



普通高等教育“九五”国家级重点教材

时间序列分析



主编 王振龙 主审 顾岚

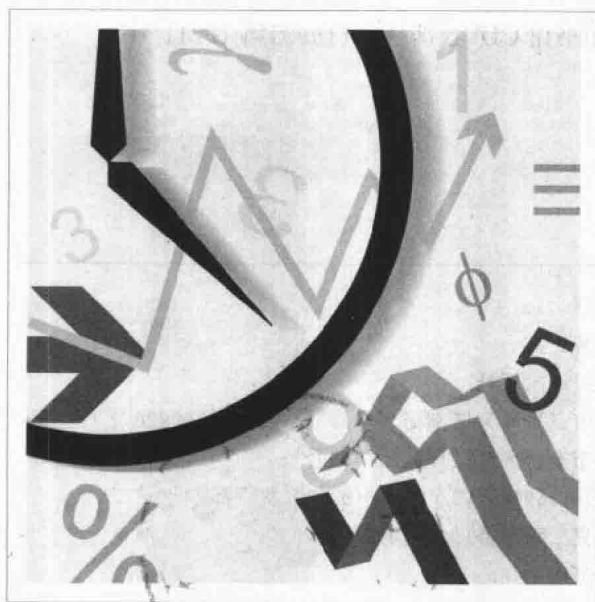


中国统计出版社



普通高等教育“九五”国家级重点教材

时间序列分析



主编 王振龙 主审 顾岚

中国统计出版社

(京)新登字 041 号

图书在版编目(CIP)数据

时间序列分析/王振龙主编.

—北京:中国统计出版社,2000.2

高等院校统计学专业规划教材

ISBN 7-5037-2926-0

I. 时… II. 王… III. 时间序列分析-高等学校-教材 IV. 0211.61

中国版本图书馆 CIP 数据核字(1999)第 10211 号

责任编辑/徐颖

责任校对/刘开颜

出版发行/中国统计出版社

通信地址/北京市三里河月坛南街 75 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

电 话/(010)63459084、63266600—22500(发行部)

印 刷/科伦克三莱印务(北京)有限公司

经 销/新华书店

开 本/787×1092mm 1/18

字 数/270 千字

印 张/14.5

印 数/1—5000 册

版 别/2000 年 2 月第 1 版

版 次/2000 年 2 月北京第 1 次印刷

书 号/ISBN 7-5037-2926-0/O. 32

定 价/24.00 元

中国统计版图书,版权所有,侵权必究。

中国统计版图书,如有印装错误,本社发行部负责调换。

出版说明

根据原国家教育委员会《普通高等教育“九五”国家级重点教材立项、管理办法》的要求,国家统计局经过专家评审,五种教材被立项,其中:三种教材是在“八五”期间规划统计教材的基础上进行修订的,二种教材属“九五”期间新编规划统计教材,这五本教材的编审工作由全国统计教材编审委员会负责审定。

“抓重点,出精品”是“九五”期间普通高等教育教材建设与改革工作的核心。按照教育部的要求,国家级重点教材都应建设成“九五”普通高等教育的精品教材。根据这一精神,这批教材力求适应我国政治、经济、科技、教育等改革的形势,充分反映改革的成果,同时适应专业目录调整及专业面拓宽以后教学改革的实际需要,科学、合理地设置教材体系和安排教材内容,努力提高教材质量。

相信通过这批教材的出版、发行,对我国普通高等教育统计教材建设工作将起到较好的示范、导向作用;对提高统计教育水平和培养面向 21 世纪的统计人才也将发挥积极的促进作用。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足之处,诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

1999 年 3 月

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够了,至于其制造原理及过程,当然熟悉更好,不了解也无关乎其使用。鉴于这样的认识,全书没有运用深奥的定理,因而也就勿需定理证明。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。

为了使读者通过建模练习来掌握时间序列分析的基本思路和方法,书中除了给出“时间序列分析软件”(TSP)的使用说明外,还搜集了近 20 个实际数据,以供读者练习之用。

当然,说本书强调应用,是相对于数学定理推证而言的,在阐述中力图做到理论严谨,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证明手段,在用描述语言和系统机理表述模型结构及其原理和特性时,可能从数学的角度看会有不准确、不严密之处,恳请读者批评指正。

本书由王振龙同志任主编,全书共八章。其中第一、二、三章由王振龙编写;第四、五、八章由胡永宏编写;第六章由胡永宏和王佐仁编写;第七章由王振龙和王佐仁编写。全书由王振龙总纂、修改定稿。

初稿形成后由顾岚教授主审,提出了很多宝贵的修改意见,同时在编写过程中参阅了大量的资料,并引用了一些现成的数据和科研成果,除书中提到的主要参考书目以外,还需提到的是美国芝加哥大学商学院刁锦寰教授、美国费城大学管理学院陈江教授和台湾辅仁大学管理学院统计系主任谢邦昌教授 1997 年 11 月在国家统计局教育中心和中国经济教育学会举办的“时间序列分析高级研究班”上的讲座,对我们有很大的启发和帮助,其它资料恕不列举。此外,国家统计局教育中心的领导和教材处同志以及我的同仁和学生给予了很大的支持和帮助,在此,一并致以真诚的感谢。

王振龙

1999 年 7 月于西安

第三节	模型参数估计	(103)
第四节	模型的适应性检验	(108)
第五节	建模的其它方法	(113)
第六节	实例	(118)
第五章	平稳时间序列预测	(126)
第一节	正交投影预测(几何预测法)	(126)
第二节	条件期望预测	(129)
第三节	适时修正预测	(137)
第四节	指数平滑预测—ARMA 模型特例	(139)
第六章	非平稳时间序列分析	(143)
第一节	非平稳性的检验	(144)
第二节	平稳化方法	(151)
第三节	齐次非平稳序列模型	(159)
第四节	非平稳时间序列的组合模型	(168)
第七章	季节性时间序列分析方法	(181)
第一节	简单随机时序模型	(181)
第二节	乘积季节模型	(184)
第三节	季节时序模型的建立	(186)
第四节	X-11 方法简介	(193)
第八章	传递函数模型	(200)
第一节	模型简介	(201)
第二节	传递函数模型的识别	(204)
第三节	传递函数模型的拟合与检验	(211)
附录 I	延伸自相关函数	(217)
附录 II	TSP 软件简介	(224)
附录 III	数据资料	(230)
附录 IV	常用统计量分布表	(243)
	主要参考书目	(250)

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 中列出的是某地电风扇 1994 年到 1996 年间各月的销售量(单位:万台),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 电风扇月销售量

单位:万台

年份 \ 月份	1	2	3	4	5	6
1994 年	10	20	41	64	148	230
1995 年	16	20	58	90	139	235
1996 年	15	23	66	91	140	253

年份 \ 月份	7	8	9	10	11	12
1994 年	203	127	41	22	16	12
1995 年	240	89	42	23	16	17
1996 年	198	96	78	50	25	19

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,例如长度、温度、速度等等。

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 (i 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

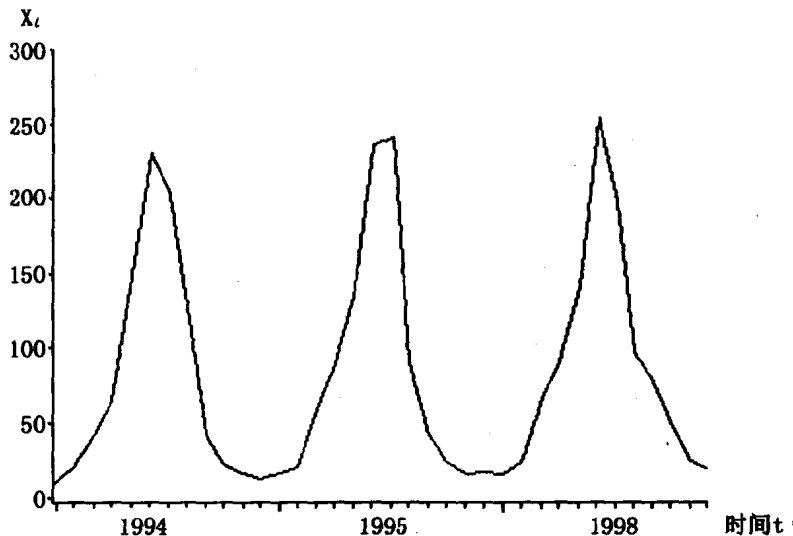


图 1.1 电风扇月销售量数据图

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，速度或其它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，也是一个时间序列（见附录Ⅲ），其散点图见图 1.2。

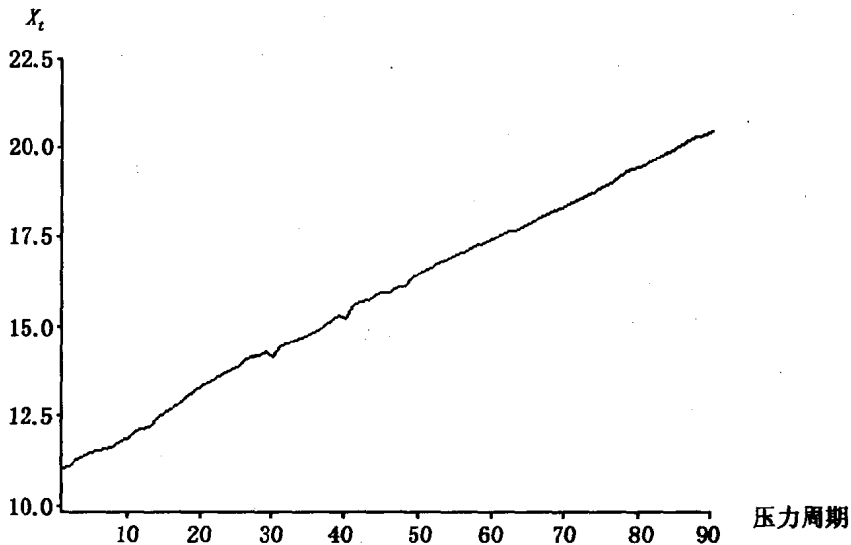


图 1.2 某材料裂纹长度

而且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种商品销售量这一数列,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。

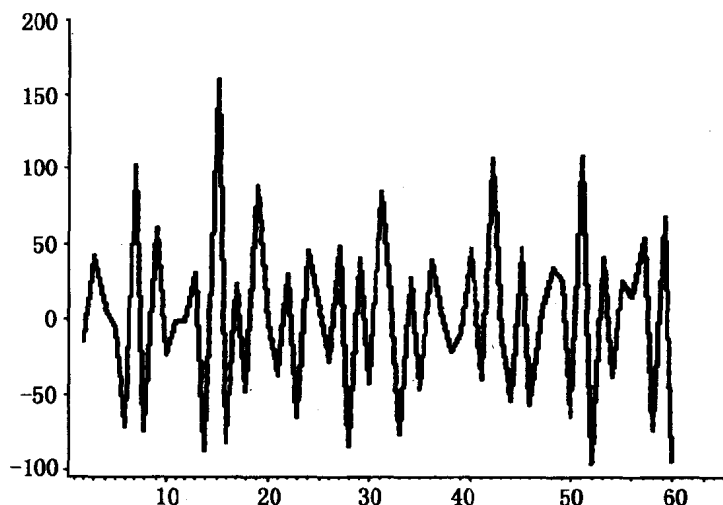


图 1.3 列车运行数量平稳化后数据图

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)的时间序列叫做高斯型时间序列,否则叫做非高斯型时间序列。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,人们便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。

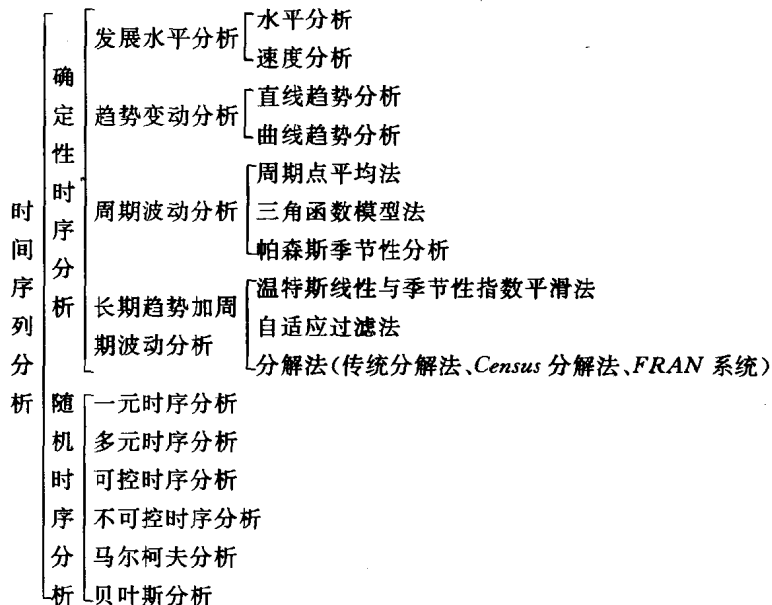
随着科学技术的不断发展,人们在实践中认识到时间序列的变动,主要是由长期趋势(随着时间的变化,按照某种规则稳步地增长、下降或保持在某一水平上)、季节变动(在一个年度内依一定周期规则性变化)、循环波动(以若干年为周期的波动变化)和随机型变动(许多不可控的偶然因素共同作用的结果)而形成的。前三种变动的一个共同特点,就是依一定的规则而变化,随机波动在综合中可以消除。基于这种认识,时间序列分析就是设法消除随机型波动,拟合确定型趋势,因而形成了长期趋势分析、季节变动分析和循环波动测定等一系列确定型时间序列分析方法。

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。自从本世纪 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。时间序列与其它变量数列不同,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它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它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它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效的工具。一元时序模型不象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. 对理论性模型与数据进行适度检验,以讨论模型是否能正确地表示所观测的现象。
2. 刻划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合的。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