 66.1491 + 4.5158t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 24.2^2)$$

$$(2) x_t = 7.3923 + 0.9193x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 7.936^2)$$

5.3.2 残差自相关检验

一、检验原理

确定性模型拟合好之后, 要对该模型的拟合效果进行检验.

如果残差序列显示出纯随机的性质, 即

$$E(\varepsilon_t \varepsilon_{t-j}) = 0, \quad \forall j \geq 1$$

就说明确定性模型拟合得非常好, 已经能够充分提取序列中的相关信息, 不需要再对残差序列进行二次信息提取, 分析结束.

反之, 如果残差序列显示出显著的自相关性, 即

$$E(\varepsilon_t \varepsilon_{t-j}) \neq 0, \quad \exists j \geq 1$$

就说明确定性模型拟合得不够精确, 序列中的相关信息没有得到充分提取, 应该对残差序列再次拟合, 提取其中残存的相关信息, 以提高模型拟合的精度.

二、Durbin-Watson 检验

Durbin-Watson 检验 (简称 DW 检验) 是由 J. Durbin 和 G. S. Watson 于 1950 年在考虑多元回归模型的残差独立性时提出的一个自相关检验统计量, 我们把它借鉴过来进行时间序列残差自相关性检验.

下面以残差 1 阶自相关性检验为例介绍 DW 检验的原理.

原假设: 残差序列不存在 1 阶自相关性, 即

$$H_0: E(\varepsilon_t \varepsilon_{t-1}) = 0 \Leftrightarrow H_0: \rho = 0$$

备择假设: 残差序列存在 1 阶自相关性, 即

$$H_1: E(\varepsilon_t \varepsilon_{t-1}) \neq 0 \Leftrightarrow H_1: \rho \neq 0$$

构造 DW 检验统计量:

$$\begin{aligned} DW &= \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \\ &= \frac{\left(\sum_{t=2}^n \varepsilon_t^2 + \sum_{t=2}^n \varepsilon_{t-1}^2 \right) - 2 \sum_{t=2}^n \varepsilon_t \cdot \varepsilon_{t-1}}{\sum_{t=1}^n \varepsilon_t^2} \end{aligned}$$

在大数场合, 有

$$\sum_{t=2}^n \varepsilon_t^2 \approx \sum_{t=2}^n \varepsilon_{t-1}^2 \approx \sum_{t=1}^n \varepsilon_t^2$$

所以 DW 检验统计量近似等于

$$DW = 2 \left(1 - \frac{\sum_{t=2}^n \varepsilon_t \cdot \varepsilon_{t-1}}{\sum_{t=1}^n \varepsilon_t^2} \right)$$

根据自相关系数的定义, 有

$$\rho = \frac{\sum_{t=2}^n \varepsilon_t \cdot \varepsilon_{t-1}}{\sum_{t=1}^n \varepsilon_t^2}$$

即 $DW \approx 2(1 - \rho)$

因为 $-1 \leq \rho \leq 1$, 所以 $0 \leq DW \leq 4$.

当 $0 < \rho \leq 1$ 时, 序列正相关, 且 $\rho \rightarrow 1$ 时, $DW \rightarrow 0$; $\rho \rightarrow 0$ 时, $DW \rightarrow 2$. 根据这个性质, 可以确定两个临界值 d_L, d_U ($0 \leq d_L, d_U \leq 2$), 当 $DW < d_L$ 时, 序列显著正相关, 当 $DW > d_U$ 时, 序列显著不相关. 当 $d_L \leq DW \leq d_U$ 时, 根据现有数据还无法做出肯定判断, 需要更多的信息支持.

当 $-1 < \rho \leq 0$ 时, 序列负相关, 且 $\rho \rightarrow -1$ 时, $DW \rightarrow 4$; $\rho \rightarrow 0$ 时, $DW \rightarrow 2$. 同理可以确定两个临界值 d'_L, d'_U ($2 \leq d'_L, d'_U \leq 4$), 当 $DW > d'_U$ 时, 序列显著负相关, 当 $DW < d'_L$ 时, 序列显著不相关. 当 $d'_L \leq DW \leq d'_U$ 时, 根据现有数据还无法做出肯定判断, 需要更多的信息支持.

根据 ρ 的对称性, 有

$$\begin{cases} 2 - d_U = d'_L - 2 \\ 2 - d_L = d'_U - 2 \end{cases} \Rightarrow \begin{cases} d'_L = 4 - d_U \\ d'_U = 4 - d_L \end{cases}$$

综上所述, 我们可以得到如下相关性判断图示, 如图 5—30 所示.

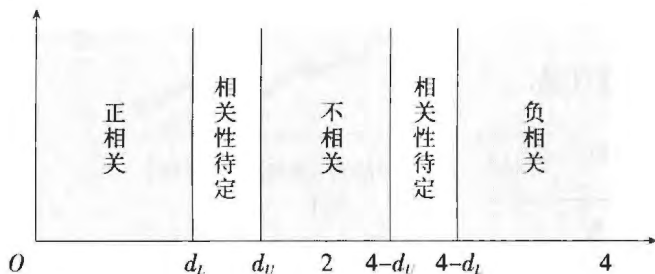


图 5—30

【例 5—6 续 (3)】 检验第一个确定性趋势模型

$$x_t = 66.149\ 1 + 4.515\ 8t + \varepsilon_t, \quad t = 1, 2, 3, \dots$$

残差序列的自相关性.

作 DW 检验之前, 需要先下载程序包 lmtest, 并用 library 函数调用该程序包. 利用该程序包中的 dwtest 函数可以进行残差自相关检验.

```
> library(lmtest)
> dwtest(x.fit1)

Durbin-Watson test

data: x.fit1
DW=0.1378, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

DW 检验结果显示, DW 统计量的值小于 2, 且 P 值极小, 这说明残差序列高度正相关. 为了充分提取相关信息, 我们需要对残差序列再次进行信息提取.

三、Durbin h 检验

值得注意的是, DW 统计量原本适用于回归模型残差自相关性的检验, 它要求自变量“独立”. Nerlove 和 Wallis 1966 年指出, 在自回归场合, 即当回归因子包含延迟因变量 (计量经济学称之为内生变量) 时, 有

$$x_t = \beta_0 + \beta_1 \cdot x_{t-1} + \cdots + \beta_k \cdot x_{t-k} + \varepsilon_t$$

残差序列 $\{\varepsilon_t\}$ 的 DW 统计量是一个有偏统计量, 当 ρ 趋于零时, $DW \neq 2$. 在这种场合下使用 DW 统计量容易产生残差序列正自相关性不显著的误判.

基于这个原因, 在 ARIMA 模型场合, 我们都是使用 Q 统计量和 LB 统计量检验残差序列的自相关性.

为了克服 DW 检验的有偏性, Durbin 在 1970 年提出了 DW 统计量的两个修正统计量: Durbin t 和 Durbin h 统计量, 这两个统计量渐近等价. Durbin h 统计量为:

$$D_h = DW \frac{n}{1 - n\sigma_\beta^2}$$

式中, n 为观察值序列长度; σ_β^2 为延迟因变量系数的最小二乘估计的方差.

通过对 DW 统计量的修正, Durbin t 和 Durbin h 统计量有效地提高了检验的精度, 代替 DW 检验统计量成为延迟因变量场合常用的自相关检验统计量.

【例 5—6 续 (4)】 检验第二个确定性趋势模型

$$x_t = 7.3923 + 0.9193x_{t-1} + \varepsilon_t, \quad t = 1, 2, 3, \cdots$$

残差序列的自相关性.

```
> dwtest(x.fit2, order.by = xlag)

Durbin-Watson test
```

```
data: x.fit2
DW=1.8153, p-value = 0.2295
alternative hypothesis: true autocorrelation is greater than 0
```

对第二个拟合模型而言, 因为自变量 x_{lag} 和因变量 x 具有自相关关系, 所以需要使用 Durbin h 检验. Durbin h 检验就是在调用 `dwtest` 函数时增加选项 `order.by`, 指定延迟因变量. 本例 Durbin h 检验结果显示, 残差序列不存在显著相关. 我们不需要对第二个趋势拟合模型的残差进行再次拟合.

5.3.3 残差自相关模型拟合

当残差自相关检验显示残差序列高度自相关时, 我们需要对残差序列进行二次拟合, 进一步提取残差序列中蕴涵的相关信息. 拟合残差自相关模型的步骤和命令与构建 ARIMA 模型的步骤和命令一样.

【例 5—6 (续 5)】 对第一个确定性拟合模型的残差序列拟合自相关模型.

- 绘制残差序列的自相关图和偏自相关图

```
> acf(x.fit1$residual)    #结果如图 5—31 所示
> pacf(x.fit1$residual)   #结果如图 5—32 所示
```

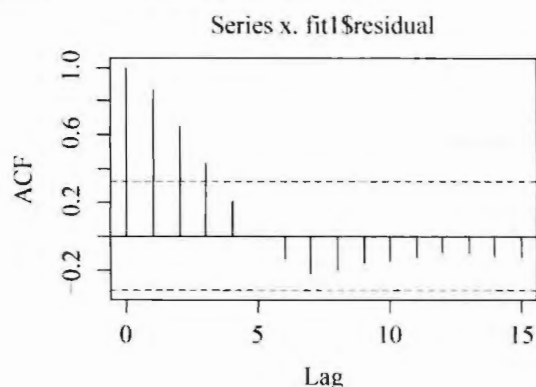


图 5—31 残差序列的自相关图

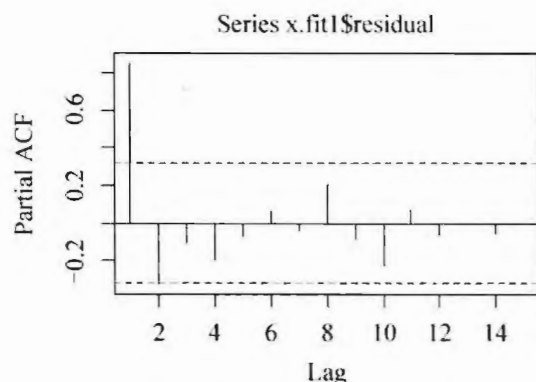


图 5—32 残差序列的偏自相关图

• 拟合 AR(2) 模型

```
> r.fit<-arima(x.fit1$residual,order=c(2,0,0),include.mean=F)
```

```
> r.fit
```

```
Series: x.fit1$residual
```

```
ARIMA(2,0,0) with zero mean
```

```
Coefficients:
```

```
      ar1      ar2
      1.4995 -0.6028
s.e.  0.1274  0.1356
```

```
sigma^2 estimated as 50.6: log likelihood=-126.59
```

```
AIC=259.17 AICc=259.9 BIC=264.01
```

• 残差自相关模型的显著性检验

```
> for(i in 1:2) print(Box.test(r.fit$residual,lag=6*i))
```

```
Box-Pierce test
```

```
data: r.fit$residual
```

```
X-squared=3.0703, df=6, p-value=0.8
```

```
Box-Pierce test
```

```
data: r.fit$residual
```

```
X-squared=10.2665, df=12, p-value=0.5926
```

根据残差序列的自相关图 (见图 5—31) 和偏自相关图 (见图 5—32) 呈现的性质, 可以判断残差序列自相关系数拖尾, 偏自相关系数 2 阶截尾, 所以对残差序列拟合 AR(2) 模型. 拟合结果为:

$$\varepsilon_t = -1.4995\varepsilon_{t-1} + 0.6028\varepsilon_{t-2} + e_t, \quad e_t \sim N(0, 50.6)$$

模型显著性检验显示该拟合模型显著成立.

综合前面的分析, 对 1952—1998 年中国农业实际国民收入指数序列, 我们可以使用如下残差自回归模型进行拟合:

$$\begin{cases} x_t = 66.1491 + 4.5158t + \varepsilon_t \\ \varepsilon_t = -1.4995\varepsilon_{t-1} + 0.6028\varepsilon_{t-2} + e_t, \quad e_t \sim N(0, 50.6) \end{cases}$$

该模型更易于直观解释原序列的波动规律: 中国农业实际国民收入指数拥有一个长期的线性递增趋势, 平均每期增长 4.5158; 同时, 它还受到诸多随机因素的影响, 导致随机波动序列具有短期自相关性.

5.4 异方差的性质

1982 年 Engle 在分析英国通货膨胀率序列时, 发现经典的 ARIMA 模型始终无法取得理想的拟合效果. 为什么会这样呢? 经过对残差序列的仔细研究, 他发现问题出在残差序列具有异方差性上.

5.4.1 异方差的影响

使用 ARIMA 模型拟合非平稳序列时, 对残差序列有一个重要假定 —— 残差序列 $\{\varepsilon_t\}$ 为零均值白噪声序列. 换言之, 残差序列要满足如下三个假定条件.

1. 零均值

$$E(\varepsilon_t) = 0$$

2. 纯随机

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-i}) = 0, \quad \forall i \geq 1$$

3. 方差齐性

$$\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$$

如果方差齐性假定不成立, 即随机误差序列的方差不再是常数, 它会随着时间的变化而变化, 可以表示为时间的某个函数:

$$\text{Var}(\varepsilon_t) = h(t)$$

这种情况称作异方差.

在残差序列的这三个假定中, 零均值假定最容易实现, 只要对序列进行中心化处理就可以实现. 所以这个假定通常无须检验.

纯随机假定一直是我们重点监控的对象. 如果这个假定不满足, 就说明残差序列中还蕴涵着值得提取的自相关信息. 为了有效检验这个假定条件是否成立, 统计学家构造了许多适用于不同场合的自相关检验统计量, 比如前面介绍的 Q 统计量、LB 统计量、DW 统计量等.

只有第三个假定 —— 方差齐性假定, 在此之前我们没有进行任何检验. 在缺省检验的情况下就默认残差序列一定满足这个条件. 但实际上, 这个假定条件并不总是满足的. 忽视异方差的存在会导致残差的方差被严重低估, 继而参数显著性检验容易犯纳伪错误, 这使得参数的显著性检验失去意义, 最终导致模型的拟合精度受影响. 为了提高模型拟合的精度, 需要对残差序列进行方差齐性检验, 并对异方差序列进行深入分析.

5.4.2 异方差的直观诊断

一、残差图

当残差序列 $\{\varepsilon_t\}$ 方差齐性时, 它应该在一个边界与均值的距离几乎相等的空间随机波动, 不带任何趋势 (见图 5—33), 否则就显示出异方差的性质了 (见图 5—34 至图 5—36).

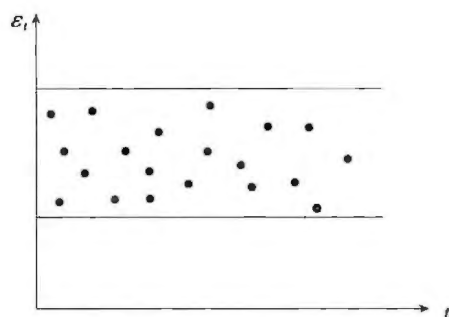


图 5—33 方差齐性残差图

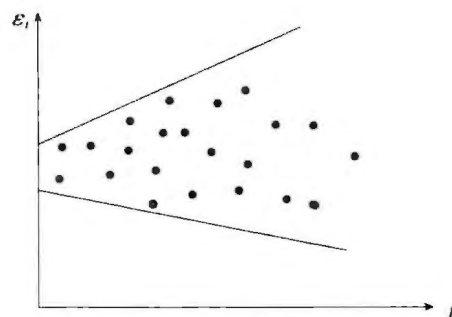


图 5—34 递增型异方差

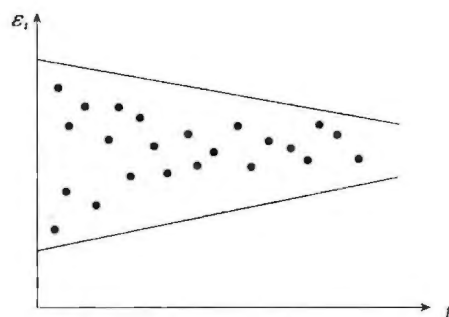


图 5—35 递减型异方差

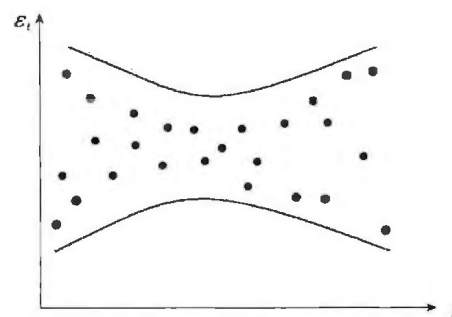


图 5—36 综合型异方差

二、残差平方图

由于残差序列的方差实际上就是它平方的期望, 即

$$\text{Var}(\varepsilon_t) = E(\varepsilon_t^2)$$

所以残差序列是否方差齐性, 主要是考察 ε_t^2 的性质. 我们可以借助残差平方图—— ε_t^2 关于 t 变化的二维坐标图, 对残差序列的方差齐性进行直观诊断.

和残差图的判断原则一样, 假设方差齐性满足, 有

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

这意味着 ε_t^2 应该在某个常数值 σ_ε^2 附近随机波动, 它不应该具有任何明显的趋势, 否则就呈现出异方差性.

【例 5—11】 直观考察美国 1963 年 4 月至 1971 年 7 月短期国库券的月度收益率序列的方差齐性 (数据见表 A1—21).

• 读入数据, 并绘制时序图

```
> h<-read.table("E:/R/data/file21.csv",sep=",",header=T)
> x<-ts(h$yield_rate,start=c(1963,4),frequency=12)
> plot(x)    #结果如图 5—37 所示
```

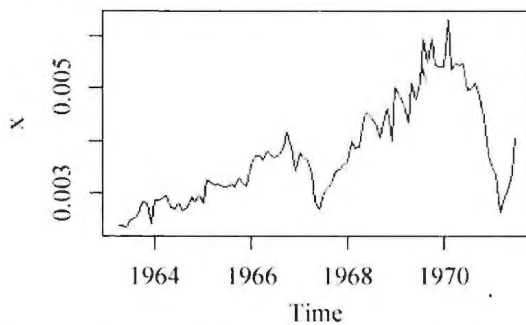


图 5—37 短期国库券月度收益率序列时序图

• 作 1 阶差分, 并绘制差分后残差序列时序图

```
> x.dif<-diff(x)
> plot(x.dif)    #结果如图 5—38 所示
```

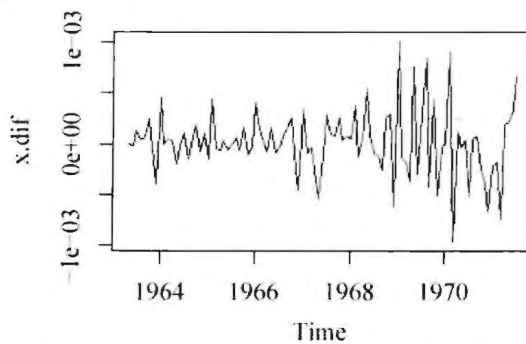


图 5—38 短期国库券月度收益率 1 阶差分后残差时序图

• 绘制 1 阶差分后残差平方图

```
> plot(x.dif^2)    #结果如图 5—39 所示
```

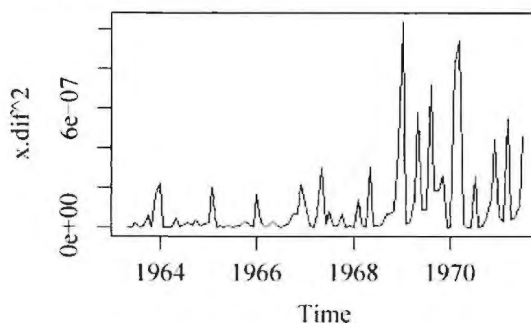


图 5—39 短期国库券月度收益率 1 阶差分后残差平方图

时序图 (见图 5—37) 显示序列显著非平稳. 1 阶差分后序列 (见图 5—38) 显示出均值平稳但方差递增的性质. 进一步观察 1 阶差分后残差平方图 (见图 5—39), 可以发现它更加明显地呈现出异方差的特征.

当残差序列异方差时, 我们需要对它进行进一步的处理, 处理思路有两种:

- (1) 假如已知异方差函数具体形式, 进行方差齐性变换.
- (2) 假如不知异方差函数的具体形式, 拟合条件异方差模型.

5.5 方差齐性变换

一、使用场合

假设序列显示出显著的异方差性, 且方差 σ_t^2 与均值 μ_t 之间具有某种函数关系:

$$\sigma_t^2 = h(\mu_t)$$

式中, $h(\cdot)$ 是某个已知函数.

在这种场合下, 我们的处理思路是尝试寻找一个转换函数 $g(\cdot)$, 使得经转换后的变量 $g(x_t)$ 满足方差齐性:

$$\text{Var}[g(x_t)] = \sigma^2$$

二、转换函数的确定

将 $g(x_t)$ 在 μ_t 附近作 1 阶泰勒展开:

$$g(x_t) \approx g(\mu_t) + (x_t - \mu_t)g'(\mu_t)$$

式中, $g'(\cdot)$ 为 $g(\cdot)$ 的 1 阶导数. 则 $g(x_t)$ 的方差

$$\begin{aligned} \text{Var}[g(x_t)] &\approx \text{Var}[g(\mu_t) + (x_t - \mu_t)g'(\mu_t)] \\ &= [g'(\mu_t)]^2 \cdot \text{Var}(x_t) \\ &= [g'(\mu_t)]^2 h(\mu_t) \end{aligned}$$

显然, 要使得 $\text{Var}[g(x_t)]$ 等于常数, 转换函数的导函数 $g'(\cdot)$ 必须与 $\sqrt{h(\cdot)}$ 具有倒函数关系:

$$g'(\mu_t) = \frac{1}{\sqrt{h(\mu_t)}}$$

在实践中, 许多金融时间序列都呈现出异方差的性质, 而且通常序列的标准差与其水平之间具有某种正比关系, 即序列的水平低时, 序列的波动范围小, 序列的水平高时, 序列的波动范围大. 对于这种异方差的性质, 最简单的假定为:

$$\sigma_t = \mu_t$$

即等价于

$$h(\mu_t) = \mu_t^2$$

要使得原序列经过适当转换后方差齐性, 就必须满足

$$g'(\mu_t) = \frac{1}{h(\mu_t)} = \frac{1}{\mu_t}$$

等价推导出

$$g(\mu_t) = \ln(\mu_t)$$

这意味着对于标准差与水平成正比的异方差序列, 对数变换可以有效地实现方差齐性.

【例 5—11 续】 对美国 1963 年 4 月至 1971 年 7 月短期国库券的月度收益率序列使用方差齐性变换方法进行分析.

- 作对数变换, 并绘制对数变换后时序图

```
> lnx<-log(x)
```

```
> plot(lnx)    #结果如图 5—40 所示
```

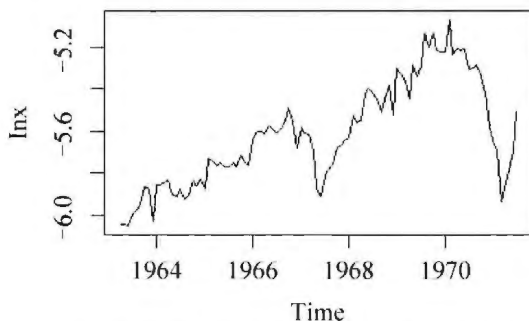


图 5—40 短期国库券月度收益率对数序列时序图

- 作 1 阶差分, 并绘制差分后序列时序图

```
> dif.lnx<-diff(lnx)
```

```
> plot(dif.lnx)    #结果如图 5—41 所示
```

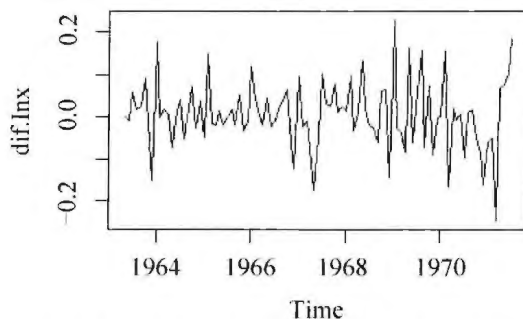


图 5—41 短期国库券月度收益率对数序列 1 阶差分后序列时序图

- 残差白噪声检验

```
> for(i in 1:2) print(Box.test(dif.lnx,lag=6*i))
```

Box-Pierce test

data: dif.lnx

X-squared=3.4118, df=6, p-value=0.7557

Box-Pierce test

data: dif.lnx

X-squared=9.8323, df=12, p-value=0.6307

在前面的分析中, 我们看到短期国库券收益率序列的波动性与序列值之间具有一定的正相关性, 假定它们之间具有正比关系, 对原序列进行对数变换.

对数序列时序图 (见图 5—40) 显示它保持了原序列的变化趋势. 对数序列 1 阶差分后时序图 (见图 5—41) 显示残差序列波动较平稳. 白噪声检验显示该残差序列可视为白噪声序列. 这说明, 通过方差齐性变换, 拟合效果不错. 我们可以得到该序列的拟合模型为:

$$\nabla \ln(x_t) = \varepsilon_t$$

式中, $\{\varepsilon_t\}$ 为零均值白噪声序列.

5.6 条件异方差模型

方差齐性变换为异方差序列的精确拟合提供了一种很好的解决方法, 但这种方法只适用于部分异方差波动序列. 因为要使用方差齐性变换必须事先知道异方差函数的形式, 而这并不是对所有的序列都可以做到的. 实践中, 我们只能根据残差图及残差平方图所显示出来的特点, 使用一些常用的函数形式估计异方差函数. 在进行金融时间序列分析时, 由于金融序列的标准差与水平之间通常具有某种正相关关系, 因此异方差函数经常被假定为:

$$h(\mu_t) = \mu_t^2$$

这导致对数变换在金融时间序列分析中普遍采用.

但大量的实践证明这种假定太单一化了, 对数变换通常只适用于方差随均值的变化而变化的部分序列. 异方差的特征有很多, 我们并不能通过对数变换将所有的异方差序列转换为方差齐性序列. 本节介绍一种在宏观经济领域和金融领域广泛采用的异方差处理方法: 条件异方差模型.

5.6.1 ARCH 模型

一、集群效应

在宏观经济领域和金融领域,经常可以看到具有如下特征的时间序列:它们在消除确定性非平稳因素的影响之后,残差序列的波动在大部分时段是平稳的,但会在某些时段波动持续偏大,在某些时段波动持续偏小,呈现出集群效应 (volatility cluster).

【例 5—12】考察 1926—1991 年标准普尔 500 股票价值加权月度收益率序列 (数据见表 A1—22) 的集群效应特征.

• 读入数据,并绘制时序图

```
> k<-read.table("E:/R/data/file22.csv",sep=";",header=T)
> x<-ts(k$returns,start=c(1926,1),frequency=12)
> plot(x) #结果如图 5—42 所示
```

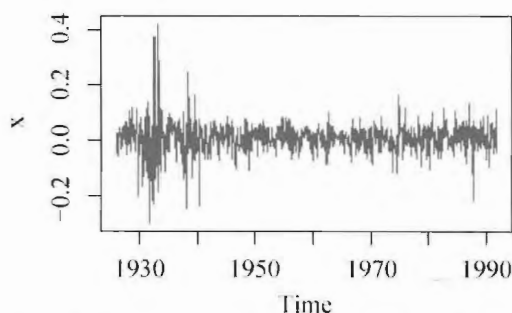


图 5—42 标准普尔 500 股票价值加权月度收益率时序图

• 绘制序列平方图

```
> plot(x^2) #结果如图 5—43 所示
```

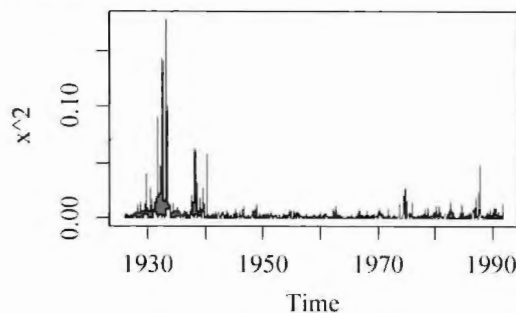


图 5—43 标准普尔 500 股票价值加权月度收益率序列平方图

考察该序列时序图 (见图 5—42). 时序图显示该序列没有显著的非平稳特征. 序列围绕在零值附近波动, 大部分时期波动范围在 $-0.1 \sim 0.1$ 之间. 但是在一些特殊时段, 比如 1930 年前后、1940 年前后、1975 年前后以及 1990 年前后, 序列的波动很大. 这就是集群效应特征.

我们通常用方差来描述序列的波动, 集群效应就意味着在整个序列观察期, 序列的方差基本是齐性的, 但在某一段或某几段时期方差却显著异于期望方差 (见图5—43). 这种序列波动特征给从事利率、汇率、通货膨胀率、股票价格指数等金融时间序列预测的研究人员带来了很大的困扰. 他们对这些变量的预测能力因时期的不同而有相当大的差异.

尤其是对于资产持有者而言, 他们并不关心资产收益率在所有时段的表现, 只关心在他们持有资产的这段时间, 资产收益率会不会有大的波动. 基于序列全程方差齐性的分析方法无法满足他们的需要, 这时需要引入条件异方差模型.

二、ARCH 模型的结构

ARCH 模型的全称是自回归条件异方差模型 (autoregressive conditional heteroskedastic), 有时简称为条件异方差模型. 它是 Engle 于 1982 年在分析英国通货膨胀率序列时提出的残差平方自回归模型.

该模型的构造原理如下: 假设在历史数据已知的情况下, 零均值、纯随机残差序列具有异方差性

$$\text{Var}(\varepsilon_t) = h_t$$

在正态分布的假定下, 有

$$\varepsilon_t / \sqrt{h_t} \sim N(0, 1)$$

异方差等价于残差平方的均值

$$E(\varepsilon_t^2) = h_t$$

使用残差平方序列的自相关系数 $\rho_k = \frac{\text{Cov}(\varepsilon_t^2, \varepsilon_{t-k}^2)}{\text{Var}(\varepsilon_t^2)}$, 可以考察异方差函数的自相关性.

考察的结果无外乎如下两种:

(1) 自相关系数恒为零, 即

$$\rho_k = 0, \quad k = 1, 2, \dots$$

这说明异方差函数是纯随机的, 此时, 历史数据对未来异方差的估计一点作用都没有. 这种情况最难分析, 至今没有有效的方法能够提取到其中的异方差信息.

(2) 存在某个自相关系数不为零, 即

$$\rho_k \neq 0, \quad k \geq 1$$

这意味着在残差平方序列中蕴涵着某种相关信息, 可以通过构造适当的模型提取这些相关信息, 以获得序列异方差波动特征. ARCH 模型就是基于这种场合构造的模型.

具有

$$h_t = E(\varepsilon_t^2) = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2$$

结构的模型称为 q 阶自回归条件异方差模型, 简记为 ARCH(q).

三、ARCH 模型的作用

我们知道构造 ARCH 模型的目的是使用自回归的方法提取误差平方序列中蕴涵的相关信息. 那么构造出这个模型有什么用? 它和之前学过的 ARIMA 模型、残差自回归模型、确定性因素分解模型之间有什么关系?

前面学过的 ARIMA 模型、残差自回归模型、确定性因素分解模型关注的重点都是序列的水平, 我们通过这些模型拟合序列的平均水平并预测它会发生怎样的变化.

比如, 在例 5—6 中我们对 1952—1988 年中国农业实际国民收入指数序列建模, 得到的拟合模型为:

$$x_t = x_{t-1} + \varepsilon_t + 0.735 \, 5\varepsilon_{t-1}, \text{Var}(\varepsilon_t) = 61.95$$

根据这个模型, 我们只要知道过去一年的历史观察值 x_{t-1} 和过去一年的估计误差 ε_{t-1} , 就能估计今年的平均水平 $\hat{x}_t$.

这个序列我们也可以用残差自回归模型或者确定性因素分解模型重新拟合, 它们得到的最终模型结构和结果可能会不一样, 但是它们做的工作性质是相同的, 都是对该序列水平的拟合.

水平只是一个点估计, 无法给出估计精度, 所以通常还需要估计出其置信区间. 之前求置信区间时, 都假定残差序列方差齐性. 这意味着, 无论过去的真实波动情况如何, 在例 5—6 中, 残差序列的方差 $\text{Var}(\varepsilon_t) = 61.95$.

这种假定在异方差场合就会出现问题, 不妨假定该序列残差具有集群效应, 如果考察过去的残差波动情况发现, 近期正处于大起大落的时期, 那么残差的真实方差通常就比 $\text{Var}(\varepsilon_t)$ 大. 这时, 按照方差齐性的假定估计的 95% 的置信区间, 它的真实置信水平一定低于 95%. 这就给投资人带来了不可评估的投资风险.

而 ARCH 模型就是要构造一个模型, 利用历史信息, 得到条件方差信息

$$\text{Var}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2$$

ARCH 模型的实质是将历史波动信息作为条件, 并采用某种自回归形式来刻画波动的变化. 对于一个时间序列而言, 在不同的时刻包含的历史信息不同, 因而相应的条件方差也不同. 利用 ARCH 模型, 可以刻画出随时间变化而变化的条件方差, 它比无条件方差更及时地反映了序列即期波动的特征, 这就是 ARCH 模型的作用. 它和前面介绍的 ARIMA 模型、残差自回归模型、确定性因素分解模型是两种性质的模型, 它关注的是序列的波动性拟合.

拿到一个观察值序列后, 完整的分析应该关注水平和波动两方面. 我们通常会首先提取序列的水平相关信息, 然后分析残差序列中蕴涵的波动相关信息. 将这两方面的信息综合起来才是比较完整和精确的分析结果.

所以使用 ARCH 模型提取异方差中蕴涵的相关信息的完整结构为:

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \end{cases}$$

式中, $f(t, x_{t-1}, x_{t-2}, \dots)$ 为 $\{x_t\}$ 的确定性信息拟合模型, $e_t \stackrel{\text{i.i.d}}{\sim} N(0, \sigma^2)$.

四、ARCH 检验

要拟合 ARCH 模型, 首先需要进行 ARCH 检验. ARCH 检验是一种特殊的异方差检验, 它不仅要求序列具有异方差性, 而且要求这种异方差性是由某种自相关关系造成的, 这种自相关关系可以用残差序列的自回归模型进行拟合. 常用的两种 ARCH 检验统计方法是 Portmanteau Q 检验和 LM 检验.

1. Portmanteau Q 检验

1983 年 McLeod 和 Li 提出了 Portmanteau Q 统计方法, 用于检验残差平方序列的自相关性, 现在它是 ARCH 检验统计方法之一.

该检验方法的构造思想是: 如果残差序列方差非齐, 且具有集群效应, 那么残差平方序列通常具有自相关性. 所以方差非齐检验可以转化为残差平方序列的自相关性检验.

Portmanteau Q 检验的假设条件为:

$$\begin{array}{ll} H_0: \text{残差平方序列纯随机} & H_1: \text{残差平方序列自相关} \\ (\text{方差齐性}) & \leftrightarrow (\text{方差非齐}) \end{array}$$

用 ρ_k 表示残差平方序列 $\{\varepsilon_t^2\}$ 的延迟 k 阶自相关系数, 则该假设条件可以等价表达为:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_q = 0 \leftrightarrow H_1: \rho_1, \rho_2, \dots, \rho_q \text{ 不全为零}$$

Portmanteau Q 检验统计量实际上就是 $\{\varepsilon_t^2\}$ 的 LB 统计量

$$Q(q) = n(n+2) \sum_{i=1}^q \frac{\rho_i^2}{n-i}$$

式中, n 为观察序列长度; ρ_i 为残差序列延迟 i 阶自相关系数, 有

$$\rho_i = \sqrt{\frac{\sum_{t=i+1}^n (\varepsilon_t^2 - \hat{\sigma}^2)(\varepsilon_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^n (\varepsilon_t^2 - \hat{\sigma}^2)^2}}, \quad \hat{\sigma}^2 = \frac{\sum_{t=1}^n \varepsilon_t^2}{n}$$

原假设成立时, Portmanteau Q 统计量近似服从自由度为 $q-1$ 的 χ^2 分布

$$Q(q) \stackrel{*}{\sim} \chi^2(q-1)$$

当 $Q(q)$ 检验统计量的 P 值小于显著性水平 α 时, 拒绝原假设, 认为该序列方差非齐且具有自相关关系.

2. 拉格朗日乘子检验

1982 年 Engle 提出了一种重要的 ARCH 检验方法: 拉格朗日乘子检验 (Lagrange multiplier test), 简记为 LM 检验.

拉格朗日乘子检验的构造思想是: 如果残差序列方差非齐, 且具有集群效应, 那么残差平方序列通常具有自相关性. 那么我们就可以尝试使用自回归模型 (ARCH(q) 模型) 拟合残差平方序列

$$\varepsilon_t^2 = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 + e_t \quad (5.7)$$

于是方差齐性检验就可以转化为这个方程是否显著成立的检验.

如果方程显著成立 (至少存在一个参数 λ_j 非零), 那就意味着残差平方序列具有自相关性, 可以用该回归方程提取自相关信息.

反之, 如果方程不能显著成立 ($\lambda_j = 0; j = 1, 2, \dots, q$), 就意味着残差平方序列不存在显著的自相关性, 不能拒绝方差齐性假定. 所以拉格朗日乘子检验实际上就是残差平方序列 ε_t^2 自回归方程的显著性检验.

拉格朗日乘子检验的假设条件为:

$$H_0: \text{残差平方序列纯随机} \leftrightarrow H_1: \text{残差平方序列具有自相关性}$$

对残差平方序列构造 q 阶自回归方程 (5.7), 假设条件等价于:

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_q = 0 \leftrightarrow H_1: \lambda_1, \lambda_2, \dots, \lambda_q \text{ 不全为零}$$

记总误差平方和为 $SST = \sum_{t=q+1}^T \varepsilon_t^2$, 自由度为 $T - q - 1$. 回归平方和为 $SSR =$

$SST - SSE$, 自由度为 q . 其中, SSE 为回归方程残差平方和, $SSE = \sum_{t=q+1}^T e_t^2$, 自由度为 $T - 2q - 1$. 则 LM 检验统计量为:

$$LM(q) = \frac{(SST - SSE)/q}{SSE/(T - 2q - 1)}$$

原假设成立时, $LM(q)$ 统计量近似服从自由度为 $q - 1$ 的 χ^2 分布

$$LM(q) \sim \chi^2(q - 1)$$

当 $LM(q)$ 检验统计量的 P 值小于显著性水平 α 时, 拒绝原假设, 认为该序列方差非齐, 并且可以用 q 阶自回归模型拟合残差平方序列中的自相关关系.

在 R 语言中, 做拉格朗日乘子检验之前, 需先下载 FinTS 程序包, 并用 library 函数调用该包. 在这个包里有一个 ArchTest 函数专门作 LM 检验. 而 Portmanteau Q 检验其实就是对残差平方序列进行纯随机性检验. 所以对残差序列进行平方运算之后, 只要调用 Box.test 函数就可以完成 Portmanteau Q 检验.

拟合 ARCH 模型可以调用 tseries 包中的 garch 函数. garch 函数的命令格式为:

garch (x, order=)

式中:

-x: 序列名.

-order: 拟合模型阶数.

(1) 拟合 ARCH(q) 模型为 order=c(0,q).

(2) 拟合 GARCH(p,q) 模型为 order=c(p,q).

【例 5—12 续】 对 1926—1991 年标准普尔 500 股票价值加权月度收益率序列进行 ARCH 检验, 并拟合该序列的波动特征.

• LM 检验

```
> for(i in 1:5) print(ArchTest(x,lag=i))
```

ARCH LM-test; Null hypothesis: no ARCH effects

data: x

Chi-squared=55.5215, df=1, p-value=9.248e-14

ARCH LM-test; Null hypothesis: no ARCH effects

data: x

Chi-squared=68.292, df=2, p-value=1.443e-15

ARCH LM-test; Null hypothesis: no ARCH effects

data: x

Chi-squared=87.1291, df=3, p-value < 2.2e-16

ARCH LM-test; Null hypothesis: no ARCH effects

data: x

Chi-squared=87.2819, df=4, p-value < 2.2e-16

ARCH LM-test; Null hypothesis: no ARCH effects

data: x

Chi-squared=87.4919, df=5, p-value < 2.2e-16

• Portmanteau Q 检验

```
> for(i in 1:5) print(Box.test(x^2,lag=i))
```

Box-Pierce test

```
data: x^2
X-squared=55.5445, df=1, p-value=9.137e-14

Box-Pierce test
data: x^2
X-squared=85.0896, df=2, p-value < 2.2e-16

Box-Pierce test
data: x^2
X-squared=126.0621, df=3, p-value < 2.2e-16

Box-Pierce test
data: x^2
X-squared=138.7394, df=4, p-value < 2.2e-16

Box-Pierce test
data: x^2
X-squared=141.8952, df=5, p-value < 2.2e-16
```

• 拟合 ARCH(3) 模型

```
> x.fit<-garch(x,order=c(0,3))
> summary(x.fit)
```

Call:

```
garch(x=x, order=c(0, 3))
```

Model:

```
GARCH(0,3)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.2420	-0.3985	0.1671	0.7501	4.4193

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
a0	1.437e-03	6.903e-05	20.809	< 2e-16 ***

```
a1 7.954e-02  2.821e-02  2.820  0.0048  **
a2 2.231e-01  2.587e-02  8.624  < 2e-16  ***
a3 2.732e-01  4.384e-02  6.232  4.61e-10  ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Diagnostic Tests:
```

```
  Jarque Bera Test
```

```
data: Residuals
```

```
X-squared=585.2712, df=2, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: Squared.Residuals
```

```
X-squared=0, df=1, p-value=0.9976
```

- 绘制条件异方差模型拟合的 95% 置信区间

```
> x.pred<-predict(x.fit)
```

```
> plot(x.pred)    #结果如图 5—44 所示
```

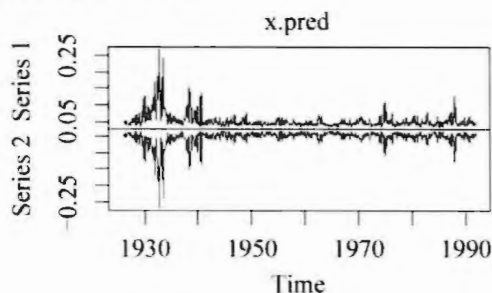


图 5—44 条件异方差模型拟合的 95% 置信区间

- 条件异方差置信区间和方差齐性置信区间比较图示

```
> plot(x)
```

```
> lines(x.pred[,1], col=2)
```

```
> lines(x.pred[,2], col=2)
```

```
> abline(h=1.96*sd(x),col=4,lty=2)
```

```
> abline(h=-1.96*sd(x),col=4,lty=2)
```

```
#结果如图 5—45 所示
```

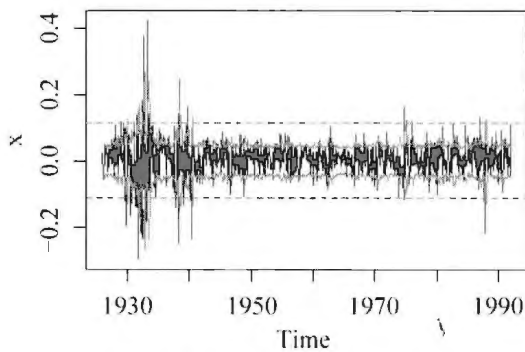


图 5—45 条件异方差置信区间和方差齐性置信区间比较图

说明：中间的波动曲线是价值加权月度收益率观察值序列，两条平行虚线是根据方差齐性得到的 95% 置信区间，两条波动实线为根据 ARCH(3) 模型拟合的 95% 的置信区间。

LM 检验和 Q 检验都显示该序列显著方差非齐，且残差平方序列具有显著自相关关系，可以用 ARCH 模型提取残差平方序列中蕴涵的相关信息。而 LM 检验和 Q 检验都显示 1 阶至 5 阶 ARCH 模型均显著成立，这说明残差平方序列具有长期相关，ARCH(q) 模型的阶数 q 将会比较大。本例尝试拟合过 ARCH(5) 和 ARCH(4)，这两个模型的参数检验结果均显示有参数不显著，直至 ARCH(3) 模型，检验结果显示模型显著，参数均显著。所以最后得到的拟合模型是 ARCH(3) 模型：

$$E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = 0.001\,437 + 0.079\,54\varepsilon_{t-1}^2 + 0.223\,1\varepsilon_{t-2}^2 + 0.273\,2\varepsilon_{t-3}^2$$

图 5—44 给出了条件异方差模型拟合的 95% 置信区间的范围，考虑到原序列的波动特征，显然条件异方差模型更好地拟合了原序列的集群效应波动特征。图 5—45 显示条件异方差模型拟合的置信区间比无条件方差两条平行线给出的 95% 的置信区间更加符合原序列的真实波动情况。这说明条件异方差模型对序列波动的预测将会更加准确。

5.6.2 GARCH 模型

一、GARCH 模型的定义

ARCH 模型的实质是使用残差平方序列的 q 阶移动平均拟合当期异方差函数值。由于移动平均模型具有自相关系数 q 阶截尾性，所以 ARCH 模型实际上只适用于异方差函数短期自相关过程。

但是在实践中，有些残差序列的异方差函数是具有长期自相关性的，这时如果使用 ARCH 模型拟合异方差函数，将会产生很高的移动平均阶数，增加参数估计的难度并最终影响 ARCH 模型的拟合精度。

为了修正这个问题，Bollerslov 在 1985 年提出了广义自回归条件异方差 (generalized autoregressive conditional heteroskedastic) 模型，它的结构如下：

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \end{cases}$$

式中, $f(t, x_{t-1}, x_{t-2}, \dots)$ 为 $\{x_t\}$ 的确定性信息拟合模型; $e_t \stackrel{\text{i.i.d}}{\sim} N(0, \sigma^2)$. 这个模型简记为 GARCH(p, q).

GARCH 模型实际上就是在 ARCH 模型的基础上, 增加考虑了异方差函数的 p 阶自相关性而形成的; 它可以有效地拟合具有长期记忆性的异方差函数. 显然 ARCH 模型是 GARCH 模型的一个特例, ARCH(q) 模型实际上就是 $p = 0$ 的 GARCH(p, q) 模型.

二、AR-GARCH 模型

对序列 $\{\varepsilon_t\}$ 拟合 GARCH 模型有一个基本要求: $\{\varepsilon_t\}$ 为零均值、纯随机、异方差序列. 有时回归函数 $f(t, x_{t-1}, x_{t-2}, \dots)$ 不能充分提取原序列中的相关信息, $\{\varepsilon_t\}$ 可能具有自相关性, 而不是纯随机的. 这时需要先对 $\{\varepsilon_t\}$ 拟合自回归模型, 再考察自回归残差序列 $\{v_t\}$ 的方差齐性, 如果 $\{v_t\}$ 异方差, 对它拟合 GARCH 模型. 这样构造的模型称为 AR(m)-GARCH(p, q) 模型

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sum_{k=1}^m \beta_k \varepsilon_{t-k} + v_t \\ v_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j v_{t-j}^2 \end{cases}$$

式中, $f(t, x_{t-1}, x_{t-2}, \dots)$ 为 $\{x_t\}$ 的确定性信息拟合模型; $e_t \stackrel{\text{i.i.d}}{\sim} N(0, 1)$.

【例 5—13】 分析拟合 1979 年 12 月 31 日至 1991 年 12 月 31 日外币对美元的日兑换率序列 (数据见表 A1—23).

• 读入数据, 并绘制时序图

```
> w<-read.table("E:/R/data/file23.csv",sep=",",header=T)
> x<-ts(w$exchange_rates,start=c(1979,12,31),frequency=365)
> plot(x)    #结果如图 5—46 所示
```

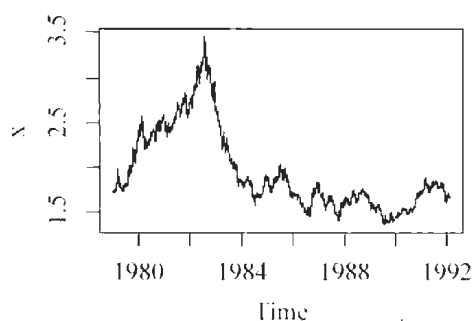



图 5 46 外币对美元的日兑换率序列时序图

● 对差分序列性质考察

```
> plot(diff(x))      #结果如图 5 47 所示
> acf(diff(x))       #结果如图 5 48 所示
> pacf(diff(x))      #结果如图 5 49 所示
> for( i in 1:2) print(Box.test(diff(x),lag=6*i))
```

Box-Pierce test

data: diff(x)

X-squared=12.9168, df=6, p-value=0.04438

Box-Pierce test

data: diff(x)

X-squared=29.7118, df=12, p-value=0.003085

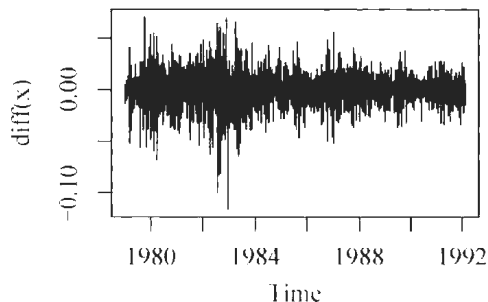


图 5—47 外币对美元的日兑换率 1 阶差分后序列时序图

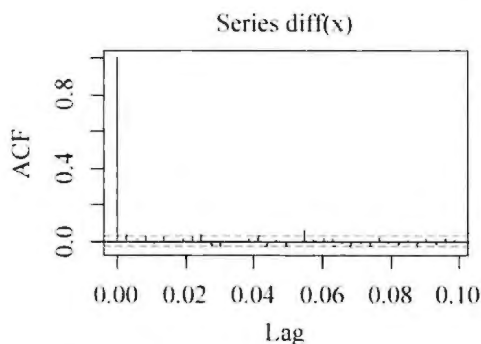


图 5—48 外币对美元的日兑换率 1 阶差分后序列自相关图

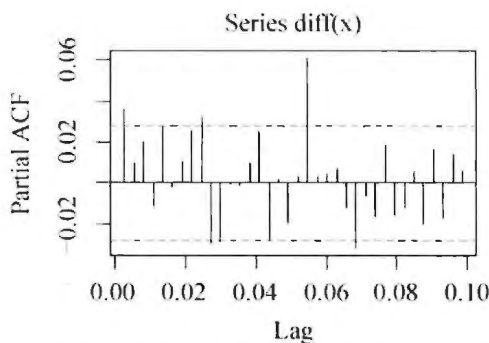


图 5—49 外币对美元的日兑换率 1 阶差分后序列偏自相关图

• 水平相关信息提取, 拟合 ARIMA(0,1,1) 模型

```
> x.fit<-arima(x,order=c(0,1,1))
> x.fit
```

Call:

```
arima(x=x, order=c(0, 1, 1))
```

Coefficients:

```
      ma1
      0.0357
s.e. 0.0143
```

sigma^2 estimated as 0.0002007: log likelihood=13545.61, aic=-27087.22

• 残差白噪声检验

```
> for( i in 1:6) print(Box.test(x.fit$residual,type="Ljung-Box",lag=i))
```

Box-Ljung test

```
data: x.fit$residual
```

```
X-squared=5e-04, df=1, p-value=0.9815
```

Box-Ljung test

data: x.fit\$residual

X-squared=0.551, df=2, p-value=0.7592

Box-Ljung test

data: x.fit\$residual

X-squared=2.6528, df=3, p-value=0.4483

Box-Ljung test

data: x.fit\$residual

X-squared=3.3062, df=4, p-value=0.5079

Box-Ljung test

data: x.fit\$residual

X-squared=6.8276, df=5, p-value=0.2338

Box-Ljung test

data: x.fit\$residual

X-squared=6.8306, df=6, p-value=0.3368

• 水平预测, 并绘制预测图

> x.fore<-forecast(x.fit,h=365)

> plot(x.fore) #结果如图 5 50 所示

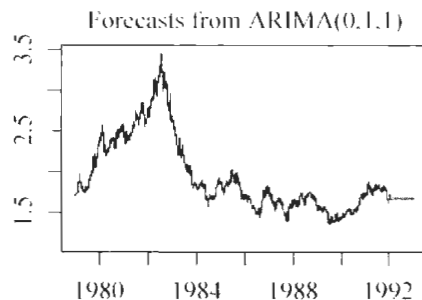


图 5—50 外币对美元的日兑换率序列预测图

- 条件异方差检验 (Portmanteau Q 检验)

```
> for( i in 1:6) print(Box.test(x.fit$residual^2,type="Ljung-Box",lag=i))
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=82.8033, df=1, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=237.9039, df=2, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=343.3273, df=3, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=490.8356, df=4, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=602.099, df=5, p-value < 2.2e-16
```

```
Box-Ljung test
```

```
data: x.fit$residual^2
```

```
X-squared=841.9569, df=6, p-value < 2.2e-16
```

- 拟合 GARCH(1,1) 模型

```
> r.fit<-garch(x.fit$residual,order=c(1,1))
```

```
> summary(r.fit)
```

```
Call:
```

```
garch(x=x.fit$residual, order=c(1,1))
```

Model:

GARCH(1,1)

Residuals:

Min	1Q	Median	3Q	Max
-4.83074	-0.58407	0.02616	0.58758	4.54060

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
a0	2.133e-06	3.014e-07	7.077	1.48e-12 ***
a1	7.623e-02	5.456e-03	13.972	< 2e-16 ***
b1	9.144e-01	6.015e-03	152.009	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:

Jarque Bera Test

data: Residuals

X-squared=319.2325, df=2, p-value < 2.2e-16

Box-Ljung test

data: Squared.Residuals

X-squared=0.2802, df=1, p-value=0.5966

• 绘制波动置信区间

```
> r.pred<-predict(r.fit)
```

```
> plot(r.pred) #结果如图 5—51 所示
```

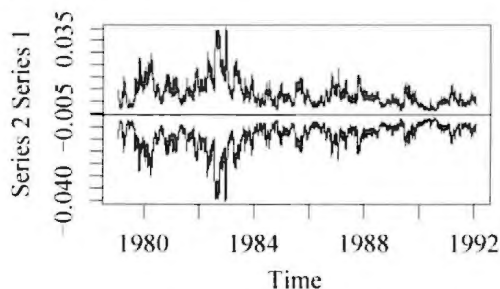


图 5—51 外币对美元的日兑换率残差序列波动置信区间

序列时序图 (见图 5—46) 显示序列非平稳, 有明显的趋势特征. 差分后序列时序图 (见图 5—47) 显示趋势消除, 但是有明显的集群效应. 所以分析该序列需要同时提

取水平相关信息和波动相关信息.

水平信息的提取是考察差分后序列的自相关图 (见图 5—48) 和偏自相关图 (见图 5—49), 拟合 ARIMA(0,1,1) 模型, 该拟合模型的残差白噪声检验显示该模型显著成立. 利用该拟合模型, 还可以预测序列未来的水平 (见图 5—50).

波动信息的提取首先是考察 ARIMA(0,1,1) 模型的残差平方序列的异方差特征. Portmanteau Q 检验显示残差序列显著方差非齐, 且具有长期相关性. 所以构造 GARCH(1,1) 模型, 并根据该模型的拟合结果绘制波动的 95% 置信区间图 (见图 5—51).

综合水平模型和波动模型, 我们得到的完整拟合模型为:

$$\begin{cases} \nabla x_t = \varepsilon_t + 0.0357\varepsilon_{t-1} + v_t, v_t \sim N(0, 0.0002) \\ v_t = \sqrt{h_t}e_t \\ h_t = 0.9144h_{t-1} + 0.07623v_{t-1}^2 \end{cases}$$

5.6.3 GARCH 的衍生模型

GARCH 模型给出了对波动性进行描述的方法, 为大量的金融序列提供了有效的分析方法, 它是迄今为止最常用、最便捷的异方差序列拟合模型. 但大量的使用经验表明, 它也存在一些不足.

一是它对参数的约束非常严格. 无条件方差必须非负的要求, 导致了参数非负的约束条件

$$\omega > 0, \eta_i \geq 0, \lambda_j \geq 0$$

同时有条件方差必须平稳的要求, 要求参数有界

$$\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j < 1$$

参数的约束条件在一定程度上限制了 GARCH 模型的适用范围.

二是它对正负扰动的反应是对称的. 扰动项是真实值与预测值之差. 如果扰动项为正, 说明真实值比预测值大, 对于投资者而言就是获得超预期收益. 如果扰动项为负, 说明真实值比预测值小, 对于投资者而言就是出现了超预期亏损. 以 ARCH(1) 模型为例, $\varepsilon_t^2 = \phi\varepsilon_{t-1}^2$, 即无论 ε_{t-1} 是正是负, 它对下一期的影响系数都是 ϕ . 这意味着无论上一期的投资是收益还是亏损, 不影响投资人的下一期投资状况. 这与实际情况不符, 大量的实践经验显示, 投资人在面对收益和亏损时的反应不是对称的. 出现收益时, 通常反应比较慢; 出现亏损时, 通常反应比较快. 忽视这种信息的不对称性, 有时会影响预测的精度.

为了拓宽 GARCH 模型的使用范围、提高 GARCH 模型的拟合精度, 统计学家从不同的角度出发, 构造了多个 GARCH 模型的衍生模型.

一、指数 GARCH 模型 (EGARCH)

Nelson 于 1991 年提出了 EGARCH 模型, 该模型结构如下:

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ \ln(h_t) = \omega + \sum_{i=1}^p \eta_i \ln(h_{t-i}) + \sum_{j=1}^q \lambda_j g(e_t) \\ g(e_t) = \theta e_t + \gamma[|e_t| - E|e_t|] \end{cases}$$

式中, $f(t, x_{t-1}, x_{t-2}, \cdots)$ 为 $\{x_t\}$ 的确定性信息拟合模型; $e_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

$\ln(h_t)$ 是条件方差的对数, 它可以是正数, 也可以是负数. 于是后面表达式 $\omega + \sum_{i=1}^p \eta_i \ln(h_{t-i}) + \sum_{j=1}^q \lambda_j g(e_t)$ 中的参数不需要任何非负假定. 可见, EGARCH 模型的一个改进是放松了对 GARCH 模型的参数非负约束.

EGARCH 模型的第二个改进是引入加权扰动函数 $g(e_t)$, 通过特殊的函数构造, 能对正、负扰动进行非对称处理.

$$\begin{aligned} g(e_t) &= \theta e_t + \gamma[|e_t| - E|e_t|] \\ &= \begin{cases} (\theta + \gamma)e_t - \gamma E|e_t|, & e_t > 0 \\ (\theta - \gamma)e_t - \gamma E|e_t|, & e_t < 0 \end{cases} \end{aligned}$$

其中, $E[g(e_t)] = 0$; $e_t \sim N(0, 1)$; $E|e_t| = \sqrt{2/\pi}$.

二、方差无穷 GARCH 模型 (IGARCH)

当 GARCH 模型平稳时, ε_t 的无条件方差为:

$$\text{Var}(\varepsilon_t) = \frac{\omega}{1 - \left(\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j \right)}$$

因而 GARCH 模型参数有界约束的目的是保证 ε_t 的无条件方差有界, GARCH 模型能实现宽平稳.

如果把这个约束条件改写为:

$$\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j = 1$$

就构成 IGARCH 模型. 对 IGARCH 模型而言, ε_t 的无条件方差无界, 所以它的无条件方差无意义.

IGARCH 模型适用于描述具有单位根特征 (随机游走) 的条件异方差. 从理论角度来看, IGARCH 现象可能是由波动率常有的水平移动引起的.

考察 IGARCH(1,1) 模型的条件异方差预测值

选择适当模型拟合该序列的发展, 并估计下一天的收盘价.

2. 我国 1949—2008 年每年铁路货运量 (单位: 万吨) 数据如表 5—2 所示. 请选择适当的模型拟合该序列, 并预测 2009—2013 年我国铁路货运量.

表 5—2

年份	铁路货运量	年份	铁路货运量	年份	铁路货运量
1949	5 589	1969	53 120	1989	151 489
1950	9 983	1970	68 132	1990	150 681
1951	11 083	1971	76 471	1991	152 893
1952	13 217	1972	80 873	1992	157 627
1953	16 131	1973	83 111	1993	162 794
1954	19 288	1974	78 772	1994	163 216
1955	19 376	1975	88 955	1995	165 982
1956	24 605	1976	84 066	1996	171 024
1957	27 421	1977	95 309	1997	172 149
1958	38 109	1978	110 119	1998	164 309
1959	54 410	1979	111 893	1999	167 554
1960	67 219	1980	111 279	2000	178 581
1961	44 988	1981	107 673	2001	193 189
1962	35 261	1982	113 495	2002	204 956
1963	36 418	1983	118 784	2003	224 248
1964	41 786	1984	124 074	2004	249 017
1965	49 100	1985	130 709	2005	269 296
1966	54 951	1986	135 635	2006	288 224
1967	43 089	1987	140 653	2007	314 237
1968	42 095	1988	144 948	2008	330 354

3. 据美国国家安全委员会统计, 1973—1978 年美国月度事故死亡数据如表 5—3 所示. 请选择适当模型拟合该序列的发展.

表 5—3

时间	死亡人数	时间	死亡人数	时间	死亡人数
1973 年 1 月	9 007	1973 年 10 月	9 938	1974 年 7 月	10 120
1973 年 2 月	8 106	1973 年 11 月	9 161	1974 年 8 月	9 823
1973 年 3 月	8 928	1973 年 12 月	8 927	1974 年 9 月	8 743
1973 年 4 月	9 137	1974 年 1 月	7 750	1974 年 10 月	9 129
1973 年 5 月	10 017	1974 年 2 月	6 981	1974 年 11 月	8 710
1973 年 6 月	10 826	1974 年 3 月	8 038	1974 年 12 月	8 680
1973 年 7 月	11 317	1974 年 4 月	8 422	1975 年 1 月	8 162
1973 年 8 月	10 744	1974 年 5 月	8 714	1975 年 2 月	7 306
1973 年 9 月	9 713	1974 年 6 月	9 512	1975 年 3 月	8 124

续前表

时间	死亡人数	时间	死亡人数	时间	死亡人数
1975 年 4 月	7 870	1976 年 7 月	10 078	1977 年 10 月	8 850
1975 年 5 月	9 387	1976 年 8 月	9 179	1977 年 11 月	8 265
1975 年 6 月	9 556	1976 年 9 月	8 037	1977 年 12 月	8 796
1975 年 7 月	10 093	1976 年 10 月	8 488	1978 年 1 月	7 836
1975 年 8 月	9 620	1976 年 11 月	7 874	1978 年 2 月	6 892
1975 年 9 月	8 285	1976 年 12 月	8 647	1978 年 3 月	7 791
1975 年 10 月	8 433	1977 年 1 月	7 792	1978 年 4 月	8 129
1975 年 11 月	8 160	1977 年 2 月	6 957	1978 年 5 月	9 115
1975 年 12 月	8 034	1977 年 3 月	7 726	1978 年 6 月	9 434
1976 年 1 月	7 717	1977 年 4 月	8 106	1978 年 7 月	10 484
1976 年 2 月	7 461	1977 年 5 月	8 890	1978 年 8 月	9 827
1976 年 3 月	7 776	1977 年 6 月	9 299	1978 年 9 月	9 110
1976 年 4 月	7 925	1977 年 7 月	10 625	1978 年 10 月	9 070
1976 年 5 月	8 634	1977 年 8 月	9 302	1978 年 11 月	8 633
1976 年 6 月	8 945	1977 年 9 月	8 314	1978 年 12 月	9 240

4. 1750—1849 年瑞典人口出生率数据如表 5—4 所示.

- (1) 请选择适当的模型拟合该序列的发展.
- (2) 检验序列的异方差性, 如果存在异方差, 请拟合条件异方差.

表 5—4

年份	出生率 (‰)	年份	出生率 (‰)	年份	出生率 (‰)	年份	出生率 (‰)
1750	9	1766	8	1782	4	1798	9
1751	12	1767	9	1783	1	1799	5
1752	8	1768	5	1784	1	1800	4
1753	12	1769	5	1785	2	1801	3
1754	10	1770	6	1786	6	1802	7
1755	10	1771	4	1787	7	1803	7
1756	8	1772	-9	1788	7	1804	6
1757	2	1773	-27	1789	-2	1805	8
1758	0	1774	12	1790	-1	1806	3
1759	7	1775	10	1791	7	1807	4
1760	10	1776	10	1792	12	1808	-5
1761	9	1777	8	1793	10	1809	-14
1762	4	1778	8	1794	10	1810	1
1763	1	1779	9	1795	4	1811	6
1764	7	1780	14	1796	9	1812	3
1765	5	1781	7	1797	10	1813	2

续前表

年份	出生率 (‰)	年份	出生率 (‰)	年份	出生率 (‰)	年份	出生率 (‰)
1814	6	1823	16	1832	7	1841	11
1815	1	1824	14	1833	12	1842	11
1816	13	1825	16	1834	8	1843	9
1817	10	1826	12	1835	14	1844	12
1818	10	1827	8	1836	11	1845	13
1819	6	1828	7	1837	5	1846	8
1820	9	1829	6	1838	5	1847	6
1821	10	1830	9	1839	5	1848	10
1822	13	1831	4	1840	10	1849	13

5. 1867—1938 年英国 (英格兰及威尔士) 绵羊数量如表 5—5 (行数据) 所示.

表 5—5

2 203	2 360	2 254	2 165	2 024	2 078	2 214	2 292	2 207	2 119	2 119	2 137
2 132	1 955	1 785	1 747	1 818	1 909	1 958	1 892	1 919	1 853	1 868	1 991
2 111	2 119	1 991	1 859	1 856	1 924	1 892	1 916	1 968	1 928	1 898	1 850
1 841	1 824	1 823	1 843	1 880	1 968	2 029	1 996	1 933	1 805	1 713	1 726
1 752	1 795	1 717	1 648	1 512	1 338	1 383	1 344	1 384	1 484	1 597	1 686
1 707	1 640	1 611	1 632	1 775	1 850	1 809	1 653	1 648	1 665	1 627	1 791

- (1) 确定该序列的平稳性.
- (2) 选择适当模型, 拟合该序列的发展.
- (3) 利用拟合模型预测 1939—1945 年英国绵羊的数量.

6. 1969 年 1 月至 1994 年 9 月澳大利亚储备银行 (Reserve Bank of Australia) 2 年期有价证券月度利率数据如表 5—6 (行数据) 所示.

表 5—6

4.99	5.00	5.03	5.03	5.25	5.26	5.30	5.45	5.49	5.52	5.70
5.68	5.65	5.80	6.50	6.45	6.48	6.45	6.35	6.40	6.43	6.43
6.44	6.45	6.48	6.40	6.35	6.40	6.30	6.32	6.35	6.13	5.70
5.58	5.18	5.18	5.17	5.15	5.21	5.23	5.05	4.65	4.65	4.60
4.67	4.69	4.68	4.62	4.63	4.90	5.44	5.56	6.04	6.06	6.06
8.07	8.07	8.10	8.05	8.06	8.07	8.06	8.11	8.60	10.80	11.00
11.00	11.00	9.48	9.18	8.62	8.3	8.47	8.44	8.44	8.46	8.49
8.54	8.54	8.50	8.44	8.49	8.40	8.46	8.50	8.50	8.47	8.47
8.47	8.48	8.48	8.54	8.56	8.39	8.89	9.91	9.89	9.91	9.91
9.90	9.88	9.86	9.86	9.74	9.42	9.27	9.26	8.99	8.83	8.83
8.83	8.82	8.83	8.83	8.79	8.79	8.69	8.66	8.67	8.72	8.77
9.00	9.61	9.70	9.94	9.94	9.94	9.95	9.94	9.96	9.97	10.83

10.75	11.20	11.40	11.54	11.50	11.34	11.50	11.50	11.58	12.42	12.85
13.10	13.12	13.10	13.15	13.10	13.20	14.20	14.75	14.60	14.60	14.45
14.50	14.80	15.85	16.20	16.50	16.40	16.40	16.35	16.10	13.70	13.50
14.00	12.30	12.00	14.35	14.60	12.50	12.75	13.70	13.45	13.55	12.60
12.00	11.00	11.60	12.05	12.35	12.70	12.45	12.55	12.20	12.10	11.15
11.85	12.10	12.50	12.90	12.50	13.20	13.65	13.65	13.50	13.45	13.35
14.45	14.30	15.05	15.55	15.65	14.65	14.15	13.30	12.65	12.70	12.80
14.50	15.10	15.15	14.30	14.25	14.05	14.70	15.05	14.05	13.80	13.25
13.00	12.85	12.60	11.80	13.00	12.35	11.45	11.35	11.55	10.85	10.90
12.30	11.70	12.05	12.30	12.90	13.05	13.30	13.85	14.65	15.05	15.15
14.85	15.70	15.40	15.10	14.80	15.80	15.80	15.00	14.40	13.80	14.30
14.15	14.45	14.10	14.05	13.75	13.30	13.00	12.55	12.25	11.85	11.50
11.10	11.15	10.70	10.25	10.55	10.25	10.30	9.60	8.40	8.20	7.25
8.35	8.25	8.30	7.40	7.15	6.35	5.65	7.40	7.20	7.05	7.10
6.85	6.50	6.25	5.95	5.65	5.85	5.45	5.30	5.20	5.55	5.15
5.40	5.35	5.10	5.80	6.35	6.50	6.95	8.05	7.85	7.75	8.60

(1) 考察该序列的方差齐性.

(2) 选择适当的模型拟合该序列的发展.

在前几章中我们介绍的都是一元时间序列的分析方法. 实际上, 很多序列的变化规律都会受到其他序列的影响. 比如说当分析居民消费支出序列时, 消费会受到收入的显著影响, 如果将收入也纳入研究范围, 就能得到更精确的消费预测. 这就涉及多元时间序列分析.

对多元时间序列的分析很早就开始了. 1970 年, Box 和 Jenkins 在 *Time Series Analysis: Forecasting and Control* 一书中就采用过将天然气的输入速率作为输入变量, 研究 CO₂ 的输出浓度, 由此将时间序列的分析领域由一元拓展到了多元的场合.

但是从技术上讲, 当时要求输入序列和响应序列都是平稳的. 显然响应序列和输入序列均平稳的要求是非常苛刻的, 这严重限制了多元时间序列分析的运用和发展. 直到 1987 年 Engle 和 Granger 提出了协整 (cointegration) 的概念.

在协整理论下, 并不要求响应序列和输入序列自身平稳, 只要求它们的回归残差序列平稳. 残差序列平稳比响应序列与输入序列均平稳容易实现多了. 这个概念的提出极大地促进了多元时间序列分析的发展, 它实际上是将多元回归分析和时间序列分析有机地结合在一起, 有效地提高了预测的精度.

本章将介绍多元时间序列建模的原理和方法.

6.1 平稳多元序列建模

1970 年, Box 和 Jenkins 采用带输入变量的 ARIMA 模型, 为平稳多元序列建模.

该模型的构造思想是: 假设响应序列 $\{y_t\}$ 和输入序列 (即自变量序列) $\{x_{1t}\}$, $\{x_{2t}\}$, $\dots$, $\{x_{kt}\}$ 均平稳, 首先构建响应序列和输入序列的回归模型:

$$y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \varepsilon_t$$

式中, $\Phi_i(B)$ 为第 i 个输入变量的自回归系数多项式; $\Theta_i(B)$ 为第 i 个输入变量的移动平均系数多项式; l_i 为第 i 个输入变量的延迟阶数; $\{\varepsilon_t\}$ 为回归残差序列.

因为 $\{y_t\}$ 和 $\{x_{1t}\}$, $\{x_{2t}\}$, $\dots$, $\{x_{kt}\}$ 均平稳, 平稳序列的线性组合仍然是平稳的, 所以残差序列 $\{\varepsilon_t\}$ 为平稳序列:

$$\varepsilon_t = y_t - \left(\mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} \right)$$

使用 ARMA 模型继续提取残差序列 $\{\varepsilon_t\}$ 中的相关信息. 最终得到的模型为:

$$\begin{cases} y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \varepsilon_t \\ \varepsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t \end{cases} \tag{6.1}$$

模型 (6.1) 称为动态回归模型, 简记为 ARIMAX. 式中, $\Phi(B)$ 为残差序列自回归系数多项式; $\Theta(B)$ 为残差序列移动平均系数多项式; a_t 为零均值白噪声序列.

【例 6—1】 在天然气炉中, 输入的是天然气, 输出的是 CO_2 , CO_2 的输出浓度与天然气的输入速率有关. 现在以中心化后的天然气输入速率为输入序列, 建立 CO_2 的输出百分浓度模型 (数据见表 A1—24).

• 读入数据, 并绘制输出序列时序图

```
> a<-read.table("E:/R/data/file24.csv",sep=";",header=T)
> y<-ts(a$output)
> plot(y) #结果如图 6—1 所示
```

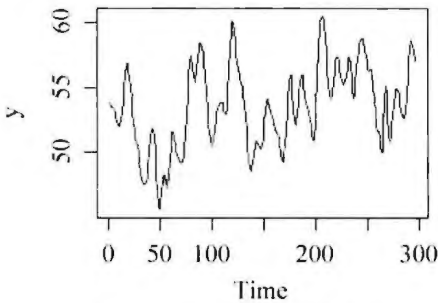


图 6—1 输出序列时序图

• 考察输出序列自相关图和偏自相关图

```
> acf(y) #结果如图 6—2 所示
> pacf(y) #结果如图 6—3 所示
```

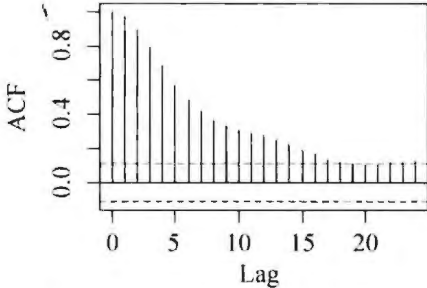


图 6—2 输出序列自相关图

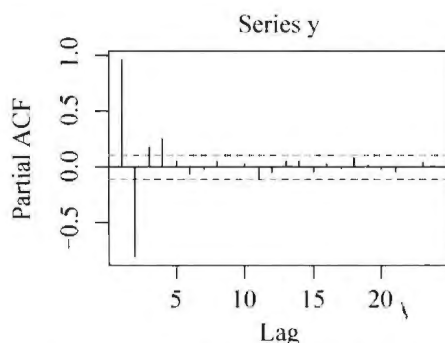


图 6—3 输出序列偏自相关图

• 对输出序列拟合模型 AR(4)

```
> y.fit<-arima(y,order=c(4,0,0))
```

```
> y.fit
```

Series: y

ARIMA(4,0,0) with non-zero mean

Coefficients:

	ar1	ar2	ar3	ar4	intercept
	2.0976	-1.3243	-0.0071	0.2124	53.6880
s.e.	0.0565	0.1344	0.1345	0.0566	0.8679

sigma^2 estimated as 0.1105: log likelihood=-97.28

AIC=206.56 AICc=206.85 BIC=228.71

• 残差白噪声检验

```
> for(i in 1:2) print(Box.test(y.fit$residual,lag=6*i))
```

Box-Pierce test

data: y.fit\$residual

X-squared=8.3004, df=6, p-value=0.2169

Box-Pierce test

data: y.fit\$residual

X-squared=12.8111, df=12, p-value=0.3829

如果不考虑输入序列和输出序列之间的相关性, 将它们作为两个独立的一元时间序列进行分析, 那么输出序列波动平稳 (见图 6—1), 自相关系数拖尾 (见图 6—2), 偏自相关系数 4 阶截尾 (见图 6—3), 所以拟合 AR(4) 模型, 模型口径为:

$$y_t = 53.688 + \frac{a_t}{1 - 2.0976B + 1.3243B^2 + 0.0071B^3 - 0.2124B^4}$$

式中, $a_t \sim N(0, 0.1105)$, 输出序列模型的 $AIC=206.56$, $BIC=228.71$.

考虑到输入天然气速率与输出 CO_2 的浓度之间有逻辑上的因果关系, 将输入天然气速率作为输入变量考虑进输出序列的模型中. 通过协相关图, 考察回归模型的结构. 其中延迟 k 协方差函数定义为:

$$Cov_k = Cov(y_t, x_{t-k}) = E[(y_t - Ey_t)(x_{t-k} - Ex_{t-k})]$$

延迟 k 协相关系数为:

$$C_{\rho k} = \frac{Cov(y_t, x_{t-k})}{\sqrt{Var(y_t)Var(x_{t-k})}}$$

R 语言中调用 ccf 函数可以直接绘制协相关图.

```
> x<-ts(a$input)
> ccf(y,x)    #结果如图 6—4 所示
```

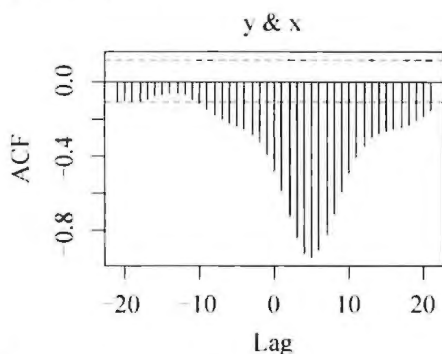


图 6—4 输出序列与输入序列的协相关图

ccf 函数中 x 序列和 y 序列的位置可以随意放置, 协相关图形状不变. $ccf(x,y)$ 表示 x 比 y 滞后的情况, $ccf(y,x)$ 则表示 y 比 x 滞后的情况. 图 6—4 显示输出序列 (y 序列) 比输入序列 (x 序列) 滞后 5 期左右.

考察协相关图 (见图 6—4) 可以看出, 输入序列和输出序列之间具有非常长期的协相关关系. 如果要建立输出序列关于输入序列的线性回归分析模型

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_k x_{t-k}$$

自变量的个数 k 会很大, 而且 $x_{t-1}, x_{t-2}, \cdots, x_{t-k}$ 分别是输入序列延迟不同阶数的延迟序列, 它们彼此之间具有高度相关性, 这导致参数估计会遇到很大的问题. 这时对输入序列采用 ARMA 模型结构 (也称为转移函数结构), 可以避免上面两个问题出现.

利用 R 语言 TSA 包中的 arimax 函数可以拟合 ARIMAX 模型, 所以在使用该函数之前, 需先下载 TSA 包, 并用 library 函数调用 TSA 包. arimax 函数的命令格式为:

```
arimax(y, order=, xreg=, xtransf=, transfer=)
```

式中:

-y: 要进行模型拟合的序列名.

-order: 指定 y 序列拟合模型阶数. $order = c(p,d,q)$, p 为自回归阶数; d 为差分阶数; q 为移动平均阶数.

-xreg: 输入变量名 (不需要做转移函数变换).
 -xtransf: 输入变量名 (需要做转移函数变换).
 -transfer: 指定转移函数的模型阶数.

(1) 假设输入变量为单变量, 且不需要做转移函数变换, 即拟合如下结构的 ARIMAX 模型:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \\ \varepsilon_t = \frac{1 + \theta_1 B + \dots + \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \end{cases}$$

命令如下:

```
> y.fit<-arimax(y, order=c(p,0,q) , xreg=x)
```

(2) 假设输入变量为多变量 (不妨假设为两变量), 且都不需要做转移函数变换, 即拟合如下结构的 ARIMAX 模型:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \\ \varepsilon_t = \frac{1 + \theta_1 B + \dots + \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \end{cases}$$

这时, 需要将多个自变量 $x_1, x_2, \dots, x_k$ 组合为一个数据框读入, 命令如下:

```
> y.fit<-arimax(y, order=c(p,0,q) , xreg=data.frame(x1, x2))
```

(3) 假设输入变量为单变量, 且需要做 ARMA(m, n) 转移函数变换, 即拟合如下结构的 ARIMAX 模型:

$$\begin{cases} y_t = \beta_0 + \frac{1 + \lambda_1 B + \dots + \lambda_n B^n}{1 - \varphi_1 B - \dots - \varphi_m B^m} x_t + \varepsilon_t \\ \varepsilon_t = \frac{1 + \theta_1 B + \dots + \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \end{cases}$$

命令如下:

```
> y.fit<-arimax(y, order=c(p,0,q) , xtransf=x,transfer=list(c(m,n)))
```

(4) 假设输入变量为多变量 (不妨假设为两变量), 且分别需要做 ARMA(m_1, n_1) 和 ARMA(m_2, n_2) 转移函数变换, 即拟合如下结构的 ARIMAX 模型:

$$\begin{cases} y_t = \beta_0 + \frac{1 + \lambda_1 B + \dots + \lambda_{n_1} B^{n_1}}{1 - \varphi_1 B - \dots - \varphi_{m_1} B^{m_1}} x_{1t} + \frac{1 + \lambda_1 B + \dots + \lambda_{n_2} B^{n_2}}{1 - \varphi_1 B - \dots - \varphi_{m_2} B^{m_2}} x_{2t} + \varepsilon_t \\ \varepsilon_t = \frac{1 + \theta_1 B + \dots + \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \end{cases}$$

命令如下:

```
> y.fit<-arimax(y, order=c(p,0,q) , xtransf=data.frame(x1,x2),
transfer=list(c(m1,n1),c(m2,n2)))
```

(5) 假设输入变量既有不需要做转换函数变换的, 又有需要做转移函数变换, 即拟合如下结构的 ARIMAX 模型:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{1t} + \frac{1 + \lambda_1 B + \cdots + \lambda_n B^n}{1 - \varphi_1 B - \cdots - \varphi_m B^m} x_{2t} + \varepsilon_t \\ \varepsilon_t = \frac{1 + \theta_1 B + \cdots + \theta_q B^q}{1 - \phi_1 B - \cdots - \phi_p B^p} a_t \end{cases}$$

命令如下:

```
> y.fit<-arimax(y, order=c(p,0,q) , xreg=x1,xtransf=x2,transfer=list(c(m,n)))
```

本例中, 假设对输入序列采用 AR(2,1) 的转移函数结构, 残差部分依然是 AR(4) 模型, 则相关命令和输出结果如下:

```
> library(TSA)
> y.fit2<-arimax(y,order=c(4,0,0),xtransf=x,transfer=list(c(2,1)))
> y.fit2
Series: y
ARIMA(4,0,0) with non-zero mean

Coefficients:
          ar1          ar2          ar3          ar4  intercept  T1-AR1   T1-AR2   T1-MA0
      1.6469   -0.8310   -0.0639   0.2095    53.5836   1.5509   -0.6667   0.3122
s.e.    0.0585    0.1119    0.1129    0.0595     0.4031   0.0275    0.0266   0.0621

      T1-MA1
     -0.5943
s.e.    0.0677

sigma^2 estimated as 0.07143: log likelihood=-31.25
AIC=80.49 AICc=81.26 BIC=117.4
```

根据系统输出结果可知, 带输入序列的 ARIMAX 模型拟合如下:

$$\begin{cases} y_t = 53.5836 + \frac{0.3122 - 0.5943B}{1 - 1.5509B + 0.6667B^2} x_t + \varepsilon_t \\ \varepsilon_t = \frac{1}{1 - 1.6469B + 0.831B^2 + 0.0639B^3 - 0.2095B^4} a_t \end{cases}$$

式中, $a_t \sim N(0, 0.07143)$.

该输出序列的 ARIMAX 模型 AIC=80.49, BIC=117.4. 显然, ARIMAX 模型比不考虑输入序列的单纯 AR(4) 模型优化得多.

6.2 虚假回归

当响应序列 $\{y_t\}$ 和输入序列 $\{x_{1t}\}, \{x_{2t}\}, \cdots, \{x_{kt}\}$ 都平稳时, 可以构建带输入变量回归的 ARIMAX 模型来拟合响应序列的变化.

$$y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \frac{\Theta(B)}{\Phi(B)} a_t$$

如果平稳性条件不满足,我们就不能大胆地构造 ARIMAX 模型,因为这时容易产生虚假回归的问题.

为了正确理解虚假回归的意义,我们考虑最简单的一元线性动态回归模型:

$$y_t = \beta_0 + \beta_1 x_t + v_t \quad (6.2)$$

为了检验模型的显著性,要对拟合模型进行检验:

$$H_0: \beta_1 = 0 \leftrightarrow H_1: \beta_1 \neq 0$$

假定响应序列 $\{y_t\}$ 和输入序列 $\{x_t\}$ 相互独立,就说明响应序列和输入序列之间没有显著的线性相关关系,理论上,检验结果应该接受 $\beta_1 = 0$ 的原假设.

如果检验结果恰好和理论结果相反,支持 β_1 显著非零的备择假设,那么就会得出响应序列和输入序列之间具有显著线性相关性的错误结论,拒绝正确的原假设并接受一个本不应该成立的回归模型 (6.2),这就犯了第一类错误 (拒真错误).

由于样本的随机性,拒真错误始终都会存在,我们使用显著性水平 α 控制犯拒真错误的概率:

$$\Pr(H_1|H_0) = \alpha$$

通常采用 t 统计量进行参数显著性检验:

$$t = \frac{\beta_1}{\sigma_\beta}$$

当响应序列和输入序列都平稳时,该统计量服从自由度为样本容量的 t 分布. 当 $|t| \leq t_{\alpha/2}(n)$ 时,可以将拒真错误发生的概率准确地控制在显著性水平 α 以内,即

$$\Pr\{|t| \leq t_{\alpha/2}(n) | \text{平稳序列}\} \leq \alpha$$

当响应序列和输入序列不平稳时,随机模拟的结果显示,检验统计量 $t = \frac{\beta_1}{\sigma_\beta}$ 将不再服从 t 分布,这时 t 统计量样本分布的方差远远大于 t 分布的方差,如果仍然采用 t 分布的临界值进行检验,拒绝原假设的概率就会大大增加,即

$$\Pr\{|t| \leq t_{\alpha/2}(n) | \text{非平稳序列}\} \geq \alpha$$

这将导致我们无法控制拒真错误,非常容易接受回归模型显著成立的错误结论,这种现象称为虚假回归(或伪回归).

1974 年,Granger 和 Newbold 进行了非平稳序列伪回归的随机模拟试验. 该模拟试验的设计思想是分别拟合两个随机游走序列:

$$(1) y_t = y_{t-1} + \omega_t$$

$$(2) x_t = x_{t-1} + v_t$$

式中, $\omega_t \stackrel{i.i.d}{\sim} N(0, \sigma_\omega^2)$, $v_t \stackrel{i.i.d}{\sim} N(0, \sigma_v^2)$, 且 $\text{Cov}(\omega_t, v_s) = 0, \forall t, s \in T$.

构建 $\{y_t\}$ 关于 $\{x_t\}$ 的回归模型: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, 并进行参数显著性检验. 由于 $\{y_t\}$ 和 $\{x_t\}$ 是两个独立的随机游走模型,因此理论上它们应该没有任何相关性,即

模型检验应该显著支持 $\beta_1 = 0$ 的假定. 如果模拟结果显示拒绝原假设的概率远远大于拒真概率 α , 即认为伪回归显著成立.

大量随机拟合的结果显示, 每 100 次回归拟合中, 平均有 75 次拒绝 $\beta_1 = 0$ 的假定, 拒真概率高达 75%. 这说明在非平稳的场合, 参数显著性检验犯拒真错误的概率远远大于 α , 伪回归显著成立.

产生伪回归的原因是在非平稳场合, 参数的 t 检验统计量不再服从 t 分布. 在样本容量 $n = 100$ 的情况下, 进行大量的随机拟合, 得到 β_1 的 t 检验统计量的样本分布 $t(\hat{\beta}_1)$ (见图 6—5). 从图中可以看到, β_1 的样本分布 $t(\hat{\beta}_1)$ 尾部肥, 方差大, 比 t 分布要扁平很多. t 分布所确定的显著性水平为 5% 的双侧拒绝域记作 $(L_{0.95}, U_{0.95})$, 由于 β_1 的样本分布并不服从 t 分布, 而是服从厚尾扁平的 $t(\hat{\beta}_1)$ 分布. 在 $t(\hat{\beta}_1)$ 分布场合 $\hat{\beta}_1$ 落在 $(L_{0.95}, U_{0.95})$ 里的概率远远大于 5% (见图 6—5 阴影部分).

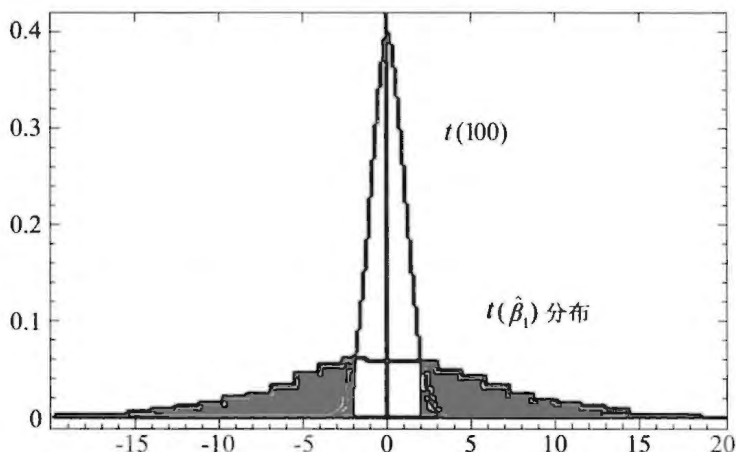


图 6—5 非平稳场合参数检验统计量的样本拟合分布

6.3 单位根检验

由于虚假回归问题的存在, 在进行动态回归模型拟合时, 必须先检验各序列的平稳性. 只有当各序列都平稳时, 才可以大胆地使用 ARIMAX 模型拟合多元序列之间的动态回归关系.

在前面我们已经介绍过序列平稳性的图检验方法, 由于图检验带有很强的主观色彩, 为了客观起见, 人们开始研究各种序列平稳性的统计检验方法, 其中应用最广的是单位根检验.

6.3.1 DF 检验

一、DF 统计量

以 1 阶自回归序列为例:

$$x_t = \phi_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2) \quad (6.3)$$

该序列的特征方程为:

$$\lambda - \phi_1 = 0$$

特征根为:

$$\lambda = \phi_1$$

当特征根在单位圆内时:

$$|\phi_1| < 1$$

该序列平稳.

当特征根在单位圆上或单位圆外时:

$$|\phi_1| \geq 1$$

该序列非平稳.

所以可以通过检验特征根是在单位圆内还是在单位圆上 (外), 来检验序列的平稳性, 这种检验就称为单位根检验.

由于现实生活中绝大多数序列都是非平稳序列, 所以单位根检验的原假设定为:

$$H_0: \text{序列 } x_t \text{ 非平稳} \Leftrightarrow H_0: |\phi_1| \geq 1$$

相应的备择假设定为:

$$H_1: \text{序列 } x_t \text{ 平稳} \Leftrightarrow H_1: |\phi_1| < 1$$

检验统计量为:

$$t(\phi_1) = \frac{\hat{\phi}_1 - \phi_1}{S(\hat{\phi}_1)} \quad (6.4)$$

式中, $\hat{\phi}_1$ 为参数 ϕ_1 的最小二乘估计值; $S(\hat{\phi}_1) = \sqrt{\frac{S_T^2}{\sum_{t=1}^T x_{t-1}^2}}$, $S_T^2 = \frac{\sum_{t=1}^T (x_t - \hat{\phi}_1 x_{t-1})^2}{T-1}$.

当 $\phi_1 = 0$ 时, $t(\phi_1)$ 的极限分布为标准正态分布:

$$t(\phi_1) = \frac{\hat{\phi}_1}{S(\hat{\phi}_1)} \xrightarrow{\text{极限}} N(0, 1)$$

当 $|\phi_1| < 1$ 时, $t(\phi_1)$ 的渐近分布为标准正态分布:

$$t(\phi_1) = \frac{\hat{\phi}_1 - \phi_1}{S(\hat{\phi}_1)} \xrightarrow{\text{渐近}} N(0, 1)$$

但当 $|\phi_1| = 1$ 时, $t(\phi_1)$ 的渐近分布将不再为正态分布. Dickey 和 Fuller 对该检验统计量的样本分布进行了研究. 为了区分于传统的 t 分布检验统计量, 记

$$\tau = \frac{|\hat{\phi}_1| - 1}{S(\hat{\phi}_1)}$$

该统计量称为 DF(Dickey-Fuller) 检验统计量, 它的极限分布为:

$$\frac{\int_0^1 W(r) dW(r)}{\sqrt{\int_0^1 [W(r)]^2 dr}}$$

式中, $W(r)$ 为自由度为 r 的维纳过程 (Weiner process). 维纳过程具有如下性质:

- (1) $W(1) \sim N(0, 1)$
- (2) $\sigma W(r) \sim N(0, \sigma^2 r)$
- (3) $[W(r)]^2 / r \sim \chi^2(1)$

随机模拟的结果显示 τ 检验统计量的极限分布为对称钟形分布, 和正态分布的形状非常相似, 但均值略有偏移.

DF 检验为单边检验, 当显著性水平取为 α 时, 记 τ_α 为 DF 检验的 α 分位点, 则当 $\tau \leq \tau_\alpha$ 时, 拒绝原假设, 认为序列 x_t 显著平稳;

当 $\tau > \tau_\alpha$ 时, 接受原假设, 认为序列 x_t 非平稳.

1979 年, Dickey 和 Fuller 使用蒙特卡洛模拟方法计算出了 DF 统计量的百分位表, 为 DF 检验扫清了最后的技术难题, 使 DF 检验成为最常用的单位根检验.

二、DF 检验的等价表达

在式 (6.3) 等号两边同时减去 x_{t-1} , 得到如下等式:

$$x_t - x_{t-1} = (\phi_1 - 1)x_{t-1} + \varepsilon_t \quad (6.5)$$

记

$$\rho = |\phi_1| - 1$$

式 (6.5) 等价于

$$\nabla x_t = \rho x_{t-1} + \varepsilon_t$$

DF 检验可以通过对参数 ρ 的检验等价进行:

$$H_0: \rho = 0 \leftrightarrow H_1: \rho < 0$$

相应的 DF 检验统计量为:

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$$

式中, $S(\hat{\rho})$ 为参数 ρ 的样本标准差.

三、DF 检验的三种类型

序列是否平稳可以采用 DF 检验进行判断. 但是只把序列分为平稳序列和非平稳

序列两大类就太笼统了, 因为非平稳的构成很复杂, 不同结构的非平稳序列选择的分析方法也不同. 通常把非平稳分成三大类型.

第一类: 无漂移项自回归过程.

$$x_t = x_{t-1} + \varepsilon_t$$

这种类型的序列称为不带漂移项的差分平稳 (difference stationary) 序列, 简记为 DS 序列. 不带漂移项的差分平稳序列也就是我们熟悉的随机游走模型. 该序列均值非平稳, 方差非齐, 但 1 阶差分平稳. 我们在例 5.5 中讨论过该序列的特征, 图 5—8 拟合了该序列时序图.

第二类: 带漂移项自回归过程.

$$x_t = \phi_0 + x_{t-1} + \varepsilon_t$$

这种类型的序列称为带漂移项的差分平稳序列. 整理该序列可以看到这是一个有趋势且波动性不断增强的非平稳序列.

$$\begin{cases} x_0 = 0 \\ x_1 = \phi_0 + x_0 + \varepsilon_1 = \phi_0 + \varepsilon_1 \\ x_2 = \phi_0 + x_1 + \varepsilon_2 = 2\phi_0 + \varepsilon_1 + \varepsilon_2 \\ \vdots \\ x_t = \phi_0 + x_{t-1} + \varepsilon_t = t\phi_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t \end{cases}$$

该序列一阶差分后平稳.

$$\nabla x_t = \phi_0 + \varepsilon_t$$

显然带漂移项的差分平稳序列, 漂移项实际是该序列的线性趋势的斜率. 但对带漂移项的差分平稳序列不建议构造线性回归模型 $\hat{x}_t = \hat{\phi}_0 t$ 提取序列中的相关关系, 因为残差序列是方差非齐的非平稳序列:

$$x_t - \hat{x}_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$

【例 6—2】 拟合 $\phi_0 = 0.1$ 差分平稳序列 $x_t = \phi_0 + x_{t-1} + \varepsilon_t, \varepsilon_t \stackrel{\text{i.i.d}}{\sim} N(0, 1)$, 考察该序列差分后残差序列和提取线性趋势后残差的波动性.

• 拟合带漂移项的差分平稳序列, 并绘制时序图和线性拟合效果图

```
> e<-rnorm(1000)
> x<-filter(0.1+e,filter=1,method="recursive")
> plot(x)
> t<-c(1:1000)
> abline(lm(x~t),col=2)
> lm(x~t)
```

#结果如图 6—6 所示

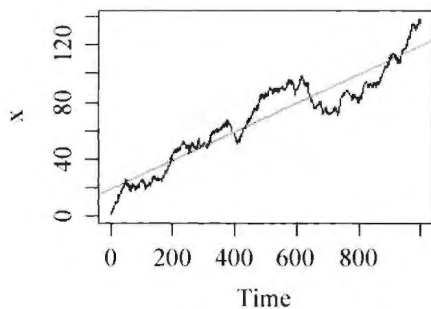


图 6—6 带漂移项的差分平稳序列时序图

说明：曲线为拟合的带漂移项差分平稳序列值；直线为线性趋势拟合值。

• 差分运算和线性拟合后残差序列比较

```
> r1<-diff(x)
> x.fit<-lm(x~t)
> r2<-ts(x.fit$residual)
> c1<-min(r1,r2)
> c2<-max(r1,r2)
> plot(r1,ylim=c(c1,c2))
> lines(r2,col=2)
```

#结果如图 6—7 所示

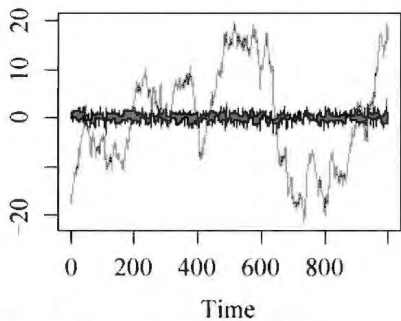


图 6—7 差分运算和线性拟合后残差序列对照图

说明：在零值附近小幅波动的是差分运算后的残差序列，大幅波动的是线性拟合后残差序列。

图 6—7 说明对于带漂移项的差分平稳序列，尽管时序图显示该序列具有显著的线性趋势，但是用差分运算提取趋势信息要比用线性拟合模型提取线性趋势更精确。

第三类：带趋势回归过程。

$$x_t = \phi_0 + \beta t + \varepsilon_t$$

这种类型的序列称为趋势平稳 (trend stationary) 序列，简记为 TS 序列。

对趋势平稳序列最好是通过拟合线性模型 $x_t = \hat{\phi}_0 + \hat{\beta}t$ 提取序列中的相关关系，实现残差序列平稳：

$$\varepsilon_t = x_t - (\hat{\phi}_0 + \hat{\beta}t)$$

对趋势平稳序列采用差分方法提取相关信息, 可以实现趋势平稳, 但增加了残差序列的方差:

$$\nabla x_t = \beta + \varepsilon_t - \varepsilon_{t-1}$$

【例 6—3】拟合 $\phi_0 = 0, \beta = 0.1$ 的趋势平稳序列 $x_t = \phi_0 + \beta t + \varepsilon_t, \varepsilon_t \stackrel{i.i.d}{\sim} N(0, 100)$, 考察该序列差分后残差序列和提取线性趋势后残差序列的波动性.

• 拟合趋势平稳序列, 并绘制时序图

```
> t<-c(1:1000)
> e<-rnorm(1000,0,10)
> x<-0.1*t+e
> x<-ts(x)
> plot(x)    #结果如图 6—8 所示
```

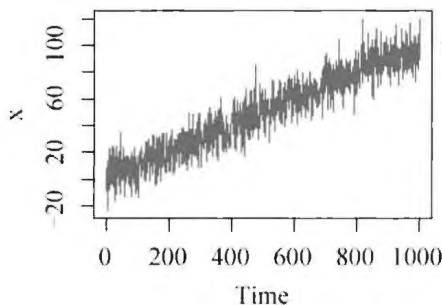


图 6—8 趋势平稳序列时序图

• 考察差分后残差序列的波动范围

```
> plot(diff(x))    #结果如图 6—9 所示
> abline(h=c(1.98*sd(diff(x)),-1.98*sd(diff(x))),col=2)
```

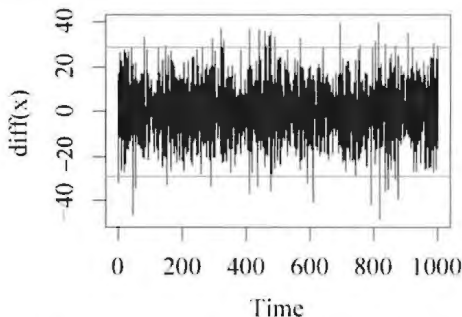


图 6—9 差分后残差序列时序图

说明: 图中两条平行线为该残差序列的 95% 置信区间.

• 考察线性拟合后残差序列的波动范围

```
> x.fit<-lm(x~t)
> r<-ts(x.fit$residual)
> plot(r)    #结果如图 6—10 所示
> abline(h=c(1.98*sd(r),-1.98*sd(r)),col=2)
```

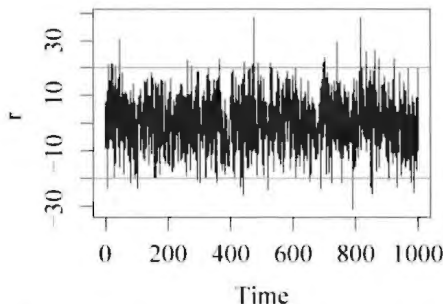


图 6—10 线性拟合后残差序列时序图

说明：图中两条平行线为该残差序列的 95% 置信区间。

显然，对于趋势平稳序列，尽管差分运算和趋势拟合都能够实现序列平稳，但是趋势拟合的残差波动范围更窄，拟合精度更高。

DF 检验不仅可以判断序列是否平稳，还可以针对非平稳的这三种形成机制，进一步判断非平稳序列到底是这三种机制中的哪一种。针对不同的结果，可以选择不同的序列分析方法。

在 R 语言中，fUnitRoots 程序包中的 `adfTest` 函数可以很便捷地进行单位根检验。在使用该函数之前，需先下载 fUnitRoots 包，并用 `library` 函数调用。`adfTest` 函数的命令格式为：

```
adfTest(x, lags=, type=)
```

式中：

-x：进行单位根检验的序列名。

-lags：延迟阶数。

(1) `lags=1`，默认设置，这时进行 DF 检验。

(2) `lags=n`， $n>1$ ，这时进行 ADF 检验。

-type：检验类型。

(1) `type='nc'`，无常数均值，无趋势类型。

(2) `type='c'`，有常数均值，无趋势类型。

(3) `type='ct'`，有常数均值，又有趋势类型。

【例 6—4】对 1978—2002 年中国农村居民家庭人均纯收入对数序列 $\{\ln x_t\}$ 和生活消费支出对数序列 $\{\ln y_t\}$ 进行 DF 检验（数据见表 A1—25）。

● 读入数据，并绘制两序列时序图

```
> b<-read.table("E:/R/data/file25.csv",sep=",",header=T)
> x<-ts(b$lnx,start=1978)
> y<-ts(b$lny,start=1978)
> c1<-min(x,y)
> c2<-max(x,y)
> plot(x,ylim=c(c1,c2))
```

```
> lines(y,lty=2,col=2)
```

#结果如图 6 11 所示

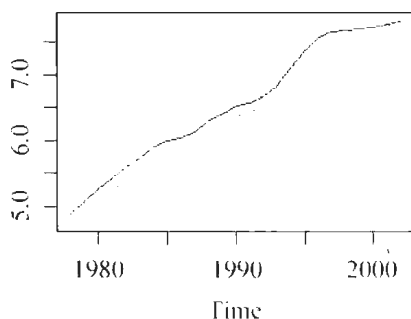


图 6—11 中国农村居民家庭人均纯收入与生活消费支出对数序列时序图

说明：图中上面的实线为人均纯收入对数序列；下面的虚线为人均生活消费支出对数序列。

● 对人均纯收入对数序列进行 DF 检验

```
> library(fBasics)
```

```
> library(fUnitRoots)
```

```
> adfTest(x,lag=1,type="nc" )
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: 1.0859

P VALUE:

0.9202

```
> adfTest(x,lag=1,type="c" )
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -1.1606

P VALUE:

0.6201

```
> adfTest(x,lag=1,type="ct" )
```

```
Title:
Augmented Dickey-Fuller Test
```

```
Test Results:
```

```
PARAMETER:
```

```
Lag Order: 1
```

```
STATISTIC:
```

```
Dickey-Fuller: -3.1381
```

```
P VALUE:
```

```
0.1388
```

- 对人均生活消费支出对数序列进行 DF 检验

```
> adfTest(y,lag=1,type="nc" )
```

```
Title:
Augmented Dickey-Fuller Test
```

```
Test Results:
```

```
PARAMETER:
```

```
Lag Order: 1
```

```
STATISTIC:
```

```
Dickey-Fuller: 1.0806
```

```
P VALUE:
```

```
0.9196
```

```
> adfTest(y,lag=1,type="c" )
```

```
Title:
Augmented Dickey-Fuller Test
```

```
Test Results:
```

```
PARAMETER:
```

```
Lag Order: 1
```

```
STATISTIC:
```

```
Dickey-Fuller: -1.4086
```

```
P VALUE:
```

```
0.5324
```

```
> adfTest(y,lag=1,type="ct" )
```

```
Title:
Augmented Dickey-Fuller Test
```

```
Test Results:
```

```
PARAMETER:
```

```
Lag Order: 1
STATISTIC:
Dickey-Fuller: -3.3125
P VALUE:
0.08993
```

在显著性水平取 0.05 时, 可以认为中国农村居民家庭人均纯收入和人均消费支出对数序列均为非平稳序列, 这与时序图 (见图 6—11) 直观显示出来的性质完全一致。

6.3.2 ADF 检验

DF 检验只适用于 1 阶自回归过程的平稳性检验, 但实际上绝大多数时间序列都不会是一个简单的 AR(1) 过程。为了使 DF 检验能适用于 AR(p) 过程的平稳性检验, 人们对其进行了一定的修正, 得到增广 DF 检验 (augmented Dickey-Fuller), 简记为 ADF 检验。

一、ADF 检验的原理

对任一 AR(p) 过程

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t \quad (6.6)$$

它的特征方程为:

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$$

如果该方程所有的特征根都在单位圆内, 即

$$|\lambda_i| < 1, i = 1, 2, \cdots, p$$

则序列 $\{x_t\}$ 平稳。

如果有一个单位根存在, 不妨设

$$\lambda_1 = 1$$

则序列 $\{x_t\}$ 非平稳, 且自回归系数之和恰好等于 1:

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0 \stackrel{\lambda=1}{\Rightarrow} 1 - \phi_1 - \cdots - \phi_p = 0 \Rightarrow \phi_1 + \phi_2 + \cdots + \phi_p = 1$$

因而, 对于 AR(p) 过程我们可以通过检验自回归系数之和是否等于 1 来考察该序列的平稳性。

为便于检验, 我们对式 (6.6) 进行等价变换:

$$\begin{aligned} x_t - x_{t-1} &= \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} - x_{t-1} + \varepsilon_t \\ &= (\phi_2 + \cdots + \phi_p) x_{t-1} + \phi_1 x_{t-1} - x_{t-1} - (\phi_2 + \cdots + \phi_p) x_{t-1} \\ &\quad + \phi_2 x_{t-2} + (\phi_3 + \cdots + \phi_p) x_{t-2} - (\phi_3 + \cdots + \phi_p) x_{t-2} \\ &\quad + \phi_3 x_{t-3} + (\phi_4 + \cdots + \phi_p) x_{t-3} + \cdots - \phi_p x_{t-p+1} + \phi_p x_{t-p} + \varepsilon_t \end{aligned}$$

整理上式, 得

$$\nabla x_t = (\phi_1 + \cdots + \phi_p - 1)x_{t-1} - (\phi_2 + \cdots + \phi_p)\nabla x_{t-1} - \cdots - \phi_p \nabla x_{t-p+1} + \varepsilon_t$$

简记为:

$$\nabla x_t = \rho x_{t-1} + \beta_1 \nabla x_{t-1} + \cdots + \beta_{p-1} \nabla x_{t-p+1} + \varepsilon_t \quad (6.7)$$

式中

$$\rho = \phi_1 + \phi_2 + \cdots + \phi_p - 1$$

$$\beta_j = -\phi_{j+1} - \phi_{j+2} - \cdots - \phi_p, j = 1, 2, \cdots, p-1$$

若序列 $\{x_t\}$ 平稳, 则

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1$$

等价于 $\rho < 0$.

若序列 $\{x_t\}$ 非平稳, 则至少存在一个单位根, 有

$$\phi_1 + \phi_2 + \cdots + \phi_p = 1$$

等价于 $\rho = 0$. 则 $AR(p)$ 过程单位根检验的假设条件可以确定为:

$$H_0: \rho = 0 (\text{序列 } \{x_t\} \text{ 非平稳}) \leftrightarrow H_1: \rho < 0 (\text{序列 } \{x_t\} \text{ 平稳})$$

构造 ADF 检验统计量:

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$$

式中, $S(\hat{\rho})$ 为参数 ρ 的样本标准差.

通过蒙特卡洛方法, 可以得到 τ 检验统计量的临界值表. 显然 DF 检验是 ADF 检验在自相关阶数为 1 时的一个特例, 所以统称为 ADF 检验.

二、ADF 检验的三种类型

和 DF 检验一样, ADF 检验也可以用于如下三种类型的单位根检验.

第一种类型: 无常数均值、无趋势的 p 阶自回归过程.

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

第二种类型: 有常数均值、无趋势的 p 阶自回归过程.

$$x_t = \mu + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

第三种类型: 既有常数均值又有线性趋势的 p 阶自回归过程.

$$x_t = \mu + \beta t + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

【例 6—4 续 (1)】对 1978—2002 年中国农村居民家庭人均纯收入对数差分后序列 $\{\nabla \ln x_t\}$ 和生活消费支出对数差分后序列 $\{\nabla \ln y_t\}$ 进行 ADF 检验.

- 差分运算

```
> dx<-diff(x)
```

```
> dy<-diff(y)
```

- 人均收入对数差分后序列 ADF 检验

```
> for(i in 1:3) print(adfTest(dx,lag=i,type="nc"))
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -1.3701

P VALUE:

0.173

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -1.3569

P VALUE:

0.1772

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 3

STATISTIC:

Dickey-Fuller: -1.3367

P VALUE:

0.1836

```
> for(i in 1:3) print(adfTest(dx,lag=i,type="c"))
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -2.2267

P VALUE:

0.2428

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -3.0963

P VALUE:

0.0427

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 3

STATISTIC:

Dickey-Fuller: -1.8967

P VALUE:

0.3596

• 人均支出对数差分后序列 ADF 检验

> for(i in 1:3) print(adfTest(dy,lag=i,type="nc"))

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -1.7967

P VALUE:

0.0719

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -1.3048

P VALUE:

0.1937

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 3

STATISTIC:

Dickey-Fuller: -1.0581

P VALUE:

0.272

```
> for(i in 1:3) print(adfTest(dy,lag=i,type="c"))
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -3.3473

P VALUE:

0.02438

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -2.809

P VALUE:

```
0.0758

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
    Lag Order: 3
STATISTIC:
    Dickey-Fuller: -1.7169
P VALUE:
0.4232
```

检验结果显示, 人均收入对数差分序列是有常数均值的平稳序列, 该平稳序列 2 阶自相关 (ADF 检验 P 值 = 0.042 7). 人均支出对数差分序列也是有常数均值的平稳序列, 该平稳序列 1 阶自相关 (ADF 检验 P 值 = 0.024 38).

6.4 协 整

6.4.1 单整与协整

一、单整的概念

在单位根检验的过程中, 如果检验结果显著拒绝原假设, 即说明序列 $\{x_t\}$ 显著平稳, 不存在单位根, 这时称序列 $\{x_t\}$ 为零阶单整 (integration) 序列, 简记为 $x_t \sim I(0)$.

假如原假设不能被显著拒绝, 说明序列 $\{x_t\}$ 为非平稳序列, 存在单位根. 这时可以考虑对该序列进行适当阶数的差分, 以消除单位根实现平稳.

假如原序列 1 阶差分后平稳, 说明原序列存在一个单位根, 这时称原序列为 1 阶单整序列, 简记为 $x_t \sim I(1)$.

假如原序列至少需要进行 d 阶差分才能实现平稳, 说明原序列存在 d 个单位根, 这时称原序列为 d 阶单整序列, 简记为 $x_t \sim I(d)$.

二、单整序列的性质

单整衡量的是单个序列的平稳性, 它具有如下重要性质:

(1) 若 $x_t \sim I(0)$, 对于任意非零实数 a, b , 有

$$a + bx_t \sim I(0)$$

(2) 若 $x_t \sim I(d)$, 对于任意非零实数 a, b , 有

$$a + bx_t \sim I(d)$$

(3) 若 $x_t \sim I(0), y_t \sim I(0)$, 对于任意非零实数 a, b , 有

$$z_t = ax_t + by_t \sim I(0)$$

(4) 若 $x_t \sim I(d), y_t \sim I(c)$, 对于任意非零实数 a, b , 有

$$z_t = ax_t + by_t \sim I(k)$$

式中, $k \leq \max(d, c)$.

三、协整的概念

在现实生活中我们会发现, 有些序列自身的变化虽然是非平稳的, 但是序列与序列之间却具有非常密切的长期均衡关系.

比如, 例 6—4 所分析的农村家庭人均纯收入对数序列 $\{\ln x_t\}$ 和人均生活消费支出对数序列 $\{\ln y_t\}$, 单整分析显示这两个序列都是非平稳的. 但是将这两个序列联合起来考虑, 通过观察它们的时序图 (见图 6—11), 我们却发现它们之间具有非常稳定的线性相关关系. 当收入增多时, 生活消费支出也增多, 它们的变化速度几乎一致. 这种稳定的同变关系让我们怀疑它们之间具有一种内在的平稳机制, 导致它们自身的变化虽然是不平稳的, 但是彼此之间却具有长期均衡关系.

为了有效地衡量序列之间是否具有长期均衡关系, Engle 和 Granger 于 1987 年提出了协整 (cointegration) 的概念.

假定自变量序列为 $\{x_1\}, \dots, \{x_k\}$, 响应变量序列为 $\{y_t\}$, 构造回归模型

$$y_t = \beta_0 + \sum_{i=1}^k \beta_i x_{it} + \varepsilon_t$$

假定回归残差序列 $\{\varepsilon_t\}$ 平稳, 称响应序列 $\{y_t\}$ 与自变量序列 $\{x_1\}, \dots, \{x_k\}$ 之间具有协整关系.

协整概念的提出有非常重要的意义, 我们之前之所以一直不敢大胆地对非平稳序列构建动态回归模型, 是因为担心非平稳序列容易产生虚假回归的问题. 而虚假回归之所以会产生, 是因为残差序列不平稳. 如果非平稳序列之间具有协整关系, 就说明残差序列平稳, 那就不会产生虚假回归问题了.

这说明要将输入变量引入响应序列建模, 不一定要所有的序列都平稳, 只需要它们的回归残差序列平稳. 这个限制条件显然要比 Box 和 Jenkins 要求所有序列都平稳的限制条件宽松许多, 这极大地拓宽了动态回归模型的适用范围.

6.4.2 协整检验

多元非平稳序列之间能否建立动态回归模型, 关键在于它们之间是否具有协整关系. 所以要对多元非平稳序列建模必须先进行协整检验, 也称为 Engle-Granger 检验, 简称 EG 检验.

一、假设条件

由于自然界中绝大多数序列之间不具有协整关系, 所以 EG 检验的假设条件确定为:

H_0 : 多元非平稳序列之间不存在协整关系

H_1 : 多元非平稳序列之间存在协整关系

由于协整关系主要是通过考察回归残差的平稳性确定的, 所以上述假设条件等价于:

H_0 : 回归残差序列 $\{\varepsilon_t\}$ 非平稳

H_1 : 回归残差序列 $\{\varepsilon_t\}$ 平稳

二、检验步骤

EG 检验也称为 EG 两步法, 它按照如下两个步骤进行.

步骤一: 建立响应序列与输入序列之间的回归模型:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \cdots + \hat{\beta}_k x_{kt} + \varepsilon_t$$

式中, $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ 是最小二乘估计值.

步骤二: 对回归残差序列 $\{\varepsilon_t\}$ 进行平稳性检验.

我们主要采用单位根检验的方法来考察回归残差序列的平稳性, 所以, 假设条件等价于

$$H_0: \varepsilon_t \sim I(k), k \geq 1 \leftrightarrow H_1: \varepsilon_t \sim I(0)$$

EG 检验的原理与计算公式和 DF 检验的原理与计算公式相同, 但是蒙特卡洛模拟的结果显示它们的临界值略有不同. EG 检验的临界值不仅与位移项、趋势项等因素有关, 而且与回归模型中非平稳变量的个数相关. MacKinnon 提供了 EG 检验的临界值表, 并将 EG 检验的临界值表与 ADF 检验的临界值表结合在一起. 当非平稳序列的个数为 1 时 ($N = 1$), 对应的就是 ADF 检验; 当非平稳序列的个数大于等于 2 时 ($N \geq 2$), 对应的就是 EG 检验.

【例 6—4 续 (2)】 对 1978—2002 年中国农村居民家庭人均纯收入对数序列 $\{\ln x_t\}$ 和生活消费支出对数序列 $\{\ln y_t\}$ 进行 EG 检验.

• 第一步: 构造回归模型

```
> y.fit<-lm(y~x)
```

```
> summary(y.fit)
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.083331	-0.043257	-0.003606	0.044968	0.083295

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.07361	0.07337	1.003	0.326
x	0.95729	0.01110	86.254	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05095 on 23 degrees of freedom

Multiple R-squared: 0.9969, Adjusted R-squared: 0.9968

F-statistic: 7440 on 1 and 23 DF, p-value: < 2.2e-16

• 第二步: 残差序列单位根检验

```
> r<-ts(y.fit$residual)
> for(i in 1:3) print(adfTest(r,lag=i,type="nc"))
```

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 1

STATISTIC:

Dickey-Fuller: -1.8644

P VALUE:

0.06224

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -1.9873

P VALUE:

0.04699

Title:

Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 3

STATISTIC:

Dickey-Fuller: -1.581

P VALUE:

0.106

输出结果显示, 尽管中国农村居民家庭人均收入对数序列和生活消费支出对数序列都是非平稳序列, 但它们之间具有协整关系 (回归残差序列属于无常数均值 2 阶自相关平稳序列, ADF 检验 P 值 = 0.046 99). 由于收入和支出之间具有协整关系, 所以可以建立如下回归模型拟合它们之间的长期均衡关系:

$$\ln y_t = 0.073\ 61 + 0.957\ 29 \ln x_t$$

6.5 误差修正模型

误差修正模型 (error correction model) 简称 ECM 模型, 最初由 Hendry 和 Anderson 于 1977 年提出, 它常常作为协整模型的补充模型出现. 由协整模型度量序列之间的长期均衡关系, 而 ECM 模型则解释序列的短期波动关系.

误差修正模型的构造原理如下:

假设非平稳响应序列 $\{y_t\}$ 与非平稳输入序列 $\{x_t\}$ 之间具有协整关系, 即

$$y_t = \beta x_t + \varepsilon_t \quad (6.8)$$

则回归残差序列为平稳序列:

$$\varepsilon_t = y_t - \beta x_t \sim I(0)$$

在式 (6.8) 等号两边同时减去 y_{t-1} , 则有

$$y_t - y_{t-1} = \beta x_t - y_{t-1} + \varepsilon_t \quad (6.9)$$

将 $y_{t-1} = \beta x_{t-1} + \varepsilon_{t-1}$ 代入式 (6.9) 等号右边, 得

$$y_t - y_{t-1} = \beta x_t - \beta x_{t-1} - \varepsilon_{t-1} + \varepsilon_t \quad (6.10)$$

假定 β 的最小二乘估计值为 $\hat{\beta}$, 则 $\hat{\varepsilon}_{t-1} = y_{t-1} - \hat{\beta}x_{t-1}$ 代表的是上一期的误差, 特别记作 ECM_{t-1} , 则式 (6.10) 可以整理成如下形式:

$$\nabla y_t = \beta \nabla x_t - ECM_{t-1} + \varepsilon_t \quad (6.11)$$

这说明响应序列的当期波动 (∇y_t) 主要受到三方面短期波动的影响:

- (1) 输入序列的当期波动 ∇x_t ;
- (2) 上一期的误差 ECM_{t-1} ;
- (3) 当期纯随机波动 ε_t .

为了定量地测定这三方面影响的大小, 尤其是为了测定上期误差 ECM_{t-1} 对当期波动 ∇y_t 的影响, 可以构建 ECM 模型, 模型结构如下:

$$\nabla y_t = \beta_0 \nabla x_t + \beta_1 ECM_{t-1} + \varepsilon_t$$

式中, β_1 称为误差修正系数, 表示误差修正项对当期波动的修正力度. 根据误差修正模型的推导原理 (式 (6.11)), 可以确定 $\beta_1 < 0$, 即误差修正机制是一个负反馈机制.

以例 6—4 的应用背景对误差修正模型的负反馈机制进行直观解释, 当 $ECM_{t-1} > 0$ 时, 等价于 $y_{t-1} > \hat{\beta}x_{t-1}$, 即上期真实支出比估计支出大, 这种信息反馈回来, 会导致下期支出适当压缩, 即 $\nabla y_t < 0$.

反之, $ECM_{t-1} < 0$, 等价于 $y_{t-1} < \hat{\beta}x_{t-1}$, 即上期真实支出比估计支出小, 这种信息反馈回来, 会导致下期支出适当增加, 即 $\nabla y_t > 0$.

【例 6—4 续 (3)】 对 1978—2002 年中国农村居民家庭人均纯收入对数序列 $\{\ln x_t\}$ 和生活消费支出对数序列 $\{\ln y_t\}$ 构造 ECM 模型.

```
> ECM<-y.fit$residual[1:24]
> dify.fit<-lm(diff(y)~0+diff(x)+ECM)
> summary(dify.fit)

Call:
lm(formula=diff(y)~0+diff(x)+ECM)

Residuals:
      Min       1Q   Median       3Q      Max
-0.049691 -0.025353  0.003606  0.024944  0.049214

Coefficients:
      Estimate Std. Error t value Pr(> |t|)
diff(x)  0.95513    0.04472   21.36 3.37e-16***
ECM     -0.17152    0.12796   -1.34  0.194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0301 on 22 degrees of freedom
Multiple R-squared:  0.9551, Adjusted R-squared:  0.951
F-statistic: 233.9 on 2 and 22 DF, p-value: 1.501e-15
```

根据系统输出, 我们得到误差修正模型:

$$\nabla y_t = 0.955\ 13\nabla x_t - 0.171\ 52ECM_{t-1}$$

从回归系数的绝对值大小可以看出, 收入的当期波动对生活消费支出当期波动的影响很大, 每增加 1 单位的对数收入, 会增加 0.955 13 单位的对数生活消费支出. 上期误差 (ECM)对生活消费支出当期波动的调整幅度不大, 单位调整比例为 -0.171 52, 而且系数显著性检验显示该系数并不显著非零.

6.6 习 题

- 1. 单位根检验的原理是什么？使用单位根检验方法重新考察例 2—1 至例 2—5 中序列的平稳性，并将检验结果与图检验方法得出的结果进行比较研究。
- 2. 某地区过去 38 年谷物产量序列如表 6—1 所示。

表 6—1

24.5	33.7	27.9	27.5	21.7	31.9	36.8	29.9	30.2	32.0	34.0
19.4	36.0	30.2	32.4	36.4	36.9	31.5	30.5	32.3	34.9	30.1
36.9	26.8	30.5	33.3	29.7	35.0	29.9	35.2	38.3	35.2	35.5
36.7	26.8	38.0	31.7	32.6						

这些年该地区相应的降雨量序列如表 6—2 所示。

表 6—2

9.6	12.9	9.9	8.7	6.8	12.5	13.0	10.1	10.1	10.1	10.8
7.8	16.2	14.1	10.6	10.0	11.5	13.6	12.1	12.0	9.3	7.7
11.0	6.9	9.5	16.5	9.3	9.4	8.7	9.5	11.6	12.1	8.0
10.7	13.9	11.3	11.6	10.4						

- (1) 使用单位根检验，分别考察这两个模型的平稳性。
 - (2) 选择适当模型，分别拟合这两个序列的发展。
 - (3) 确定这两个序列之间是否具有协整关系。
 - (4) 如果这两个序列之间具有协整关系，请建立适当的模型拟合谷物产量序列的发展。
3. 在一定浓度的溶液中 (CC=0.5)，考察草履虫 (*aramecium aurelia*) 和某种草履虫掠食动物 (*didinium nasutum*) 之间的动态数量变化，相关数据如表 6—3 所示。

表 6—3

时间 day	被掠食者 ind/ml	掠食者 ind/ml	时间 day	被掠食者 ind/ml	掠食者 ind/ml	时间 day	被掠食者 ind/ml	掠食者 ind/ml
0.00	15.65	5.76	4.00	56.70	18.77	8.00	65.59	26.00
0.50	53.57	9.05	4.50	86.37	31.11	8.50	81.23	24.32
1.00	73.34	17.26	5.00	121.00	58.31	9.00	143.90	21.00
1.50	93.93	41.97	5.50	71.48	73.13	9.50	237.90	33.35
2.00	115.40	55.97	6.00	55.78	63.21	10.00	276.60	64.67
2.50	76.57	74.91	6.50	31.84	52.46	10.50	222.20	94.34
3.00	32.83	62.52	7.00	26.87	40.07	11.00	137.20	103.40
3.50	23.74	27.04	7.50	53.24	27.67	11.50	46.45	82.74

续前表

时间 day	被掠食者 ind/ml	掠食者 ind/ml	时间 day	被掠食者 ind/ml	掠食者 ind/ml	时间 day	被掠食者 ind/ml	掠食者 ind/ml
12.00	27.46	65.40	20.00	206.90	41.10	28.00	87.73	32.48
12.50	41.46	51.35	20.50	309.90	52.62	28.50	136.40	24.21
13.00	44.73	28.24	21.00	156.50	120.20	29.00	290.60	35.73
13.50	88.42	23.27	21.50	63.30	112.80	29.50	345.80	55.50
14.00	105.70	38.09	22.00	77.29	92.14	30.00	271.60	93.41
14.50	155.20	14.97	22.50	45.11	65.72	30.50	156.10	117.30
15.00	205.50	24.84	23.00	57.45	33.54	31.00	71.10	95.02
15.50	312.70	49.56	23.50	69.80	21.14	31.50	43.86	85.92
16.00	213.70	75.93	24.00	121.70	17.82	32.00	30.64	82.60
16.50	163.40	104.00	24.50	185.20	26.04	32.50	35.56	66.08
17.00	85.78	106.40	25.00	175.30	65.61	33.00	52.03	63.58
17.50	48.64	100.60	25.50	139.00	76.30	33.50	37.99	37.99
18.00	44.49	84.08	26.00	77.11	96.07	34.00	62.71	25.60
18.50	63.44	45.30	26.50	57.29	68.84	34.50	103.90	23.10
19.00	71.66	35.37	27.00	54.79	54.79	35.00	187.20	37.09
19.50	127.70	35.35	27.50	75.38	35.80			

(1) 考察这两种生物之间的动态关系, 检验它们之间是否具有协整关系.

(2) 选择适当的模型拟合这两种生物之间的动态互动关系, 并预测未来一周这两种生物的浓度.

4. 我国 1950—2008 年进出口总额数据 (单位: 亿元) 如表 6—4 所示.

表 6—4

年份	出口	进口	年份	出口	进口	年份	出口	进口
1950	20.0	21.3	1963	50.0	35.7	1976	134.8	129.3
1951	24.2	35.3	1964	55.4	42.1	1977	139.7	132.8
1952	27.1	37.5	1965	63.1	55.3	1978	167.6	187.4
1953	34.8	46.1	1966	66.0	61.1	1979	211.7	242.9
1954	40.0	44.7	1967	58.8	53.4	1980	271.2	298.8
1955	48.7	61.1	1968	57.6	50.9	1981	367.6	367.7
1956	55.7	53.0	1969	59.8	47.2	1982	413.8	357.5
1957	54.5	50.0	1970	56.8	56.1	1983	438.3	421.8
1958	67.0	61.7	1971	68.5	52.4	1984	580.5	620.5
1959	78.1	71.2	1972	82.9	64.0	1985	808.9	1 257.8
1960	63.3	65.1	1973	116.9	103.6	1986	1 082.1	1 498.3
1961	47.7	43.0	1974	139.4	152.8	1987	1 470.0	1 614.2
1962	47.1	33.8	1975	143.0	147.4	1988	1 766.7	2 055.1

续前表

年份	出口	进口	年份	出口	进口	年份	出口	进口
1989	1 956.0	2 199.9	1996	12 576.4	11 557.4	2003	36 287.9	34 195.6
1990	2 985.8	2 574.3	1997	15 160.7	11 806.5	2004	49 103.3	46 435.8
1991	3 827.1	3 398.7	1998	15 223.6	11 626.1	2005	62 648.1	54 273.7
1992	4 676.3	4 443.3	1999	16 159.8	13 736.5	2006	77 594.6	63 376.9
1993	5 284.8	5 986.2	2000	20 634.4	18 638.8	2007	93 455.6	73 284.6
1994	10 421.8	9 960.1	2001	22 024.4	20 159.2	2008	100 394.9	79 526.5
1995	12 451.8	11 048.1	2002	26 947.9	24 430.3			

- (1) 使用单位根检验, 分别考察进口总额和出口总额序列的平稳性.
- (2) 分别对进口总额序列和出口总额序列拟合模型.
- (3) 考察这两个序列是否具有协整关系.
- (4) 如果这两个序列具有协整关系, 请建立适当模型拟合它们之间的相关关系.
- (5) 构造该协整模型的误差修正模型.

附 录

表 A1—1 至表 A1—25 是一些统计数据列表.

表 A1—1 1884—1939 年英格兰和威尔士地区小麦的平均亩产量

年份	产量	年份	产量	年份	产量	年份	产量
1884	15.2	1898	16.9	1912	14.2	1926	16
1885	16.9	1899	16.4	1913	15.8	1927	16.4
1886	15.3	1900	14.9	1914	15.7	1928	17.2
1887	14.9	1901	14.5	1915	14.1	1929	17.8
1888	15.7	1902	16.6	1916	14.8	1930	14.4
1889	15.1	1903	15.1	1917	14.4	1931	15
1890	16.7	1904	14.6	1918	15.6	1932	16
1891	16.3	1905	16	1919	13.9	1933	16.8
1892	16.5	1906	16.8	1920	14.7	1934	16.9
1893	13.3	1907	16.8	1921	14.3	1935	16.6
1894	16.5	1908	15.5	1922	14	1936	16.2
1895	15	1909	17.3	1923	14.5	1937	14
1896	15.9	1910	15.5	1924	15.4	1938	18.1
1897	15.5	1911	15.5	1925	15.3	1939	17.5

资料来源: Time Series Data Library (citing: Kendall & Ord (1990)).

表 A1—2 1500—1869 年 Beveridge 小麦价格指数序列 (行数据)

17	19	20	15	13	14	14	14	14	11
16	19	23	18	17	20	20	18	14	16
21	24	15	16	20	14	16	25.5	25.8	26
29	20	18	16	22	22	16	19	17	17
19	20	24	28	36	20	14	18	27	29

36	29	27	30	38	50	24	25	30	31
37	41	36	32	47	42	37	34	36	43
55	64	79	59	47	48	49	45	53	55
55	54	56	52	76	113	68	59	74	78
69	78	73	88	98	109	106	87	77	77
63	70	70	63	61	66	78	93	97	77
83	81	82	78	75	80	87	72	65	74
91	115	99	99	115	101	90	95	108	147
112	108	99	96	102	105	114	103	98	103
101	110	109	98	84	90	120	124	136	120
135	100	70	60	72	70	71	94	95	110
154	116	99	82	76	64	63	68	64	67
71	72	89	114	102	85	88	97	94	88
79	74	79	95	70	72	63	60	74	75
91	126	161	109	108	110	130	166	143	103
89	76	93	82	71	69	75	134	183	113
108	121	139	109	90	88	88	93	106	89
79	91	96	111	112	104	94	98	88	94
81	77	84	92	96	102	95	98	125	162
113	94	85	89	109	110	109	120	116	101
113	109	105	94	102	141	135	118	115	111
127	124	113	122	130	137	148	142	143	176
184	164	146	147	124	119	135	125	116	132
133	144	145	146	138	139	154	181	185	151
139	157	155	191	248	185	168	176	243	289
251	232	207	276	250	216	205	206	208	226
302	261	207	209	280	381	266	197	177	170
152	156	141	142	137	161	189	226	194	217
199	151	144	138	145	156	184	216	204	186
197	183	175	183	230	278	179	161	150	159
180	223	294	300	297	232	179	180	215	258
236	202	174	179	210	268	267	208	224	

资料来源: Time Series Data Library (citing: Newton (1988)).

表 A1—3 1770—1869 年太阳黑子年度数据

年份	黑子数	年份	黑子数	年份	黑子数	年份	黑子数
1820	16	1823	2	1826	36	1829	67
1821	7	1824	8	1827	50	1830	71
1822	4	1825	17	1828	62	1831	48

表 A1—5 1962 年 1 月至 1975 年 12 月平均每头奶牛月产奶量序列 单位: 磅

589	561	640	656	727	697	640	599	568	577
553	582	600	566	653	673	742	716	660	617
583	587	565	598	628	618	688	705	770	736
678	639	604	611	594	634	658	622	709	722
782	756	702	653	615	621	602	635	677	635
736	755	811	798	735	697	661	667	645	688
713	667	762	784	837	817	767	722	681	687
660	698	717	696	775	796	858	826	783	740
701	706	677	711	734	690	785	805	871	845
801	764	725	723	690	734	750	707	807	824
886	859	819	783	740	747	711	751	804	756
860	878	942	913	869	834	790	800	763	800
826	799	890	900	961	935	894	855	809	810
766	805	821	773	883	898	957	924	881	837
784	791	760	802	828	778	889	902	969	947
908	867	815	812	773	813	834	782	892	903
966	937	896	858	817	827	797	843		

资料来源: Cryer (1986).

表 A1—6 1949—1998 年北京市每年最高气温序列 单位: 摄氏度

年份	温度	年份	温度
1949	38.8	1967	35.8
1950	35.6	1968	40.1
1951	38.3	1969	35.9
1952	39.6	1970	35.3
1953	37.0	1971	35.2
1954	33.4	1972	39.5
1955	39.6	1973	37.5
1956	34.6	1974	35.8
1957	36.2	1975	38.4
1958	37.6	1976	35.0
1959	36.8	1977	34.1
1960	38.1	1978	37.5
1961	40.6	1979	35.9
1962	37.1	1980	35.1
1963	39.0	1981	38.1
1964	37.5	1982	37.3
1965	38.5	1983	37.2
1966	37.5	1984	36.1

续前表

年份	温度	年份	温度
1985	35.1	1992	37.5
1986	38.5	1993	35.8
1987	36.1	1994	37.2
1988	38.1	1995	35.0
1989	35.8	1996	36.0
1990	37.5	1997	38.2
1991	35.7	1998	37.2

资料来源: 北京市统计局计算中心: 《北京五十年》.

表 A1—7 1950—1998 年北京市城乡居民定期储蓄所占比例序列 (%)

年份	定期储蓄	年份	定期储蓄
1950	83.5	1975	83.8
1951	63.1	1976	84.5
1952	71.0	1977	84.8
1953	76.3	1978	83.9
1954	70.5	1979	83.9
1955	80.5	1980	81.0
1956	73.6	1981	82.2
1957	75.2	1982	82.7
1958	69.1	1983	82.3
1959	71.4	1984	80.9
1960	73.6	1985	80.3
1961	78.8	1986	81.3
1962	84.4	1987	81.6
1963	84.1	1988	83.4
1964	83.3	1989	88.2
1965	83.1	1990	89.6
1966	81.6	1991	90.1
1967	81.4	1992	88.2
1968	84.0	1993	87.0
1969	82.9	1994	87.0
1970	83.5	1995	88.3
1971	83.2	1996	87.8
1972	82.2	1997	84.7
1973	83.2	1998	80.2
1974	83.5		

资料来源: 北京市统计局计算中心: 《北京五十年》.

表 A1—8 1950—2008 年我国邮路及农村投递线路每年新增里程数序列 单位: 万公里

年份	新增里程	年份	新增里程	年份	新增里程
1950	15.71	1970	26.39	1990	6.76
1951	24.43	1971	31.09	1991	-0.83
1952	18.23	1972	19.78	1992	4.67
1953	22.50	1973	2.56	1993	11.68
1954	12.53	1974	12.95	1994	0.82
1955	9.94	1975	15.54	1995	8.54
1956	7.19	1976	3.97	1996	24.51
1957	41.13	1977	2.42	1997	28.91
1958	79.03	1978	0.31	1998	44.94
1959	119.32	1979	-5.10	1999	11.16
1960	-12.10	1980	-7.52	2000	11.08
1961	-89.71	1981	-7.69	2001	15.75
1962	-52.26	1982	1.61	2002	-0.31
1963	20.01	1983	4.46	2003	20.99
1964	19.92	1984	10.97	2004	6.50
1965	42.81	1985	15.15	2005	10.45
1966	18.78	1986	6.00	2006	-3.51
1967	-0.75	1987	-0.90	2007	23.42
1968	-1.08	1988	-3.22	2008	17.99
1969	5.09	1989	-8.54		

资料来源: 国家统计局国民经济综合统计司编:《新中国 60 年统计资料汇编》, 北京, 中国统计出版社, 2010, 1—49 全国邮电业务量, 整理后获得.

表 A1—9 某加油站连续 57 天的 OVERSHORT

78	-58	53	-63	13	-6	-16	-14
3	-74	89	-48	-14	32	56	-86
-66	50	26	59	-47	-83	2	-1
124	-106	113	-76	-47	-32	39	-30
6	-73	18	2	-24	23	-38	91
-56	-58	1	14	-4	77	-127	97
10	-28	-17	23	-2	48	-131	65
-17							

资料来源: Brockwell and Davis.

表 A1—10 1880—1985 年全球气表平均温度改变值序列 单位: 摄氏度

-0.40	-0.37	-0.43	-0.47	-0.72	-0.54	-0.47	-0.54	-0.39	-0.19
-0.40	-0.44	-0.44	-0.49	-0.38	-0.41	-0.27	-0.18	-0.38	-0.22
-0.03	-0.09	-0.28	-0.36	-0.49	-0.25	-0.17	-0.45	-0.32	-0.33
-0.32	-0.29	-0.32	-0.25	-0.05	-0.01	-0.26	-0.48	-0.37	-0.20
-0.15	-0.08	-0.14	-0.13	-0.12	-0.10	0.13	-0.01	0.06	-0.17

-0.01	0.09	0.05	-0.16	0.05	-0.02	0.04	0.17	0.19	0.05
0.15	0.13	0.09	0.04	0.11	-0.03	0.03	0.15	0.04	-0.02
-0.13	0.02	0.07	0.20	-0.03	-0.07	-0.19	0.09	0.11	0.06
0.01	0.08	0.02	0.02	-0.27	-0.18	-0.09	-0.02	-0.13	0.02
0.03	-0.12	-0.08	0.17	-0.09	-0.04	-0.24	-0.16	-0.09	0.12
0.27	0.42	0.02	0.30	0.09	0.05				

说明: 平均温度为零点.
资料来源: James Hansen and Sergej Lebedeff (1987).

表 A1—11 连续读取 70 个化学反应数据

47	64	23	71	38	64	55	41	59	48	71	35	57	40
58	44	80	55	37	74	51	57	50	60	45	57	50	45
25	59	50	71	56	74	50	58	45	54	36	54	48	55
45	57	50	62	44	64	43	52	38	59	55	41	53	49
34	35	54	45	68	38	50	60	39	59	40	57	54	23

资料来源: O'Donovan, Consec. Readings Batch Chemical Process, R. B. Miller et al.

表 A1—12 1949—2008 年我国化肥产量序列 单位: 万吨

年份	化肥产量	年份	化肥产量	年份	化肥产量
1949	0.6	1969	174.9	1989	1 802.5
1950	1.5	1970	243.5	1990	1 879.7
1951	2.8	1971	299.4	1991	1 979.5
1952	3.9	1972	370.1	1992	2 047.9
1953	5	1973	459.2	1993	1 956.3
1954	6.7	1974	422.2	1994	2 272.8
1955	7.9	1975	524.7	1995	2 548.1
1956	11.1	1976	524.4	1996	2 809
1957	15.1	1977	723.8	1997	2 821
1958	19.4	1978	869.3	1998	3 010
1959	26.6	1979	1 065.4	1999	3 251
1960	40.5	1980	1 232.1	2000	3 186
1961	29.7	1981	1 239	2001	3 383
1962	46.4	1982	1 278.1	2002	3 791
1963	64.8	1983	1 378.9	2003	3 881.31
1964	100.8	1984	1 460.2	2004	4 804.8
1965	172.6	1985	1 322.2	2005	5 177.9
1966	240.9	1986	1 395.7	2006	5 345.1
1967	164.1	1987	1 672.2	2007	5 825
1968	110.9	1988	1 740.2	2008	6 012.7

资料来源: 国家统计局国民经济综合统计司编:《新中国 60 年统计资料汇编》, 1—37 全国主要工业产品产量.

表 A1—13

北京市 1995—2000 年月平均气温序列

单位: 摄氏度

年 月	1995	1996	1997	1998	1999	2000
1	-0.7	-2.2	-3.8	-3.9	-1.6	-6.4
2	2.1	-0.4	1.3	2.4	2.2	-1.5
3	7.7	6.2	8.7	7.6	4.8	8.1
4	14.7	14.3	14.5	15.0	14.4	14.6
5	19.8	21.6	20.0	19.9	19.5	20.4
6	24.3	25.4	24.6	23.6	25.4	26.7
7	25.9	25.5	28.2	26.5	28.1	29.6
8	25.4	23.9	26.6	25.1	25.6	25.7
9	19.0	20.7	18.6	22.2	20.9	21.8
10	14.5	12.8	14.0	14.8	13.0	12.6
11	7.7	4.2	5.4	4.0	5.9	3.0
12	-0.4	0.9	-1.5	0.1	-0.6	-0.6

表 A1—14

1993—2000 年中国社会消费品零售总额序列

单位: 亿元

月份	1993 年	1994 年	1995 年	1996 年	1997 年	1998 年	1999 年	2000 年
1 月	977.5	1 192.2	1 602.2	1 909.1	2 288.5	2 549.5	2 662.1	2 774.7
2 月	892.5	1 162.7	1 491.5	1 911.2	2 213.5	2 306.4	2 538.4	2 805.0
3 月	942.3	1 167.5	1 533.3	1 860.1	2 130.9	2 279.7	2 403.1	2 627.0
4 月	941.3	1 170.4	1 548.7	1 854.8	2 100.5	2 252.7	2 356.8	2 572.0
5 月	962.2	1 213.7	1 585.4	1 898.3	2 108.2	2 265.2	2 364.0	2 637.0
6 月	1 005.7	1 281.1	1 639.7	1 966.0	2 164.7	2 326.0	2 428.8	2 645.0
7 月	963.8	1 251.5	1 623.6	1 888.7	2 102.5	2 286.1	2 380.3	2 597.0
8 月	959.8	1 286.0	1 637.1	1 916.4	2 104.4	2 314.6	2 410.9	2 636.0
9 月	1 023.3	1 396.2	1 756.0	2 083.5	2 239.6	2 443.1	2 604.3	2 854.0
10 月	1 051.1	1 444.1	1 818.0	2 148.3	2 348.0	2 536.0	2 743.9	3 029.0
11 月	1 102.0	1 553.8	1 935.2	2 290.1	2 454.9	2 652.2	2 781.5	3 108.0
12 月	1 415.5	1 932.2	2 389.5	2 848.6	2 881.7	3 131.4	3 405.7	3 680.0

资料来源: 中国经济信息网.

表 A1—15 1950—1999 年北京市民用车辆拥有量序列 单位: 万辆

年份	车辆拥有量	年份	车辆拥有量
1950	5.43	1975	91.71
1951	6.19	1976	106.70
1952	6.63	1977	119.93
1953	7.18	1978	135.84
1954	8.95	1979	155.49
1955	10.14	1980	178.29
1956	11.74	1981	199.14
1957	12.60	1982	215.75
1958	17.26	1983	232.63
1959	21.07	1984	260.41
1960	22.38	1985	321.12
1961	24.00	1986	361.95
1962	24.80	1987	408.07
1963	26.13	1988	464.38
1964	27.61	1989	511.32
1965	29.95	1990	551.36
1966	33.92	1991	606.11
1967	33.21	1992	691.74
1968	34.80	1993	817.58
1969	37.16	1994	941.95
1970	42.41	1995	1 040.00
1971	49.44	1996	1 100.08
1972	57.74	1997	1 219.09
1973	67.27	1998	1 319.30
1974	78.57	1999	1 452.94

资料来源: 北京市统计局计算中心: 《北京五十年》.

表 A1—16 2001 年 1 月至 2014 年 11 月中国月度出口额 (行数据) 单位: 千美元

21 986 149	16 926 471	19 198 624	23 139 003	22 792 181	20 813 507
22 081 032	22 887 692	23 540 123	23 989 367	22 790 526	24 000 007
24 502 579	21 703 291	19 137 164	23 817 766	26 723 688	24 644 182
26 012 762	29 209 981	29 437 423	31 910 755	29 946 376	31 209 190
31 889 489	29 775 788	24 455 585	32 091 277	35 617 074	33 842 228
34 475 859	38 109 608	37 415 036	41 940 698	40 925 906	41 761 968
48 061 529	35 712 196	34 157 738	45 847 915	47 124 115	44 867 889
50 496 344	51 003 919	51 400 194	55 803 041	52 523 879	60 925 461
63 784 483	50 755 189	44 273 164	60 869 119	62 151 250	58 426 919
65 959 601	65 584 266	68 399 603	70 185 509	68 092 010	72 217 762

75 412 367	64 988 815	54 108 294	78 045 837	76 948 177	73 113 881
81 311 219	80 337 153	90 766 001	91 629 488	88 127 371	95 854 497
94 094 055	86 619 731	82 095 593	83 424 155	97 449 955	94 065 041
103 268 286	107 744 104	111 354 781	112 373 049	107 723 989	117 620 807
114 415 986	109 640 030	87 367 796	108 962 891	118 771 084	120 496 491
121 179 554	136 675 046	134 872 675	136 432 088	128 327 028	114 987 285
111 157 222	90 453 600	64 894 649	90 290 547	91 934 722	88 757 886
95 411 208	105 420 240	103 707 009	115 937 994	110 762 383	113 653 429
130 723 881	109 475 299	94 522 714	112 111 793	119 920 666	131 760 663
13 739 5578	145 519 322	139 302 230	144 985 285	135 978 419	153 325 762
154 148 895	150 734 151	96 736 127	152 199 071	155 684 084	157 157 191
161 980 774	175 128 036	173 315 697	169 673 012	157 490 772	174 464 123
174 717 651	149 939 113	114 470 693	165 658 100	163 251 712	181 140 762
180 204 470	176 940 032	177 973 297	186 349 509	175 571 102	179 380 767
199 229 682	187 365 572	139 367 235	182 189 999	187 061 474	182 766 116
174 315 796	185 991 317	190 608 317	185 643 576	185 405 576	202 204 803
207 742 423	207 132 123	114 093 694	170 107 988	188 540 548	195 472 920
186 788 537	212 891 063	208 465 863	213 687 462	206 866 715	

表 A1—17

1952—1988 年中国农业实际国民收入指数序列

(以 1952 年农业国民收入总额为基数 100)

年份	农业	年份	农业
1952	100.0	1971	142.0
1953	101.6	1972	140.5
1954	103.3	1973	153.1
1955	111.5	1974	159.2
1956	116.5	1975	162.3
1957	120.1	1976	159.1
1958	120.3	1977	155.1
1959	100.6	1978	161.2
1960	83.6	1979	171.5
1961	84.7	1980	168.4
1962	88.7	1981	180.4
1963	98.9	1982	201.6
1964	111.9	1983	218.7
1965	122.9	1984	247.0
1966	131.9	1985	253.7
1967	134.2	1986	261.4
1968	131.6	1987	273.2
1969	132.2	1988	279.4
1970	139.8		

资料来源: <http://www.cs.few.eur.nl/few/people/franses>.

表 A1—18 1917—1975 年美国 23 岁妇女每万人生育率序列

年份	每万人生育率	年份	每万人生育率
1917	183.1	1947	212.0
1918	183.9	1948	200.4
1919	163.1	1949	201.8
1920	179.5	1950	200.7
1921	181.4	1951	215.6
1922	173.4	1952	222.5
1923	167.6	1953	231.5
1924	177.4	1954	237.9
1925	171.7	1955	244.0
1926	170.1	1956	259.4
1927	163.7	1957	268.8
1928	151.9	1958	264.3
1929	145.4	1959	264.5
1930	145.0	1960	268.1
1931	138.9	1961	264.0
1932	131.5	1962	252.8
1933	125.7	1963	240.0
1934	129.5	1964	229.1
1935	129.6	1965	204.8
1936	129.5	1966	193.3
1937	132.2	1967	179.0
1938	134.1	1968	178.1
1939	132.1	1969	181.1
1940	137.4	1970	165.6
1941	148.1	1971	159.8
1942	174.1	1972	136.1
1943	174.7	1973	126.3
1944	156.7	1974	123.3
1945	143.3	1975	118.5
1946	189.7		

资料来源: Hipel and Mcleod (1994).

表 A1—19 1962—1991 年德国工人季度失业率序列 (行数据 (%))

1.1	0.5	0.4	0.7	1.6	0.6	0.5	0.7
1.3	0.6	0.5	0.7	1.2	0.5	0.4	0.6
0.9	0.5	0.5	1.1	2.9	2.1	1.7	2.0
2.7	1.3	0.9	1.0	1.6	0.6	0.5	0.7
1.1	0.5	0.5	0.6	1.2	0.7	0.7	1.0
1.5	1.0	0.9	1.1	1.5	1.0	1.0	1.6
2.6	2.1	2.3	3.6	5.0	4.5	4.5	4.9
5.7	4.3	4.0	4.4	5.2	4.3	4.2	4.5
5.2	4.1	3.9	4.1	4.8	3.5	3.4	3.5
4.2	3.4	3.6	4.3	5.5	4.8	5.4	6.5
8.0	7.0	7.4	8.5	10.1	8.9	8.8	9.0
10.0	8.7	8.8	8.9	10.4	8.9	8.9	9.0
10.2	8.6	8.4	8.4	9.9	8.5	8.6	8.7
9.8	8.6	8.4	8.2	8.8	7.6	7.5	7.6
8.1	7.1	6.9	6.6	6.8	6.0	6.2	6.2

资料来源: P. H. Franses, *Time Series Models for Business and Economic Forecasting*, Cambridge University, 1998.

表 A1—20 1948—1981 年美国女性 (20 岁以上) 月度失业率序列

446	650	592	561	491	592	604	635	580	510
553	554	628	708	629	724	820	865	1 007	1 025
955	889	965	878	1 103	1 092	978	823	827	928
838	720	756	658	838	684	779	754	794	681
658	644	622	588	720	670	746	616	646	678
552	560	578	514	541	576	522	530	564	442
520	484	538	454	404	424	432	458	556	506
633	708	1 013	1 031	1 101	1 061	1 048	1 005	987	1 006
1 075	854	1 008	777	982	894	795	799	781	776
761	839	842	811	843	753	848	756	848	828
857	838	986	847	801	739	865	767	941	846
768	709	798	831	833	798	806	771	951	799
1 156	1 332	1 276	1 373	1 325	1 326	1 314	1 343	1 225	1 133
1 075	1 023	1 266	1 237	1 180	1 046	1 010	1 010	1 046	985
971	1 037	1 026	947	1 097	1 018	1 054	978	955	1 067
1 132	1 092	1 019	1 110	1 262	1 174	1 391	1 533	1 479	1 411
1 370	1 486	1 451	1 309	1 316	1 319	1 233	1 113	1 363	1 245
1 205	1 084	1 048	1 131	1 138	1 271	1 244	1 139	1 205	1 030
1 300	1 319	1 198	1 147	1 140	1 216	1 200	1 271	1 254	1 203
1 272	1 073	1 375	1 400	1 322	1 214	1 096	1 198	1 132	1 193

1 163	1 120	1 164	966	1 154	1 306	1 123	1 033	940	1 151
1 013	1 105	1 011	963	1 040	838	1 012	963	888	840
880	939	868	1 001	956	966	896	843	1 180	1 103
1 044	972	897	1 103	1 056	1 055	1 287	1 231	1 076	929
1 105	1 127	988	903	845	1 020	994	1 036	1 050	977
956	818	1 031	1 061	964	967	867	1 058	987	1 119
1 202	1 097	994	840	1 086	1 238	1 264	1 171	1 206	1 303
1 393	1 463	1 601	1 495	1 561	1 404	1 705	1 739	1 667	1 599
1 516	1 625	1 629	1 809	1 831	1 665	1 659	1 457	1 707	1 607
1 616	1 522	1 585	1 657	1 717	1 789	1 814	1 698	1 481	1 330
1 646	1 596	1 496	1 386	1 302	1 524	1 547	1 632	1 668	1 421
1 475	1 396	1 706	1 715	1 586	1 477	1 500	1 648	1 745	1 856
2 067	1 856	2 104	2 061	2 809	2 783	2 748	2 642	2 628	2 714
2 699	2 776	2 795	2 673	2 558	2 394	2 784	2 751	2 521	2 372
2 202	2 469	2 686	2 815	2 831	2 661	2 590	2 383	2 670	2 771
2 628	2 381	2 224	2 556	2 512	2 690	2 726	2 493	2 544	2 232
2 494	2 315	2 217	2 100	2 116	2 319	2 491	2 432	2 470	2 191
2 241	2 117	2 370	2 392	2 255	2 077	2 047	2 255	2 233	2 539
2 394	2 341	2 231	2 171	2 487	2 449	2 300	2 387	2 474	2 667
2 791	2 904	2 737	2 849	2 723	2 613	2 950	2 825	2 717	2 593
2 703	2 836	2 938	2 975	3 064	3 092	3 063	2 991		

资料来源: Andrews & Herzberg (1985).

表 A1—21 1963 年 4 月至 1971 年 7 月美国短期国库券的月度收益率序列

0.002 38	0.002 38	0.002 36	0.002 5	0.002 54	0.002 6	0.002 85	0.002 81
0.002 41	0.002 88	0.002 87	0.002 92	0.002 94	0.002 73	0.002 71	0.002 82
0.002 67	0.002 73	0.002 93	0.002 85	0.002 96	0.002 81	0.003 26	0.003 21
0.003 15	0.003 19	0.003 13	0.003 13	0.003 19	0.003 13	0.003 30	0.003 19
0.003 15	0.003 55	0.003 70	0.003 71	0.003 64	0.003 81	0.003 72	0.003 68
0.003 74	0.003 89	0.004 15	0.003 89	0.003 43	0.003 77	0.003 68	0.003 64
0.003 38	0.002 83	0.002 71	0.003 00	0.003 09	0.003 17	0.003 43	0.003 47
0.003 55	0.003 60	0.003 98	0.003 85	0.003 89	0.004 44	0.004 53	0.004 44
0.004 32	0.004 06	0.004 32	0.004 61	0.003 98	0.005 00	0.004 87	0.004 70
0.004 32	0.005 08	0.004 78	0.005 08	0.005 93	0.005 50	0.005 93	0.005 42
0.005 40	0.005 40	0.006 31	0.005 34	0.005 46	0.005 42	0.005 46	0.004 95
0.005 00	0.005 08	0.004 83	0.004 44	0.003 77	0.003 55	0.003 38	0.002 64
0.002 83	0.003 05	0.003 38	0.004 06				

资料来源: Hipel and Mcleod (1994).

表 A1—22 1926—1991 年标准普尔 500 股票价值加权月度收益率序列 (行数据)

0.022 5	-0.044	-0.059 1	0.022 7	0.007 7	0.043 2	0.045 5	0.017 1	0.022 9	-0.031 3
0.022 3	0.016 6	-0.020 8	0.047 7	0.006 5	0.017 2	0.052 2	-0.009 4	0.065	0.044 5
0.043 2	-0.053 1	0.067 8	0.019	-0.005 1	-0.017 6	0.108 3	0.032 4	0.012 7	-0.040 5
0.012 5	0.074 1	0.024	0.014 5	0.119 9	0.002 9	0.057 1	-0.005 8	-0.002 3	0.016 1
-0.042 8	0.112 4	0.045 6	0.098	-0.048 9	-0.199 3	-0.133 7	0.025 3	0.062 5	0.021 5
0.079 9	-0.009 5	-0.016 5	-0.164 6	0.036 7	0.007 5	-0.130 1	-0.088 8	-0.021 8	-0.074 2
0.048 9	0.114 4	-0.069 2	-0.095 9	-0.137 2	0.139	-0.074 2	0.009 5	-0.299 4	0.084 4
-0.097 8	-0.145 3	-0.028 3	0.050 7	-0.118 2	-0.202 5	-0.233 3	-0.008 9	0.377	0.375 4
-0.036 9	-0.138 6	-0.058 9	0.051 9	0.007 3	-0.184 4	0.033 6	0.422 2	0.158 7	0.131 7
-0.088	0.114 6	-0.113 6	-0.088 5	0.102 7	0.022 3	0.105 9	-0.036 7	-0.000 9	-0.027
-0.081 3	0.020 8	-0.115 2	0.054 1	-0.005 5	-0.031 9	0.082 9	-0.004 2	-0.042 1	-0.039 6
-0.030 9	0.095 6	0.032 3	0.067 8	0.083 1	0.021 7	0.023 9	0.075 1	0.039 3	0.037 1
0.065 5	0.016 8	0.025 4	-0.077 1	0.045 8	0.030 6	0.068 1	0.008 8	0.001 3	0.075
0.004 1	-0.005 8	0.037 8	0.014 6	-0.009 4	-0.083 1	-0.010 3	-0.052 9	0.102 6	-0.055 4
-0.142 1	-0.101 7	-0.101 1	-0.050 4	0.013 3	0.060 8	-0.250 4	0.141 2	-0.044 3	0.247
0.072 7	-0.027 4	0.014 9	0.076	-0.033 4	0.037 7	-0.068 9	0.032 5	-0.135 4	-0.005 5
0.062 3	-0.063 8	0.108 7	-0.071 4	0.164 6	-0.014 6	-0.049 1	0.023 8	-0.035 2	0.006 6
0.009 9	-0.004 9	-0.239 5	0.076 6	0.031 1	0.026 2	0.009 5	0.039 4	-0.042 4	-0.002 8
-0.048 2	-0.014 9	0.004	-0.065 3	0.004 3	0.053 5	0.054 8	-0.008 7	-0.009 7	-0.068 6
-0.042 1	-0.045 1	0.013 8	-0.025	-0.067 5	-0.043 7	0.064	0.018 4	0.031 3	0.007
0.026 7	0.064 4	-0.013 8	0.051 7	0.071 6	0.050 6	0.052 7	0.000 9	0.044 9	0.019 8
-0.054 3	0.010 3	0.023 7	-0.013 2	-0.075 5	0.059	0.015 4	-0.002 5	0.016 9	-0.012 5
0.040 4	0.051	-0.020 8	0.008 7	-0.003 1	0	0.003 9	0.035 1	0.014 3	0.061 6
-0.046 2	0.088	0.011 5	-0.003 3	-0.020 1	0.058	0.041 9	0.030 3	0.032 4	0.009 9
0.069 7	-0.069 5	0.046 3	0.037 6	0.022 4	-0.039 1	-0.025 5	-0.072 9	-0.101 5	-0.008
-0.011 5	0.042 9	0.023 5	-0.014 7	-0.016 9	-0.038 9	-0.008 9	0.052 6	0.036 2	-0.027 9
-0.013 7	0.021 2	-0.028 5	0.020 7	-0.039 9	-0.047	0.077 1	0.026 5	0.078 2	0.003
-0.053 2	0.007 6	-0.030 1	0.067 8	-0.108 2	0.030 5	0.001 3	-0.039 4	0.030 1	-0.021 2
-0.037 3	-0.002 1	0.062 1	0.012	0.023 7	0.029 5	0.001 2	0.043 6	0.017 3	0.01
0.004 1	0.045 1	0.039 3	-0.058	0.008 5	0.032 5	0.055 9	0.004 1	-0.001	0.046 1
0.061 2	0.006 5	-0.018 3	0.048 1	-0.040 6	-0.026	0.068 7	0.039 3	-0.000 9	-0.013 8
-0.002 6	0.038 9	0.015 6	-0.036 5	0.047 7	-0.043 1	0.023 2	0.046 1	0.017 6	-0.014 6
-0.019 6	-0.000 8	0.046 5	0.035 5	-0.007 2	-0.018 2	-0.023 6	-0.026 5	-0.003 2	-0.016 3
0.025 3	-0.057 8	0.001 3	0.051	0.009	0.002	0.051 2	0.002 7	0.030 2	0.049
0.032 9	0.000 7	0.057 2	-0.034	0.083 1	-0.019 5	0.080 8	0.050 8	0.018 1	0.003 5
-0.004 9	0.037 7	-0.001 3	0.082 3	0.060 7	-0.007 8	0.011 3	-0.030 5	0.074 9	-0.000 7
-0.036 5	0.034 7	0.069 3	-0.002 1	-0.065 7	0.039 2	0.051 5	-0.038 1	-0.045 5	0.005 1
-0.011	0.035 3	-0.041 8	-0.032 6	0.019 6	0.037	0.036 9	-0.001 3	0.011 4	-0.056 1
-0.061 9	-0.032 1	0.016 1	-0.041 5	0.042 8	-0.020 6	0.030 9	0.031 8	0.015	0.026 1
0.043 1	0.011 9	0.048 4	0.025 4	0.022 4	0.052	0.003 8	-0.000 2	0.000 5	0.038 8

0.018 9	-0.003 6	0.034 9	-0.015	-0.045 6	0.011 3	0.013 2	0.027 6	-0.071 5	0.009 2
-0.013 9	-0.017 5	0.026 9	0.019 5	-0.024 8	0.026 1	-0.060 4	-0.002 4	0.040 3	0.046 3
0.063 2	0.026 9	0.025 5	0.003 8	0.019 1	-0.028 8	0.032 8	0.019 6	-0.019 7	0.028 3
0.039 3	0.003 2	-0.037 9	0.016 3	-0.005 9	-0.062	-0.086	-0.081 8	0.063 6	0.015 3
-0.048 2	0.004 4	0.101 6	0.013 5	0.049 1	-0.028 9	0.035 5	0.048 5	0.014 3	-0.020 2
-0.003 5	0.048 7	-0.011	0.032 2	-0.010 5	0.024 4	0.026 9	0.009 9	0.015 2	0.006 1
0.011 5	0.016 4	0.018 2	-0.016 2	0.028 7	0.008 1	-0.005 2	0.003 9	0.033 2	-0.001 5
-0.014 5	0.034 2	-0.007 7	-0.048 6	0.013 4	0.022 5	0.032	0.027 3	-0.008 8	0.009
0.004 9	-0.017 9	-0.021 8	0.020 5	-0.054 1	-0.016 1	-0.013 5	-0.077 8	-0.007	0.047 5
0.003 1	-0.001 5	0.078 2	0.002	0.039 4	0.042 2	-0.052 4	0.017 5	0.045 3	-0.011 7
0.032 8	-0.029 1	0.001 1	0.026 3	-0.043 8	-0.031 2	0.009 4	0.081 9	0.011 2	0.009 1
-0.018 5	0.011 5	0.038 5	0.007 2	0.048	-0.041 6	-0.008 2	-0.047 4	0.034 4	0.021 5
-0.002 2	-0.055 6	-0.060 2	0.040 1	-0.025	0.044 2	-0.035 3	-0.018 7	-0.076 5	0.052 7
0.001 5	-0.090 5	-0.061	-0.05	0.073 3	0.044 5	0.033	-0.011 4	0.047 4	0.056 8
0.040 5	0.009 1	0.036 8	0.036 3	-0.041 6	0.000 7	-0.041 3	0.036 1	-0.007	-0.041 8
-0.002 5	0.086 2	0.018 1	0.025 3	0.005 9	0.004 4	0.017 3	-0.021 8	0.002 3	0.034 5
-0.004 9	0.009 3	0.045 6	0.011 8	-0.017 1	-0.037 5	-0.001 4	-0.040 8	-0.018 9	-0.006 6
0.038	-0.036 7	0.040 1	-0.001 3	-0.113 9	0.016 6	-0.01	-0.003 6	-0.023 3	-0.039 1
-0.033 6	-0.014 7	-0.077 8	-0.090 3	-0.119 3	0.163	-0.053 2	-0.020 2	0.122 8	0.059 9
0.021 7	0.047 3	0.044 1	0.044 3	-0.067 7	-0.021 1	-0.034 6	0.061 6	0.024 7	-0.011 5
0.032 8	-0.029 1	0.001 1	0.026 3	-0.043 8	-0.031 2	0.009 4	0.081 9	0.011 2	0.009 1
0.118 3	-0.011 4	0.030 7	-0.011	-0.014 4	0.040 9	-0.008 1	-0.005 1	0.022 6	-0.022 2
-0.007 8	0.052 5	-0.050 5	-0.021 7	-0.014	0.000 2	-0.023 6	0.045 4	-0.016 2	-0.021
-0.002 5	-0.043 4	0.027	0.002 8	-0.061 5	-0.024 8	0.024 9	0.085 4	0.004 2	-0.017 6
0.053 9	0.025 9	-0.007 3	-0.091 6	0.016 6	0.014 9	0.039 7	-0.036 5	0.055 2	0.001 7
-0.026 3	0.038 7	0.008 7	0.053 1	0	-0.068 6	0.042 6	0.016 8	0.057 6	-0.004 4
-0.101 8	0.041 1	0.046 6	0.027	0.065	0.005 8	0.025 2	0.016	0.102 4	-0.033 9
-0.045 7	0.013 3	0.036	-0.023 5	-0.001 7	-0.010 4	-0.002 2	-0.062 1	-0.053 8	0.049 1
0.036 6	-0.030 1	-0.017 5	-0.060 5	-0.010 2	0.04	-0.039 2	-0.020 3	-0.023	0.116
0.007 6	0.110 4	0.036 1	0.015 2	0.033 1	0.019	0.033 1	0.074 9	-0.012 3	0.035 2
-0.033	0.011 3	0.010 2	-0.015 2	0.017 4	-0.008 8	-0.009 2	-0.038 9	0.013 5	0.005 5
-0.059 4	0.017 5	-0.016 5	0.106 3	-0.003 5	-0.000 1	-0.015 1	0.022 4	0.074 1	0.008 7
-0.002 9	-0.004 6	0.054 1	0.012 1	-0.004 8	-0.012	-0.034 7	0.042 5	0.065 1	0.045 1
0.002 4	0.071 5	0.052 8	-0.014 1	0.050 2	0.014 1	-0.058 7	0.071 2	-0.085 4	0.054 7
0.021 5	-0.028 3	0.131 8	0.036 9	0.026 4	-0.011 5	0.006	0.047 9	0.048 2	0.035
-0.024 2	-0.217 6	-0.085 3	0.072 9	0.040 4	0.041 8	-0.033 3	0.009 4	0.003 2	0.043 3
-0.005 4	-0.038 6	0.039 7	0.026	-0.018 9	0.014 7	0.071 1	-0.028 9	0.020 8	0.050 1
0.035 1	-0.007 9	0.088 4	0.015 5	-0.006 5	-0.025 2	0.016 5	0.021 4	-0.068 8	0.008 5
0.024 3	-0.026 9	0.092	-0.008 9	-0.005 2	-0.094 3	-0.051 2	-0.006 7	0.059 9	0.024 8
0.041 5	0.067 3	0.022 2	0.000 3	0.038 6	-0.047 9	0.044 9	0.019 6	-0.019 1	0.011 9

资料来源: Time Series Data Library (citing: Pena, Tiao and Tsay (2001)).

表 A1—23 1979 年 12 月 31 日至 1991 年 12 月 31 日外币对美元的日兑换率序列 (列数据)

1.726	2.23	2.603	2.9	1.89	1.928 5	1.697 5	1.599 2	1.413 4	1.644
1.713 5	2.231 5	2.584 2	2.882	1.894 2	1.89	1.713	1.605	1.385 5	1.648 3
1.710 8	2.216	2.585	2.892	1.897 5	1.890 5	1.720 8	1.605 5	1.397 2	1.651 7
1.712 6	2.213 8	2.588 8	2.906 5	1.892 5	1.884	1.739 7	1.583	1.385 5	1.674 8
1.708	2.224	2.582 8	2.901	1.886 8	1.858	1.746	1.567 5	1.401 4	1.656 3
1.712 9	2.218	2.61	2.847 5	1.876	1.880 5	1.739	1.58	1.416 5	1.673 1
1.712 2	2.232	2.604 2	2.732	1.877	1.876	1.749	1.584 8	1.406 3	1.685 8
1.716 8	2.226 5	2.617 5	2.716	1.844 5	1.882	1.753	1.582	1.407 3	1.686 9
1.721 4	2.241	2.619	2.68	1.841	1.903 5	1.77	1.587 6	1.418 5	1.685 9
1.719 1	2.249	2.620 5	2.661 5	1.827 7	1.907	1.770 5	1.583 5	1.405 7	1.688 5
1.728	2.241	2.631 5	2.678	1.818	1.892	1.773	1.571 5	1.391 5	1.696 7
1.721 5	2.254	2.654 5	2.679	1.816 5	1.902	1.792	1.58	1.384	1.686 3
1.723 6	2.261	2.645 2	2.652 5	1.823	1.901	1.797 5	1.578 5	1.411 5	1.688 6
1.723 5	2.283	2.658 5	2.649	1.825 5	1.906	1.789 5	1.583 5	1.378 5	1.673
1.729 4	2.268	2.670 8	2.62	1.813 5	1.909 2	1.798	1.604	1.374	1.677 2
1.733 8	2.277 5	2.679 5	2.634 5	1.81	1.895 8	1.814 2	1.608 5	1.384 1	1.687 3
1.729 7	2.276 5	2.667 5	2.650 5	1.810 8	1.874 5	1.79	1.612	1.380 5	1.688 5
1.731 8	2.292	2.658	2.647	1.813	1.869	1.797 7	1.614 5	1.376	1.687 2
1.736 5	2.284	2.676 3	2.648	1.803	1.887 5	1.798 5	1.601 5	1.384	1.696 7
1.738 8	2.272 5	2.677 5	2.656 5	1.793 5	1.865 8	1.784 5	1.617 5	1.409 5	1.713 4
1.736 8	2.275	2.686 5	2.658	1.796	1.865 1	1.799	1.624 5	1.402	1.712 8
1.737 3	2.268	2.697 5	2.662 5	1.792 7	1.851 5	1.785 3	1.623 7	1.404 3	1.717 2
1.743 8	2.269 5	2.704 5	2.685	1.797	1.866 3	1.795 5	1.613	1.389 8	1.719
1.744	2.271 5	2.701	2.648 5	1.806 5	1.897	1.813 8	1.627	1.391 3	1.703 1
1.741 8	2.251	2.688 5	2.637 5	1.799 8	1.905 5	1.821 5	1.636 5	1.369 5	1.703 6
1.742 5	2.238	2.702 5	2.638	1.817 5	1.890 5	1.835	1.629	1.365	1.702
1.733	2.241 3	2.712	2.644 5	1.816	1.900 5	1.830 5	1.632 4	1.356 5	1.705 5
1.738 5	2.261 5	2.712 2	2.638	1.816	1.891 7	1.826 5	1.613	1.381	1.696 2
1.739 3	2.246	2.715 2	2.650 5	1.817	1.939	1.811	1.588	1.375	1.690 3
1.739	2.261 3	2.699	2.653 5	1.806	1.939	1.817 5	1.592	1.374 5	1.673 1
1.735	2.258 5	2.702	2.642 8	1.815	1.947	1.814 5	1.597	1.368	1.682 9
1.736 1	2.284	2.711 5	2.622	1.820 3	1.941 5	1.834	1.600 5	1.377 5	1.689
1.741 5	2.287	2.744	2.623 5	1.821 8	1.969	1.789	1.623 5	1.376	1.688 2
1.748	2.293	2.735	2.619	1.821 7	1.950 2	1.788 3	1.620 5	1.385 6	1.687 7
1.748	2.306	2.735 5	2.602	1.821	1.952	1.799 5	1.624 5	1.39	1.692 1
1.747	2.312	2.734 5	2.597 5	1.824 7	1.941	1.789 5	1.638	1.379	1.691 7
1.759	2.297 5	2.752	2.613 8	1.84	1.956 5	1.771	1.634	1.374 2	1.675 3
1.761 6	2.294 8	2.753	2.6	1.842 8	1.959	1.748 5	1.622	1.368 8	1.678 8
1.763 5	2.308 5	2.771	2.628	1.846 8	1.954	1.759	1.626 2	1.372 1	1.667 8

1.758	2.297 2	2.766 5	2.623 5	1.842 5	1.959	1.756 8	1.637 2	1.363 5	1.671
1.763 8	2.308 5	2.772	2.623 5	1.842 7	1.937 5	1.736 5	1.658 8	1.372 5	1.675 1
1.777	2.338	2.776 7	2.614 5	1.835 5	1.946 2	1.755 8	1.656 5	1.391 5	1.669 3
1.782 9	2.314	2.772	2.616	1.826	1.958 2	1.738 5	1.68	1.429	1.680 2
1.790 7	2.322 5	2.767 5	2.623 5	1.822 5	1.978	1.753	1.689 5	1.446 5	1.708 6
1.788 5	2.327 5	2.769	2.614 5	1.814	1.990 5	1.751 5	1.697 5	1.444 5	1.711 5
1.795	2.316 5	2.768 5	2.598 7	1.821	1.975	1.746 5	1.694	1.436	1.724 3
1.796	2.35	2.759	2.612 5	1.811 3	1.985 2	1.758 3	1.708 8	1.437	1.718 1
1.809 5	2.352 5	2.755 8	2.594	1.801	1.980 5	1.735	1.703 5	1.457	1.723 8
1.812	2.359 5	2.751 5	2.578	1.801 5	1.987 3	1.719 8	1.7	1.440 2	1.724 5
1.807 6	2.341 5	2.728 5	2.566 3	1.777 7	1.988 2	1.712 5	1.693	1.440 3	1.733 9
1.821 5	2.345	2.725	2.551	1.807 8	1.980 5	1.709	1.706	1.446 5	1.729
1.833	2.357 7	2.760 5	2.543	1.813 3	1.969	1.723 5	1.701 8	1.438 5	1.724 5
1.874	2.375	2.788	2.512 5	1.814	2	1.728	1.696	1.404	1.712 5
1.877	2.366 7	2.798	2.521 7	1.787	1.952 5	1.724 8	1.7	1.379 3	1.703 2
1.873	2.366 5	2.810 7	2.533 5	1.774 5	1.949 5	1.731	1.708	1.387 8	1.714 7
1.865 5	2.395 5	2.829	2.519 5	1.763 5	1.941 5	1.735	1.704 5	1.414	1.712 3
1.878	2.399	2.814	2.518 5	1.753 1	1.944	1.733 2	1.721 5	1.417 6	1.716 5
1.888	2.366	2.839 5	2.525 5	1.728 5	1.95	1.761	1.728 5	1.418 7	1.724 6
1.897 4	2.364 5	2.839 5	2.535	1.728 3	1.906	1.816 5	1.72	1.405	1.728 9
1.897	2.351 5	2.808	2.54	1.712 5	1.889	1.804	1.717	1.416 2	1.722 7
1.919 5	2.373	2.798	2.543	1.715	1.882 5	1.753	1.724 3	1.405	1.733
1.932	2.367	2.808 5	2.515	1.710 5	1.881 5	1.728	1.723	1.415 2	1.720 7
1.951 5	2.369 7	2.803	2.522	1.669	1.872 5	1.752 5	1.705 8	1.403 5	1.727
1.961 5	2.388	2.792 5	2.523 5	1.674 5	1.875	1.748	1.709 5	1.401	1.731 7
1.954 5	2.387 5	2.820 5	2.511 5	1.661 5	1.886 5	1.749 2	1.697 5	1.406	1.722 5
1.963 5	2.376 8	2.812	2.518	1.665	1.878	1.739 5	1.707 5	1.401	1.721 5
1.97	2.371 9	2.826	2.517	1.698 5	1.886	1.747 2	1.719	1.412 5	1.707 9
1.975	2.364 5	2.823	2.512	1.688	1.892 5	1.748 5	1.724 3	1.401 5	1.690 3
1.941	2.343 5	2.811	2.505	1.703 5	1.896 5	1.740 2	1.726 5	1.404 3	1.703 8
1.89	2.343	2.821	2.505 5	1.693 2	1.917 2	1.736 5	1.719	1.394 3	1.697 2
1.877	2.367 5	2.811 7	2.505	1.687 5	1.907 8	1.74	1.736	1.391 5	1.700 1
1.876 5	2.358 3	2.816 5	2.473	1.680 5	1.891	1.732 5	1.740 5	1.399 1	1.698 7
1.871 5	2.371	2.792	2.465	1.682	1.866	1.694 5	1.716	1.387 8	1.692 9
1.895	2.379 5	2.76	2.447	1.667 8	1.852 2	1.694	1.704	1.392 7	1.710 5
1.86	2.373 5	2.735 5	2.438	1.687	1.854	1.697 2	1.710 5	1.383 5	1.696 1
1.857	2.377 5	2.743	2.461 3	1.668 6	1.851 5	1.686 5	1.716	1.401 5	1.691 8
1.854	2.373	2.765 5	2.448	1.654 5	1.854 3	1.684 5	1.697 5	1.384	1.694 3
1.848 5	2.375 4	2.753 5	2.444 5	1.640 5	1.855 7	1.673 8	1.697 3	1.383 5	1.692 7

1.862	2.383 8	2.737 5	2.456 5	1.648 5	1.839	1.670 5	1.707 5	1.380 5	1.699 3
1.823	2.391	2.741 5	2.448	1.653 5	1.840 9	1.682	1.718	1.380 3	1.693 9
1.819 5	2.385	2.722	2.455 5	1.659 5	1.838	1.689 8	1.718 3	1.382 2	1.699 5
1.813	2.386 5	2.667 2	2.462 5	1.669	1.83	1.687 5	1.711	1.392 7	1.709
1.798 3	2.393 5	2.673 5	2.467 8	1.671 5	1.836 5	1.673	1.687	1.396 5	1.728 3
1.796 5	2.396	2.683 3	2.462 5	1.668 8	1.841 3	1.678	1.696 5	1.403 7	1.728 4
1.8	2.415	2.708	2.467 5	1.664 2	1.837 7	1.682 5	1.683 5	1.402 5	1.731 2
1.802	2.408	2.662 5	2.464 5	1.636 5	1.847 5	1.683 8	1.686 3	1.392	1.73
1.823 5	2.417 5	2.661 5	2.457 5	1.631 5	1.848	1.676	1.678 5	1.391 9	1.73
1.796	2.402 8	2.625	2.448 5	1.631	1.851 8	1.662 5	1.689 5	1.394 8	1.708 6
1.793	2.398	2.636 2	2.454 5	1.630 5	1.844 8	1.665 7	1.682 5	1.387 6	1.72
1.785 5	2.408 2	2.604 4	2.417	1.630 8	1.845 5	1.663	1.688 7	1.387 3	1.717 1
1.793 3	2.413 8	2.598 3	2.399 2	1.622 5	1.848 5	1.666 2	1.668 7	1.379 8	1.724 8
1.812	2.413 5	2.611	2.398	1.634	1.863 3	1.678	1.667	1.386 8	1.730 5
1.790 7	2.422 5	2.585 5	2.389	1.634 5	1.862	1.680 5	1.673 6	1.384	1.730 8
1.794 2	2.422 5	2.579 5	2.379	1.628 5	1.848 5	1.696	1.678	1.394 5	1.734
1.789 2	2.419 5	2.551 8	2.388	1.636 5	1.838 5	1.691 8	1.659 5	1.388	1.732 9
1.791	2.412	2.555	2.401	1.629	1.835 5	1.694	1.647 7	1.382	1.724 4
1.793 8	2.419 2	2.587 5	2.414 5	1.596 5	1.842 3	1.692 5	1.622	1.379 8	1.729 8
1.802 5	2.424 5	2.582	2.386	1.595 8	1.836	1.711 8	1.610 7	1.385 2	1.720 4
1.803 1	2.422 2	2.603	2.391 5	1.594	1.820 5	1.702 4	1.62	1.381 6	1.726 4
1.788 3	2.410 4	2.572	2.41	1.570 8	1.813 7	1.703 2	1.6	1.398 5	1.725 8
1.789 5	2.398 8	2.565 5	2.385 5	1.584 5	1.805 3	1.689	1.597 7	1.39	1.727 2
1.778 5	2.394 8	2.583 5	2.375	1.627	1.79	1.690 5	1.612	1.393 5	1.737 6
1.762 4	2.396	2.627 5	2.367 8	1.649 2	1.788	1.694 7	1.607	1.406 3	1.743 8
1.772 3	2.387 5	2.629	2.349	1.639	1.783	1.704	1.593 5	1.409 8	1.742 1
1.778 5	2.373 5	2.637	2.347	1.650 5	1.78	1.699 5	1.606 5	1.407	1.752 9
1.778 5	2.368 5	2.647	2.344	1.637	1.782 3	1.701	1.610 7	1.422	1.749 8
1.782 8	2.355 5	2.635 5	2.307	1.635	1.781	1.710 8	1.613 5	1.434 5	1.749 1
1.775 3	2.339	2.628 5	2.307 5	1.633 3	1.777	1.690 5	1.641	1.435 5	1.758 7
1.773 8	2.331	2.599 5	2.309	1.637	1.766	1.672	1.628 5	1.474	1.761 5
1.774 8	2.338 5	2.588 5	2.282 5	1.679 5	1.772 7	1.668	1.644 5	1.475	1.751 8
1.771 6	2.345 3	2.591 5	2.245	1.675 8	1.772 2	1.661 5	1.648 5	1.478 2	1.770 4
1.762 2	2.321	2.585 5	2.227 5	1.661 8	1.765 9	1.632 2	1.631 2	1.473	1.788 5
1.761 8	2.309 5	2.594	2.222	1.664	1.734	1.643	1.617 3	1.471 7	1.802 9
1.765 5	2.289 7	2.598	2.230 8	1.675 5	1.74	1.640 7	1.613 5	1.484 2	1.798 7
1.763 3	2.286	2.622 5	2.220 5	1.677 5	1.737	1.631	1.634	1.481 5	1.792 2
1.755	2.282	2.621 5	2.201	1.676 5	1.723	1.642 5	1.630 5	1.476 5	1.788 5
1.767 7	2.283	2.631	2.267	1.671 2	1.726	1.636 5	1.624 5	1.479 1	1.795

1.763	2.293	2.622 5	2.242 5	1.667	1.737 8	1.639 1	1.623 5	1.462 5	1.819 3
1.763 6	2.306 5	2.630 5	2.238	1.678 5	1.740 8	1.630 8	1.623	1.471 5	1.824 1
1.767	2.305	2.618 5	2.280 5	1.692	1.725 5	1.638	1.623	1.475 8	1.829 8
1.767 8	2.317	2.609 5	2.265 5	1.682	1.708	1.606	1.605 5	1.467 9	1.834 8
1.767 3	2.318 5	2.618	2.284	1.688	1.699 5	1.603	1.597	1.465 5	1.848 2
1.77	2.316 5	2.625 5	2.295 5	1.690 3	1.69	1.602 5	1.603 5	1.463 3	1.832 1
1.768 5	2.306	2.635	2.263	1.698 5	1.687 5	1.597	1.614	1.478 7	1.837 4
1.764 5	2.300 5	2.642	2.25	1.695	1.689 5	1.586	1.614 9	1.476 2	1.837 9
1.763 5	2.314 5	2.646	2.248	1.700 5	1.708 8	1.583	1.627 5	1.480 3	1.864 6
1.764 1	2.329	2.66	2.261 5	1.690 5	1.720 8	1.580 2	1.638 3	1.471 8	1.864 5
1.761 5	2.346	2.666	2.256 5	1.691 5	1.683 5	1.616	1.645 5	1.470 8	1.881
1.758 5	2.347 5	2.662	2.242	1.713 5	1.680 5	1.627	1.664	1.493	1.878 6
1.756	2.376 3	2.683	2.288	1.708 1	1.668 8	1.608 5	1.675	1.485 4	1.870 2
1.738 9	2.369	2.680 5	2.339	1.706 3	1.682 1	1.614 3	1.672	1.487 5	1.845 8
1.739 7	2.363	2.697	2.33	1.710 5	1.68	1.615 3	1.684	1.486	1.856 8
1.741	2.365	2.704 5	2.326 5	1.707 5	1.682	1.590 3	1.681 5	1.485 4	1.867 5
1.737 7	2.382	2.721 5	2.335	1.695 5	1.679 5	1.572 5	1.685 5	1.467 2	1.835
1.738 5	2.400 5	2.728 5	2.345	1.691 8	1.694 8	1.574 8	1.677 1	1.435	1.841 7
1.743 7	2.414 5	2.726 5	2.343	1.695 3	1.690 5	1.578	1.694	1.411	1.823 7
1.747 8	2.413	2.704 5	2.374	1.697	1.708 5	1.576	1.695	1.427 7	1.823 6
1.743 7	2.394 5	2.723 5	2.393	1.687 5	1.704 5	1.580 9	1.693 5	1.437 8	1.842 3
1.736	2.422	2.773	2.388 5	1.688 3	1.722 5	1.583	1.697	1.427 1	1.856 2
1.747	2.419 5	2.781 5	2.396 5	1.685 5	1.696 8	1.578 8	1.692	1.421 9	1.842 6
1.741 5	2.433 5	2.768	2.339	1.693 5	1.687	1.576 7	1.685 7	1.429 5	1.814 5
1.738 5	2.445 5	2.778 5	2.342 5	1.696 6	1.687 5	1.573 5	1.695	1.429 1	1.820 5
1.739 7	2.460 5	2.773 5	2.328	1.689 5	1.693 3	1.558 7	1.692	1.441	1.797 1
1.735 3	2.478	2.755 5	2.307	1.677 2	1.679	1.534 7	1.693	1.433 2	1.810 4
1.738 6	2.454	2.743	2.313	1.669	1.681 5	1.514 8	1.690 5	1.427	1.797 5
1.734	2.477	2.746	2.313 5	1.668	1.688	1.525	1.705	1.426	1.805
1.758 7	2.473 5	2.76	2.249 5	1.665 2	1.681 5	1.510 7	1.706	1.418 8	1.833
1.776 5	2.487	2.777	2.218	1.662	1.686	1.517	1.711 7	1.425 5	1.814 8
1.780 5	2.493 3	2.755 5	2.218	1.663 5	1.669 7	1.519 8	1.716	1.421 8	1.818
1.793 2	2.462 5	2.742	2.206 5	1.665 8	1.658	1.52	1.703 5	1.430 8	1.800 1
1.785	2.459 5	2.758 5	2.202 5	1.671 5	1.655	1.528 3	1.700 5	1.420 6	1.808 4
1.776 2	2.469 5	2.743	2.196	1.691 3	1.673	1.547	1.702 2	1.414 5	1.812
1.768 7	2.477	2.728	2.201	1.696 5	1.675	1.518 8	1.711 6	1.422	1.797 8
1.782 3	2.489 7	2.732	2.176	1.691 3	1.67	1.520 7	1.708 6	1.411 3	1.781 5
1.783 2	2.511	2.736	2.164 5	1.693 7	1.677	1.515 2	1.717 5	1.398 8	1.771 5
1.782 8	2.514	2.731 8	2.176	1.691 7	1.672 3	1.546 5	1.717 5	1.385 5	1.760 4

1.780 2	2.493 5	2.684 5	2.165	1.679 5	1.680 8	1.583	1.722	1.386 9	1.768
1.782 4	2.477 5	2.666	2.205	1.676 3	1.679 5	1.572 5	1.721 5	1.393 5	1.772 1
1.786	2.497	2.686	2.215	1.660 6	1.675	1.591	1.706 5	1.402 7	1.778
1.786 5	2.495 8	2.678	2.187 5	1.663 2	1.666	1.621 5	1.705 8	1.400 8	1.771 7
1.798 2	2.492 5	2.695	2.207 8	1.663 7	1.670 3	1.618	1.705 5	1.404	1.795 5
1.8	2.476 5	2.697 5	2.196 5	1.655 8	1.678	1.607 5	1.706	1.407 4	1.795 6
1.796 9	2.474	2.714 5	2.18	1.656	1.691 5	1.584 3	1.698	1.415	1.775 9
1.795 7	2.451	2.725	2.178	1.654	1.685	1.587	1.705	1.418	1.768 6
1.807 2	2.430 5	2.715 5	2.174	1.669 5	1.696 5	1.595 5	1.714 2	1.415 1	1.761 9
1.804 7	2.432 8	2.721	2.198	1.674	1.715	1.586	1.716 3	1.414	1.758 4
1.802	2.396	2.741	2.184 5	1.68	1.72	1.616 5	1.708 5	1.417 6	1.767 1
1.800 2	2.409	2.756	2.197	1.672 7	1.704	1.595 5	1.705	1.417 5	1.775 4
1.787 3	2.427 5	2.758 5	2.211	1.686 5	1.701	1.590 3	1.713 6	1.412	1.772 1
1.792	2.45	2.779	2.224	1.688	1.707	1.622	1.709	1.412 5	1.759
1.778	2.46	2.783	2.241	1.693 5	1.702	1.607 5	1.701 5	1.418 6	1.762 9
1.776	2.466	2.782 5	2.242 5	1.663	1.705	1.597 5	1.694	1.414	1.753 9
1.776 9	2.434	2.803 5	2.279	1.661	1.711	1.592 5	1.697 5	1.402 5	1.753 3
1.782	2.442	2.786 5	2.276	1.657 5	1.716 5	1.589 5	1.700 5	1.41	1.743 6
1.777	2.486	2.796 5	2.274	1.663	1.713 3	1.577	1.701	1.402	1.753 4
1.78	2.486	2.783 5	2.279 5	1.664	1.702 9	1.56	1.718	1.41	1.754 5
1.778 7	2.496 5	2.781 5	2.305 5	1.664 2	1.695	1.571 5	1.735	1.409 8	1.749 5
1.777 5	2.517 7	2.791 5	2.325	1.672 5	1.682 8	1.592	1.739 5	1.405 8	1.744 7
1.78	2.516	2.820 5	2.324 5	1.677 5	1.696 5	1.606	1.742	1.418 3	1.773 4
1.783	2.523	2.829 2	2.273 5	1.675 2	1.708 5	1.623	1.738	1.437	1.772 4
1.784 2	2.518	2.838 5	2.279	1.675	1.705	1.628 5	1.741	1.431 5	1.789 2
1.781 3	2.51	2.840 5	2.238	1.668 7	1.710 5	1.649	1.742 5	1.435 7	1.783 5
1.791 9	2.52	2.824	2.23	1.674 6	1.711 5	1.646 5	1.731 7	1.446 8	1.773 2
1.794 2	2.507	2.841	2.236	1.678 6	1.711 6	1.647	1.734 3	1.444	1.774 1
1.806 9	2.48	2.849	2.22	1.677 2	1.710 8	1.650 2	1.739 5	1.438 5	1.762 5
1.811 5	2.486	2.832	2.213 5	1.679 9	1.698 3	1.649	1.735	1.437	1.729 1
1.795	2.453	2.822	2.21 ₁	1.679 2	1.688 3	1.640 5	1.750 8	1.444 9	1.736 6
1.798 3	2.455	2.848 7	2.208 5	1.680 5	1.704	1.653	1.751	1.443 5	1.726
1.802	2.418	2.863	2.201	1.682	1.701	1.644	1.746 5	1.446 5	1.722
1.813 5	2.427 5	2.844 5	2.240 5	1.678 7	1.700 2	1.637 3	1.746 5	1.444	1.737 3
1.812 7	2.438 5	2.859 5	2.240 5	1.677 5	1.699 5	1.641	1.736 8	1.448 8	1.723 7
1.808 5	2.471	2.876 5	2.233	1.679 2	1.696	1.656 5	1.747	1.450 5	1.721 8
1.812 1	2.507	2.868	2.246	1.683 2	1.692 7	1.672 3	1.750 5	1.440 7	1.72
1.811 3	2.497	2.852	2.249	1.689	1.691	1.672 2	1.753 8	1.442 5	1.708
1.799	2.491	2.846 5	2.222	1.704 3	1.675 3	1.67	1.746 8	1.431	1.712

1.804 8	2.473 5	2.885 5	2.216	1.705 4	1.673	1.664 5	1.745 3	1.438 5	1.723 6
1.803 7	2.476	2.894	2.221	1.700 6	1.672 8	1.672	1.733	1.441 9	1.720 8
1.805	2.473 5	2.906 3	2.204	1.703 4	1.677 5	1.664 8	1.739 9	1.439 9	1.732
1.804 2	2.488	2.918	2.201	1.701 5	1.678 2	1.673	1.731 5	1.438 6	1.733 8
1.817 4	2.481	2.897 5	2.176	1.702 5	1.672 5	1.666 7	1.733 5	1.433 5	1.725 2
1.825 8	2.505	2.860 5	2.189	1.705 8	1.677	1.665 8	1.737	1.431	1.735
1.828	2.518 5	2.885 5	2.173 5	1.707 9	1.689	1.652 5	1.751 5	1.442	1.733 8
1.844	2.513	2.915	2.183	1.718 5	1.697 5	1.649	1.760 5	1.434 5	1.746 6
1.865	2.514 5	2.905 5	2.181	1.729	1.694 5	1.67	1.762 7	1.435 4	1.737
1.861 7	2.513	2.9	2.187	1.725 8	1.683 4	1.677 7	1.758	1.441	1.744 4
1.863 7	2.501	2.897 5	2.175	1.724 8	1.682 5	1.669 5	1.754	1.454 2	1.760 5
1.862 7	2.503	2.915 2	2.189	1.727 2	1.675 6	1.664 5	1.753 1	1.438 6	1.765
1.866 8	2.504 5	2.888	2.180 5	1.717	1.679	1.651	1.743 1	1.439 6	1.777
1.879	2.496 8	2.878 5	2.166	1.715 4	1.681	1.662	1.727 5	1.442 2	1.775 3
1.893 5	2.509	2.857 2	2.15	1.718 5	1.682 6	1.644 5	1.724 2	1.441 1	1.770 1
1.884 8	2.513 5	2.861 1	2.147 5	1.72	1.682 5	1.649 3	1.730 9	1.44	1.770 6
1.898	2.540 5	2.87	2.131	1.719 5	1.672	1.646 5	1.718 5	1.440 5	1.782 4
1.905 8	2.543	2.885 3	2.121	1.725 3	1.655 5	1.651	1.725 5	1.459 3	1.786 8
1.916	2.527 8	2.88	2.137 5	1.747 5	1.658 4	1.645 5	1.729 2	1.466	1.791 6
1.943 3	2.529	2.865	2.133	1.753 2	1.637 5	1.624	1.721	1.469 5	1.785 4
1.945 8	2.525 5	2.874	2.138	1.754	1.636 5	1.621 5	1.709	1.481 5	1.770 2
1.958	2.554	2.883 5	2.146	1.753 2	1.630 5	1.624 5	1.703 8	1.477	1.772 8
1.901 5	2.552	2.879 5	2.114	1.760 3	1.644 5	1.626	1.712	1.481 5	1.772 2
1.900 5	2.544 5	2.877	2.111 5	1.757 8	1.647 1	1.621	1.703	1.485	1.783 6
1.893	2.522 5	2.886 5	2.114	1.778 5	1.643	1.637 5	1.707 5	1.487 5	1.769 7
1.905	2.527	2.889 5	2.093 5	1.786	1.647 5	1.655 2	1.718	1.490 9	1.775 3
1.931	2.489	2.93	2.084	1.815	1.648	1.645 1	1.717 5	1.484 5	1.768 7
1.919	2.510 5	2.954	2.084	1.818 5	1.663 7	1.661 8	1.716 5	1.489	1.780 2
1.902 5	2.517 2	2.947 5	2.09	1.809	1.669 7	1.666	1.708 5	1.489 5	1.774 3
1.914 8	2.517	2.987	2.085	1.829 5	1.672 5	1.669	1.687 5	1.493	1.773 2
1.917	2.514 5	2.992	2.067	1.817	1.678	1.674 2	1.676 7	1.486	1.774 9
1.917 3	2.516	3.008	2.07	1.820 3	1.682 5	1.67	1.690 9	1.466 3	1.787 5
1.911	2.525 3	2.998	2.061	1.823 8	1.672 1	1.663 3	1.694 3	1.473 2	1.792
1.921 3	2.529 5	3.04	2.078 3	1.821 8	1.682	1.662	1.688 5	1.478 9	1.799 1
1.926 9	2.528 5	3.033 5	2.070 5	1.816 5	1.699	1.651 5	1.684 5	1.475 7	1.803
1.940 5	2.554 5	3.073	2.062 5	1.838	1.693 5	1.651	1.690 8	1.479	1.822 3
1.948	2.548	3.093 5	2.063	1.846 3	1.693 3	1.655 5	1.697 8	1.472 8	1.821
1.939 5	2.543	3.096	2.071	1.837 7	1.691 5	1.657 4	1.687 2	1.477 9	1.823 5
1.943 5	2.561	3.128	2.058 5	1.847	1.687 5	1.652 5	1.685 8	1.467 4	1.820 3

1.961	2.56	3.065	2.048 5	1.847 5	1.691 5	1.641	1.668	1.47	1.818 3
1.963	2.552 5	3.027	2.042 5	1.876 5	1.696 8	1.643 5	1.667	1.46	1.817 7
1.983	2.557 5	3.074 5	2.044 5	1.881 5	1.691 7	1.641 5	1.672 5	1.453	1.820 3
2.007 5	2.576	3.042	2.042 5	1.864 3	1.692 5	1.632	1.675	1.453 9	1.822 6
2.018 5	2.579	3.035	2.05	1.850 7	1.692 5	1.636 2	1.674	1.450 6	1.831
2.007	2.596	3.072	2.045 5	1.838	1.695	1.647	1.668	1.445	1.831
1.99	2.581 5	3.061 5	2.05	1.826 3	1.694	1.641 5	1.695	1.448 3	1.843 1
2.018 5	2.576 8	3.042	2.035 5	1.845 2	1.677 3	1.622 2	1.697 5	1.451 1	1.827 5
1.978 5	2.591	3.016	2.032	1.860 7	1.669	1.621 8	1.711	1.465 6	1.803
1.993	2.586 5	3.027 5	2.026 5	1.849	1.678 5	1.612 6	1.715 2	1.470 3	1.780 8
1.973 5	2.571 8	3.05	2.029 8	1.865	1.680 5	1.615	1.718 5	1.478 5	1.794 4
1.957	2.573 5	3.088	2.048	1.875	1.674 5	1.595	1.711	1.476 5	1.789 8
1.952 5	2.568 5	3.078	2.072	1.876 7	1.678	1.590 5	1.711 5	1.475 2	1.807
1.943	2.543	3.095 5	2.06	1.876 3	1.671 7	1.601 3	1.719 8	1.476 2	1.821 2
1.942 5	2.545 5	3.119	2.063 8	1.875 5	1.663 6	1.614 5	1.711 5	1.479 3	1.831 5
1.955	2.549 2	3.135	2.092	1.880 5	1.668 7	1.618 8	1.709	1.487 7	1.821
1.952	2.527	3.14	2.061	1.893 2	1.665 8	1.608 3	1.711 7	1.480 5	1.817 4
1.973 5	2.506 5	3.121	2.051 5	1.902 2	1.657 5	1.633 8	1.719	1.477 9	1.801
1.975	2.484 5	3.119	2.049	1.916	1.651 9	1.626 9	1.700 8	1.474 8	1.788
1.939	2.465	3.072	2.028 5	1.916	1.649 3	1.606 5	1.695	1.474 2	1.803 8
1.935	2.466 5	3.073 5	1.989 8	1.899 5	1.654	1.610 5	1.69	1.474 2	1.815 8
1.947 5	2.448	3.040 5	1.992	1.880 5	1.648 8	1.607	1.693 5	1.476 5	1.811 5
1.963 2	2.426	3.023 2	2.031 5	1.881	1.646 2	1.604 4	1.678 8	1.475 7	1.818 9
1.978 5	2.428 5	3.011	2.051 5	1.871 5	1.651	1.6	1.678 5	1.473 3	1.809
1.985 2	2.425	3.032	2.048	1.913	1.657 5	1.588 4	1.670 5	1.477 6	1.823 7
2.000 5	2.433 5	3.072	2.046	1.898	1.643	1.589 8	1.667 1	1.476 5	1.826 6
1.992 5	2.459	3.044 5	2.046	1.891 5	1.653 3	1.590 5	1.661	1.478 7	1.82
2.000 5	2.452 5	3.030 5	2.025 5	1.919 7	1.646	1.593 7	1.647 3	1.476 6	1.823 7
2.007	2.478	2.989	2.027 2	1.906 5	1.643 8	1.579 3	1.640 5	1.486 8	1.823 1
2.009	2.439	2.948	2.021 5	1.891 5	1.644	1.577	1.659	1.476	1.791 7
2	2.451	2.929 5	2.006 5	1.863 5	1.641 5	1.571	1.666 6	1.476 6	1.800 4
1.993 5	2.416 5	2.938 5	1.999 5	1.857	1.628	1.567 5	1.663	1.482 5	1.811 6
2.024 5	2.420 5	2.94	1.990 5	1.868 7	1.618 5	1.577 5	1.654	1.481 4	1.811 6
2.035	2.417 5	2.947 5	2.001 5	1.868	1.619	1.565	1.671 5	1.479 5	1.814 1
2.051 8	2.407	2.943	1.997	1.878 7	1.622 8	1.571	1.671 5	1.480 7	1.812 7
2.071	2.405	2.973	2.004	1.874	1.615 3	1.566	1.668 2	1.484 3	1.809
2.085	2.395 5	2.961 5	1.987	1.859 2	1.609 2	1.565 2	1.67	1.477 2	1.815 1
2.094	2.39	2.972 5	1.975	1.842 8	1.587 5	1.551 7	1.673 9	1.492 7	1.826
2.132	2.381 5	2.994 5	1.974	1.843 7	1.592 5	1.542 2	1.665 5	1.500 7	1.835 7

2.121	2.370 5	3.008 5	1.975 5	1.848 5	1.595	1.538 5	1.657	1.503	1.829
2.135	2.373	3.018 5	1.975	1.849	1.592 5	1.517	1.654 8	1.504	1.831 5
2.114	2.382	3.031	1.988	1.845 2	1.576 8	1.524	1.647	1.510 5	1.835 2
2.146	2.378	3.062	1.988 5	1.850 3	1.580 7	1.517 5	1.645 5	1.510 9	1.827 4
2.140 5	2.374	3.054 7	1.985 5	1.879	1.594 5	1.514	1.653 5	1.506 5	1.819 5
2.138	2.365 5	3.071 5	1.995	1.875 7	1.592 2	1.517	1.645	1.507 7	1.821 3
2.148	2.349	3.078 5	2.032	1.876 7	1.59	1.512 7	1.643 5	1.507	1.817 9
2.149 5	2.349 5	3.11	2.047 5	1.868 8	1.577	1.499 8	1.644	1.515 8	1.828 7
2.22	2.338 3	3.115	2.033 5	1.864 2	1.569 5	1.497	1.644 8	1.519 4	1.829 6
2.196	2.331 5	3.059 5	2.029 5	1.874 7	1.563	1.517 5	1.644 5	1.521 9	1.833 4
2.187 5	2.344 5	3.079	2.052 5	1.879 5	1.564 3	1.495 8	1.653 7	1.529 3	1.826 6
2.132	2.360 5	3.062	2.061	1.875 5	1.550 8	1.478 5	1.668 4	1.528 5	1.827 6
2.096	2.342	3.087	2.064	1.883 5	1.558	1.490 5	1.667 5	1.523	1.827 6
2.112	2.355	3.093	2.058	1.879 3	1.555 3	1.477 5	1.666 5	1.530 2	1.822 3
2.116	2.358 3	3.087	2.064	1.878	1.553 5	1.475 8	1.671	1.534 5	1.826 5
2.107	2.393	3.086 5	2.061 5	1.882 5	1.548 3	1.459 5	1.675	1.529 8	1.846 6
2.116	2.414	3.105	2.065 5	1.873	1.557 5	1.482	1.665 5	1.526 7	1.85
2.131	2.422 5	3.091	2.046	1.861	1.551 5	1.489 2	1.647	1.526 7	1.854 2
2.152	2.442	3.09	2.022 5	1.861 2	1.552 5	1.488 7	1.643 5	1.521 7	1.851 7
2.152 5	2.464 5	3.087	2.011	1.863	1.557 5	1.483 5	1.635 7	1.516 2	1.851 5
2.138	2.421 5	3.1	2.002 5	1.864 9	1.566	1.501 8	1.631 7	1.516 5	1.842 5
2.138 5	2.426	3.129 5	2.013 5	1.861 8	1.576	1.483 1	1.611 5	1.527 5	1.840 8
2.136	2.426 5	3.116	2.013	1.852 3	1.583 8	1.474	1.601	1.532 3	1.820 5
2.108 5	2.432	3.124	2.008 3	1.836	1.563 5	1.483 7	1.593	1.535	1.820 2
2.107 5	2.465	3.126	2.009	1.821	1.558	1.481 8	1.606	1.532 4	1.823 8
2.116	2.464	3.152	2.0215	1.813	1.566	1.480 7	1.600 8	1.534 8	1.824 4
2.113	2.494 5	3.144	2.019 5	1.810 5	1.582	1.477 8	1.586 5	1.5279	1.803 5
2.109 5	2.466	3.155	1.993 5	1.804 4	1.591 9	1.475 3	1.578	1.534 3	1.798 5
2.091 8	2.476	3.173	1.991 5	1.811	1.587 5	1.477 5	1.581	1.541 5	1.807
2.088	2.450 5	3.162 5	1.973	1.802 5	1.574 8	1.482	1.583 7	1.539 5	1.808 8
2.055 5	2.443	3.168	1.965 5	1.780 5	1.561 8	1.465	1.585	1.540 5	1.799 2
2.05	2.431	3.177 5	1.972 7	1.788 5	1.558	1.467 7	1.592 5	1.544	1.788 7
2.068	2.410 5	3.154 5	1.972 5	1.788	1.548 2	1.471	1.583 5	1.547 7	1.792
2.087 3	2.406	3.163	1.992 5	1.774 5	1.569	1.464 8	1.569 2	1.544 5	1.801
2.082 3	2.413	3.140 5	2.001	1.775 8	1.58	1.459	1.574	1.532 5	1.791 8
2.085 5	2.41	3.155	2.01	1.777 5	1.577 8	1.467 7	1.566	1.526 5	1.787 5
2.113	2.388	3.193	2.019 5	1.782	1.556 8	1.463 2	1.535 8	1.528 5	1.794 2
2.125	2.397	3.185 5	2.012	1.786	1.564	1.458 3	1.521 5	1.532 7	1.794 8
2.097 5	2.397 2	3.19	2.016 5	1.778 8	1.572	1.453 4	1.535 5	1.531 3	1.794 4

2.110 5	2.409 5	3.183	2.017	1.771 8	1.570 5	1.448 3	1.549	1.532 5	1.778 3
2.097 5	2.418	3.183 5	2.018	1.794 2	1.566 5	1.437 5	1.556 5	1.534 4	1.780 2
2.111 5	2.408 5	3.171	2.019	1.784	1.551 7	1.403 5	1.546	1.534 3	1.772 1
2.127	2.414 5	3.182	2.016 3	1.789 5	1.555	1.399 5	1.565 8	1.536 9	1.765 6
2.156 5	2.434	3.158	2.009 5	1.761	1.544 5	1.407	1.562	1.535 6	1.763 6
2.119	2.442 5	3.166	2.006 5	1.757 5	1.534	1.411 3	1.575 2	1.533 2	1.770 7
2.133	2.431 5	3.167	1.979	1.745	1.547	1.406 7	1.592 3	1.524 7	1.779 6
2.144	2.425	3.176 5	1.966	1.745 5	1.526	1.401 8	1.590 7	1.516 8	1.774 5
2.153	2.402 5	3.169 5	1.968	1.724 3	1.528 3	1.394	1.587	1.515 8	1.778
2.172	2.388 5	3.163	1.963	1.730 7	1.532	1.390 7	1.575 5	1.521 3	1.781 7
2.165 5	2.406	3.18	1.942 5	1.735 7	1.517 7	1.412	1.582 2	1.522 9	1.786 1
2.179	2.406 5	3.214	1.946 5	1.727 5	1.523	1.404 5	1.587 5	1.531 9	1.787 1
2.187 5	2.393 7	3.226	1.923 5	1.720 7	1.516	1.391 7	1.577 5	1.530 4	1.783 8
2.19	2.399	3.223	1.918	1.717	1.510 5	1.411 5	1.579 3	1.529 9	1.770 4
2.184 5	2.384	3.245 5	1.931	1.715 5	1.510 8	1.411	1.581	1.527 7	1.759
2.181 5	2.388	3.257 5	1.920 2	1.733 5	1.503 5	1.452	1.581 5	1.522 3	1.759 3
2.163	2.381	3.275 5	1.928	1.727 3	1.517	1.486 2	1.580 3	1.521	1.768 4
2.175	2.387 5	3.302	1.921	1.736	1.517 5	1.487 2	1.580 3	1.524	1.781 6
2.177 5	2.396	3.29	1.918 5	1.732 5	1.508	1.518	1.582	1.525 5	1.781 7
2.181 5	2.417	3.266 5	1.891 5	1.731	1.513 2	1.481 5	1.577 2	1.522 7	1.784
2.186 5	2.418	3.315	1.876	1.724	1.516 5	1.502	1.577 5	1.532	1.779 4
2.213	2.424 5	3.337	1.836	1.732 6	1.516 8	1.481 5	1.558 5	1.528 8	1.780 8
2.210 5	2.421 8	3.358 2	1.842	1.76	1.518 8	1.490 5	1.551	1.524 5	1.772 5
2.203	2.417	3.388	1.845	1.741 5	1.515 7	1.501 5	1.550 7	1.526	1.764 8
2.244 6	2.434	3.452 5	1.838 5	1.733 5	1.508	1.485 5	1.555 7	1.521	1.774 5
2.266	2.431	3.435	1.847 5	1.742 5	1.502 3	1.485 5	1.547	1.521 5	1.779 9
2.258	2.427 5	3.335	1.814 5	1.747 3	1.487	1.452	1.54	1.520 2	1.780 8
2.274 5	2.431	3.344	1.826	1.737 7	1.487	1.425 5	1.530 5	1.489 5	1.796 7
2.255	2.427	3.354	1.823 5	1.748 5	1.481 5	1.414 7	1.539 1	1.484 8	1.804 3
2.284	2.418 5	3.375 5	1.805 5	1.766	1.488 2	1.426	1.546 2	1.485 5	1.805 1
2.295 5	2.418	3.432	1.788 5	1.762 5	1.488	1.409	1.547	1.489 2	1.814 2
2.296	2.412 5	3.385	1.787	1.776 5	1.48	1.405	1.567	1.483 5	1.797 8
2.311	2.416 5	3.397	1.833	1.77	1.471 5	1.43	1.578 9	1.49	1.787 4
2.309	2.427	3.415	1.807	1.774	1.477 5	1.442 5	1.577 5	1.49	1.792 7
2.304 5	2.422 5	3.338 5	1.798 5	1.776 5	1.470 8	1.472	1.579 5	1.479 9	1.785 6
2.309	2.424 7	3.346	1.814	1.777 8	1.474 9	1.486 8	1.574 2	1.483 3	1.794 4
2.298	2.441 5	3.357	1.840 5	1.787	1.477	1.472 2	1.557 2	1.480 4	1.803 1
2.323	2.442 5	3.385 5	1.856	1.788 5	1.481 7	1.463	1.544 1	1.479 2	1.801 4
2.315 5	2.437	3.378	1.817 5	1.773 5	1.486 2	1.449 7	1.548 7	1.472 5	1.796

2.333 9	2.459	3.341	1.798	1.764	1.492	1.477 3	1.558 4	1.473 5	1.812 7
2.346	2.449 5	3.276 5	1.817 3	1.787 9	1.48	1.484 5	1.539	1.478 2	1.814 3
2.324 5	2.461 2	3.258	1.825	1.795	1.484 5	1.52	1.540 5	1.483 2	1.803 3
2.33	2.455 5	3.21	1.83	1.814	1.496 7	1.508 5	1.547 5	1.483	1.817 4
2.335	2.447 2	3.227 5	1.814 5	1.828 8	1.5	1.510 5	1.539 5	1.485 2	1.822 9
2.371	2.444 9	3.237	1.828 3	1.835	1.513 5	1.531 7	1.547 7	1.483 8	1.818 8
2.360 5	2.450 7	3.209	1.835	1.828 7	1.503 1	1.539	1.534 5	1.478 5	1.811 6
2.414	2.448 5	3.127	1.828	1.825 2	1.502 5	1.535 6	1.552 1	1.477	1.815 2
2.411	2.464	3.130 5	1.830 5	1.834	1.489 7	1.539 5	1.555	1.476 3	1.818
2.405	2.466 5	3.083	1.834 5	1.850 5	1.478 5	1.541	1.548 1	1.487	1.830 5
2.403	2.461 5	3.118 5	1.829 5	1.871	1.473	1.543	1.547 3	1.485 6	1.818 9
2.372	2.458	3.153 5	1.823	1.866 8	1.480 5	1.567	1.544 1	1.493 8	1.8
2.404 5	2.447	3.122	1.827 5	1.843 8	1.478 2	1.563	1.555 8	1.487	1.802 3
2.395	2.443 5	3.178	1.832 4	1.837 8	1.480 8	1.568	1.549 3	1.488 7	1.800 8
2.346	2.441 5	3.175	1.834	1.842 5	1.485	1.578	1.549 5	1.483 2	1.790 4
2.347 5	2.435 2	3.164	1.838 5	1.843	1.486 8	1.594 5	1.547 8	1.491 3	1.783 9
2.348 7	2.443 5	3.141	1.832	1.844 5	1.484	1.588 9	1.550 8	1.484 8	1.779 7
2.372 2	2.437	3.135	1.839 8	1.863	1.478 5	1.597 2	1.554 7	1.479 6	1.79
2.377 5	2.440 5	3.073 5	1.854 5	1.859	1.502 5	1.588	1.550 8	1.479	1.788 3
2.363	2.443	3.06	1.857	1.878 5	1.526 5	1.566 2	1.542 5	1.476 4	1.792 4
2.355 5	2.462	3.027	1.871	1.864	1.545 5	1.592 5	1.544 5	1.481 5	1.778 7
2.37	2.461 7	3.028	1.861	1.869	1.532	1.599 5	1.545 3	1.480 1	1.784
2.392 5	2.464 5	3.037	1.853	1.879 5	1.53	1.596 8	1.548	1.484 4	1.773 3
2.388	2.478 5	2.98	1.837 5	1.878 7	1.497	1.579 6	1.542 5	1.482 6	1.773 5
2.393	2.477 5	2.979	1.839	1.867 8	1.497	1.593 3	1.540 5	1.484 7	1.776 3
2.394	2.485	3.02	1.837 1	1.872	1.49	1.602 4	1.52	1.491 5	1.779
2.413	2.494	3.08	1.834	1.859 6	1.492 7	1.602	1.500 5	1.494 3	1.784 4
2.411 8	2.489 7	3.119	1.834 5	1.861 5	1.507 5	1.593 5	1.501 5	1.508 4	1.780 2
2.415 5	2.494 7	3.15	1.820 7	1.87	1.533 3	1.598 5	1.502 8	1.509	1.772 4
2.448 8	2.514 5	3.132 5	1.822	1.849 5	1.531 3	1.593 5	1.501 5	1.514 7	1.768
2.443 8	2.524 5	3.11	1.826 5	1.844	1.514 1	1.582	1.498 8	1.512 3	1.779 8
2.476	2.548	3.098	1.828 5	1.842 3	1.532 2	1.571 8	1.497	1.510 7	1.781
2.465 5	2.540 5	3.149	1.823 5	1.837	1.533	1.582	1.492	1.515 5	1.781 2
2.438 5	2.543	3.177 5	1.803 3	1.842	1.544 5	1.597 5	1.495 5	1.509 7	1.775
2.423 3	2.565	3.219	1.802	1.836	1.543 8	1.568 2	1.497 4	1.514 6	1.785 4
2.447 9	2.568 5	3.222	1.821 2	1.821 5	1.544	1.556	1.509 7	1.514 7	1.803 9
2.408	2.567 7	3.168	1.814 5	1.824	1.513 5	1.570 5	1.504 3	1.511 3	1.792 8
2.410 5	2.565	3.193 5	1.826	1.817 5	1.504	1.581 5	1.496 5	1.513 5	1.799 1
2.426 5	2.556	3.112	1.825 7	1.823	1.495	1.576 5	1.504	1.512 4	1.798 4

2.446 8	2.553 5	3.121	1.826 5	1.836 5	1.488	1.569 5	1.524	1.523 3	1.797 9
2.47	2.565	3.072 5	1.839 5	1.841 5	1.485 7	1.567 5	1.521 5	1.524 8	1.796 9
2.440 5	2.567 5	3.044 5	1.824	1.840 5	1.49	1.555	1.513 7	1.527	1.794 7
2.446	2.555 5	3.077	1.806 5	1.852 8	1.489 5	1.561 5	1.507	1.526 3	1.798 5
2.438	2.558 5	3.09	1.811	1.853 5	1.487 5	1.566 5	1.528 7	1.527 9	1.806 6
2.435 5	2.545 5	3.068	1.803	1.858 5	1.492 5	1.567 2	1.533 3	1.528 7	1.798 3
2.441	2.549 5	3.036	1.806 5	1.862	1.478 5	1.587	1.540 5	1.530 2	1.771 9
2.457 5	2.521 7	3.067	1.811 5	1.862 8	1.467 5	1.594 5	1.545 3	1.531 1	1.758 4
2.457	2.524	3.079	1.815 5	1.867 5	1.464 7	1.599 2	1.551 5	1.527 2	1.750 2
2.471	2.526	3.083	1.813 3	1.858 2	1.462 5	1.607 7	1.545 8	1.528 6	1.750 7
2.527	2.532 5	3.094	1.817	1.871 5	1.451 2	1.613	1.555 3	1.529 9	1.727 9
2.535	2.54	3.113	1.824 5	1.872	1.451 5	1.614	1.559 5	1.531	1.735
2.509 5	2.544	3.079	1.813	1.876	1.457	1.619 7	1.555 1	1.540 4	1.725 8
2.521 5	2.541	3.081	1.791	1.875 5	1.445	1.637 5	1.551 8	1.538	1.725 5
2.531 5	2.541 3	3.059	1.783	1.873	1.453 3	1.624 8	1.559 6	1.543 5	1.691 6
2.570 5	2.568 5	3.043	1.796 3	1.869 3	1.457 4	1.634 5	1.564 7	1.541 7	1.688 9
2.559	2.567 6	3.055	1.792 5	1.871 5	1.466 3	1.636 5	1.566 5	1.535 7	1.702 4
2.499	2.577	3.074	1.798 5	1.873 5	1.476 5	1.638 5	1.568 5	1.534 5	1.689 5
2.516 7	2.578	3.057 5	1.782	1.882 8	1.493	1.63	1.574	1.522 7	1.692 3
2.527	2.585 7	3.08	1.774 2	1.893 7	1.497 3	1.632	1.576 5	1.521 7	1.685
2.497	2.588 2	3.102	1.768 9	1.891 5	1.491	1.633	1.572 3	1.520 1	1.689 3
2.498	2.584 8	3.088 5	1.773 5	1.897 5	1.505 5	1.619 5	1.574 5	1.517 8	1.680 5
2.498	2.578 2	3.089 5	1.778	1.896 5	1.5225	1.634	1.571 7	1.509	1.682 8
2.471	2.598 5	3.1	1.789	1.882 5	1.52	1.611 5	1.577	1.509 8	1.687 4
2.449	2.596 8	3.058	1.796 5	1.870 5	1.522	1.606 5	1.578 3	1.514 1	1.677 5
2.473 5	2.592	3.062	1.789 5	1.872	1.526	1.615 5	1.573 5	1.515	1.673 1
2.503 5	2.577	3.03	1.79	1.868 5	1.537 5	1.59	1.571	1.514 4	1.678 1
2.464 5	2.582	3.02	1.787	1.875	1.537	1.573 2	1.571	1.520 6	1.674 6
2.475 5	2.592	3.074 5	1.779	1.88	1.534	1.576 5	1.570 5	1.518 2	1.670 2
2.451	2.607	3.064	1.778 8	1.887	1.541 5	1.583 3	1.572 5	1.507 2	1.650 4
2.447 5	2.610 8	3.07	1.778	1.886 2	1.547 5	1.586	1.573 5	1.505 5	1.642 8
2.456	2.624 5	3.071	1.774	1.873 5	1.562 5	1.611	1.572 5	1.501 1	1.635 2
2.44	2.629 3	3.044	1.777 5	1.870 6	1.58	1.637 5	1.570 7	1.505 3	1.633 5
2.425 5	2.648 5	3.06	1.779	1.856	1.571 6	1.641 8	1.578	1.509 6	1.613 7
2.411	2.679	3.029	1.807 5	1.852 9	1.577 3	1.646	1.58	1.508 5	1.606
2.436 5	2.661	3.042	1.82	1.861 7	1.579	1.654	1.576 5	1.500 7	1.632 8
2.43	2.661 3	3.045 5	1.816 4	1.846 5	1.599 2	1.654 8	1.571 4	1.504 3	1.638 1
2.399	2.67	3.033 5	1.824	1.849 2	1.633	1.655 3	1.549 5	1.496 3	1.640 8
2.401	2.682	3.007	1.834	1.861 2	1.654	1.651 5	1.549 5	1.498 3	1.626 4

2.376	2.695 6	2.983	1.804	1.865 5	1.649 5	1.655 4	1.554 7	1.504 5	1.618 5
2.337 5	2.689 5	2.960 3	1.807 5	1.879	1.637 5	1.658	1.561 2	1.522 4	1.627 9
2.307	2.724 8	2.938	1.818 5	1.876 8	1.644 5	1.663	1.552 2	1.525 2	1.642 6
2.28	2.734 8	2.918 5	1.814	1.880 2	1.671	1.632 7	1.560 3	1.528 5	1.644 2
2.27	2.724	2.890 5	1.804 5	1.890 8	1.689 5	1.629	1.545 5	1.537 7	1.640 9
2.23	2.711	2.878 5	1.791	1.886 5	1.705 8	1.633 7	1.539 5	1.543 8	1.637 3
2.282	2.662	2.885 5	1.796 3	1.889 5	1.706	1.636	1.535 5	1.565 5	1.650 9
2.3	2.640 5	2.837	1.793 5	1.891 5	1.697 5	1.628 5	1.527	1.562 6	1.654 4
2.316 5	2.634 8	2.874	1.812	1.899 2	1.684 5	1.613 5	1.536	1.552 5	1.653 7
2.323	2.655 5	2.889	1.824 9	1.907 5	1.677 5	1.618 7	1.531 1	1.546 9	1.653 5
2.34	2.635	2.898	1.827	1.916	1.668	1.634	1.529 5	1.557 3	1.656 5
2.310 5	2.628 5	2.85	1.826 2	1.905	1.665	1.642 3	1.522 7	1.553 8	1.652 7
2.325	2.636	2.888 5	1.824 5	1.913 5	1.681	1.647	1.513 1	1.548 8	1.659 9
2.323	2.639 5	2.863 5	1.829 7	1.92	1.689	1.643	1.512	1.544 1	1.666 8
2.294	2.663	2.863	1.841	1.942	1.668 3	1.648	1.512 5	1.545 8	1.656 9
2.225	2.68	2.823	1.833 7	1.943 5	1.678 8	1.636	1.515 5	1.552 8	1.669
2.252 5	2.688	2.816	1.822 5	1.973	1.670 5	1.668	1.513 5	1.545 6	1.690 9
2.219	2.697	2.799 5	1.828 5	1.967	1.684	1.662 7	1.515 5	1.553 9	1.687 2
2.206	2.696	2.823	1.826 5	1.969 8	1.676 5	1.668 1	1.504 3	1.556 5	1.692 5
2.190 5	2.695 5	2.825	1.830 5	2.009	1.665 2	1.665 5	1.523 2	1.550 6	1.688 9
2.221 5	2.680 5	2.826	1.826 2	2.008 4	1.673	1.661 8	1.523 5	1.556 5	1.667
2.222 5	2.688	2.85	1.828 9	2.012	1.688 5	1.665 5	1.518 5	1.556	1.669 5
2.229 5	2.677	2.847	1.831 9	1.994 5	1.737	1.662	1.525 5	1.557	1.67
2.238	2.672 5	2.829	1.84	1.989 5	1.767	1.663 5	1.530 5	1.556	1.684 5
2.233	2.660 7	2.831 5	1.839 6	2.009 3	1.741 5	1.665	1.536 2	1.557 7	1.689 2
2.237	2.68	2.798	1.839 5	1.976	1.753 5	1.654	1.532 5	1.538 7	1.705 7
2.273	2.686 7	2.787	1.841 5	1.983 8	1.747 5	1.634	1.528 2	1.541 3	1.707 2
2.281 5	2.679 5	2.791	1.838 5	1.958	1.755	1.637 5	1.522 2	1.562 9	1.703 6
2.275	2.667	2.764	1.844 7	1.962 5	1.769	1.631 5	1.520 8	1.567 2	1.713 2
2.31	2.658	2.758	1.842 9	1.979 5	1.718 5	1.628 5	1.514	1.563 2	1.696 6
2.305 5	2.668 5	2.764	1.849 7	1.962 2	1.706	1.641 5	1.511 5	1.576 5	1.680 8
2.291	2.666 5	2.778 5	1.830 8	1.977 7	1.724 2	1.63	1.491 5	1.577 9	1.672
2.27	2.682 5	2.767	1.837 2	2.003 5	1.748	1.629 7	1.488 6	1.585 6	1.672 5
2.243 5	2.652 5	2.742 5	1.858 5	2.025	1.738 3	1.618	1.476 5	1.591	1.679 4
2.223	2.646	2.753	1.862 3	2.026 5	1.721	1.610 5	1.471	1.594 5	1.677 3
2.212	2.650 5	2.772 5	1.865 3	2.029 5	1.733 3	1.597	1.469 5	1.591 6	1.671 8
2.224 5	2.645 5	2.792	1.863 8	2.024 3	1.743 5	1.610 5	1.467 4	1.595 4	1.669 5
2.222	2.641	2.772	1.855	1.992	1.727	1.595 5	1.456 7	1.612 2	1.655
2.205 5	2.631 5	2.778 5	1.853 5	1.974	1.714	1.613 3	1.456 7	1.624	1.649 1

2.229	2.623	2.811	1.855 2	1.978	1.701 5	1.615 5	1.464 7	1.637	1.649 6
2.211 5	2.627 5	2.854	1.857 3	1.97	1.690 5	1.607	1.458 5	1.640 6	1.659 1
2.222	2.596	2.851	1.858	1.959 5	1.697 5	1.603 9	1.435 3	1.63	1.666 9
2.241	2.580 5	2.862	1.856	1.940 5	1.738	1.593	1.398 5	1.642 4	1.666 2
2.251 5	2.567 3	2.931 5	1.859	1.957 5	1.727 5	1.579 5	1.374	1.648 5	1.665 4
2.236	2.587 3	2.943 5	1.864 5	1.956 7	1.712 5	1.591	1.393 5	1.642 7	1.673 7
2.244	2.613 5	2.940 5	1.891 5	1.964 7	1.722	1.605	1.393 5	1.637 5	1.674 5
2.259	2.604 6	2.960 5	1.877 5	1.955 5	1.714	1.614 5	1.412 3	1.636 2	1.675
2.253	2.618	2.944	1.884	1.954 5	1.708 1	1.601	1.407 5	1.639 9	1.684

资料来源: <http://datamarket.com/data/set/237v/#!display=line&ds=237v!2jx6=6>.

表 A1—24

天然气炉输入 — 输出数据

输出序列 (在输出气体中 CO ₂ 的百分浓度)							
53.8	53.6	53.5	53.5	53.4	53.1	52.7	52.4
52.2	52.0	52.0	52.4	53.0	54.0	54.9	56.0
56.8	56.8	56.4	55.7	55.0	54.3	53.2	52.3
51.6	51.2	50.8	50.5	50.0	49.2	48.4	47.9
47.6	47.5	47.5	47.6	48.1	49.0	50.0	51.1
51.8	51.9	51.7	51.2	50.0	48.3	47.0	45.8
45.6	46.0	46.9	47.8	48.2	48.3	47.9	47.2
47.2	48.1	49.4	50.6	51.5	51.6	51.2	50.5
50.1	49.8	49.6	49.4	49.3	49.2	49.3	49.7
50.3	51.3	52.8	54.4	56.0	56.9	57.5	57.3
56.6	56.0	55.4	55.4	56.4	57.2	58.0	58.4
58.4	58.1	57.7	57.0	56.0	54.7	53.2	52.1
51.6	51.0	50.5	50.4	51.0	51.8	52.4	53.0
53.4	53.6	53.7	53.8	53.8	53.8	53.3	53.0
52.9	53.4	54.6	56.4	58.0	59.4	60.2	60.0
59.4	58.4	57.6	56.9	56.4	56.0	55.7	55.3
55.0	54.4	53.7	52.8	51.6	50.6	49.4	48.8
48.5	48.7	49.2	49.8	50.4	50.7	50.9	50.7
50.5	50.4	50.2	50.4	51.2	52.3	53.2	53.9
54.1	54.0	53.6	53.2	53.0	52.8	52.3	51.9
51.6	51.6	51.4	51.2	50.7	50.0	49.4	49.3
49.7	50.6	51.8	53.0	54.0	55.3	55.9	55.9
54.6	53.5	52.4	52.1	52.3	53.0	53.8	54.6
55.4	55.9	55.9	55.2	54.4	53.7	53.6	53.6
53.2	52.5	52.0	51.4	51.0	50.9	52.4	53.5
55.6	58.0	59.5	60.0	60.4	60.5	60.2	59.7
59.0	57.6	56.4	55.2	54.5	54.1	54.1	54.4

55.5	56.2	57.0	57.3	57.4	57.0	56.4	55.9
55.5	55.3	55.2	55.4	56.0	56.5	57.1	57.3
56.8	55.6	55.0	54.1	54.3	55.3	56.4	57.2
57.8	58.3	58.6	58.8	58.8	58.6	58.0	57.4
57.0	56.4	56.3	56.4	56.4	56.0	55.2	54.0
53.0	52.0	51.6	51.6	51.1	50.4	50.0	50.0
52.0	54.0	55.1	54.5	52.8	51.4	50.8	51.2
52.0	52.8	53.8	54.5	54.9	54.9	54.8	54.4
53.7	53.3	52.8	52.6	52.6	53.0	54.3	56.0
57.0	58.0	58.6	58.5	58.3	57.8	57.3	57.0
输入序列 (输入天然气速率 (ft. / min.))							
-0.109	0	0.178	0.339	0.373	0.441	0.461	0.348
0.127	-0.180	-0.588	-1.055	-1.421	-1.520	-1.302	-0.814
-0.475	-0.193	0.088	0.435	0.771	0.866	0.875	0.891
0.987	1.263	1.775	1.976	1.934	1.866	1.832	1.767
1.608	1.265	0.790	0.360	0.115	0.088	0.331	0.645
0.960	1.409	2.670	2.834	2.812	2.483	1.929	1.485
1.214	1.239	1.608	1.905	2.023	1.815	0.535	0.122
0.009	0.164	0.671	1.019	1.146	1.155	1.112	1.121
1.223	1.257	1.157	0.913	0.620	0.255	-0.280	-1.080
-1.551	-1.799	-1.825	-1.456	-0.944	-0.570	-0.431	-0.577
-0.960	-1.616	-1.875	-1.891	-1.746	-1.474	-1.201	-0.927
-0.524	0.040	0.788	0.943	0.930	1.006	1.137	1.198
1.054	0.595	-0.080	-0.314	-0.288	-0.153	-0.109	-0.187
-0.255	-0.229	-0.007	0.254	0.330	0.102	-0.423	-1.139
-2.275	-2.594	-2.716	-2.510	-1.790	-1.346	-1.081	-0.910
-0.876	-0.885	-0.800	-0.544	-0.416	-0.271	0	0.403
0.841	1.285	1.607	1.746	1.683	1.485	0.993	0.648
0.577	0.577	0.632	0.747	0.900	0.993	0.968	0.790
0.399	-0.161	-0.553	-0.603	-0.424	-0.194	-0.049	0.060
0.161	0.301	0.517	0.566	0.560	0.573	0.592	0.671
0.933	1.337	1.460	1.353	0.772	0.218	-0.237	-0.714
-1.099	-1.269	-1.175	-0.676	0.033	0.556	0.643	0.484
0.109	-0.310	-0.697	-1.047	-1.218	-1.183	-0.873	-0.336
0.063	0.084	0	0.001	0.209	0.556	0.782	0.858
0.918	0.862	0.416	-0.336	-0.959	-1.813	-2.378	-2.499
-2.473	-2.330	-2.053	-1.739	-1.261	-0.569	-0.137	-0.024
-0.050	-0.135	-0.276	-0.534	-0.871	-1.243	-1.439	-1.422
-1.175	-0.813	-0.634	-0.582	-0.625	-0.713	-0.848	-1.039
-1.346	-1.628	-1.619	-1.149	-0.488	-0.160	-0.007	-0.092
-0.620	-1.086	-1.525	-1.858	-2.029	-2.024	-1.961	-1.952

-1.794	-1.302	-1.030	-0.918	-0.798	-0.867	-1.047	-1.123
-0.876	-0.395	0.185	0.662	0.709	0.605	0.501	0.603
0.943	1.223	1.249	0.824	0.102	0.025	0.382	0.922
1.032	0.866	0.527	0.093	-0.458	-0.748	-0.947	-1.029
-0.928	-0.645	-0.424	-0.276	-0.158	-0.033	0.102	0.251
0.280	0	-0.493	-0.759	-0.824	-0.740	-0.528	-0.204
0.034	0.204	0.253	0.195	0.131	0.017	-0.182	-0.262

资料来源: Cox&Jenkins, Time Series Analysis: Forecasting and Control 数据集 J, 1976.

表 A1—25 1978—2002 年中国农村居民家庭人均纯收入序列 $\{x_t\}$,
生活消费支出序列 $\{y_t\}$ 及纯收入对数序列 $\{\ln x_t\}$, 生活消费支出对数序列 $\{\ln y_t\}$

年份	纯收入		生活消费支出	
	x_t	$\ln x_t$	y_t	$\ln y_t$
1978	133.6	4.894 85	116.1	4.754 45
1979	160.7	5.079 54	134.5	4.901 56
1980	191.3	5.253 84	162.2	5.088 83
1981	223.4	5.408 96	190.8	5.251 23
1982	270.1	5.598 79	220.2	5.394 54
1983	309.8	5.735 93	248.3	5.514 64
1984	355.3	5.872 96	273.8	5.612 40
1985	397.6	5.985 45	317.4	5.760 16
1986	423.8	6.049 26	357.0	5.877 74
1987	462.6	6.136 86	398.3	5.987 21
1988	544.9	6.300 60	476.7	6.166 89
1989	601.5	6.399 43	535.4	6.283 01
1990	686.3	6.531 31	584.6	6.370 93
1991	708.6	6.563 29	619.8	6.429 40
1992	784.0	6.664 41	659.8	6.491 94
1993	921.6	6.826 11	769.7	6.646 00
1994	1 221.0	7.107 43	1 016.8	6.924 42
1995	1 577.7	7.363 72	1 310.4	7.178 09
1996	1 926.1	7.563 25	1 572.1	7.360 17
1997	2 090.1	7.644 97	1 617.2	7.388 45
1998	2 162.0	7.678 79	1 590.3	7.371 68
1999	2 210.3	7.700 88	1 577.4	7.363 53
2000	2 253.4	7.720 20	1 670.1	7.420 64
2001	2 366.4	7.769 13	1 741.0	7.462 21
2002	2 476.0	7.814 40	1 834.0	7.514 25