TIME 时间序列

X-12-ARIMA

季节调整——原理与方法

中国人民银行调查统计司

中国金融出版社

时间序列 X-12-ARIMA 季节调整
——原理与方法

中国人民银行调查统计司

中国金融出版社

责任编辑：古炳鸿
责任校对：李俊英
责任印制：张　莉

图书在版编目（CIP）数据

时间序列 X-12-ARIMA 季节调整——原理与方法（Shijian Xulie X-12-ARIMA Jijie Tiaozheng——Yuanli yu Fangfa）/中国人民银行调查统计司. —北京：中国金融出版社，2006.10
ISBN 7-5049-4151-4

Ⅰ.时…　Ⅱ.中…　Ⅲ.金融—经济统计　Ⅳ.F830.2

中国版本图书馆 CIP 数据核字（2006）第 095115 号

出版发行 中国金融出版社
社址　北京市广安门外小红庙南里 3 号
市场开发部　(010)63272190，66070804（传真）
网 上 书 店　http://www.chinafph.com
(010)63286832，63365686（传真）
读者服务部　(010)66070833，82672183
邮编　100055
经销　新华书店
印刷　保利达印务有限公司
尺寸　169 毫米×239 毫米
印张　27.5
字数　491 千
版次　2006 年 10 月第 1 版
印次　2006 年 10 月第 1 次印刷
印数　1—2000
定价　50.00 元

序

随着中国社会主义市场经济的不断发展以及融入世界经济全球化进程的不断加快，经济运行的复杂性日趋增加。及时察觉经济运行的趋势性变化，尽早采取应对政策，已成为加强和改善宏观调控的重要前提。用传统的思维方式和分析手段去分析判断经济问题，往往难以找到合理的答案。如何运用先进的分析方法和分析工具，从纷繁复杂的经济现象中，探寻其内在规律和发展趋势，进而制定出科学、合理、有效、前瞻的货币政策，促进中国经济继续沿着健康、平稳、快速的轨道运行，成为新时期中央银行宏观经济分析工作责无旁贷的第一要务。

长期以来，我们对经济数据的分析和表述习惯用同期相比。我们用月度同期相比数据反映 12 个月之间总的变化，但在经济波动较快的情况下很难反映当期经济的变化，例如，我们常说的“物价翘尾因素”就是指当期虽然没有调价，但因前期调价反映的同期相比的物价指数上涨。因此，在分析短期经济波动时，必须重视对相邻两月、两季的环比数据以及固定基期的定基比数据的分析。但环比数据、定基比数据往往存在明显的季节性波动，直接使用会有很大误差。例如春节所在月份和前后月份相比，生产下降，消费增加，物价可能上涨；节后月份与春节所在月份相比，则反之，生产增长，消费减少，物价可能降低。这些并不反映经济运行趋势的变化，仅仅是经济在不同月份的规律性波动，我们称之为季节波动。要分析时间序列指标的发展趋势，判断经济运行的拐点，必须考虑季节性因素对数据波动的影响。

要去掉季节因素的影响，就需要对时间序列数据进行季节调整。季节调整的方法很多，其中 X-11 方法作为国际上比较通用的时间序列季节调整方法，自 20 世纪 60 年代由加拿大统计局和美国经济调查局开发以来，在美国、加拿大、德国、日本等国家的中央银行和统计部门的积极推动下，经过多个版本演变，不断融入现代时间序列分析的计量新方法，逐步形成了目前的 X-12-ARIMA版本。该程序作为时间序列分析的有力工具，主要完成对时间序列数据中周期性因素（C）、趋势性因素（L）、季节性因素（S）、节日性因素

（H）和不规则因素（I）等的分解和消除，在消除以上影响后，对历史数据序列进行科学对比，并在此基础之上利用 ARIMA 模型进行进一步计量、预测和分析。

不少国家都存在移动节假日，我国春节有时在 1 月份，有时在 2 月份，也属于移动节假日。由于各国传统习惯不同，移动节假日也存在差异，各国中央银行、统计部门的分析人员在运用 X-11 方法时，并非采取简单的拿来主义，而是结合自己的国情，特别是根据自己国家移动假日的需要，进行本地化改造。二十多年前，我还在国家计委工作，当时去美国访问，美国经济调查局赠送给我们 X-11 软件，后来在我们的时间序列分析中获得广泛应用，当时我们就有将其进行本地化改造的想法。目前，中国人民银行调查统计司会同中国人民大学、南开大学和武汉理工大学的专家学者研究了 X-11 方法的原理，并对软件进行了中国本地化改造，提升了软件的功能，对此，我甚感欣慰。

数据处理是计量经济分析的基础性工作，运用 X-12-ARIMA 软件剔除时间序列的季节波动性，更是数据处理的基础。我相信，符合中国国情的 X-12-ARIMA 软件不但对中国人民银行的经济分析人员有用，而且对其他领域时间序列数据的处理也是有用的。而正是由于数据处理只是经济分析的第一步，我希望能通过 X-11 方法的运用，进一步加强计量理论的学习，加强与世界发达国家之间的业务交流，掌握并使用国内外先进的数量分析方法和工具，提高我们的经济分析质量，为货币政策决策和宏观金融调控提供强有力的支撑与保障。

苏宁

2006 年 8 月 8 日

前　言

对时间序列 X-12-ARIMA 季节调整的原理进行研究并对其软件进行中国本地化改造是经苏宁同志提议，由中国人民银行立项支持的科研项目。此原理手册是该项目的成果之一。另外，尚有一套时间序列 X-12-ARIMA 季节调整软件以及与之配套的软件操作指南。

“时间序列 X-12-ARIMA 季节调整项目”由中国人民银行调查统计司承担，唐思宁同志是项目总负责人，王毅、章丽盛等同志负责具体组织实施。以南开大学张晓峒教授、栾惠德博士为主的研究小组具体负责原理部分的研究；以中国人民大学谭荣华教授、谢波峰博士以及武汉理工大学童恒庆教授为主的研究小组具体承担了此项目的软件本地化改造及扩展功能的实现。

在项目实施过程中，调查统计司陈志理、蒋万进、李跃、阮健弘、向晓岚等同志从最终使用的角度提出了有价值的意见，科技司文四立、杨竑、陆书春等同志也给予了极大的帮助，中国科学院数学与系统科学研究院的郑桂环博士对此原理手册进行了审定，在此一并致谢。

“时间序列 X-12-ARIMA 季节调整项目”得到了美国普查局（U. S. Census Bureau）X-11 项目小组的帮助，Brian. C. Monsell 先生及时解答了在原理及软件研究中的问题，提供了 X-12 的 DOS 版源程序并授权中国人民银行对之进行改造。

此原理手册中引用的相关资料得到了施普林格出版集团（Springer）和美国统计学会（the American Statistical Association）主办的《商业与经济统计》（Journal of Business and Economic Statistics）的版权许可。

X-12-ARIMA 是一种应用广泛的时间序列季节调整方法，公布我们对其中国本地化的改造成果，不仅希望使用者受益，更希望不足之处得到大方指教。

中国人民银行调查统计司

2005 年 5 月

目 录

第1章

时间序列（ARIMA和SARIMA）模型

时间序列分析方法由Box-Jenkins于1965年提出。它适用于各种领域的时间序列分析，研究对象为非季节和季节时间序列，建立的模型分别用ARIMA和SARIMA表示。本章的时间序列特指随机时间序列。

时间序列模型不同于经典回归模型的两个特点是：（1）这种建模方法不以经济理论为依据，而是依据变量自身的变化规律，利用外推机制描述时间序列的变化。（2）明确考虑时间序列的非平稳性。如果时间序列非平稳，建立模型之前应先通过差分把它变换成平稳的时间序列，再考虑建模问题。

1.1 随机过程、时间序列

为什么在研究时间序列建模之前先要介绍随机过程？这是因为要从理论高度来认识时间序列。时间序列不是无源之水，它是由相应随机过程产生的。只有从随机过程的高度认识其一般规律，对时间序列的研究才会有指导意义，对时间序列的认识也才会更深刻。在实际中，是通过一个具体的时间序列来推断相应随机过程的形式的。

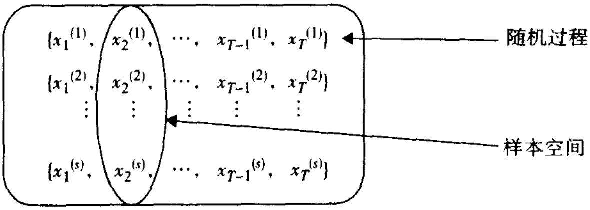
自然界中事物变化的过程可以分成两类：一类是确定型过程，一类是非确定型过程。

确定型过程是可以用关于时间t的函数描述的过程。例如，真空中的自由落体运动过程，电容器通过电阻的放电过程，行星的运动过程等。

非确定型过程是不能用一个（或几个）关于时间t的确定性函数描述的过程。换句话说，对同一事物的变化过程独立、重复地进行多次观测而得到的结果不相同。例如，对河流水位的测量。其中，每一时刻的水位值都是一个随机变量。如果以一年的水位记录作为实验结果，便得到一个水位关于时间的函数x_t，$t = 1, 2, \cdots, T$。这个水位函数是预先不可确知的，只有通过测

量才能得到，而在每年中同一时刻的水位记录是不相同的。

随机过程定义：由随机变量组成的一个有序序列称为随机过程，记为 $\{x(s, t), s \in S, t \in T\}$。其中，$S$ 表示样本空间，T 表示序数集，通常指时间。对于每一个 t，$t \in T$，$x(\cdot, t)$ 是样本空间 S 中的一个随机变量。对于每一个 s，$s \in S$，$x(s, \cdot)$ 是随机过程在序数集 T 中的一次实现。随机过程图示如下：



随机过程简记为 $\{x_t\}$ 或 x_t，$t = 1, 2, \cdots, T$。

随机过程一般分为两类：一类是离散型的，一类是连续型的。如果一个随机过程 $\{x_t\}$ 对任意的 $t \in T$ 都是一个连续型随机变量，则称此随机过程为连续型随机过程。如果一个随机过程 $\{x_t\}$ 对任意的 $t \in T$ 都是一个离散型随机变量，则称此随机过程为离散型随机过程。本书只考虑离散型随机过程。随机过程分类如下：

$$\text{随机过程}\begin{cases}\text{连续型} \\ \text{离散型}\begin{cases}\text{平稳的过程}\begin{cases}\text{严平稳过程} \\ \text{宽平稳过程}\end{cases} \\ \text{非平稳的过程}\end{cases}\end{cases}$$

严（强）平稳过程定义：一个随机过程中若随机变量的任意子集的联合分布函数与时间无关，即无论对 T 的任何时间子集 $(t_1, t_2, \cdots, t_n)$ 以及任何实数 k，$(t_i + k) \in T$，$i = 1, 2, \cdots, n$ 都有

$$F(x(t_1), x(t_2), \cdots, x(t_n)) = F(x(t_1 + k), x(t_2 + k), \cdots, x(t_n + k))$$

成立，其中 $F(\cdot)$ 表示 n 个随机变量的联合分布函数，则称其为严平稳过程或强平稳过程。

严平稳意味着随机过程所有存在的矩都不随时间的变化而变化。严平稳的条件是非常严格的，而且对于一个随机过程，上述联合分布函数不便于分析和使用，因此希望给出不像强平稳那样严格的条件。若放松条件，则可以只要求分布的主要参数相同。如只要求从一阶到某阶的矩函数相同，这就引出了宽平稳概念。

如果一个随机过程 m 阶矩以下的矩的取值全部与时间无关，则称该过程为 m 阶平稳过程。比如对于随机过程 $\{x_t\}$，如果有

$$E[x(t_i)] = E[x(t_i + k)] = \mu < \infty, (t_i + k) \in T$$

$$\mathrm{Var}[x(t_i)] = \mathrm{Var}[x(t_i + k)] = \sigma^2 < \infty, (t_i + k) \in T$$

$$\mathrm{Cov}[x(t_i), x(t_j)] = \mathrm{Cov}[x(t_i + k), x(t_j + k)] = {\sigma_{ij}}^2 < \infty, (t_i + k) \in T$$

其中 μ，σ^2 和 ${\sigma_{ij}}^2$ 为常数，不随 t，$(t \in T)$；k，$((t_r + k) \in T, r = i, j)$ 变化而变化，则称该随机过程 $\{x_t\}$ 为二阶平稳过程（协方差平稳过程）。该过程属于宽平稳过程。

如果严平稳过程的二阶矩为有限常数值，则其一定是宽平稳过程。反之，一个宽平稳过程不一定是严平稳过程。但对于正态随机过程而言，严平稳与宽平稳是一致的。这是因为正态随机过程的联合分布函数完全由均值、方差和协方差所唯一确定。以下简称二阶平稳过程为平稳过程。

时间序列定义：随机过程的一次实现称为时间序列，也用 $\{x_t\}$ 或 x_t 表示。

与随机过程相对应，时间序列分类如下：

$$\text{时间序列}\begin{cases}\text{连续型（如心电图记录，水位记录，温度记录等）}\\ \text{离散型}\begin{cases}\text{从相同的时间间隔点取自连续变化的序列（如人口序列）}\\ \text{相同时间间隔内的累积值（如年粮食产量，进出口额序列）}\end{cases}\end{cases}$$

时间序列中的元素称为观测值。$\{x_t\}$ 既表示随机过程，也表示时间序列。x_t 既表示随机过程中的元素——随机变量，也表示时间序列中的元素——观测值。在不致引起混淆的情况下，为方便起见，x_t 也直接用来表示随机过程和时间序列。

随机过程与时间序列的关系如下所示：

随机过程：$\{x_1, x_2, \cdots, x_{T-1}, x_T\}$

第1次观测：$\{{x_1}^{(1)}, {x_2}^{(1)}, \cdots, {x_{T-1}}^{(1)}, {x_T}^{(1)}\}$

第2次观测：$\{{x_1}^{(2)}, {x_2}^{(2)}, \cdots, {x_{T-1}}^{(2)}, {x_T}^{(2)}\}$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

第 n 次观测：$\{{x_1}^{(n)}, {x_2}^{(n)}, \cdots, {x_{T-1}}^{(n)}, {x_T}^{(n)}\}$

某河流在某观测点每年的水位值 $\{x_1, x_2, \cdots, x_{T-1}, x_T\}$，作为研究对象可以看做一个随机过程。而每一年的具体水位记录，比如第一年的水位记录 $\{{x_1}^{(1)}, {x_2}^{(1)}, \cdots, {x_{T-1}}^{(1)}, {x_T}^{(1)}\}$，就是一个时间序列。而在每年中同一时刻（如 $t=2$ 时）的水位记录是不相同的，$\{{x_2}^{(1)}, {x_2}^{(2)}, \cdots, {x_2}^{(n)}\}$ 构成了 x_2 取值的样本空间。

比如，要记录某市日电力消耗量，则每日的电力消耗量就是一个随机变量，于是得到一个日电力消耗量关于天数 t 的函数。而这些以年为单位的函数族构成了一个随机过程 $\{x_t\}$，$t=1, 2, \cdots, 365$。因为时间以天为单位，是离散的，所以这个随机过程是离散型随机过程。而一年的日电力消耗量的实际观测值序列就是一个时间序列。

自然科学领域中的许多时间序列常常是平稳的。如工业生产中对液面、压力、温度的控制过程，某地的气温变化过程，某地 100 年的水文资料，单位时间内路口通过的车辆数等。但经济领域中多数宏观经济时间序列却都是非平稳的。如一个国家的年 GDP 序列，年投资额序列，年进出口额序列等。

为便于下面的分析，首先给出差分定义。

差分定义：时间序列变量的本期值与其滞后值相减的运算叫差分。先给出差分和滞后符号。对于时间序列 x_t，如下运算称为一阶差分：

$$x_t - x_{t-1} = \Delta x_t = (1 - L)x_t = x_t - Lx_t \tag{1.1}$$

其中，Δ 称为一阶差分算子；L 称为滞后算子，其定义是 $L^n x_t = x_{t-n}$。

二次一阶差分表示为：

$$\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2} \tag{1.2}$$

或

$$\Delta^2 x_t = (1 - L)^2 x_t = (1 - 2L + L^2)\, x_t = x_t - 2\, x_{t-1} + x_{t-2}$$

上两式的计算结果相同，可见差分算子和滞后算子都可以直接参与运算。当与其 k 期滞后变量相减时，称做 k 阶差分，可表示为：

$$x_t - x_{t-k} = \Delta_k x_t = (1 - L^k)x_t = x_t - L^k x_t$$

k 阶差分常用于季节性数据的差分。

下面介绍两种基本的随机过程。

(1) 白噪声（white noise）过程

白噪声源于物理学与电学，原指音频和电信号在一定频带中的一种强度不变的干扰声。

白噪声过程定义：对于随机过程 $\{x_t, t \in T\}$，如果 $E(x_t)=0$，$\mathrm{Var}(x_t)=\sigma^2<\infty$，$t \in T$；$\mathrm{Cov}(x_t, x_{t+k})=0$，$(t+k) \in T$，$k \neq 0$，则称 $\{x_t\}$ 为白噪声过程。

图 1.1（a）是由计算机生成的白噪声序列；图 1.1（b）是日元兑美元序列的一阶差分序列，其近似于一个白噪声序列。

白噪声是平稳的随机过程，因其均值为零，方差不变，随机变量之间非

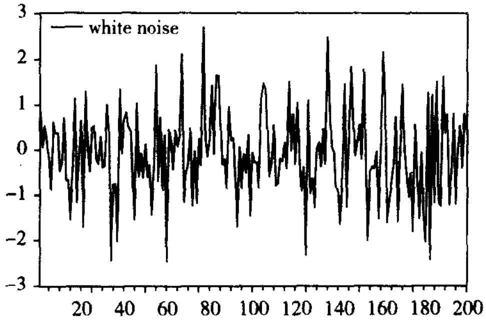


图 1.1（a）　由白噪声过程产生的时间序列（nrnd）

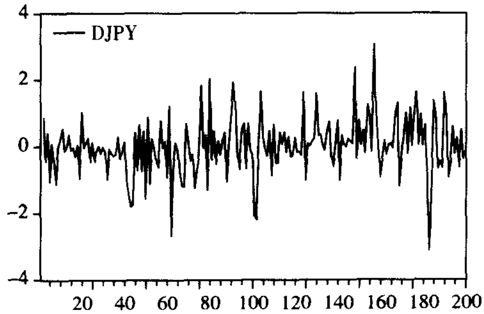


图 1.1（b）　日元兑美元汇率的收益率序列

相关。显然，上述白噪声是二阶宽平稳随机过程。如果 $\{x_t\}$ 同时还服从正态分布，则它就是一个强平稳的随机过程。

（2）随机游走（random walk）过程

对于下面的表达式

$$x_t = x_{t-1} + u_t \tag{1.3}$$

如果 u_t 为白噪声过程，则称 x_t 为随机游走过程。

“随机游走”一词首次出现于 1905 年《自然》（Nature）杂志第 72 卷 Pearson K. 和 Rayleigh L. 的一篇通信中。该信件的题目是“随机游走问题”。文中讨论寻找一个被放在野地中央的醉汉的最佳策略是从投放点开始搜索。

随机游走过程的均值为零，方差为无限大。对（1.3）式进行迭代运算，得：

$$x_t = x_{t-1} + u_t = u_t + u_{t-1} + x_{t-2} = u_t + u_{t-1} + u_{t-2} + \cdots$$

对上式分别求期望与方差，得

$$E(x_t) = E(u_t + u_{t-1} + u_{t-2} + \cdots) = 0$$

$$\mathrm{Var}(x_t) = \mathrm{Var}(u_t + u_{t-1} + u_{t-2} + \cdots) = \sum_{-\infty}^{t} {\sigma_u}^2 \rightarrow \infty$$

所以随机游走过程是非平稳的随机过程。

图 1.2（a）是由计算机生成的随机游走序列；图 1.2（b）是 300 天的上海股票综合指数序列。可以看到，这种序列的方差随时间越来越大。

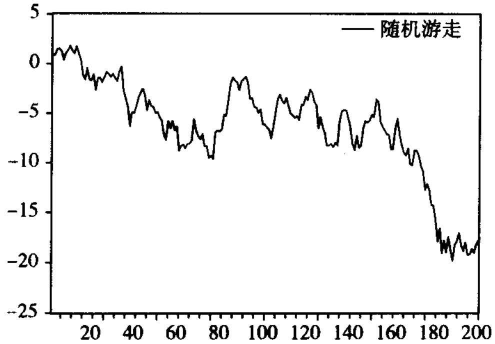


图 1.2（a） 由随机游走过程产生的时间序列

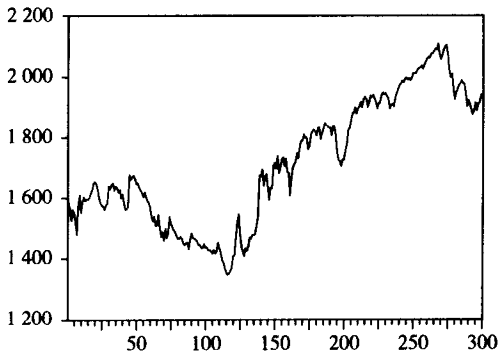


图 1.2（b） 上海股票综合指数序列

1.2 时间序列模型的分类

时间序列模型可分为四类：自回归（AR）模型、移动平均（MA）模型、自回归移动平均（ARMA）模型和单积自回归移动平均（ARIMA）模型。下面逐一介绍。

1.2.1　自回归过程

如果一个剔除均值和确定性成分的线性过程可表达为：

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{1.4}$$

其中，ϕ_i（$i=1$，…，p）是自回归参数，u_t 是白噪声过程，则称 x_t 为 p 阶自回归过程，用 AR（p）表示。x_t 是由它的 p 个滞后变量的加权和以及 u_t 相加而成。

若用滞后算子表示（1.4）式，则：

$$(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p)\ x_t = \Phi(L)\ x_t = u_t \tag{1.5}$$

其中，$\Phi(L)=1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p$ 称为特征多项式或自回归算子。

与自回归模型常联系在一起的是平稳性问题。对于自回归过程 AR（p），如果其特征方程

$$\Phi(z) = 1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p = (1-G_1 z)(1-G_2 z)\cdots(1-G_p z) = 0 \tag{1.6}$$

其中，z 表示变量的所有根的绝对值都大于 1，则 AR（p）是一个平稳的随机过程。

AR（p）过程中最常用的是 AR（1）、AR（2）过程。图 1.3 是由计算机生成的 AR（1）序列。

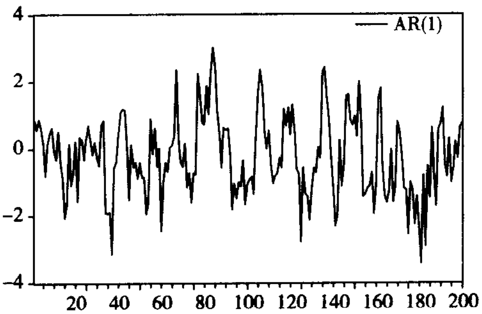


图 1.3　AR（1）序列

对于 AR（1）过程

$$x_t = \phi_1 x_{t-1} + u_t \tag{1.7}$$

保持其平稳性的条件是特征方程

$$(1-\phi_1 z) = 0$$

其中，z 表示变量的根的绝对值必须大于 1，即满足

$$|1/\phi_1| > 1$$

也就是

$$|\phi_1| < 1$$

解释如下：一阶自回归过程 $x_t = \phi_1 x_{t-1} + u_t$ 可写为

$$(1 - \phi_1 L)x_t = u_t$$

$$x_t = (1 - \phi_1 L)^{-1} u_t$$

在 $|\phi_1| < 1$ 条件下，有

$$x_t = (1 + \phi_1 L + (\phi_1 L)^2 + (\phi_1 L)^3 + \cdots)\, u_t$$

若保证 AR（1）具有平稳性，$\sum_{i=0}^{\infty} \phi_1{}^i L^i$ 必须收敛，即在 $|L| \leqslant 1$ 的条件下，ϕ_1 必须满足 $|\phi_1| < 1$。这是容易理解的：如果 $|\phi_1| \geqslant 1$，$\sum_{i=0}^{\infty} \phi_1{}^i L^i$ 发散，于是 x_t 变成一个非平稳随机过程。

由（1.7）式有：

$$x_t = u_t + \phi_1 u_{t-1} + \phi_1{}^2 x_{t-2} = u_t + \phi_1 u_{t-1} + \phi_1{}^2 u_{t-2} + \cdots \quad \text{（短记忆过程）}$$

因为 u_t 是一个白噪声过程，所以对于平稳的 AR（1）过程

$$E(x_t) = 0$$

$$\mathrm{Var}(x_t) = \sigma_u{}^2 + \phi_1{}^2 \sigma_u{}^2 + \phi_1{}^4 \sigma_u{}^2 + \cdots = \frac{1}{1 - \phi_1{}^2} \sigma_u{}^2$$

上式也说明，若保证 x_t 平稳，必须满足 $|\phi_1| < 1$（当 $\phi_1 = 1$ 时，Var（x_t）$\to \infty$，x_t 不再具有平稳性）。举例说明如下：

例 1 有 AR（1）过程 $x_t = 0.6\, x_{t-1} + u_t$，则

$$(1 - 0.6L)\, x_t = u_t$$

$$x_t = \frac{1}{1 - 0.6L} u_t = (1 + 0.6L + 0.36L^2 + 0.216L^3 + \cdots)\, u_t$$

$$= u_t + 0.6u_{t-1} + 0.36u_{t-2} + 0.216u_{t-3} + \cdots$$

上式变换为一个无限阶的移动平均过程。由于 $\phi_1 = 0.6 < 1$，所以 x_t 是一个平稳过程。其均值为零，方差为 $1.56\sigma_u{}^2$。

下面分析 AR（2）过程 $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t$ 具有平稳性的条件。其特征方程式是

$$1 - \phi_1 z - \phi_2 z^2 = 0 \tag{1.8}$$

其中，z 表示变量。上式的两个根是

$$z_1, z_2 = \frac{\phi_1 \pm \sqrt{\phi_1{}^2 + 4\phi_2}}{-2\phi_2} \tag{1.9}$$