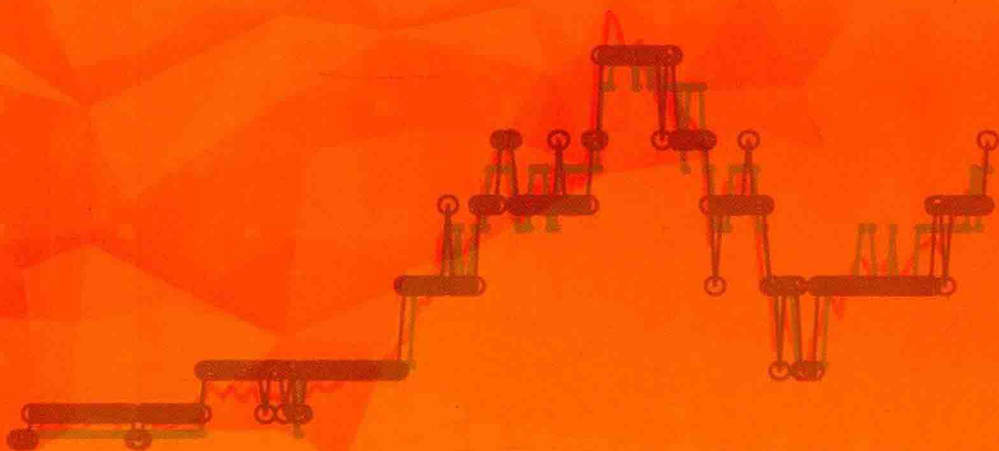




沈阳师范大学学术文库系列丛书

时间序列模型及预测

王立柱 著



科学出版社



沈阳师范大学学术文库系列丛书

时间序列模型及预测

王立柱 著

科学出版社

北京

内 容 简 介

本书介绍时间序列分析的主要理论和方法,既涵盖时间序列分析的经典理论内容,又反映 21 世纪以来时间序列分析理论的一些新进展。本书主要包括经典的时间序列分析理论及相关内容、基于计算智能的时间序列模型及预测、模糊时间序列模型及相关理论三大部分。本书在借鉴国内外相关教材优点的基础上,总结作者多年从事时间序列分析方向研究和课程教学的心得和体会;既重视概念、理论和方法的严密性、准确性和前瞻性,同时又注重从实际需要出发,兼顾对运用这些理论方法分析研究乃至最终解决实际经济、金融、管理类问题的能力。

本书可作为应用统计类、金融工程类本科专业的教材或参考书以及经济统计类相关专业研究生的教材,也可供自学时间序列分析的读者阅读。

图书在版编目(CIP)数据

时间序列模型及预测/王立柱著. —北京:科学出版社, 2017.4
(沈阳师范大学学术文库系列丛书)

ISBN 978-7-03-052528-4

I. ①时… II. ①王… III. ①时间序列分析 IV. ①O211.61

中国版本图书馆 CIP 数据核字(2017) 第 062684 号

责任编辑:王丽平/责任校对:邹慧卿

责任印制:张 伟/封面设计:陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018 年 1 月第 一 版 开本:720×1000 B5

2018 年 1 月第一次印刷 印张:10 1/4

字数:204 000

定价 68.00 元

(如有印装质量问题,我社负责调换)

前 言

随着我国经济的不断发展, 经济和金融活动更加频繁. 人们常常会遇到大量的时间序列数据. 例如, 经济领域中, 国内生产总值、居民消费价格指数、人均投资、人均收入、人均支出等指标可以形成时间序列数据; 金融领域中, 股票、基金、期货、黄金及其他金融衍生品的成交价格和成交量也可以构成时间序列. 无论是规避风险, 还是巨大的商业利益都促使人们分析并发现时间序列数据背后的内在运行规律, 进而对其发展趋势进行科学的预测. 因此, 学习时间序列模型及预测的人越来越多. 虽然现在有关时间序列模型及预测的书籍并不少, 但找到一本系统的、具有前瞻性的、便于学习的时间序列模型及预测的书却并不容易. 一方面, 这类书籍大部分都侧重于理论讲授, 缺乏实战能力的培养. 读者从这一类的书中学到的是一些理论知识, 对于解决实际的时间序列分析预测问题束手无策. 另一方面, 虽然有一些书中实例较多, 但缺乏理论和方法的严密性, 以及叙述的逻辑性.

为了对时间序列理论及应用做深入的研究, 把握前沿理论及发展动态, 我们决定编写一本理论联系实际、能够把握前沿动态的、适于科学研究及教材的时间序列模型及预测的书.

本书分 7 章.

第 1 章简要地介绍什么是时间序列、时间序列分析的产生与发展及时间序列分析软件.

第 2 章介绍随机过程的基本理论和相关概念, 包括随机变量的一些常用数字统计特征, 随机过程的平稳性、正态性及独立性检验及动态数据预处理方法.

第 3 章从经典的时间序列模型及预测、神经网络的时间序列预测模型、模糊时间序列模型及预测三个方面来介绍分析时间序列模型及预测的相关工作.

第 4 章总结模糊时间序列模型及预测的基本框架; 模糊时间序列预测是建立在模糊理论和模糊逻辑概念基础上的一种时间序列预测方法.

第 5 章是基于信息粒度的模糊时间序列模型及预测, 并结合两个案例进行详细说明.

第 6 章是包含时域信息的模糊时间序列模型及预测, 并给出了实验供读者参考.

第 7 章介绍 Granger 相关性的相关知识, 包括相关性检验, 给出基于 Granger 相关性的时间序列模型及预测, 并通过实验说明建模的方法和结果分析比较.

为了使正文中的章节主题突出, 一些常用统计表以附录的形式给出.

为了使本书易读易懂易学, 我们力求做到: 重要的概念和方法用简单明了的例子说明; 注意前后的逻辑关系; 重视概念、理论和方法的前瞻性、严密性、准确性。

本书初稿形成后, 作者为统计学的硕士研究生和本科生各讲授了一遍。同学们提出了很多宝贵的建议, 作者在此基础上进行了修改, 然后定稿。

本书介绍了一些时间序列分析及预测的理论、方法和技术, 可作为培养应用型人才的高等学校相关专业的硕士研究生和本本科生的教材或参考书, 也可供从事金融活动的管理人员及时间序列爱好者阅读。

在编写过程中, 作者深受书后所列参考文献的启发, 在此对文献的作者表示诚挚的感谢。

本书得到了沈阳师范大学重大孵化项目 (基于计算智能技术的时间序列模型及预测研究, ZD201617) 资助, 是辽宁省自然科学基金指导计划项目 (模糊系统理论在股指技术面分析中的应用, 20170540821) 的部分成果。

由于作者的水平有限, 难免有疏漏和不妥之处, 恳请读者批评指正。

王立柱

2017 年 2 月 5 日于沈阳

目 录

第 1 章 时间序列分析概述	1
1.1 关于时间序列分析	1
1.1.1 引例	1
1.1.2 时间序列分析及预测的产生与发展	7
1.2 时间序列分析的基本步骤	9
1.2.1 特征分析	9
1.2.2 模型识别	10
1.2.3 模型参数估计	11
1.2.4 模型检验	11
1.2.5 模型应用	12
1.3 时间序列分析软件	12
本章小结	13
思考题	14
第 2 章 时间序列分析的基本概论	15
2.1 随机过程	15
2.1.1 随机过程的相关概念	15
2.1.2 随机过程的分布及其特征	16
2.1.3 几种重要的随机过程	18
2.2 平稳随机过程	18
2.3 时间序列的平稳性与非平稳性	20
2.4 时间序列数据的预处理	21
2.4.1 平稳性检验	21
2.4.2 正态性检验	24
2.4.3 独立性检验	26
2.4.4 异常点的检验与处理	27
本章小结	27
思考题	28
第 3 章 时间序列模型及预测方法	29
3.1 几种典型的时间序列模型	29
3.1.1 向量自回归模型	29

3.1.2	滑动平均模型	32
3.1.3	向量自回归滑动平均模型	33
3.1.4	向量误差修正模型	35
3.1.5	自回归条件异方差模型	37
3.2	最小均方误差预测	39
3.3	区间预测	41
3.4	基于计算智能技术的时间序列模型及预测方法	42
3.4.1	人工神经网络模型及预测方法	43
3.4.2	贝叶斯网络模型及预测方法	46
3.4.3	支持向量机预测方法	47
3.5	平稳时间序列模型的建立	48
3.5.1	模型识别与定阶	48
3.5.2	模型参数估计	54
	本章小结	58
	思考题	59
第 4 章	模糊时间序列模型及预测方法	61
4.1	引言	61
4.2	模糊集合的基础知识	62
4.2.1	模糊集合的概念	62
4.2.2	模糊集合的运算	66
4.2.3	λ 水平截集及分解定理	68
4.3	模糊时间序列预测模型	72
4.3.1	模糊时间序列模型简介	72
4.3.2	模糊时间序列模型及其基本框架	74
	本章小结	80
	思考题	80
第 5 章	基于信息粒度的模糊时间序列模型及预测	82
5.1	引言	82
5.2	基于信息粒度和模糊聚类的论域划分方法	82
5.2.1	方法的理论知识	82
5.2.2	本节提出的论域划分方法	85
5.3	基于信息粒度的模糊时间序列模型	87
5.4	仿真实验与分析	89
5.4.1	Alabama 大学入学人数预测	89
5.4.2	DAX 股票月值指数预测	94

本章小结	97
思考题	97
第 6 章 包含时域信息的模糊时间序列模型及预测	99
6.1 引言	99
6.2 包含时间变量信息的论域划分方法	100
6.2.1 方法的理论知识	100
6.2.2 本节提出的论域划分方法	102
6.3 包含时间信息的模糊时间序列模型	105
6.4 实验结果与分析	106
6.4.1 Alabama 大学入学人数预测	106
6.4.2 TAIEX 预测	112
本章小结	116
思考题	116
第 7 章 基于 Granger 相关性的时间序列模型及预测	118
7.1 引言	118
7.2 基础知识	119
7.2.1 Granger 相关性及相关假设检验	119
7.2.2 利用神经网络识别函数关系	121
7.3 基于 Granger 相关性的时间序列模型	121
7.4 实验结果与分析	123
7.4.1 人工合成数据预测	124
7.4.2 真实数据预测	128
本章小结	132
思考题	132
参考文献	134
附录 A 标准正态分布表	139
附录 B t 分布的分位点	141
附录 C F 分布上侧 α 分位数表	143
附录 D χ^2 分布表	149
附录 E 协整检验临界值表	151

第 1 章 时间序列分析概述

现实生活中的各个领域, 常常会遇到大量的时间序列数据. 例如, 金融领域中, 股票、基金、期货、黄金及其他金融衍生品的成交价格和成交量可以构成时间序列; 经济领域中, 国内生产总值、居民消费价格指数、人均投资、人均收入、人均支出等指标也可以形成时间序列数据; 自然、社会领域中, 不同时段某旅游景点的人数、某公路上不同时段汽车的数量、某地区不同时段的 PM2.5 值、某地区不同年份的降水量、某房间不同时段室内的光线强度、湿度、温度等都形成了时间序列. 大量的时间序列数据, 特别是来自各种传感器、数据仓库及商业记录等时间序列数据, 促使学者对时间序列进行不断的探索和研究, 希望能够分析并发现时间序列数据背后的内在运行规律, 进而对其发展趋向进行科学的预测.

本章内容安排如下: 1.1 节是关于时间序列分析, 包括什么是时间序列、时间序列分析及预测的产生与发展; 1.2 节介绍时间序列分析的基本步骤; 1.3 节简要介绍时间序列分析软件, 包括 Eviews, MATLAB, R 等; 最后是本章小结和思考题.

1.1 关于时间序列分析

1.1.1 引例

所谓时间序列是指在不同时刻所关心的指标变量的数值按照先后顺序排列而成的序列. 在实际应用中, 一些数据往往并不是按时间先后顺序获得或根本与时间没有关系, 但如果能够按照它们的某一共同属性构成一个序关系 (如密度、重要性、来源等), 则这些数据也可以看成时间序列. 在飞速发展的大数据时代, 各领域都积累了海量的时间序列数据. 例如, 表 1.1 和图 1.1 分别为 1846—1869 年太阳黑子数据及图像.

例 1.1 1846—1869 年太阳黑子数据 (单位: 个) 列于表 1.1; 1846—1869 年的太阳黑子数依照时间画在图 1.1 中. 图中, 横轴是时间指标 t (单位: 年), 纵轴表示在时间 t 内太阳黑子个数的观测值 X_t , 这种图称为时间序列图.

例 1.2 1960—1982 年联邦德国的季度消费支出、固定投资及可支配收入数据列于表 1.2. 此数据反映了 20 世纪 70 年代联邦德国的经济运行情况. 将 1960—1982 年联邦德国的季度消费支出、固定投资及可支配收入数据取对数后, 再进行一阶差分使之平稳化, 并画在图 1.2 中.

表 1.1 1846—1869 年太阳黑子数据

年份	黑子数	年份	黑子数	年份	黑子数
1846	62	1854	21	1862	59
1847	98	1855	7	1863	44
1848	124	1856	4	1864	47
1849	96	1857	23	1865	30
1850	66	1858	55	1866	16
1851	64	1859	94	1867	7
1852	54	1860	96	1868	37
1853	39	1861	77	1869	74

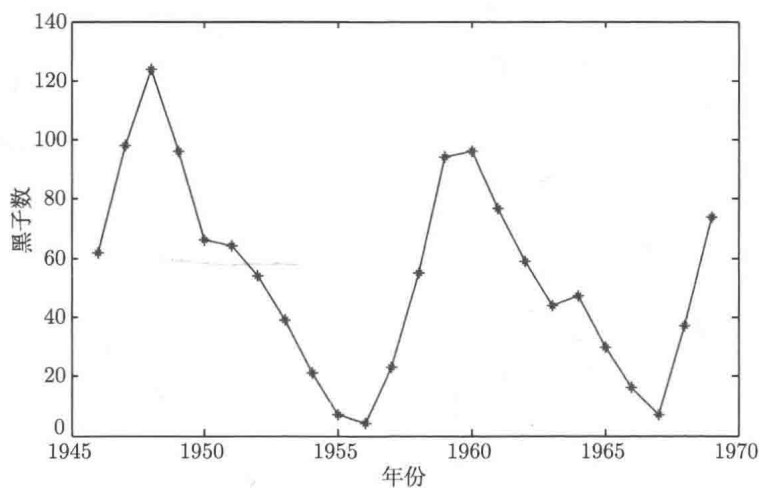


图 1.1 1846—1869 年太阳黑子数据

表 1.2 1960—1982 年联邦德国的季度消费支出、固定资产投资及可支配收入数据

年/季度	消费支出/亿马克	固定资产投资/亿马克	可支配收入/亿马克
1960/1	180	451	415
1960/2	179	465	421
1960/3	185	485	434
1960/4	192	493	448
1961/1	211	509	459
1961/2	202	520	458
1961/3	207	521	479
1861/4	214	540	487
⋮	⋮	⋮	⋮

续表

年/季度	消费支出/亿马克	固定投资/亿马克	可支配收入/亿马克
1982/1	801	2639	2235
1982/2	824	2618	2237
1982/3	831	2628	2250
1982/4	830	2651	2271

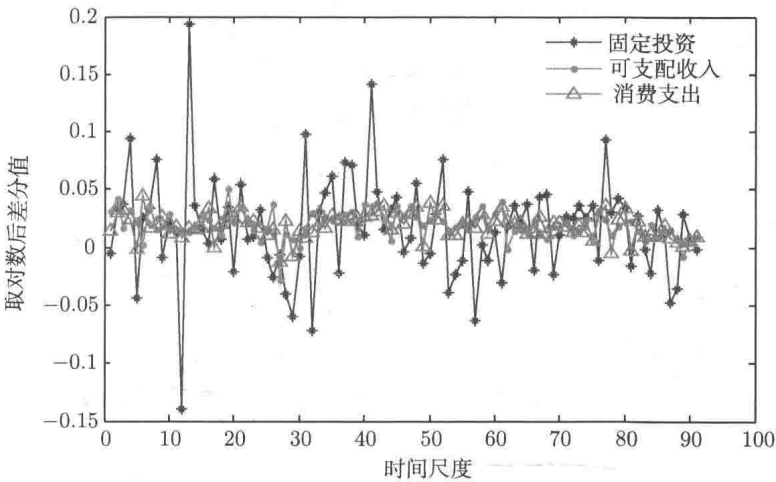


图 1.2 1960—1982 年联邦德国的季度消费支出、固定投资及可支配收入数据

例 1.3 1992 年台湾加权股价指数 (Taiwan Stock Exchange Capitalization Weighted Stock Index, TAIEX) 列于表 1.3. 股票数据是一类典型的金融时间序列数据, 常常用于实验数据. 将数据画在图 1.3 中, 图中横轴是时间指标 t (单位: 天), 纵轴表示在时间指标 t 内台湾加权股价指数观测值.

表 1.3 1992 年台湾加权股价指数

年/月/日	TAIEX	年/月/日	TAIEX	年/月/日	TAIEX
1992/01/04	4613.0	1992/02/25	4987.7	1992/12/01	3646.7
1992/01/05	4692.2	1992/02/26	4951.5	1992/12/02	3635.7
1992/01/07	4718.8	1992/02/27	5019.2	1992/12/03	3614.0
1992/01/08	4772.6	1992/02/28	5031.6	1992/12/04	3651.3
⋮	⋮	⋮	⋮	⋮	⋮
1992/02/20	4995.1	1992/11/26	3676.5	1992/12/23	3448.1
1992/02/21	4950.0	1992/11/27	3681.9	1992/12/24	3456.0
1992/02/22	5074.6	1992/11/28	3686.1	1992/12/28	3327.6
1992/02/24	5070.7	1992/11/30	3675.0	1992/12/29	3377.0

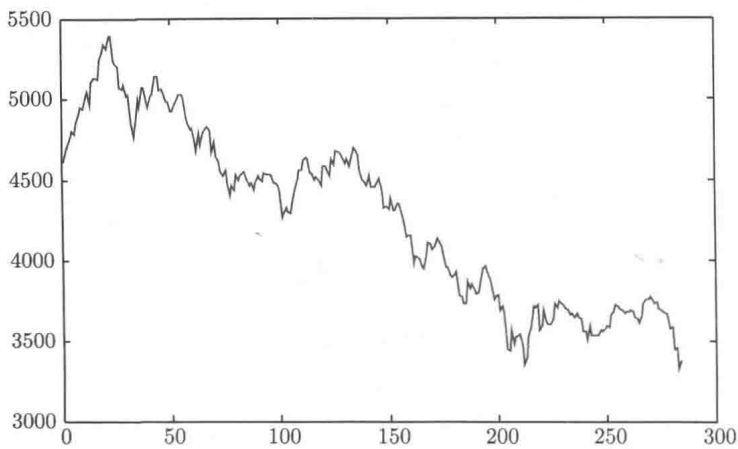


图 1.3 1992 年台湾加权股价指数

表 1.4 1958—1960 年每月美国加利福尼亚州特里尼蒂河水量

年/月	水流量/(cm ³ /s)	年/月	水流量/(cm ³ /s)	年/月	水流量/(cm ³ /s)
1958/01	11.2135	1959/01	81.6092	1960/01	92.4262
1958/02	91.1803	1959/02	330.4576	1960/02	62.0422
1958/03	91.5767	1959/03	106.2165	1960/03	71.1319
1958/04	76.3139	1959/04	146.568	1960/04	106.0749
1958/05	129.8611	1959/05	256.6073	1960/05	76.0591
1958/06	55.5577	1959/06	141.1878	1960/06	38.8224
1958/07	13.0824	1959/07	42.9567	1960/07	9.8543
1958/08	5.6351	1959/08	14.9513	1960/08	4.5307
1958/09	6.6261	1959/09	7.1075	1960/09	5.8616
1958/10	54.0002	1959/10	6.0032	1960/10	4.8422
1958/11	46.5246	1959/11	9.0614	1960/11	4.559
1958/12	55.1895	1959/12	9.1463	1960/12	5.805

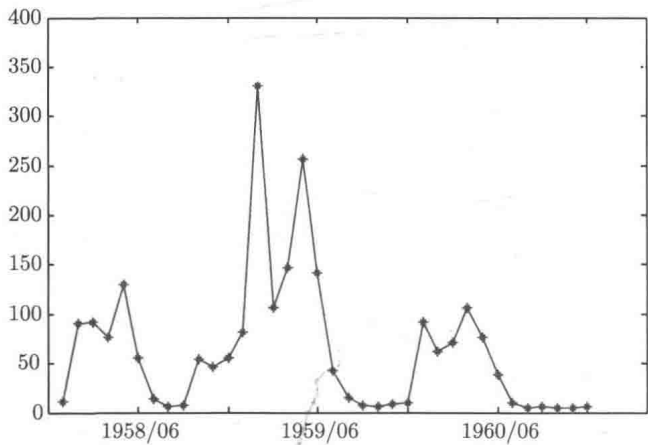


图 1.4 1958—1960 年每月美国加利福尼亚州特里尼蒂河水量

例 1.4 1958—1960 年每月美国加利福尼亚州特里尼蒂河水流量列于表 1.4 中. 将数据画在图 1.4 中, 图中横轴是时间指标 t (单位: 月), 纵轴表示在时间指标 t 内水流量观测值 (单位: m^3/s).

从上述例子可以看出, 时间序列具有如下特点.

(1) 时间序列中数据的位置与时间对应, 即数据的取值随着时间的变化而不同. 时间通常用 t 表示, 它可以是年、季度、月、日或其他时间单位. 究竟取何种时间单位要根据具体的问题的观测频率或观测周期确定. 例如, 例 1.1 中时间单位是年, 例 1.2 中的时间单位是季度, 例 1.3 中的时间单位是日, 例 1.4 中的时间单位是月. 时间单位也不仅局限于年、季度、月、日, 在某些资本市场的高频时间序列分析中, 就会用到时间单位是小时、分甚至是秒、微秒等时间单位.

(2) 时间序列在时间前后上通常存在着相互依存关系, 这也是进行时间序列分析及建模的前提假设. 这种依存关系从整体上看, 往往表现为时间序列呈现出某种趋势性或周期性变化, 此依存关系被认为是系统的动态规律性, 也是进行时间序列分析的基础.

(3) 时间序列是对某一个或某几个变量指标在不同时间进行观察所得到的结果序列. 由于观测指标通常受到众多因素的影响, 所以指标变量在每一时刻的取值都伴有一定的随机性. 我们所观测到的时间序列在某一时间的数值, 实际上是时间序列随机变量在某一时间的实现值.

(4) 时间序列中的观测值可以是一个周期内的存量, 也可以是周期节点上的瞬时流量值. 在许多金融及经济统计中, 通常将经济指标划分为两类, 即存量和流量. 经济存量是指在某一观测周期内经济指标的总量值, 流量是指在某一观测周期节点上的数据值. 例如, 人民币汇率、期货价格、股票交易指数等则属于一个时点上的流量指标, 相应的数据就是流量数据, 而人口总数、财产额等则属于某一观测周期内的存量值, 相应的数据就是存量数据.

在实际问题中, 存在着的不同类别的时间序列. 根据所研究的问题不同, 可以对时间序列进行如下分类.

(1) 按研究对象的多少, 可以将时间序列分为一元时间序列和多元时间序列. 有些研究对象可以用一个指标描述, 这时对应的时间序列就是一元时间序列, 如例 1.1, 例 1.3 和例 1.4. 但在实际问题中, 单纯地观测一个指标不容易深刻认识事物的本质, 这就需要对多个指标进行观测, 得到的就是多元时间序列. 这样可以深入了解时间序列变量指标间的相互关系, 多元时间序列变量不仅描述了各变量的变化情况, 而且蕴含了各变量间的依存关系. 例如, 例 1.2 中, 要考察某个国家的经济运行状况, 就需要对消费支出、固定资产投资及可支配收入等多个指标进行分析. 随着研究的深入, 多元时间序列分析是研究的热点, 它可以深入地挖掘指标变量间的深

层依存关系.

(2) 按照观测时间的连续与否, 可将时间序列分为连续时间序列和离散时间序列. 如果时间序列的观测值所对应的时间参数为离散间断点, 则该序列就为离散时间序列; 如果某一时间序列观测值所对应的时间参数为连续函数, 则该序列就是一个连续时间序列. 通常研究较多的是离散时间序列, 在经济金融等领域主要研究离散时间序列, 对于连续时间序列, 可以通过等周期或非周期取样的方法使之转化为离散时间序列进行研究.

(3) 按时间序列的统计特征, 可以将时间序列分为平稳时间序列和非平稳时间序列. 当观测样本数据的一、二阶矩不随时间发生变化, 则称为平稳时间序列; 否则, 称为非平稳时间序列. 本书将主要涉及多元时间序列的分析研究.

实际问题中, 除了时间序列数据外, 我们常常能遇到一些截面数据及面板数据. 时间序列数据与截面数据及面板数据的区别如下:

时间序列数据是在不同时间点上收集到的某一个统计对象的数据, 这类数据反映某一事物、现象随时间的变化状态.

例如: 2000—2004 年各年北京市的年降水量分别为 37.1, 33.8, 37.0, 44.4, 48.3(单位: cm). 这就是时间序列, 选一个城市, 看各个样本时间点的不同就是时间序列.

截面数据是在同一时间点、不同统计对象相同统计指标组成的数据列. 截面数据是按照统计单位排列的. 因此, 截面数据不要求统计对象及其范围相同, 但要求统计的时间相同. 也就是说必须是同一时间截面上的数据.

例如: 北京、上海、重庆、天津的 2000 年降水量分别为 37.1, 179.3, 458.8, 102.4(单位: cm). 这就是截面数据, 在一个时间点 2000 年处切开, 看各个城市的不同年降水量就是截面数据.

面板数据是截面数据与时间序列数据综合起来的一种数据类型. 其有时间序列和截面两个维度, 当这类数据按两个维度排列时, 是排在一个平面上, 与只有一个维度的数据排在一条线上有着明显的不同, 整个表格像是一个面板, 所以称为面板数据.

例如, 2000—2004 年各年中国所有直辖市的年降水量 (单位: cm):

北京市分别为 37.1, 33.8, 37.0, 44.4, 48.3;

上海市分别为 179.3, 159.4, 110.0, 112.8, 92.9;

天津市分别为 458.8, 502.9, 383.1, 647.5, 490.5;

重庆市分别为 102.4, 84.3, 110.8, 92.9, 105.8.

这就是面板数据 (表 1.5).

表 1.5 降水量面板数据

	2000 年	2001 年	2002 年	2003 年	2004 年
北京市	37.1	33.8	37.0	44.4	48.3
上海市	179.3	159.4	110.0	112.8	92.9
天津市	458.8	502.9	383.1	647.5	490.5
重庆市	102.4	84.3	110.8	92.9	105.8

时间序列分析所讨论的内容就是对时间序列观测数据进行研究,在飞速发展的
大数据时代,各领域都积累了海量的数据资源,面对庞大的信息资源人们急需从中
获取有价值的信息,数据分析技术由此产生.时间序列分析作为一种数据分析技术,
有其独特的、自成体系的数据分析处理方法.时间序列分析是以随机过程理论和
数理统计方法为工具,挖掘时间序列现象随时间变化而演变的统计规律,为人们正
确、科学地理解时间序列表象提供理论支持,为应对时间序列未来发展变化状况提
供决策依据.时间序列预测是以连续性假设原理为前提,应用已有的统计方法和技
术,从序列变量的观测数据中找出其内在演变机理及产生模式,从而建立描述系统
的依存关系模型,最终对预测变量的变化趋势进行定量的估计.

1.1.2 时间序列分析及预测的产生与发展

时间序列分析的发展历史可以追溯到远古时代人们对一些自然现象的记录和
绘图观测.例如,古埃及人就根据每年尼罗河泛滥的历史数据,绘制出历史数据图
进行分析比较,以便寻找洪灾的发生规律,这种分析方法被认为是时间序列分析方
法的雏形.此种分析方法也称为描述性时间序列分析方法.该方法操作简单、直
观有效,从古至今一直被人们广泛应用.时间序列分析的应用领域涉及天文、地
理、生物、经济、人口等自然和社会科学领域.例如,19 世纪中后期,德国天文学家
Schwabe 就采用描述性时间序列分析方法发现了太阳黑子的活动具有周期性.

随着研究的不断深入,人们发现简单地依靠描述性时间序列分析方法很难准确
地估计出时间序列发展变化规律.在经济、金融、物理等领域,科学家发现时间序
列变量的发展变化趋势呈现出较强的随机性,要想准确地预测时间变量的未来走势
单纯地依靠观察和描述分析方法是很难实现的.因此,学术界利用数理统计学的分
析原理来分析时间序列,从而逐步发展起了一套专门研究时间序列随机规律性的方
法,这就是现代人们通常所说的时间序列分析.

现代时间序列的研究工作可以追溯到英国统计学家 Yule 于 1927 年提出的
AR(2) 模型.在此基础上,瑞典统计学家 H. Wold 于 1948 年给出了平稳时间序
列的线性自回归滑动平均 (autoregressive moving average, ARMA) 模型,并提出
了著名的 Wold 分解定理.1970 年,Box 和 Jenkins 出版了 *Time Series Analysis:
Forecasting and Control* 一书^[1].此著作对时间序列的分析研究起到了奠基性作用,

并提出了处理非平稳时间序列的自回归滑动平均模型, 即 ARIMA 模型. 它是更一般、更具代表性、应用最广泛的线性模型. 自回归整合滑动平均 (autoregressive integrated moving average, ARIMA) 模型是解决时间序列预测问题的一个有力工具. ARIMA 模型是由 Box 和 Jenkins 提出的一种时间序列模型, 所以又称为 Box-Jenkins 模型. 之后, 很多学者在此模型基础上从各个方面进行了深入的研究, 并提出了很多具有巨大影响力的模型. 例如, 向量误差修正 (vector error correction, VEC) 模型、条件异方差 (auto-regressive conditional heteroscedasticity, ARCH) 模型、门限自回归 (threshold auto-regressive, TAR) 模型等. 虽然这些经典的时间序列模型可以解决大量的预测问题, 但是仍然存在不足之处. 主要表现为如下三个方面.

(1) 经典的时间序列模型是建立在时间序列数据具有线性结构的假设条件下, 而现实生活中的时间序列数据通常具有较强的非线性结构, 这就使得经典的 ARIMA 模型不再适合, 利用此类模型进行预测得到的结果相对粗糙.

(2) 对于模糊、缺失和不完整的时间序列数据, 经典的时间序列分析模型较难处理, 由于信息利用不完全, 即使建立了时间序列模型也不能很好地反映数据产生的内在机理. 另外, 由于经典的时间序列模型是建立在统计方法基础上的, 当时间序列历史数据较少时, 建立的模型存在一定的缺欠. 因为信息量的缺失, 不能真实反映情况并完成真正意义上的时间序列建模, 所以, 依靠此模型进行预测的结果偏差较大.

(3) 依靠经典时间序列模型进行预测得到的是定量的结果. 有些现象或事实不仅具有随机性, 同时也具有模糊性, 人们往往更倾向于用模糊概念来描述. 例如, 人们对股票指数的变化常用“上涨”“震荡”“下跌”等模糊语言来描述. 非专业领域的人员无法深刻理解精确数值的含义, 而相对于模糊语义更容易理解. 这时依靠经典时间序列模型得到的精确预测值将无法满足人们的现实需求.

经典时间序列模型的上述不足, 在某种程度上激发了更科学、更合理的时间序列建模预测方法的出现. 这其中包括两大类. 一类是利用计算智能技术对时间序列数据进行建模预测. 计算智能包括人工神经网络、遗传优化算法、贝叶斯网、支持向量机、信息粒度等软计算技术. 这些技术的最大优势在于能够很好地捕捉到时间序列数据中存在的非线性特征. 甄别时间序列数据的非线性结构, 可以帮助人们对时间序列数据的产生有一个更为全面、深刻的理解. 截止到 2012 年 12 月, 在谷歌学术搜索引擎中搜索 “Time Series and Computational Intelligence” 可以找到约 16800 条结果, 而截止到 2017 年 9 月, 可以找到约 1390000 条结果. 短短 5 年时间增加一百多万条结果, 这足以证明计算智能技术在时间序列分析这一研究领域的发展空间及发挥的重要作用.

另一类是运用模糊集合理论对时间序列数据建立模糊时间序列模型进行预测.

1965 年美国加利福尼亚大学模糊数学创始人扎德 (Zadeh) 教授提出了模糊集合理论, 并建立了处理具有模糊性问题的模型. 这也为建立模糊时间序列模型提供了理论基础和思想方法. 以模糊集合理论为基础, Song 和 Chissom 于 1994 年首次提出了模糊时间序列预测模型^[2], 自此开启了模糊时间序列理论和应用的时代. 此后, 大量的学者致力于模糊时间序列的理论与应用研究, 并取得了丰硕的成果. 有时对某些经济时间序列数据, 不仅要考虑其随机性还要考虑其模糊性, 对时间序列数据的某些指标进行适当的模糊化处理可以得到更好的预测质量. 模糊时间序列模型的优点在于对数据进行建模时考虑了模糊性, 对数据处理更具柔性, 得到的预测结果有明确的语义解释.

经过国内外专家学者几十年的共同努力, 现已在时间序列模型及预测方面取得了丰硕的成果. 按照预测所依赖的时间序列模型, 可将研究成果划分为以下三个主要方面: 一是基于经典的线性及非线性时间序列模型的预测; 二是基于计算智能技术的时间序列预测; 三是基于模糊集合理论的模糊时间序列模型预测.

1.2 时间序列分析的基本步骤

时间序列分析依赖于不同的应用背景, 有不同的目的. 一般地, 时间序列被看作一个随机过程的实现. 分析的主要任务是揭示观测时间序列的数据的内在随机规律, 通过所了解的这个随机规律, 我们希望把握系统的动态结构, 预测未来的事件, 以期通过干预来控制未来事件.

Box 和 Jenkins 的专著 *Time Series Analysis: Forecasting and Control* 是时间序列分析发展的里程碑, 他们的工作为实际工作者提供了对时间序列进行分析、预测以及对模型进行识别、估计和诊断的体系统方法, 这使得 ARIMA 模型成为一套完整的、结构化的建模方法, 并且有坚实的统计理论基础作为保障, 这种对 ARIMA 模型识别、估计和检验的统计方法简称为 Box-Jenkins 方法. 对 ARIMA 建模的过程具体分为五个步骤, 特征分析、模型识别、模型参数估计、模型检验、模型应用.

1.2.1 特征分析

在进行时间序列建模的过程中, 首先要对时间序列的特征有所了解. 一般地, 从时间序列的随机性、平稳性和季节性三个方面进行考虑, 其中平稳性尤为重要. 对于一个非平稳时间序列, 通常需进行平稳化处理后再进行建模, 也可以根据特性直接建模.

单位根检验是指判断时间序列中是否存在单位根, 即对时间序列的平稳性进行检验. 可以证明若存在单位根, 则序列是不平稳的^[1]. 常用的单位根检验方法包括: ADF(augmented Dickey Fuller) 检验、PP(Phillips Person) 检验、NP(Nelson