SHIJIAN XULIE FENXI
JIQI YINGYONG

时间序列分析及其应用

王沁/编著

西南交通大学出版社
Http://press.swjtu.edu.cn

时间序列分析及其应用

王　沁　编著

西南交通大学出版社
·成　都·

图书在版编目（CIP）数据

时间序列分析及其应用 / 王沁编著. —成都：西南交通大学出版社，2008.11
ISBN 978-7-5643-0111-8

Ⅰ. 时… Ⅱ. 王… Ⅲ. 时间序列分析 Ⅳ. 0211.61
中国版本图书馆 CIP 数据核字（2008）第 162854 号

时间序列分析及其应用

王　沁　编著

责任编辑	张宝华
封面设计	本格设计
出版发行	西南交通大学出版社 （成都二环路北一段 111 号）
发行部电话	028-87600564　028-87600533
邮　　编	610031
网　　址	http: //press.swjtu.edu.cn
印　　刷	四川森林印务有限责任公司
成品尺寸	146 mm×208 mm
印　　张	7.937 5
字　　数	220 千字
印　　数	1—3 000 册
版　　次	2008 年 11 月第 1 版
印　　次	2008 年 11 月第 1 次印刷
书　　号	ISBN 978-7-5643-0111-8
定　　价	18.00 元

前　言

《时间序列分析及其应用》一书是作者多年来在西南交通大学从事该门课程的教学与研究的基础上，专为高等院校学生编写的教材.

时间序列分析是概率统计学科中应用性较强的一个分支，在金融经济、气象水文、信号处理、工程技术领域等众多领域有着广泛的应用. 近几年来，时间序列分析引起了国内外学者及科研和管理人员的极大兴趣，特别是随着统计软件的开发和应用，广大工程技术与管理人员希望掌握时间序列分析方法，并利用时间序列分析的方法探索社会经济现象的动态结构，预测发展规律，从而对未来状态进行控制.

本书以时间序列的统计特征和建模步骤为主线，共分八章，内容包括平稳时间序列模型的统计特征、平稳时间序列模型的建立与预测、时间序列的确定性分析与随机性分析、波动率模型等，并系统介绍了时间序列的基本理论、建模和预测方法以及应用. 本书旨在将实际应用与理论推导联系起来，通过详细的建模步骤和流程图、习题、软件操作步骤以及实际时间序列分析实例，尽量体现时间序列分析的综合性和整体性，从而使得统计专业和其他工程专业、管理专业的学生掌握时间序列分析的理论和应用.

本书力求突出以下几点:

（1）着眼于用概率论与数理统计的观点研究与探讨若干基本模型 的统计特征，将实际应用与理论推导联系起来;

（2）着眼于建模步骤所蕴含的基本思想和技巧，使用大量时间序列分析实例，尽量体现时间序列分析的应用性;

（3）着重于揭示传统时间序列模型的统计特征，便于读者领悟时间序列分析的思想. 现代时间序列分析包括向量自回归模型，门

限自回归模型、动量计量模型，以及一些非线性模型等内容，但本书立足于基础的和传统的时间序列模型，立足于读者领悟传统时间序列模型的特性及其具体应用的方法. 关于现代时间序列分析的相关知识内容没有作介绍.

(4) 对于基本概念，一方面深入浅出，强调从直观的角度来理解；另一方面从统计的角度理解，尽量体现时间序列分析的综合性和整体性.

(5) 详细介绍了使用 Eviews 软件的操作方法，强调如何利用软件对时间序列拟合最佳数学模型并进行相应的预测.

本书是应用时间序列分析的入门教材，仅以初等概率论及高等数学、线性代数为基础，可作为综合性大学、工科大学和高等师范院校本科生的时间序列分析课程的教材，也可以作为信息工程、通信工程、控制工程、电子信息类等相关专业的本科教材或教学参考用书.

本书的编写与修改得到了同仁和学生的鼓励、帮助与支持，特别是马洪教授、李裕奇教授、何平教授等给予了很多的鼓励、关心与支持. 研究生陈志杰为本书的录入做了很多工作. 本书在编写和出版的过程中，得到了西南交通大学数学学院概率统计系、教务处教材科与西南交通大学出版社的大力支持与帮助，作者在此表示由衷的感谢.

限于作者水平，本书的错误与不妥在所难免，恳请同行及读者批评指正.

目　　录

第一章 绪 论

在工农业生产、科学技术和社会经济生活的许多领域，普遍存在着按时间顺序发生的具有概率特征的各种随机现象. 按照时间顺序把随机现象变化发展的过程记录下来就构成了时间序列的一次观察.

对时间序列进行观察、研究，提取有用的信息，以便找出客观事物发展的规律，预测其发展趋势，并进行必要的控制就是时间序列分析. 时间序列分析是数理统计这一学科中应用性较强的一个分支，在金融经济、气象水文、信号处理、机械振动等众多领域有着广泛的应用.

第一节 时间序列

一、随机过程与时间序列

随机过程是对一族随机变量动态关系的定量描述，是研究随“时间”变化的、“动态”的、“整体”的随机现象的统计规律性.

从数学意义上来讲，Ω 为随机试验 E 的样本空间，T 为实数集的子集，如果对每个参数 $t \in T$，$X(e,t)$ 为样本空间 Ω 上的一个随机变量，对每一个 $e \in \Omega$，$X(e,t)$ 为 t 的函数，那么，$\{X(e,t), t \in T, e \in \Omega\}$ 称为随机过程，简记为 $\{X(t), t \in T\}$.

参数 t 的变化范围 T，称为随机过程的参数集. 对于一切 $e \in \Omega, t \in T$，$X(e,t)$ 的全部可能的取值的集合，称为随机过程的状态集，记为 I.

随机过程的参数集T可以分为离散集与连续集，状态集I亦可分为离散集与连续集. 这样一来，随机过程可分为以下4类：

（1）连续参数集，连续状态集随机过程.

（2）连续参数集，离散状态集随机过程.

（3）离散参数集，连续状态集随机过程.

（4）离散参数集，离散状态集随机过程.

一般地，称状态空间离散的随机过程为链，参数空间离散的随机过程为随机序列. 由于参数集T通常表示时间，所以，随机序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$通常又称为时间序列. 通常意义下的时间序列是指参数集为离散的随机过程，是指能用有限维参数模型来描述的一类特殊随机序列.

对时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$，取一系列时间点$t_1<t_2<\cdots<t_N$，$t_i\in T=\{0,\pm1,\pm2,\cdots\}$进行观察，观察值按时间先后顺序排列得到$\{x(i),i=1,2,\cdots,N\}$，这样就形成了时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的一次观察（或实现）. 实际工作中常见的按季度、月、周、日统计的商品销量、销售额或库存量，按年统计的一个省、市或国家的国民生产总值、人口出生率等都是时间序列的一次观察. 从统计意义上讲，时间序列是变量在某一时间段内不同时间点上观测值的集合，而且这些观测值按时间先后顺序排列. 从经济到工程技术，从天文到地理和气象，几乎在各种领域都会遇到时间序列.

值得注意的是，时间序列中“时间”可以指时间，也可以指长度、温度等具有顺序的其他物理量.

时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的一次观察$\{x(i),i=1,2,\cdots,N\}$所得到的数据，实际上是$N$维随机变量$\{X(t_1),X(t_2),\cdots,X(t_N)\}$的一次观察. 这些数据依赖于时间点和时间序列的统计特征而变化，并按时间先后顺序排列，呈现一定的相关性，而且数据的相关性在整体上呈现某种趋势性或周期性变化，反映了时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$随“时间”变化的、“动态”的、“整体”的统计规律性，包含了产生该时间序列的系统的历史行为的全部信息.

二、时间序列的分布与数字特征

由于时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$在任意时刻$t$的状态是随机变量，所以，可以利用随机变量的一维、多维分布和数字特征来描述时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的统计特征.

（1）有限维分布函数：$\{X(t),t=0,\pm1,\pm2,\cdots\}$为实的时间序列（本书只讨论实序列，今后不再申明），参数集$T=\{0,\pm1,\pm2,\cdots\}$，对任意n个时刻$t_1,t_2,\cdots,t_n\in T$，及实数$x_1,x_2,\cdots,x_n\in \mathbf{R}$，称

$$F_n(x_1,x_2,\cdots,x_n,t_1,t_2,\cdots,t_n)=P(X(t_1)\leqslant x_1,X(t_2)\leqslant x_2,\cdots,X(t_n)\leqslant x_n)$$

为时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的$n$维分布函数.

关于时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的所有有限维分布函数的集合：

$$\{F_n(x_1,x_2,\cdots,x_n,t_1,t_2,\cdots,t_n),t_1,t_2,\cdots,t_n\in T=\{0,\pm1,\pm2,\cdots\},n\geqslant 1\}$$

为时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的有限维分布函数族. 时间序列的有限维分布函数族完全刻画了时间序列的统计特征.

（2）均值函数，记为$m_X(t)$：若对于任意$t\in T=0,\pm1,\pm2,\cdots$，$EX(t)$存在，则

$$m_X(t)=EX(t)$$

（3）均方值函数，记为$\psi_X^2(t)$：若对于任意$t\in T=0,\pm1,\pm2,\cdots$，$EX^2(t)$存在，则

$$\psi_X^2(t)=EX^2(t)$$

（4）方差函数，记为$D_X(t)$或$\mathrm{Var}(X)$：若对于任意$t\in T=0,\pm1,\pm2,\cdots$，$E(X(t)-m_X(t))^2$存在，则

$$D_X(t)=E(X(t)-m_X(t))^2=EX^2(t)-m_X^2(t)$$

（5）自相关函数，记为$\gamma_X(t_1,t_2)$：若对于任意$t_1,t_2\in T=0,\pm1,\pm2,\cdots$，$E[X(t_1)X(t_2)]$存在，则

$$\gamma_X(t_1,t_2)=E[X(t_1)X(t_2)]$$

（6）自协方差函数，记为$C_X(t_1,t_2)$：若对于任意$t_1,t_2\in T=0,\pm1,\pm2,\cdots$，$E[(X(t_1)-m_X(t_1))(X(t_2)-m_X(t_2))]$存在，则

$$\begin{aligned}C_X(t_1,t_2)&=E[(X(t_1)-m_X(t_1))(X(t_2)-m_X(t_2))]\\&=\gamma_X(t_1,t_2)-m_X(t_1)m_X(t_2)\end{aligned}$$

（7）自相关系数，记为$\rho_X(t_1,t_2)$：若对于任意$t_1,t_2\in T=0,\pm1,\pm2,\cdots$，$EX^2(t)<+\infty$，则

$$\rho_X(t_1,t_2)=\frac{C_X(t_1,t_2)}{\sqrt{D_X(t_1)D_X(t_2)}}$$

如果时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的任意$n(n\geqslant1)$维随机变量$(X(t_1),X(t_2),\cdots,X(t_n))$都服从高斯（正态）分布，则称为高斯型时间序列. 此时，n维随机变量$(X(t_1),X(t_2),\cdots,X(t_n))$的概率密度为

$$f(x_1,x_2,\cdots,x_n)=(2\pi)^{-\frac{n}{2}}|\boldsymbol{C}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

其中

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)',\quad x_1,x_2,\cdots,x_n\in\mathbf{R}$$
$$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)',\quad \mu_i=EX(t_i)\quad(i=1,2,\cdots,n)$$
$$\boldsymbol{C}=(C_{ij})_{n\times n},\quad C_{ij}=\mathrm{Cov}(X(t_i),X(t_j))\quad(i,j=1,2,\cdots,n)$$

如果时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$的概率分布不随时间的变化而变化，即对任意$\varepsilon$，任意$n\in N$，任意$t_1,t_2,\cdots,t_n\in T$，任意$x_1,x_2,\cdots,x_n\in\mathbf{R}$，有

$$\begin{aligned}&P(X(t_1)\leqslant x_1,X(t_2)\leqslant x_2,\cdots,X(t_n)\leqslant x_n)\\&=P(X(t_1+\varepsilon)\leqslant x_1,X(t_2+\varepsilon)\leqslant x_2,\cdots,X(t_n+\varepsilon)\leqslant x_n)\end{aligned}$$

则称该时间序列为严平稳时间序列.

如果时间序列$\{X(t),t=0,\pm1,\pm2,\cdots\}$满足以下三个条件：

（1）时间序列的均方值函数是存在的，即对任意$t\in T=0,\pm1,$

$\pm 2, \cdots$，有

$$\psi_X^2(t) = EX^2(t) < \infty$$

（2）时间序列的均值函数恒为常数，即对任意 $t \in T = 0, \pm 1, \pm 2, \cdots$，有

$$E[X(t)] = \mu$$

（3）时间序列的自协方差函数是时间间隔的函数，即对任意 $t, s \in T = 0, \pm 1, \pm 2, \cdots$，$\tau = s - t$，有

$$\mathrm{Cov}(X_t, X_s) = E(X_t - \mu)(X_s - \mu) = c(\tau)$$

则称该时间序列为宽平稳时间序列.

如果时间序列的均方值函数存在，则严平稳时间序列一定是宽平稳时间序列. 通常意义上的平稳时间序列都是指宽平稳时间序列.

若时间序列 $\{\varepsilon(t), 0, \pm 1, \pm 2, \cdots\}$ 满足：

$$E\varepsilon_t = 0$$

$$\mathrm{Cov}\,(\varepsilon_t, \varepsilon_s) = E\varepsilon_t\varepsilon_s = \begin{cases} \sigma_\varepsilon^2, & t = s \\ 0, & t \neq s \end{cases}$$

则称 $\{\varepsilon(t), t = 0, \pm 1, \pm 2, \cdots\}$ 为白噪声，表示为 $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$. 若 $\{\varepsilon_t\}$ 是独立同分布，均值为零、有限方差为 σ_ε^2 的白噪声，则表示为 $\{\varepsilon_t\} \sim \mathrm{IID}(0, \sigma_\varepsilon^2)$. 若 $\{Z_t\}$ 是独立同正态分布、均值为零、有限方差为 σ_ε^2 的白噪声，则表示为 $\{\varepsilon_t\} \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$.

白噪声是一类典型的平稳时间序列，其本质特点是：时刻 t 的随机变量 ε_t 与另一时刻 s 的随机变量 ε_s 是互不相关，不存在线性关系.

设 $\{X(t), t = 0, \pm 1, \pm 2, \cdots\}$ 与 $\{Y(t), t = 0, \pm 1, \pm 2, \cdots\}$ 为两个时间序列，则称 $\{(X(t), Y(t)), t = 0, \pm 1, \pm 2, \cdots\}$ 为二元时间序列；设 $\{X_i(t), t = 0, \pm 1, \pm 2, \cdots\}$ $(i = 1, 2, \cdots, n)$ 为 n 个时间序列，则称 $\{(X_1(t), X_2(t), \cdots, X_n(t)), t = 0, \pm 1, \pm 2, \cdots\}$ 为 n 元时间序列. 所以，时间序列通常分为一元时间序列和多元时间序列. 相应的，有一元时间序列分析和多元时间序列分析. 本书只讨论一元时间序列分析，今后不再申明.

三、时间序列的变动因素

随着科学技术的不断发展，人们在实践中逐渐认识到时间序列的变动，主要是由长期趋势变动、季节性变动、循环变动和不规则变动而形成的.

长期趋势变动（T：secular trend）是指时间序列在较长持续期内受某种基本因素的影响，数据依时间变化时展现出来的总态势. 具体表现为不断增加或不断减少的基本趋势，也可以表现为只围绕某一常数值波动而无明显增减变化的水平趋势. 例如，每年死亡率，因为医疗技术的改进及生活水平的提高而有长期递减的趋势.

季节性变动（S：seasonal variation）是指时间序列的观察值，由于受自然季节因素或节假日的影响，在一年中或固定时间内，呈现固定的规则变动. 例如，医院住院患者人数，每逢星期一都出现一个高峰，而星期六将出现一个低谷，呈现类似于 7 天的周期性规律. 季节性变动的周期小于或等于一年，通常为一年、一月、一周等.

循环变动（C：cyclical variation）是指时间序列的观察值以若干年、十几年、甚至几十年为周期，呈上升与下降交替出现的循环往复运动. 循环变动的周期为 2～15 年. 其变动的原因甚多，周期的长短与幅度亦不一致. 通常一个时间序列的循环是由其他多个小的循环组合而成. 例如，总体经济指标的循环往往是由各个产业的循环组合而成；经济膨胀往往在循环的顶点，而经济萧条则在循环的谷底.

不规则变动（或随机变动）（I：irregular variation）是指时间序列由于受偶然不可控因素的影响，而表现出的不规则波动.

总之，时间序列是上述四种或其中几种变动因素的综合作用的结果.

第二节 时间序列分析

一、时间序列分析的方法

时间序列中，一个典型的本质特征就是相邻观察值之间的相关性. 时间序列观察值之间的这种相关性具有很多的实际意义. 通过对时间序列 $\{X(t), t=0,\pm1,\pm2,\cdots\}$ 的一次观察 $\{x(i), i=1,2,\cdots,N\}$ 的研究，可以认识所研究的时间序列的统计特征和结构特征，进而揭示时间序列的运行规律，预测其发展趋势并进行必要的控制. 用来实现上述目的的整个方法称为时间序列分析（time series analysis）.

传统时间序列分析认为长期趋势变动、季节性变动、循环变动是依一定的规则而变化的，不规则变动因素在综合中可以消除. 基于这种认识，形成了确定性时间序列分析. 确定性时间序列分析就是设法消除不规则变动因素，拟合长期变动趋势，分析季节变动和循环波动因素等一系列确定性的方法.

随着科学技术的不断发展，人们逐渐认识到，虽然长期趋势的分析、季节变动的分析和循环波动的分析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的全貌，而且用随机过程理论和统计理论来考察长期趋势变动、季节性变动等许多因素的共同作用的时间序列更具有合理性和优越性. 根据随机过程理论和统计理论，对时间序列进行分析，进而形成了统计时间序列分析.

从所采用的数学工具和理论来看，通常将统计时间序列分析分为时域（time domain）分析和频域（frequency domain）分析两大类.

（1）时域（time domain）分析主要是从序列的自相关角度来揭示时间序列的发展规律的. 常用的数学工具有自相关系数、偏相关系数、差分方程理论等，其理论基础扎实，操作步骤规范，分析结果易于解释，是时间序列分析的主流方法.

时域分析又称为随机时间序列分析，常采用的手段有数据图法、指标法和模型法三类. 数据图法是以时间为横轴，以时间序列在 t 时

刻的观察值为纵轴，并在平面坐标系中绘出曲线图，再根据图形直接观察序列的总趋势和周期变化以及变异点、升降转折点等的方法. 这种方法简单、直观、易懂易用，但是获取的信息少而且肤浅和粗略. 指标法是通过计算一系列核心指标，进而反映研究系统动态特征的方法. 例如，分别计算 1994—1996 年的平均销售量和 1996—1998 年的平均销售量，并比较两者，从而分析平均销售量是否保持某一水平或者是否具有某种趋势等. 这种方法也比较粗略. 模型法是根据数理统计理论和数学方法，建立描述该序列的适应性或者最优化统计模型，并据以进行预测或控制的方法. 本书主要介绍模型法.

随机性时间序列分析包括平稳时间序列分析、非平稳时间序列分析、可控时间序列分析等，常见的模型有：自回归模型（auto regressive model）、移动平均模型（moving average model）、自回归移动平均模型（auto regressive moving average model）、求和自回归移动平均模型（auto regressive integrated moving average model）、乘积季节模型（multiplicative seasonality model）等.

就数学方法而言，平稳时间序列分析是用一种比较简单的有限参数模型——自回归移动平均模型（简称 ARMA 模型）近似地代替一类相当广泛的平稳随机序列. 理论上，平稳时间序列分析已比较成熟，因此可作为随机时间序列分析的基础.

（2）频域（frequency domain）分析是将时间序列分解成各种周期扰动的叠加，以确定各周期振动能量的分配. 这种分配称为谱或功率谱，因此频域分析又称谱分析（spectral analysis）. 常用的数学工具有傅里叶变换、功率谱密度、最大熵谱估计理论等，是一种非常有用的动态数据分析方法. 但是由于分析方法复杂、结果抽象、又有一定的使用局限性，故本书仅介绍时域分析.

确定性时序分析是一种因素分解方法，其侧重点在于确定性信息快速、便捷地提取；随机性时间序列分析是利用随机理论，研究时间序列的统计规律和各因素之间的相互影响. 自从 19 世纪 20 年代以来，随机性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，其理论和方法得到了不断发展和广泛的应用. 虽然随机

性时间序列分析大大丰富和发展了时间序列分析的理论和方法，成为时间序列分析的主流，但它并不能完全取代确定性时间序列分析．通常意义上的时间序列分析都是指随机性时间序列分析，即时域分析．

二、时间序列分析的特点

作为数理统计学的一个专业分支，时间序列分析遵循数理统计学的基本原理，即利用观察信息估计总体的性质．但是由于时间的不可重复性，使得在任意一个时刻只能获得唯一的一个序列观察值．这种特殊的数据结构导致时间序列分析有它非常特殊的、自成体系的一套分析方法．近年来，时序分析已普遍应用于工农业生产、科学技术和社会经济生活的许多领域．

从数据的形成来看，多元统计分析所处理的数据是横剖面数据．所谓横剖面数据是指由若干现象在某一时间点上所处的状态所形成的数据，它反映的是一定时间、地点等客观条件下各个现象之间所存在的内在数值关系．而时间序列分析所处理的数据是纵剖面数据．所谓纵剖面数据就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的基本特征．这也是时间序列分析区别与其他统计分析的重要特征之一．

时间序列分析是研究序列随时间动态发展的变化模式，其序列的观测值需要按照一定顺序取得，并保持其顺序不变．观测值之间顺序的重要性，是时间序列分析区别于其他统计分析的又一重要特征．

时间序列的观测值是某一随机序列的一次样本实现．观测值之间存在相关性，且不满足所谓“各观测值为独立”的必要假设，这是时间序列分析区别于其他统计分析的又一重要特征．

时间序列分析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．其基本思想是：根据系统的有限长度的观察数据，建立能够比较精确地反映序列中所包含动态依存关系的数学模型，并借以对系统的未来进行预报．所以，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上刻画若干现象之间