

经济与管理类统计学系列教材

应用时间序列分析

王振龙 胡永宏 主编

科学出版社

北 京

内 容 简 介

本书是根据教育部统计学教学指导分委员会新制定的《统计学专业教学规范(授经济学学位)》所设计的课程体系和教学内容编写的。本书遵循“通俗、易懂、实用”的原则,试图借助计算机的存储功能和计算功能来抽象掉时序分析方法的深奥数学理论和复杂运算,从而使具有一般数学知识的读者可轻松掌握和运用时间序列分析方法。在阐述中,尽可能回避严格的数学推导和证明,而从系统运动的惯性(即记忆性)加以解释和展开,或者说,本书把时序分析看作是一种统计分析工具,而不是数学的一个分支理论。全书分10章系统地介绍了时间序列分析的基本理论、基本思想、基本方法及其应用,各章均附有思考与练习,书后还附有例题用的数据。本书配有教学光盘,光盘中备有例题的SAS程序、PowerPoint教学课件和部分思考与练习答案,便于教师组织教学和学生进行学习。

本书主要是作为经济与管理类统计学专业本科生的基础教材,也可用作经济与管理类研究生的教学参考书。对于自学时间序列分析方法的读者来说,更是一本必备的入门教材。

图书在版编目(CIP)数据

应用时间序列分析/王振龙,胡永宏主编. —北京:科学出版社,2007
(经济与管理类统计学系列教材)

ISBN 978-7-03-018885-4

I. 应… II. ①王…②胡… III. 时间序列分析-高等学校-教材
IV. O211.61

中国版本图书馆 CIP 数据核字(2007)第 057787 号

责任编辑:林 建 贾瑞娜/责任校对:邹慧卿
责任印制:张克忠/封面设计:耕者设计工作室

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>

印刷

科学出版社发行 各地新华书店经销

*

2007年5月第 一 版 开本:B5(720×1000)

2007年5月第一次印刷 印张:16 1/2

印数:1—4 000 字数:306 000

定价:22.00 元

(如有印装质量问题,我社负责调换(环伟))

经济与管理类统计学系列教材编委会

主任

曾五一 教育部高等学校统计学专业教学指导分委员会副主任委员、厦门大学教授、博导

委员(按姓氏笔画排序)

王艳明 山东工商学院教授

王振龙 教育部高等学校统计学专业教学指导分委员会委员、西安财经学院教授

朱建平 厦门大学教授、博导

刘建平 暨南大学教授、博导

刘 洪 中南财经政法大学教授

许 鹏 湖南大学教授、博导

李金昌 浙江工商大学教授、博导

李宝瑜 山西财经大学教授、博导

杨 灿 厦门大学教授、博导

肖红叶 教育部高等学校经济学类学科教学指导委员会委员、天津财经大学教授、博导

张润楚 教育部高等学校统计学专业教学指导分委员会委员、南开大学教授、博导

茆诗松 华东师范大学教授、博导

杭 斌 山西财经大学教授、博导

罗良清 教育部高等学校经济学类学科教学指导委员会委员、江西财经大学教授、博导

周恒彤 天津财经大学教授、博导

庞 皓 国家级教学名师、西南财经大学教授、博导

郑 明 教育部高等学校统计学专业教学指导分委员会委员、复旦大学教授、博导

徐国祥 上海财经大学教授、博导

蒋 萍 东北财经大学教授、博导

雷钦礼 暨南大学教授、博导

黎 实 西南财经大学教授、博导



总 序

统计学是有关如何测定、收集和分析反映客观总体数量的数据，以便给出正确认识的方法论科学。随着社会经济的发展和科学技术的进步，统计应用的领域越来越广，统计已经成为人们认识世界不可或缺的重要工具。

现代统计学可以分为两大类：一类是以抽象的数量为研究对象，研究一般的收集数据、分析数据方法的理论统计学；另一类是以各个不同领域的具体数量为研究对象的应用统计学。前一类统计学具有通用方法论的理学性质，其特点是计量不计质；后一类统计学则与各不同领域的实质性学科有着非常密切的联系，是有具体对象的方法论，因而具有复合性学科和边缘学科的性质。所谓应用，既包括一般统计方法的应用，也包括各自领域实质性科学理论的应用。经济与管理统计学是以社会经济数量为对象的应用统计学。要在经济和管理领域应用统计方法，必须解决如何科学地测定经济现象即如何科学地设置指标的问题，这就离不开对有关经济现象的质的研究。要对经济和管理问题进行统计分析，也必须以有关经济和管理理论为指导。因此，经济与管理统计学的特点是在质与量的紧密联系中，研究事物的数量特征和数量表现。不仅如此，由于社会经济现象所具有的复杂性和特殊性，经济与管理统计学除了要应用一般的统计方法外，还需要研究自己独特的方法，如核算的方法、综合评价的方法等。

从历史和现状看，我国统计学专业的办学也有两种模式：一是强调各类统计学所具有的共性。这种模式主要培养学生掌握通用的统计方法和理论。它肯定统计学的“理学性质”，按照理学类学科的特点设置课程，概率论和各种数理统计方法等通用的统计方法论在课程中占有较大分量。其培养目标是有良好的数学基础、熟练掌握统计学基本理论与各种方法，同时有一定的专门领域的知识，能够适应各个不同领域的统计工作和统计研究的统计人才。二是强调各类统计学的个

性,对于经济与管理类统计学来说,就是强调其与经济学和管理学等其他学科的密切联系,按照经济与管理类学科的特点设置课程,除统计学本身的专业课外,经济管理类的课程占相当大的比重。其培养目标是所谓的“复合型人才”,即具有坚实的经济与管理理论功底;既懂数理统计方法,又懂经济统计方法,并能熟练掌握现代计算手段的经济与管理统计人才。这种人才既是统计人才又是经济管理人才,不仅能胜任基层企事业单位和政府部门的日常统计业务,而且能从事市场调查、经济预测、信息分析和其他经济管理工作。上述两种办学模式,各有特色,同时也各有一定的社会需求。从我国的国情看,现阶段后一种模式培养的人才市场需求要更大一些。应该根据“百花齐放,百家争鸣”的方针,允许多种办学模式同时并存,由各院校根据自己的特色和市场对有关人才需求的大小,自主选择合适的办学模式。

为了更好地满足 21 世纪对统计人才的需要,无论是理学类统计学专业还是经济管理类统计学专业都有一个如何面向未来、面向世界、加强自身建设、更好地与国际接轨的问题。但是,这两类专业的培养目标不同,知识体系也有相当大的差异,难以完全统一或互相取代。2003 年 11 月,教育部高等学校统计学专业教学指导分委员会在厦门召开年会,会上各方面的专家达成共识,为了促进统计学的学科建设和发展,有必要按授予学位的不同,分别制定指导性的教学规范。2004 年 1 月,全国经济与管理类统计学专业的部分专家和学者在天津财经学院讨论了《统计学专业教学规范(授经济学学士学位)》征求意见稿,对初稿进行修改与补充,又经过教育部高等学校统计学专业教学指导分委员会研究和审定,最终形成了正式的教学规范(以下简称新规范),并已上报国家教育部。

根据新规范的设计,经济管理类统计学专业应开设的统计学专业主干课程包括以下 10 门:①统计学导论;②数理统计学;③应用多元统计分析;④应用时间序列分析;⑤应用抽样技术;⑥计量经济学;⑦国民经济统计学;⑧企业经营统计学;⑨证券投资分析;⑩货币与金融统计学。为了进一步提高统计教材的质量,更好地满足 21 世纪培养经济管理类统计人才的需要,我们成立了经济与管理类统计学系列教材编委会,组织全国高校同行分工协作,根据新规范设计的课程体系和教学内容,编辑出版一套经济与管理类统计学系列教材。本系列教材比较适合作为高等院校经济与管理类统计学专业的教材,其中《统计学导论》和《计量经济学》还可作为一般经济与管理类专业的核心课程教材。

参加本系列教材建设的有厦门大学、西南财经大学、天津财经大学、上海财经大学、浙江工商大学、山西财经大学、湖南大学、西安财经学院、南开大学、东北财经大学、中南财经政法大学、暨南大学、华东师范大学、江西财经大学、山东工商学院、复旦大学等 10 多所院校的教师。本系列教材实行主编负责制,担任主编和主审的老师都是曾经主编和出版过相关统计教材的国内知名教授,不

仅具有在第一线教学的经验，而且对相关学科的发展趋势和学科前沿也比较熟悉。本系列教材的编写，力求体现以下五个特点。

（一）与时俱进，构建与培养目标相适应的教学内容体系

教材建设的关键在于构建与培养目标相适应的教学内容体系。为此，要根据时代的发展，不断补充和引进新的教学内容。作为 21 世纪经济与管理类统计专业的教材，不能只是简单地向理学类统计专业的教学内容靠拢，而应该根据自身的特点，努力贯彻“少而精”和“学以致用”的原则。在大胆吸收国外优秀教材特点的基础上，对原有的体系重新进行整理和完善，既适当增加一些在经济社会分析中有良好应用前景的数理统计理论与方法的内容，又适当增补经济社会统计方面的最新进展。同时删除过时的和不再适用的内容。尽可能做到既反映本门学科的先进水平，又比较简明易懂、便于教学。

（二）统筹兼顾，防止低水平重复，发挥系列教材的整体功能

适应未来需要的经济管理类统计人才必须掌握多方面的知识和能力，各种知识是相互联系的，各门课程在内容上难免有所交叉。为了提高学习的效率，更好地发挥系列教材的整体功能，在编写本系列教材的过程中，我们作了必要的协调和适当的分工，尽可能做到统筹兼顾，防止低水平重复。同时，各本教材采用相同的版式、体例和统一规范的学术用语。

（三）与计算机结合，培养学生的动手能力

为了提高学生运用统计方法解决实际问题的能力，本系列教材的编写注意与计算机的紧密结合。本系列教材中统计方法类的教材均根据教材的内容，结合常用的计算机统计软件，并给出相应的案例和数据。从而使学生不仅可以从中学习统计学理论和方法，而且可以上机操作，培养实际动手的能力。

（四）编写体例新颖，提高学生学习的兴趣和效率

为了便于师生教学互动，提高学生学习的兴趣和学习效率，本系列教材在编写体例上也作了一些新的尝试。各章开篇有内容要点和教学要求提示，章末附有小结，对有关教学内容和计算公式作扼要的总结。教材中尽可能使用我国的真实数据作为案例。各教材的“思考与练习”不设一般常见的名词解释型的简答题和论述题，而是通过判断题、选择题、计算题和有趣味的思考题，来帮助学生掌握有关概念和计算方法。为便于学生自学，同时又为其留有独立思考和独立完成作业的余地，各教材均给出编号为奇数的习题的详解。

（五）配套出版教学课件和光盘，便于教师组织教学

本系列教材在出版纸质出版物的同时配套出版相应的光盘。内容包括：用 PowerPoint 制作的教学课件，教材中有关案例的数据，常用的统计表，编号为奇数的习题的详解等。

经济与管理类统计学专业系列教材的建设是一个复杂艰巨的系统工程，完成这一工程需要全国统计教育工作者的共同努力。感谢参与本系列教材编写的全国各高校的专家和学者，感谢为本系列教材的出版提供帮助的科学出版社的领导和编辑。衷心祝愿大家的辛勤劳动能够结出丰硕的果实，能够为我国统计学的普及和提高作出更大的贡献。

曾五一

2005 年 8 月于厦门

前言

《应用时间序列分析》是教育部统计学教学指导分委员会新制定的《统计学专业教学规范（授经济学学位）》所设计的经济与管理类统计学专业 10 门主干课程之一。本书是根据新规范设计的课程体系和教学内容规划编写的《经济与管理类统计学系列教材》之一。因此，本书主要是作为经济与管理类统计学专业本科生的基础教材而编写的，也可用作经济与管理类研究生的教学参考书。

基于上述定位，考虑到课程体系中各门课程之间部分内容的交叉和重复，本教材中关于确定性时间序列分析的内容没有作详细介绍，这里要特别强调的是应完整地理解和掌握时间序列分析思想和各种方法的分析思路，要把确定性时间序列分析方法和随机时间序列分析方法结合起来学习和应用。

另外，时间序列分析方法发展很快，近几年来理论研究工作者和实际应用工作者提出了不少新方法和新模型，但是考虑到本科生教材对“基本理论、基本方法、基本技术”的要求，本教材虽然不可能对这些新方法和新模型进行深入的介绍但也尽量予以涉及，以开阔学生眼界与思路。

本书遵循“通俗、易懂、实用”的原则，试图借助计算机的存储功能和计算功能来抽象掉时序分析方法的深奥数学理论和复杂运算，从而使具有一般数学知识的读者可轻松掌握和运用时间序列分析方法。在阐述中，尽可能回避严格的数学推导和证明，而从系统运动的惯性（即记忆性）加以解释和展开，或者说，本书把时序分析看作是一种统计分析工具，而不是数学的一个分支理论。对于“工具”来说，使用者只要知道其特性、功能和使用方法以及使用过程中应注意的有关事项就足矣，至于其制造原理及工序过程，熟悉当然更好，不了解也无关乎其使用。鉴于这样的认识，全书没有使用深奥的定理，因而也就无需定理的证明。模型的形成来自于对系统记忆性的长短及其特性的剖析，一些数学推导也只涉及

高等数学、线性代数和概率论与数理统计的一般知识。

当然，说本书强调应用，仅仅是相对于数学定理的推证而言，在阐述中仍力图做到理论严谨、逻辑严密，以形成相应于读者认知结构的时间序列分析理论和方法体系。不过，这仅是我们的愿望而已，由于才疏学浅，水平有限，书中一定会有错误和不妥之处，尤其是为了尽量避免使用数学语言和定理及其证明手段，在用描述性语言和系统机理表述模型结构及其原理和特性时，可能从数学的角度来看会有不准确、不严格之处，恳请读者批评指正。

本书计算由 SAS 软件完成，为了方便读者学习与使用，本书配备的教学光盘中给出了书中例题的 SAS 数据和 SAS 程序以及各章奇数号思考与练习的参考答案，同时给出了各章的 PowerPoint 教学课件，以方便教师授课。

参加本书编写的有西安财经学院王振龙教授（第一、二、三章），中央财经大学统计学院胡永宏副教授（第四、五、八章），西南财经大学统计学院谢小燕教授（第七、九章），江西财经大学统计学院罗良清教授和万兆泉副教授（第六、十章）。全书由王振龙和胡永宏同志担任主编，负责全书的设计、修改、总纂和定稿等工作。

在本书的编写过程中，编者参考了国内外大量的相关文献资料，除在书中提到的主要参考文献外，还参阅和吸收了许多同仁的文献和资料，恕不一一列举，这里一并致谢！

本书的出版得到科学出版社和诸位编者所在单位的大力支持，在此我们对所有为本书的编写和出版作出贡献的单位和个人致以真诚的感谢！

王振龙

2006 年 10 月于西安



总序
前言

第一章

| | |
|-------------------------|----|
| 绪论 | 1 |
| 第一节 时间序列分析的一般问题..... | 1 |
| 第二节 时间序列的建立..... | 8 |
| 第三节 确定性时序分析方法概述 | 13 |
| 第四节 随机时序分析的几个基本概念 | 19 |
| 本章小结 | 27 |
| 思考与练习 | 27 |

第二章

| | |
|---------------------|----|
| 平稳时间序列模型..... | 28 |
| 第一节 一阶自回归模型 | 28 |
| 第二节 一般自回归模型 | 32 |
| 第三节 移动平均模型 | 34 |
| 第四节 自回归移动平均模型 | 36 |
| 本章小结 | 40 |

思考与练习 41

第三章

ARMA 模型的特性 42

第一节 格林函数和平稳性 42

第二节 逆函数和可逆性 65

第三节 自协方差函数 71

第四节 自谱 79

本章小结 86

思考与练习 86

第四章

平稳时间序列模型的建立 88

第一节 模型识别 89

第二节 模型定阶 92

第三节 模型参数估计 98

第四节 模型的适应性检验 101

第五节 Pandit-Wu 建模方法 104

第六节 建模实例 106

本章小结 111

思考与练习 111

第五章

平稳时间序列预测 113

第一节 条件期望预测 114

第二节 预测的三种形式 114

第三节 预测值的适时修正 122

本章小结 124

思考与练习 124

第六章

| | | |
|-------|-------------------|-----|
| | 趋势模型 | 126 |
| 第一节 | 趋势性时间序列的重要特征..... | 126 |
| 第二节 | 随机时间序列的趋势性检验..... | 128 |
| 第三节 | 平稳化方法..... | 131 |
| 第四节 | 趋势模型..... | 133 |
| 本章小结 | | 140 |
| 思考与练习 | | 140 |

第七章

| | | |
|-------|-------------------|-----|
| | 季节模型 | 142 |
| 第一节 | 季节时间序列的重要特征..... | 142 |
| 第二节 | 季节性时间序列模型..... | 144 |
| 第三节 | 季节性检验..... | 146 |
| 第四节 | 季节时间序列模型的建立..... | 157 |
| 第五节 | X-11 方法简介 | 159 |
| 本章小结 | | 173 |
| 思考与练习 | | 175 |

第八章

| | | |
|-------|----------------------|-----|
| | 条件异方差模型 | 177 |
| 第一节 | 条件异方差模型..... | 178 |
| 第二节 | 条件异方差模型的建立..... | 182 |
| 第三节 | 几种扩展模型..... | 191 |
| 本章小结 | | 196 |
| 思考与练习 | | 197 |

第九章

| | | |
|-------|-----------------|-----|
| | 传递函数模型 | 198 |
| 第一节 | 模型简介..... | 198 |
| 第二节 | 传递函数模型的识别..... | 202 |
| 第三节 | 传递函数的拟合与检验..... | 212 |
| 第四节 | 干预模型..... | 217 |
| 本章小结 | | 224 |
| 思考与练习 | | 225 |

第十章

| | | |
|-------|----------------------|-----|
| | 异常值分析 | 227 |
| 第一节 | 含异常值的 ARIMA 模型 | 228 |
| 第二节 | 异常值的检测..... | 230 |
| 第三节 | 异常值分析的实例..... | 233 |
| 本章小结 | | 234 |
| 思考与练习 | | 234 |
| 参考文献 | | 235 |

附录

| | | |
|------|------------------|-----|
| 附录 A | X-11 报表与说明 | 236 |
| 附录 B | 数据资料..... | 239 |
| 附录 C | 统计表 | 246 |

第一章

绪 论

本章介绍时间序列分析的基本思想、基本概念以及确定性时间序列分析的基本方法,使学生对时间序列分析方法在整体上有个初步了解,为学习和掌握运用时间序列分析方法认识问题的思路奠定基础。具体要求:①理解时间序列的特征;②掌握时间序列预处理的基本方法;③了解时间序列分析的基本思想;④熟悉确定性时间序列分析方法;⑤掌握随机时间序列分析中的基本概念的含义。

人们的一切活动,其根本目的无不在于认识和改造客观世界。时间序列分析不仅可以从数量上揭示某一现象的发展变化规律或从动态的角度刻画某一现象与其他现象之间的内在关系及其变化规律性,达到认识客观世界之目的,而且运用时间序列模型还可以预测和控制现象的未来行为,修正或重新设计系统以达到利用和改造客观之目的。从统计学的内容来看,统计所研究和处理的是一批有“实际背景”的数据,尽管数据的背景和类型各不相同,但从数据的形成来看,无非是横剖面数据和纵剖面数据两类(或者叫作静态数据和动态数据)。横剖面数据是由若干相关现象在某一时点上所处的状态组成的,它反映一定时间、地点等客观条件下诸相关现象之间存在的内在数值联系。研究这种数据结构的统计方法是多元统计分析。纵剖面数据是由某一现象或若干现象在不同时刻上的状态所形成的数据,它反映的是现象以及现象之间关系的发展变化规律性。研究这种数据的统计方法就是时间序列分析。由此足以看出时间序列分析的重要性及其应用的广泛性。

第一节 时间序列分析的一般问题

一、时间序列的含义

从统计意义上讲,所谓时间序列就是将某一个指标在不同时间上的不同数值,

按照时间的先后顺序排列而成的数列。这种数列由于受到各种偶然因素的影响，往往表现出某种随机性，彼此之间存在着统计上的依赖关系。如表 1.1 中列出的是某企业 1998~2002 年各月销售额数据(单位:百万元)，按时间顺序排成一个数列就是一个时间序列。相对于时间的数据图如图 1.1 所示。

表 1.1 某企业 1998~2002 年各月销售额数据(单位:百万元)

| 年份 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|-------|
| 1998 | 67.68 | 68.83 | 62.67 | 63.16 | 61.96 | 68.44 | 72.08 | 68.76 | 70.98 | 71.81 | 68.36 | 62.73 |
| 1999 | 69.59 | 75.00 | 70.08 | 68.14 | 68.97 | 77.88 | 77.40 | 78.73 | 78.40 | 80.69 | 72.46 | 73.21 |
| 2000 | 78.21 | 82.27 | 77.18 | 75.03 | 77.68 | 80.97 | 85.07 | 88.33 | 83.34 | 85.70 | 80.50 | 77.18 |
| 2001 | 86.72 | 90.20 | 85.22 | 80.38 | 82.78 | 90.55 | 92.07 | 92.74 | 91.77 | 92.96 | 89.69 | 83.62 |
| 2002 | 94.28 | 98.89 | 91.09 | 93.84 | 94.17 | 103.06 | 102.29 | 102.31 | 100.15 | 101.03 | 101.27 | 97.94 |

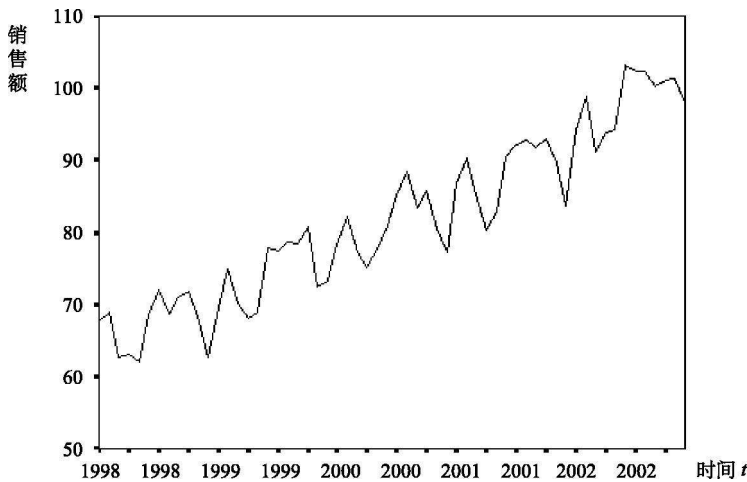


图 1.1 销售额数据图

值得注意的是,时间顺序中的“时间”也可以具有不同的物理意义,如长度、温度、速度等。

从数学意义上讲,如果我们对某一过程中的某一个变量或一组变量 $X(t)$ 进行观察测量,在一系列时刻 t_1, t_2, \dots, t_N (t 为自变量,且 $t_1 < t_2 < \dots < t_N$) 得到的离散有序数集合 $X_{t_1}, X_{t_2}, \dots, X_{t_i}, \dots, X_{t_N}$ 称为离散数字时间序列,即随机过程的一次样本实现。设 $X(t; t \in T)$ 是一个随机过程, $X_{t_i} (i=1, 2, \dots)$ 是在时刻 i 对过程 $X(t)$ 的观察值,则 $X_{t_i} (i=1, 2, \dots)$ 称为一次样本实现,也就是一个时间序列。

从系统意义上看,时间序列就是某一系统在不同时间(地点、条件等)的响应。

这个定义从系统运行的观点出发,不仅指出时间序列是按一定顺序排列而成的;这里的“一定顺序”既可以是时间顺序,也可以是具有各种不同意义的物理量,如代表温度、速度或其他单调递增地取值的物理量。可见,时间序列只强调顺序的重要性,而并非强调必须以时间顺序排列。例如,材料裂纹长度与其承受的压力有关,将材料裂纹长度按其所受压力周期数排列如下:

10.98 11.08 11.24 11.33 11.46 11.51 ...

共 90 个数据,这也是一个时间序列,其数据图见图 1.2。

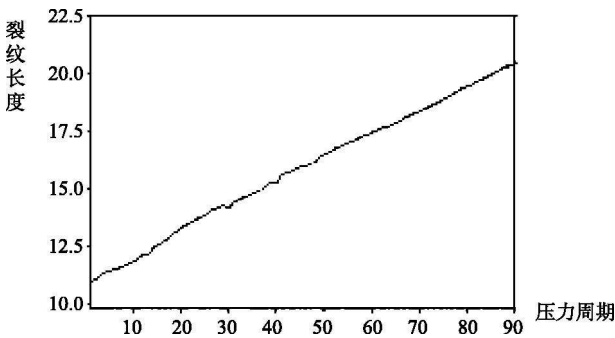


图 1.2 材料裂纹长度数据图

时间序列是所研究系统的历史行为的客观记录,因而它包含了系统结构特征及其运行规律。所以我们可以通过对时间序列的研究来认识所研究系统的结构特征(如周期波动的周期、振幅、趋势的种类等);揭示其运行规律,进而用以预测、控制其未来行为;修正和重新设计系统(如改变其周期、参数),使之按照新的结构运行。

综上所述,时间序列具有如下特点:首先,序列中的数据或数据点的位置依赖于时间,即数据的取值依赖于时间的变化,但不一定是时间 t 的严格函数。其次,每一时刻上的取值或数据点的位置具有一定的随机性,不可能完全准确地用历史值预测。再次,前后时刻(不一定是相邻时刻)的数值或数据点的位置有一定的相关性,这种相关性就是系统的动态规律性。最后,从整体上看,时间序列往往呈现某种趋势性或出现周期性变化的现象。

二、时间序列的主要分类

时间序列根据所研究的依据不同,可有不同的分类。

(1)按所研究的对象的多少分,有一元时间序列和多元时间序列。表 1.1 中,我们所研究的只是某企业月销售额数据这一数列,即为一元时间序列。但是,如果我们所研究的对象不仅仅是这一数列,而是多个变量,如是按年、月顺序排列的气

温、气压、雨量数据,每个时刻 t 对应着多个变量,则这种序列为多元时间序列。多元时间序列不仅描述了各个变量的变化规律,而且还揭示了各变量间相互依存关系的动态规律性。

(2)按时间的连续性可将时间序列分为离散时间序列和连续时间序列两种。如果某一序列中的每一个序列值所对应的时间参数为间断点,则该序列就是一个离散时间序列;如果某一序列中的每个序列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。我们主要研究离散时间序列,并用 X_t 表示,对于连续时间序列,可通过等间隔采样使之转化为离散时间序列后加以研究。

(3)按序列的统计特性分,有平稳时间序列和非平稳时间序列两类。如果一个时间序列的概率分布与时间 t 无关,则称该序列为严格的(狭义的)平稳时间序列。如果序列的一、二阶矩存在,而且对任意时刻 t 满足:

- ① 均值为常数;
- ② 协方差为时间间隔 τ 的函数。

则称该序列为宽平稳时间序列,也叫广义平稳时间序列,如图 1.3 所示。我们以后所研究的时间序列主要是宽平稳时间序列。如果不明确提出严平稳,所谓的平稳即指宽平稳。反之,不具有平稳性即序列均值或协方差与时间有关的序列称之为非平稳序列,如图 1.1、图 1.2 所示。

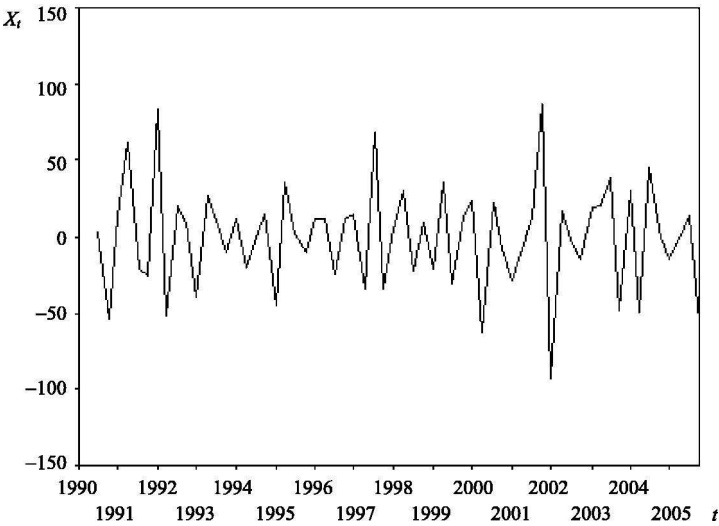


图 1.3 平稳化后的某地区个人实际消费支出

(4)按序列的分布规律来分,有高斯型(Gaussian)时间序列和非高斯型(non-Gaussian)时间序列。服从高斯分布(正态分布)的时间序列叫作高斯型时间序列,

否则叫作非高斯型时间序列。本书所介绍的模型多数是假设服从高斯分布的高斯型时序模型。对于一些非高斯序列,往往通过适当变换,则可近似地看成是高斯型时间序列。

三、时间序列分析

不论是经济领域中每年的产值、国民收入、某一商品在某一市场上的销量、某一商品在某一市场上的价格变动等,或是社会领域中某一地区的人口数、医院患者人数、铁路客流量等,还是自然领域的太阳黑子数、月降水量、河流流量,等等,都形成了一个时间序列。所有这些序列的基本点就是每一个序列包含了产生该序列的系统的历史行为的全部信息。问题在于怎样才能根据这些时间序列,较精确地找出相应系统的内在统计特性和发展规律性,尽可能多地从中提取出我们所需要的准确信息。用来实现上述目的整个方法称为时间序列分析。它是一种根据动态数据揭示系统动态结构和规律的统计方法,是统计学科的一个分支。其基本思想是根据系统的有限长度的运行记录(观察数据),建立能够比较精确地反映时间序列中所包含的动态依存关系的数学模型,并借以对系统的未来行为进行预报。

人们为了根据时间序列揭示所研究现象的动态规律性,在认识—实践—再认识的不断循环过程中,产生了一系列分析研究时间序列的方法。最朴素的动态思想认为现象的未来行为与现在的行为有关,于是,人们使用现象的现在值作为其下一时刻的预测值。这种方法对于平稳发展变化的现象来说是可行的。例如,要预测下周某种销售额较稳的产品的需求量,就用最近一周的实际需求量作为下周的预测值。按照这种思想,对于具有季节性的序列也可以先将最近的观察值去掉季节性,然后再赋予要预测时刻的季节性进行预测。遗憾的是,观察值含有偶然因素的随机影响,因而,据这种思想进行预测,受随机影响较大。为了弥补这一不足,一种有效的方法就是取一段时间上的观察值的平均数。于是,产生了移动平均法和指数平滑法。

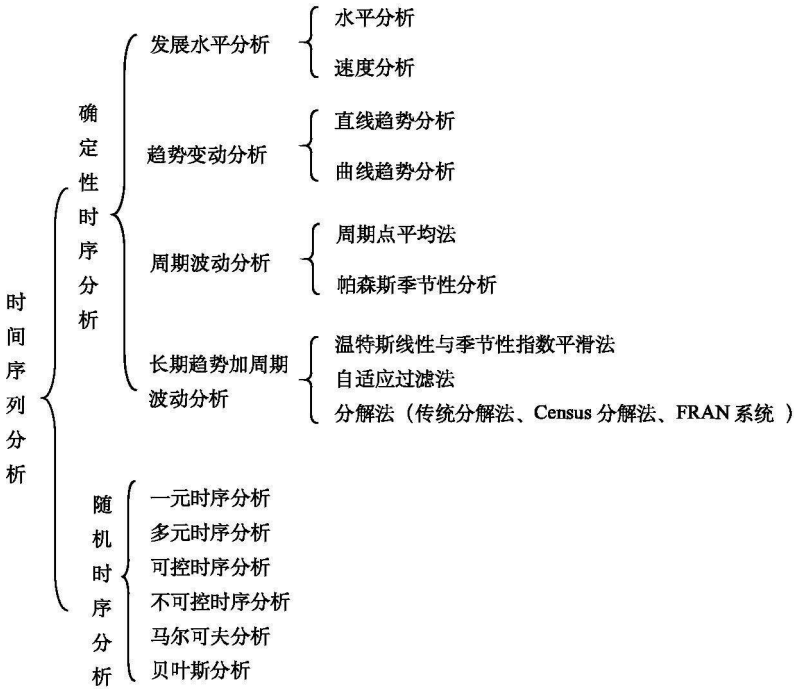
随着科学技术的不断发展,人们在实践中认识到时间序列的变动,主要是由长期趋势(随着时间的变化,按照某种规则稳步地增长、下降或保持在某一水平上)、季节变动(在一个年度内依一定周期规则性变化)、循环波动(以若干年为周期的波动变化)和随机型变动(许多不可控的偶然因素共同作用的结果)而形成的。前三种变动的一个共同特点,就是依一定的规则而变化,随机波动在综合中可以消除。基于这种认识,时间序列分析就是设法消除随机型波动,拟合确定型趋势,因而形成了长期趋势分析、季节变动分析和循环波动测定等一系列确定型时间序列分析方法。

虽然确定型趋势控制着时间序列变动的基本样式,但毕竟不是时间序列变动模式的全貌。另一方面,用随机理论来考察,许多偶然因素共同作用的随机型波

动,其实也并非完全杂乱无章,而有一定的规律性。人们根据随机理论,对随机时间序列进行分析,就叫作随机时间序列分析,其相应的方法,称之为随机时间序列分析方法。自从 20 世纪 20 年代问世以来,随机时间序列分析的理论和方法,引起了广大理论研究和实际工作者的极大重视,其理论和方法不断发展且得到了广泛的应用。

随机时间序列分析丰富和发展了时间序列分析的理论和方法。但不能取代确定型时序分析,也不能认为时间序列分析因此而就完善了。从系统的观点来看,影响时间序列总变动的各个要素并非独立地发生作用,而是在相互影响中共同发生作用,因而,分别测定的各种因素的变动的简单综合,不等于时间序列的总变动。应该将时间序列看作某一系统的动态行为的客观记录,从整体上来考察其动态结构和变动规律性。这是当前和今后时间序列分析研究的一个重大课题。此外,对一组(多维)相关随机变量的动态过程同时进行观测,并将其作为整体加以研究的多元时间序列分析,也是时序分析研究的一个广阔领域。

为了对时间序列分析方法有一个比较全面的了解,便于进一步学习和研究,现将时间序列分析的主要方法归纳如下:



时间序列分析方法如果按其采用的手段不同可概括为数据图法、指标法和模型法三类。数据图法是将时间序列在平面坐标系中绘出坐标图,根据图形直接观

察序列的总趋势和周期变化以及异常点,升降转折点等。这种方法简单、直观、易懂易用;但获取的信息少且肤浅和粗略;需要有相当丰富的分析经验,否则难以获得更深层次的信息;分析结果的主观性较大。指标法是指通过计算一系列核心指标来反映所研究系统的动态特征,如反映变化率的发展速度和增长速度,反映均衡性和节奏性的动态平均指标和变异指标等。虽然指标法较数据图法客观,但它所提取的信息仍是肤浅的、有限的。模型法是对给定的时间序列,根据统计理论和数学方法,建立描述该序列的适应或最优统计模型,并进而据以进行预测或控制。这是本书将要介绍的主要内容。

时间序列分析的基本特征就是研究序列随时间发展的模式。其区别于其他统计分析的重要特征之一,就是明确重视顺序的重要性。时间序列与其他变量数列不同,序列中的观察值是按照一定顺序取得的,并保持其顺序不变。只有这样,才能保证所研究现象的历史发展过程不致改变。其次,时间序列中的观察值之间存在着一定的依存关系。这是时序分析区别于其他统计分析的又一重要特征。数理统计研究的其他变量数列,一般要求每一变量各自独立。

但是,由于任何现象的发展,一般都具有一定的惯性(延续性),因而,相应的时间序列中各时刻的观察值之间就体现为一定的依存关系。从某种意义上说,时间序列分析就是要定量地描述这种依存关系。再次,时间序列分析区别于其他统计分析的重要特征还表现在对所研究系统未来行为的分析即预测推断的依据不同。时序分析不是根据某一变量与其他变量之间的静态相关关系来预测该变量的未来变化,而是根据预测变量本身或其他相关变量过去的变化规律来预测未来的变化。就一元时间序列分析而言,它就是根据该变量本身的变化规律来预测未来变化的。

时间序列分析与数理统计学的主要区别表现在如下三个方面,首先数理统计学的样本值是对同一随机变量进行 n 次独立重复试验的结果,或是 n 个相互独立、同分布的随机变量序列的一个实现;而时间序列则是某一随机过程的一次样本实现。其次,在数理统计学中,进行统计推断的目的主要是对一个随机变量的分布参数进行估计或假设检验;而在时间序列分析中,则是对某一时间序列建立统计模型。最后,数理统计学中的回归模型描述的是因变量与其他自变量之间的统计静态依存关系;而时序分析中的自回归模型描述的是某一变量自身变化的统计规律性,是某一系统的现在的行为与其历史行为之间的统计动态依存关系。

时间序列分析是一种重要的现代统计分析方法,广泛地应用于自然领域、社会领域、科学研究和人类思维。不论是自然现象,还是社会经济现象,都是一个有规律的辩证发展过程。任何运动都有一定的惯性,这种惯性就表现为系统的动态性即记忆性。时间序列是系统历史行为的客观记录,它包含了系统动态特征的全部信息。这些信息,具体地表现为时间序列中观察值之间的统计相关性。因而,人们可以通过研究时间序列中数值上的统计相关关系,来揭示相应系统的动态结构特

征及其发展变化规律。基于上述观点,时间序列分析就是用历史的观点,通过量的手段揭示所研究现象的动态结构和动态规律。这不仅是可能的,而且也是合理的,科学的。

系统的动态特性即事物运动的惯性,决定了时序分析理论和方法的客观性。同时,时序分析也是其他统计分析方法的客观要求。一方面,利用变量之间的因果关系预测某一变量的未来行为时,这些自变量或前导(领先)变量本身的未来值需要预测;另一方面,在许多情况下,很难或不可能利用变量之间的因果关系模型(即结构模型)来说明某一变量的变化。如某商品的销售量,它随着价格、个人收入和利率的变化而上下起伏,但是,影响其变化的其他许多因素如天气变化、顾客爱好的变化等,则可能是无法说明的。对于这种情况,时序模型则是用于预测的一个很有用的工具。一元时序模型不像回归模型那样是根据因果关系,而是根据被预测变量过去的变化规律性来建立模型,然后利用这个模型来预测该变量未来的变化。因此,时序模型是一种先进的统计方法。利用时序模型不需要知道影响预测变量的因果关系,在系统的动态性较强,关于影响预测变量的决定性因素的信息很少,且有足够多的数据量可以用来构成一个合理长度的时间序列的情况下,运用时序分析往往可达到事半功倍之功效。

概括起来,时序分析的作用主要有以下几个方面:

- (1)对理论性模式与数据进行适度检验,以讨论模式是否能正确地表示所观测的现象。
- (2)刻画系统所处的状态及其结构性,从而达到认识和解释系统之目的。
- (3)描述系统的运行规律性,从而达到认识规律和掌握规律性之目的。
- (4)预测系统的未来行为,从而达到利用规律之目的。
- (5)控制系统的未来行为,从而达到利用和支配系统之目的。

第二节 时间序列的建立

时间序列分析处理加工的是来自所研究系统的时间序列数据。因此,建立一个时间序列是时间序列分析的第一步。一般来说,研究者是运用记录仪或通过观察测量来获取所研究系统的真实有限的数据集的。有时也可以直接运用次级资料。但是,不论是通过哪一种途径获得的时间序列,在进行分析处理前,必须对所依据的资料进行认真的检查、整理,有时还需要进行适当的预处理(如缺少数据的补足等)。我们把获取时间序列以及对其进行检查、整理和预处理等工作,称为时间序列的建立。

一、时间序列数据的采集

对于所研究系统来说,相应于时间的连续性,系统在不同时刻上的响应常常是时间 t 的连续函数。为了数字计算处理上的方便,往往只按照一定的时间间隔对所研究系统的响应进行记录和观察,我们称之为采样。相应地把记录和观察的时间间隔称为采样间隔,并用 Δ 表示。例如,对某市场某种水果的价格每隔一日进行一次记录,则采样间隔为一日($\Delta=1$ 日或 24 小时),每个观察值也称为采样值,第 k 个采样值为 X_k ,也就是连续函数 $X(t)$ 的值 $X(t_0 + k\Delta)$ (图 1.4)。

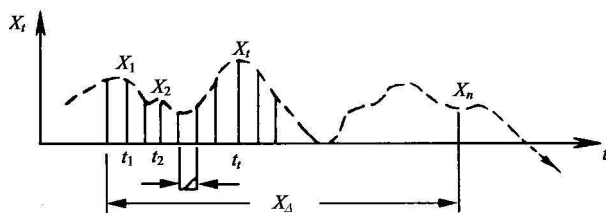


图 1.4 以 Δ 为采样间隔对 X_t 的采样

在对时间序列 $X(t)$ 的采样过程中,取不同的采样间隔 Δ 可以得到不同的数字时间序列 $X_t, t=1, 2, \dots$ 。采样间隔既可以相等,也可以不等。我们这里只讨论等间隔采样,因此,在本书中将“采样间隔相等”作为建立时间序列的一条重要原则。

如前所述,时间序列包含了系统的全部信息,因而人们可以借以研究系统的动态结构和运行规律。但是,采样得到的离散时间序列 X_t 失去了 $X(t)$ 在 $t_0 + (i-1)\Delta$ 和 $t_0 + i\Delta$ 之间的值,即这个区间内系统的信息在采样之后丢失了。显然,在合理的范围内,采样间隔 Δ 越小,采样值就越多,信息损失就越小,数据处理量越大,处理时间、人力、财力消耗越大。相反,采样间隔越大,采样值越少,处理时间、人力、财力消耗越小,但信息损失越大。因此,选择合适的采样间隔是建立时间序列的关键,理想的采样间隔就是既没有损失信息,也没有出现信息冗余。在实际中,研究者只能根据所研究系统的具体特性和经验,在不过分减少信息损失和不过分增加数据量之间作出合理选择。

二、离群点的检验与处理

离群点(outlier)是指一个时间序列中,远离序列一般水平的极端大值和极端小值。因此,也称为奇异值,有时也称其为野值。图 1.5 所描述的是 1949~2003 年我国的人口自然增长率(X_t)。从图中可以直观地看出,1960 年的人口自然增长率大大低于其他年份的自然增长率。从序列来看, X_{12} (即 1960 年的观察值)远离了序列 X_t ,我们称 X_{12} 为离群点。

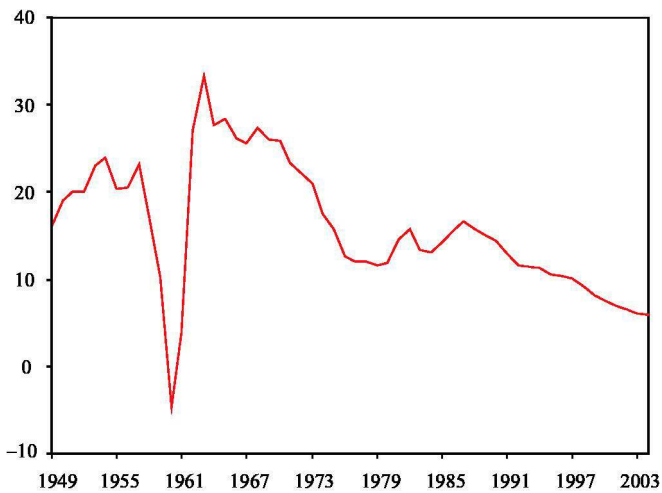


图 1.5 我国人口自然增长率(‰)数据图

概括地说,离群点是由于系统受外部干扰而造成的。但是,形成离群点的系统外部干扰是多种多样的。首先可能是采样中的误差,如记录仪出现偏误、工作人员出现笔误、计算差错等,都有可能产生极端大值或极端小值。其次可能是被研究现象本身由于受各种偶然非正常的因素影响而引起的,例如,在人口死亡序列中,由于某年发生了地震,使得该年的死亡人数急剧增加,形成了离群点;在股票价格序列中,由于受某项政策出台或某种谣传的刺激,都会出现极增、极减现象,表现为序列中的离群点。

不论是何种原因引起的离群点,对以后的时间序列分析都会造成一定的影响。从造成分析中的困难来看,统计分析人员是不希望序列中出现离群点的,离群点会直接影响模型的拟合精度,甚至会造成一些虚假的信息。例如,两个相距很近的离群点将在谱分析中产生许多虚假的频率。因此,离群点往往被分析人员看作是一个“坏值”。但是,从获得信息来看,离群点提供了很重要的信息,它不仅提示我们认真检查采样中是否有差错,在进行时间序列分析前,认真确认序列,而且,当确认离群点是由于系统受外部突发因素刺激而引起的时候,它会提供关于系统稳定性、灵敏性等重要信息。

在时间序列分析中,通常把各种各样的离群点分为四种类型进行处理。第一种是加性离群点(additive outlier),造成这种离群点的干扰,只影响该干扰发生的那一个时刻 T 上的序列值,即 X_T ,而不影响该时刻以后的序列值 X_{T+1}, X_{T+2}, \dots ; 第二种是更新离群点(innovational outlier),造成这种离群点的干扰不仅作用于 X_T ,而且影响 T 时刻以后序列的所有观察值 X_{T+1}, X_{T+2}, \dots 它的出现意味着一个

外部干扰作用于系统的开始,并且其作用方式与系统的动态模型有关;第三种是水平移位离群点(level shift outlier),造成这种离群点的干扰是在某一时刻 T ,系统的结构发生了变化,并持续影响 T 时刻以后的所有行为,在数列上往往表现出 T 时刻前后的序列均值发生水平位移;第四种是暂时变更离群点(temporary change outlier),造成这种离群点的干扰是在 T 时刻干扰发生时具有一定初始效应,以后随时间根据衰减因子 δ 的大小呈指数衰减的一类干扰事件。

显然,在得到时间序列以后,首先要检验是否存在离群点,如果存在的话,还需要进一步判明在何时出现了离群点,以及所出现的离群点属于何种类型。检验离群点并对其进行处理的方法很多,但归纳起来大致有两类:一是根据数据取值进行检查,如果某一时刻的数值超出了一定的范围,则认为该点是一个离群点,并用一定的方法进行剔点(剔除离群点)处理;另一类是对数据进行模型分析,然后根据拟合模型后的剩余序列计算特定的统计量,测出显著的离群点及其类型,并用相应的模型进行修正,然后再对修正模型的剩余序列重复上述程序,依次测出各个离群点。关于后一类分析方法将在本书后面章节介绍,下面仅就前一类方法作一些简单介绍。

一种方法是将序列值与平滑值进行比较,检测其是否显著地大(或小)。这种方法是假定正常的序列值是平滑的,而离群点是突变的。我们用 \bar{X}_i 表示先对序列 X_i 进行平滑再平方得到的数值, $\overline{x_i^2}$ 表示先对序列 X_i 取平方再作平滑而得到的数值,并用 S_i^2 表示样本方差,有 $S_i^2 = \overline{x_i^2} - \bar{X}_i^2$, S_i 表示标准差,如果

$$\bar{X}_i - kS_i < X_{i+1} < \bar{X}_i + kS_i \quad (1.1)$$

则认为 X_{i+1} 是正常的,否则,认为 X_{i+1} 是一个离群点。 k 是统计分析研究人员根据序列和研究目的设定的一个常数,一般取 3~9 的整数,开始时不妨可取 $k=6$ 。

如果 X_{i+1} 是一个离群点,则可用 \hat{X}_{i+1} 来代替,即

$$\hat{X}_{i+1} = 2X_i - X_{i-1} \quad (1.2)$$

这实际上是线性外推。这种方法需要事先规定连续外推的次数,以免出现无休止的外推。因为连续检测到一些离群点后,最终的外推结果可能偏差很远,以致会排除本来是正常的数据点。

另一种方法是检测序列值与其相应的曲线平滑估计值的绝对离差是否大于某一预先设定值 k 。其具体检测步骤如下:首先根据序列 X_i 从首项开始取 5 项移动中位数生成一个新序列 X'_i ,即取 X_1, X_2, X_3, X_4, X_5 的中位数作为 X'_3 ,然后舍去 X_1 ,加入 X_6 取中位数得 X'_4 ,以此类推,直到加入序列最后一个数 X_N 。其次,用同样的方法根据序列 X'_i 从首项开始取 3 项移动中位数构成序列 X''_i 。再次,由序列 X''_i 按照如下公式构成序列 X'''_i :

$$X'''_i = \frac{1}{4} X''_{i-1} + \frac{1}{2} X''_i + \frac{1}{4} X''_{i+1} \quad (1.3)$$

最后,检测是否存在 $|X_t - X_t''| > k$ (其中 k 为预先设定值),若有,则该点是一个离群点。

三、缺损值的补足

在采集时间序列时,有时会由于仪器故障、操作失误、观测问题等种种原因,引起在某些观测点上未能记录下来观测值,这种缺少的观测值称为缺损值(missing value)。这种缺损值在使用次级资料时也会出现。当序列中存在缺损值时,就破坏了系统运行的连续性,违背了时间序列“顺序的重要性”原则。严格地说,我们不能依据一个“残缺”的序列进行分析,即使强制性地进行了分析,其结果也是无意义的,可是,由于时间的不可逆性,我们又无法重新观测,因此,我们所能做的就是依据其运动轨迹或变化趋势,运用一定的方法对缺损值进行估计、推测,以补足缺损的数值。具体估算、推测方法很多,如增长量推算法、发展速度推算法、比例推算法、平滑法、插值估算法等,分析人员可根据具体序列选择运用。鉴于这些方法已在先修课程中作过详细介绍,这里不再赘述。

此外,在对一序列进行分析之前,还要对序列中的每一个数据的指标口径、计算范围、计算方法、计量单位等进行认真检查。对经济时间序列来说,还必须检查计算价格等方面是否一致。若存在不一致,则要运用科学的方法进行调整,使整个序列中的每一个数据除时间属性不同之外,其所代表的实际意义完全一致。

时间序列分析不是做数学游戏,也不是做数学练习,因此,不能盲目地对一个序列建立模型进行分析。在建模分析之前,对所研究的时间序列进行认真检查和科学地预处理,也就是建立一个规范的时间序列是十分重要的,是做好时间序列分析的基础,分析研究人员千万不可忽视。

下面我们以前述我国人口自然增长率为例,简要说明建立时间序列的一般过程。就人口增长率来说,它是报告期人口的自然增减数与报告期内平均人口数之比。显然,相应于时间的连续性,它是时间 t 的连续函数,但是我们没有必要对每时每刻的人口变动都进行了解和掌握。从一般管理的需要和人口变动往往含有周期性这一特点考虑,该序列用自然年度(即1~12月)为采样间隔,也叫年度序列。既然是自然年度序列,那么序列中不仅不允许出现其他间隔(如半年、五年等)的序列值,而且年间隔也必须是自然年度,而不能是其他年度。

时间间隔确定之后,指标的计算范围在同一序列中也必须保持一致。该序列习惯叫“我国人口数”,实际上是以境内为核算范围的,而不是以领土内为核算范围的。也就是说,只是中国内地人口数,而不包括港、澳、台地区的人口数。因此,虽然1997年中国香港已经回归祖国,但是为了保持序列值计算范围的一致性,在计算1997年的人口自然增长率时,不能包含中国香港特别行政区的人口情况。就指标的计算方法来说,计算平均人口数的方法很多,用各种方法计算的结果也不完全

一致,在同一序列中,各年的平均人口数必须采用相同的计算方法,本序列是用(年初人口数+年末人口数) $\div 2$ 来计算的。

至于指标口径,也必须一致。如出生人数是指活产婴儿数,即离开母体有生命现象的活婴数总和。凡出生后有呼吸、心跳、脐带蠕动,随意肌抽动等一种现象,皆称为活产。出生后有上述一种生命现象但随后死亡者,也列入出生人数之中,同时也计入死亡人数之中。

这样,使序列中的每个数据除时间属性不同之外,其所代表的实际意义完全一致。

根据图 1.5 可以看出,1960 年的人口自然增长率极端异常,从数值上看为 -4.57% ,下面我们用第一种方法来检测该序列是否存在离群点。

我们对原序列作一次指数平滑。设平滑常数为 $\alpha=0.3$,并取 $k=6$,根据式 (1.1), $\bar{X}_t - kS_t$, $\bar{X} + kS_t$ 及 X_{t+1} 的数据如图 1.6 所示。

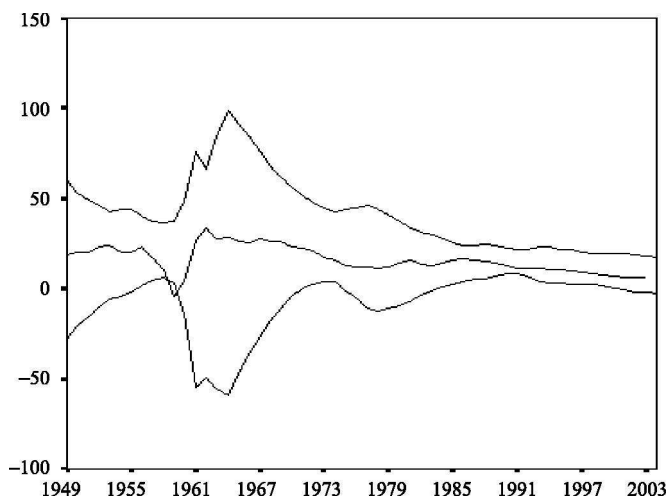


图 1.6 人口自然增长率的离群点检测图

可见,在 X_{12} 处, $X_{12} < \bar{X}_{11} - 6S_{11}$ 即 $-4.57\% < 2.96\%$,除此之外,所有的观测值均满足式 (1.1)。所以,我们认为 X_{12} 为一个离群点,根据式 (1.2),我们可以用

$$2X_{11} - X_{10} = 2 \times 10.19\% - 17.24\% = 3.14\%$$

代替原 X_{12} 序列的观测值 -4.57% 。

第三节 确定性时序分析方法概述

时间序列预测技术就是通过对预测目标自身时间序列的处理,来研究其变化

趋势的。当我们刚接触到某一个观测序列时,会觉得它是杂乱无章,无规律可循的。其实不然,大量事实表明,一个时间序列往往是以下几类变化形式的叠加或耦合:

(1)趋势变动。它是指时间序列朝着一定的方向持续上升或下降,或停留在某一水平上的倾向,它反映了客观事物的主要变化趋势。

(2)季节变动。即指一年或更短的时间之内,由于受某种固定周期性因素的影响而呈现出有规律的周期性波动。

(3)循环变动。通常是指周期为一年以上的有规律的波动。

(4)不规则变动。通常它分为突然变动和随机变动。所谓突然变动是指战争、自然灾害或是其他社会因素等意外事件引起的变动。随机变动是指由于大量的随机因素产生的宏观影响。根据中心极限定理,通常认为随机变动近似服从正态分布。

通常用 T_t 表示长期趋势项, S_t 表示季节变动趋势项, C_t 表示循环变动趋势项, R_t 表示随机干扰项。常见的确定性时间序列模型有以下几种类型:

(1)加法模型

$$y_t = T_t + S_t + C_t + R_t \quad (1.4)$$

(2)乘法模型

$$y_t = T_t \cdot S_t \cdot C_t \cdot R_t \quad (1.5)$$

(3)混合模式

$$y_t = T_t \cdot S_t + R_t \quad (1.6)$$

$$y_t = S_t + T_t \cdot C_t \cdot R_t$$

其中, y_t 是观测目标的观测记录; $E(R_t) = 0$, $E(R_t^2) = \sigma^2$ 。

如果在预测时间范围以内,无突然变动且随机变动的方差 σ^2 较小;并且有理由认为过去和现在的历史演变趋势将继续发展到未来时,可用一些经验方法进行预测,具体方法如下:

1. 移动平均法

设观测序列为 y_1, \dots, y_T , 正整数 $N < T$ 。一次移动平均值计算公式为

$$\begin{aligned} M_t^{(1)} &= \frac{1}{N}(y_t + y_{t-1} + \dots + y_{t-N+1}) \\ &= \frac{1}{N}(y_{t-1} + \dots + y_{t-N}) + \frac{1}{N}(y_t - y_{t-N}) \\ &= M_{t-1}^{(1)} + \frac{1}{N}(y_t - y_{t-N}) \end{aligned} \quad (1.7)$$

二次移动平均值计算公式为

$$\begin{aligned} M_t^{(2)} &= \frac{1}{M} (M_t^{(1)} + \cdots + M_{t-N+1}^{(1)}) \\ &= M_{t-1}^{(2)} + \frac{1}{N} (M_t^{(1)} - M_{t-N}^{(1)}) \end{aligned} \quad (1.8)$$

当预测目标的基本趋势是在某一水平上上下波动时,可用一次移动平均方法建立预测模型

$$\hat{y}_{t+m} = M_t^{(1)} = \frac{1}{N} (\hat{y}_t + \cdots + \hat{y}_{t-N+1}), \quad m = 1, 2, \cdots \quad (1.9)$$

它表明最近 N 期序列值的平均值作为未来各期的预测结果。一般 N 的取值范围是: $5 \leq N \leq 200$ 。当历史序列的基本趋势变化不大且序列中随机变动成分较多时 N 的取值应较大一些,否则 N 的取值应小一些。在有确定的季节变动周期的资料中,移动平均的项数应取周期长度。选择最佳 N 值的一个有效方法是,比较若干模型的预测误差,均方预测误差最小者为好。

当预测目标的基本趋势与某一线性模型相吻合时,常用二次移动平均法,但序列同时存在线性趋势与周期波动时,可用趋势移动平均法建立预测模型

$$\hat{y}_{t+m} = a_r + b_r m, \quad m = 1, 2, \cdots \quad (1.10)$$

其中

$$\begin{aligned} a_r &= 2M_r^{(1)} - M_r^{(2)} \\ b_r &= \frac{2}{N-1} (M_r^{(1)} - M_r^{(2)}) \end{aligned}$$

上述移动平均法在数据处理中常用它作为预处理,消除周期波动(取 N 为周期长度)和减弱随机干扰的影响往往是很有效的。

2. 指数平滑法

一次移动平均实际上认为最近 N 期数据对未来值影响相同,都加权 $\frac{1}{N}$;而 N 期以前的数据对未来值没有影响,加权为 0。但是,二次及更高次移动平均数的权数却不是 $\frac{1}{N}$,且次数越高,权数的结构越复杂,但永远保持对称的权数,即两端项权数小,中间项权数大,不符合一般系统的动态性。一般来说历史数据对未来值的影响是随时间间隔的增长而递减的。所以,更切合实际的方法应是对各期观测值依时间顺序进行加权平均作为预测值。指数平滑法可满足这一要求,而且具有简单的递推形式。

设观测序列 y_1, \cdots, y_t , α 为加权系数, $0 < \alpha < 1$, 一次指数平滑公式为

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha(y_t - S_{t-1}^{(1)}) \quad (1.11)$$

假定历史序列无限长,则式(1.11)可写为

$$\begin{aligned} S_t^{(1)} &= \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha) S_{t-2}^{(1)}] \\ &= \dots \\ &= \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j y_{t-j} \end{aligned} \quad (1.12)$$

式(1.12)表明 $S_t^{(1)}$ 是全部历史数据的加权平均,加权系数分别为 α , $\alpha(1-\alpha)$, $\alpha(1-\alpha)^2, \dots$,显然有

$$\sum_{j=0}^{\infty} \alpha(1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1$$

由于加权系数序列呈指数函数衰减,加权平均又能消除或减弱随机干扰的影响,所以式(1.11)称为指数平滑,类似地,二次指数平滑公式为

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)} \quad (1.13)$$

同理,三次指数平滑公式为

$$S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha) S_{t-1}^{(3)} \quad (1.14)$$

一般 p 次指数平滑公式为

$$S_t^{(p)} = \alpha S_t^{(p-1)} + (1 - \alpha) S_{t-1}^{(p)} \quad (1.15)$$

利用指数平滑公式可以建立指数平滑预测模型。原则上说,不管序列的基本趋势多么复杂,总可以利用高次指数平滑公式建立一个逼近很好的模型,但计算量很大。因此用的较多的是几个低阶指数平滑预测模型。

(1)水平趋势预测模型

$$\hat{y}_{T+m} = S_T^{(1)}, \quad m = 1, 2, \dots \quad (1.16)$$

(2)线性趋势预测模型——Brown 单系数线性平滑预测

$$\hat{y}_{T+m} = a_T + b_T m, \quad m = 1, 2, \dots \quad (1.17)$$

其中, $a_T = 2S_T^{(1)} - S_T^{(2)}$, $b_T = \frac{\alpha}{1-\alpha}(S_T^{(1)} - S_T^{(2)})$ 。

(3)二次曲线趋势预测模型——Brown 单系数二次式平滑预测

$$\hat{y}_{T+m} = a_T + b_T m + \frac{1}{2} c_T m^2, \quad m = 1, 2, \dots \quad (1.18)$$

其中

$$a_T = 3S_T^{(1)} - 3S_T^{(2)} + S_T^{(3)}$$

$$b_T = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_T^{(1)} - 2(5-4\alpha)S_T^{(2)} + (4-3\alpha)S_T^{(3)}]$$

$$c_T = \frac{\alpha^2}{(1-\alpha)^2} [S_T^{(1)} - 2S_T^{(2)} + S_T^{(3)}]$$

指数平滑预测模型是以当前时刻 T 为起点,综合历史序列的信息,对未来进行预测的。选择合适的加权系数 α 是提高预测精度的关键环节。根据实践经验, α 的取值范围一般以 $0.1 \sim 0.3$ 为宜。如何进一步确定 α 的最佳取值,通常要结合理论分析和模型对比的方法来进行。由式(1.11)不难看出

$$\begin{aligned}\hat{y}_{t+1} &= S_t^{(1)} = S_{t-1}^{(1)} + \alpha(y_t - S_{t-1}^{(1)}) \\ &= \hat{y}_t + \alpha y_t'\end{aligned}$$

其中, $y_t' = y_t - \hat{y}_t$ 它表明指数平滑预测是用预测误差对上期预测值的修正, α 的大小体现了修正的幅度。由式(1.9)可知, α 值越大,加权系数序列衰减速度越快,所以实际上 α 取值大小起着控制参加平均的历史数据的个数的作用。 α 值越大意味着采用的数据越少。由此,可以得到选择 α 值的一些基本准则:

(1)如果序列的基本趋势比较稳,预测偏差由随机因素造成,则 α 值应取小一些,以减少修正幅度,使预测模型能包含更多历史数据的信息。

(2)如果预测目标的基本趋势已发生系统地变化,则 α 值应取得大一些。这样,可以偏重新数据的信息对原模型进行大幅度修正,以使预测模型适应预测目标的新变化。

上述原则可结合模型对比方法来进行。通常将历史数据分成两段,前一段 y_1, \dots, y_k 用于建立预测模型,第二段 y_{k+1}, \dots, y_T 用于事后预测,以事后预测误差平方和为评价标准,确定最佳 α 值。

另外,由于指数平滑公式是递推计算公式,所以必须确定初始值 $S_0^{(1)}, S_0^{(2)}, S_0^{(3)}$ 。初始值实质上应该是序列起始点 $t=0$ 以前所有历史数据的加权平均值。但在实际工作中,由于获得历史数据多少的不同,往往采用经验方法来确定。我们可以通过在最初预测时,选择较高的 α 值来减小由初始值选择不当所造成的预测偏差,以使预测模型迅速调整到当前水平。

例如, $\alpha=0.5$, 一般取前 $3 \sim 5$ 个数据的算术平均值作为初始值

$$S_0^{(1)} = S_0^{(2)} = S_0^{(3)} = \frac{1}{3}(y_1 + y_2 + y_3)$$

为了克服初始值选择不当,最初预测选择较大 α 值

$$\alpha = \frac{1}{3}, \alpha = \frac{1}{4}, \alpha = \frac{1}{5}, \alpha = \frac{1}{6}, \alpha = 0.15, \alpha \equiv 0.15 \quad (t > 5)$$

3. 时间回归法

虽然指数平滑法应用广泛,但当序列具有其他类型变化形式时,指数平滑法计