

第二讲

时间序列分析

1 时间序列成分分析

1.1 时间序列的构成因素

时间序列中的数据（也称为观测值），总是由各种不同的影响因素共同作用所至；换一句话说，时间序列中的数据，总是包含着不同的影响因素。我们可以将这些影响因素合并归类为几种不同的类型，并对各种类型因素的影响作用加以测定。对时间序列影响因素的归类，最常见的是归为 3 类：

- 长期趋势（SPSS 的名称为 Smoothed Trend-Cycle，

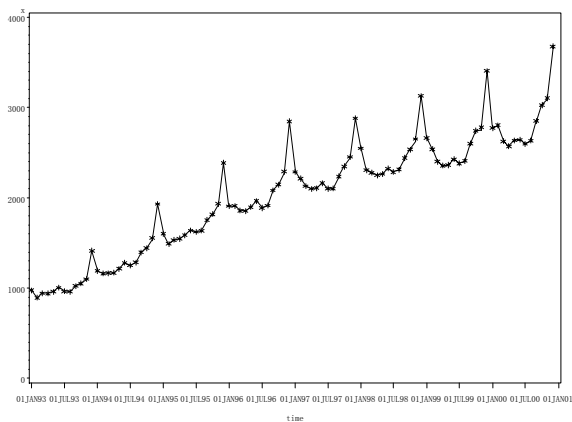
缩写 stc), 长期趋势是一种对事物的发展普遍和长期起作用的基本因素。受长期趋势因素的影响, 事物表现出在一段相当长的时期内沿着某一方向的持续发展变化。这种变化最常见的是是一种向上的发展, 对于经济现象而言, 通常由各种经济投入 (如技术进步、劳动力、资金等) 所引起, 因此, 长期趋势有时也可视作经济成长的因素。

- 季节周期因子（SPSS 的名称为 Season Factors Component），缩写 saf，季节周期也称为季节变动，是一种现象以一定时期（如一年、一月、一周等）为一周期呈现较有规律的上升、下降交替运动的影响因素。通常表现为现象在一年内随着自然季节的更替而发生的较有规律的增减变化（如某些季节性商品的销售额、旅游客流量、各月的降雨量等）。形成季节周期的原因，

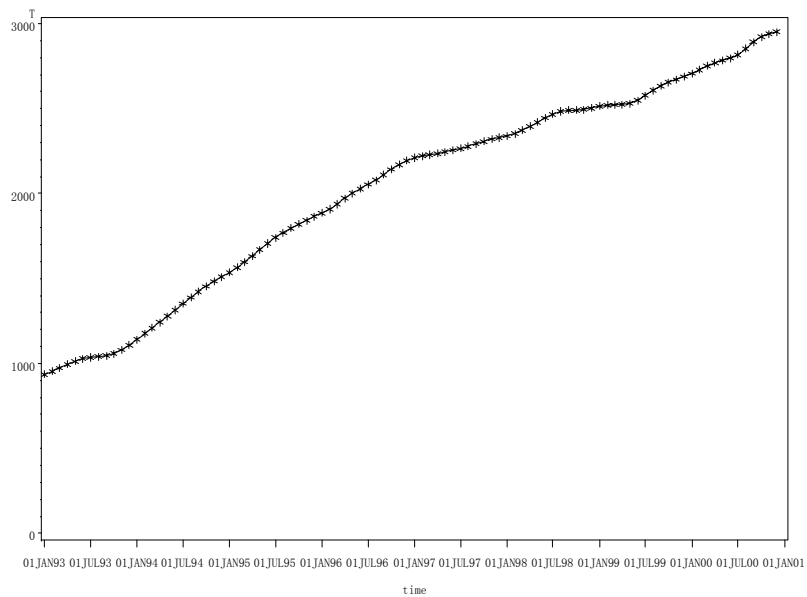
除了自然因素，也有人为和社会因素。

- 不规则变动因子(SPSS 的名称为 Irregular Component, 缩写 err)。不规则变动是一种偶然性、随机性、突发性因素。受这种因素影响，现象呈现时大时小、时起时伏、方向不定、难以把握的变动。这种变动不同于前三种变动，它完全无规律可循，无法控制和消除，例如战争、自然灾害等。

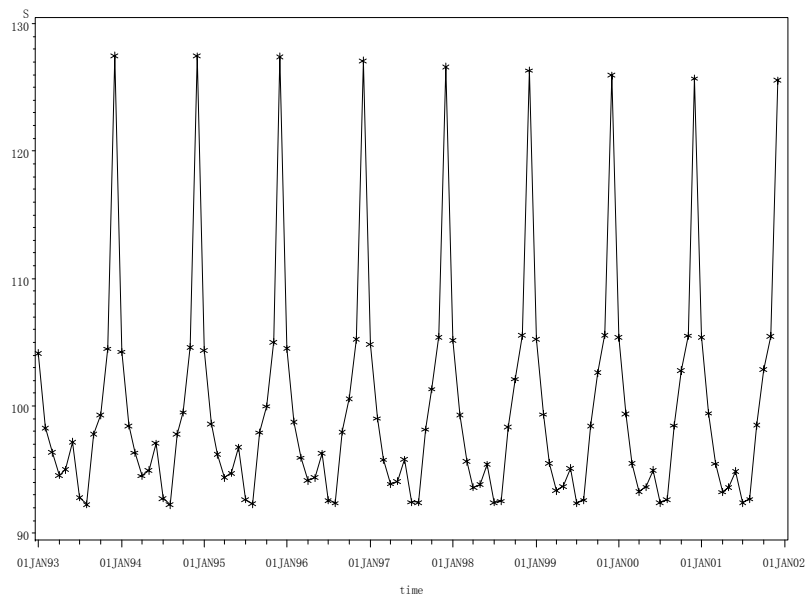
【例】1993 年 1 月至 2000 年 12 月社会消费品月零售总额的各成分图如下。



1993 年 1 月至 2000 年 12 月社会消费品月零售总额曲线图



长期趋势成分



季节因子图

1.2 时间序列的组合模型

若以 Y 代表时间序列中的数据（观测值），则 Y 由上述四类因素所决定的组合模型为：

$$Y = T + S + I \quad (\text{加法模型})$$

在加法模型中，各种影响因素是相互独立的，均为与 Y 同计量单位的绝对量。加法模型中，各因素的分解是根据减法进行（如 $Y - T = S + I$ ）。

$$Y = T \times S \times I \quad (\text{乘法模型})$$

在乘法模型中，只有长期趋势是与 Y 同计量单位的绝对量；其余因素均为以长期趋势为基础的比率，表现为对于长期趋势的一种相对变化幅度，通常以百分数表示。乘法模型中，各因素的分解是根据除法进行（如 $Y / T = S \times I$ ）。乘法模型是时间序列构成因素分析的主要模型形式。

1.3 SPSS 时间序列成分分解的实现及输出结果

（一）SPSS 时间序列成分分解的实现

为了简单起见，我们先来看一个简单的时间序列例子。表

1 是 1984 年到 1988 年某机场每个季度通过安全检测门的人数，单位：万人。

第一步：将数据输入 SPSS 的表格，记住现在只有一个变量序列，按时间顺序输成一行；

第二步：定义时间。通过 DATA 的菜单，选择 Define Dates 定义时间变量（图 1）。选中后得如下的对话框（图 2），选择时间序列的频率，如年度数据，季度数据和月度数据等。

表 1 84 到 88 年某机场季度过安全检测门的人数

t	一季度	二季度	三季度	四季度
1984	318	380	358	423
1985	379	394	412	439
1986	413	458	492	493
1987	461	468	529	575
1988	441	548	561	620

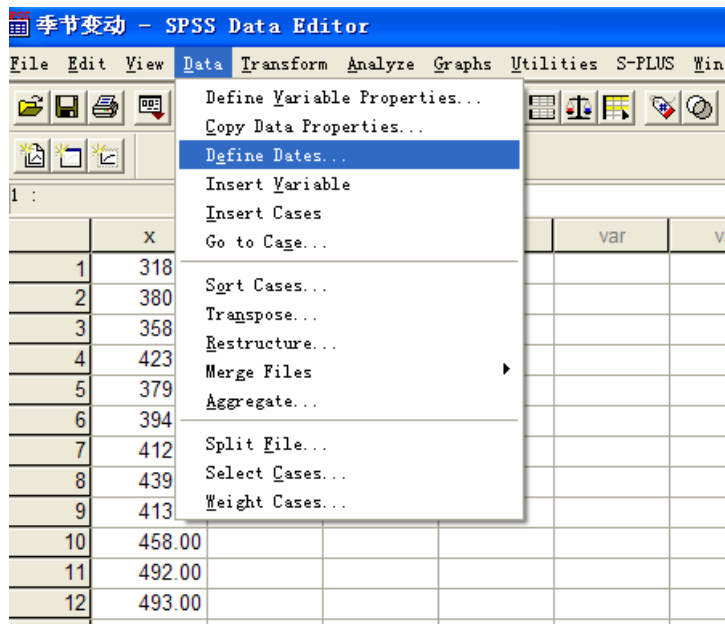


图 1 操作图

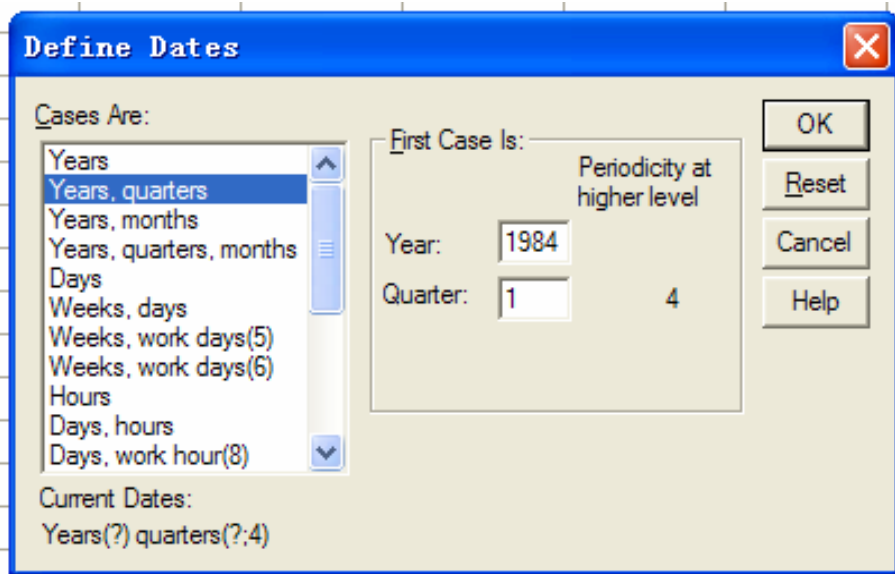


图 2 操作图

第三步：进行时间序列的成分分解。通过 Analyze(分析) 的菜单, 选择 Time-Series(时间序列), 再在 Time-Series 的菜单选择 Seasonal Decomposition(季节分解)。如图 3。

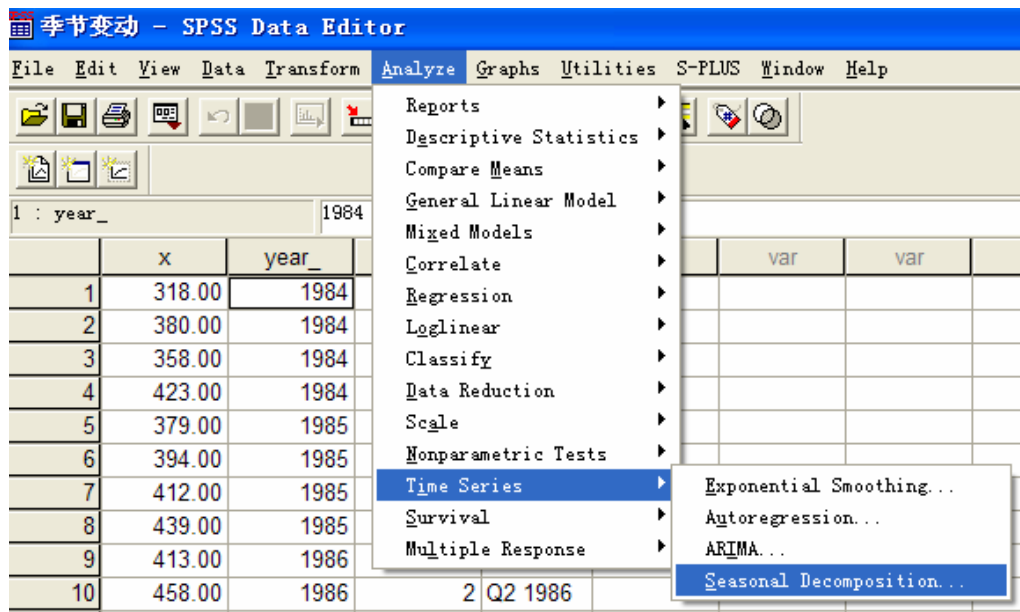


图3 操作图

选中后有如下的对话框出现，如图 4。

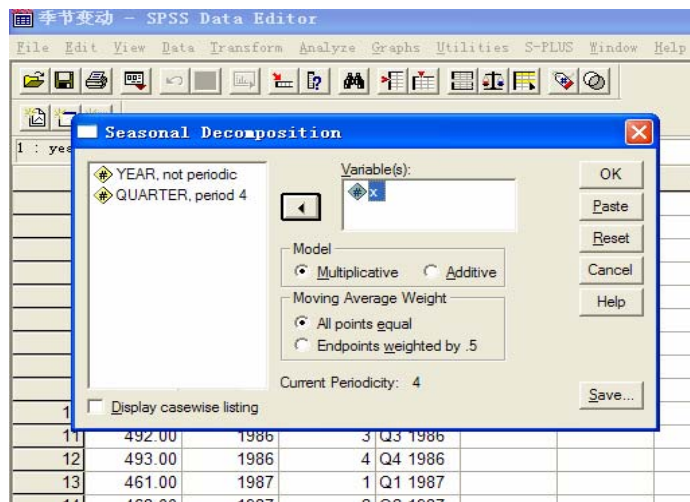


图 4 操作图

最后,

在【Variable(s)】(变量)处选择要分析的变量

在【Model】(模型)选择 Multiplicative(乘法模型)或 Additive(加法模型)

在【Moving Average Weight】选择 All points equal(等权移动平均)和 Endpoints Weighted by .5(端点为 0.5 为权数的移动平均)

在【Display cesewise listing】处选中, 要求列出中间

计算结构。

完成后，数据文件增加了一些附加变量，如图 5。

季节变动 - SPSS Data Editor						
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help						
1 : err_1		D. 950140164775553				
	x	err_1	sas_1	saf_1	stc_1	
1	318.00	.95014	336.44173	.94519	354.09695	
2	380.00	1.07824	387.14358	.98155	359.04997	
3	358.00	.94425	353.56460	1.01254	374.44139	
4	423.00	1.02721	398.78525	1.06072	388.22155	
5	379.00	1.01264	400.97930	.94519	395.97579	
6	394.00	.99450	401.40677	.98155	403.62496	
7	412.00	.99265	406.89557	1.01254	409.90770	
8	439.00	.98091	413.86933	1.06072	421.92421	
9	413.00	.99192	436.95106	.94519	440.51240	
10	458.00	1.01826	466.60990	.98155	458.24312	
11	492.00	1.03014	485.90442	1.01254	471.68611	
12	493.00	.97619	464.77809	1.06072	476.11337	
13	461.00	1.00802	487.73472	.94519	483.85628	
14	468.00	.96266	476.79789	.98155	495.29080	
15	529.00	1.03128	522.44601	1.01254	506.60126	
16	575.00	1.05159	542.08397	1.06072	515.48815	
17	441.00	.89784	466.57486	.94519	519.66574	
18	548.00	1.03758	558.30180	.98155	538.08290	
19	561.00	.97954	554.04955	1.01254	565.61976	
20	620.00	1.00251	584.50794	1.06072	583.04718	

图 5

（二）输出结果的解释和展示

4 个新的附加变量序列分别是不规则成分（err_1）、季节调整后的序列（SAS_1）、季节因子（saf_1）和去掉季节和不规则变动的趋势循环成分（stc_1）。

（1）saf_1 是用 12×2 的移动平均方法求出长期趋势的估计，然后用长期趋势去除 X ，得到的季节因子¹；

（2）sas_1 等于 x 除以 saf_1 ($x / \text{saf_1}$)；

¹ 方法可以参看统计学的书籍。统计学书中的时间序列一章均会介绍该方法。

(3) stc_1 是由如下的公式给出²

$$(stc)_t = \frac{1}{9}[(sas)_{t-2} + 2(sas)_{t-1} + 3(sas)_t + 2(sas)_{t+1} + (sas)_{t+2}] \quad ,$$

$$t = 3, 4, \dots, n-2$$

$$(stc)_2 = \frac{1}{3}[(sas)_1 + (sas)_2 + (sas)_3]$$

$$(stc)_{n-1} = \frac{1}{3}[(sas)_{n-2} + (sas)_{n-1} + (sas)_n]$$

² 公式比较复杂, 作为资料的完整性的需要, 给出这些公式, 初学的者可以不看。

$$(stc)_1 = (stc)_2 + \frac{1}{2}[(stc)_2 - (stc)_3]$$

$$(stc)_n = (stc)_{n-1} + \frac{1}{2}[(stc)_{n-1} - (stc)_{n-2}]$$

(4) err_1 等于 SAS_1 除以 stc_1 (SAS_1/stc_1)。

(5) 作图

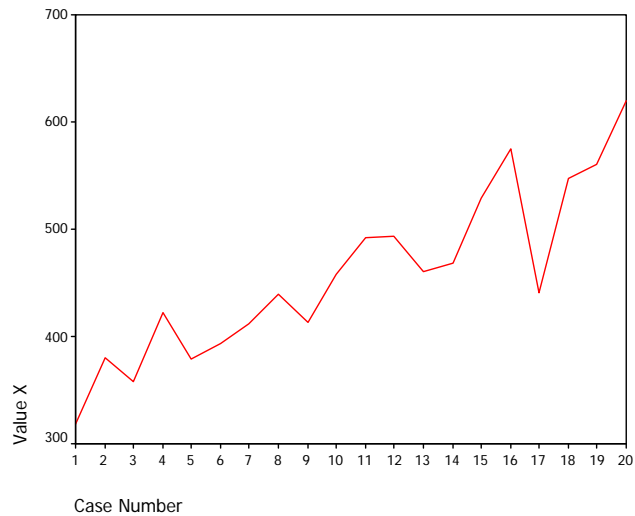


图6 时间序列原始数据x图

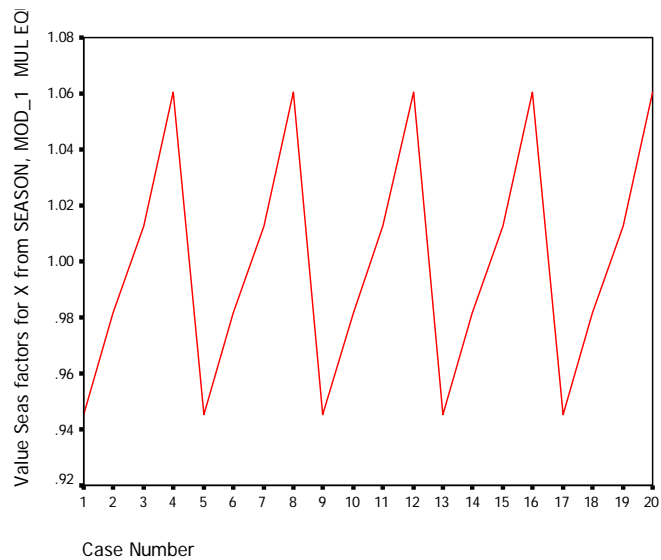


图7 季节因子saf图

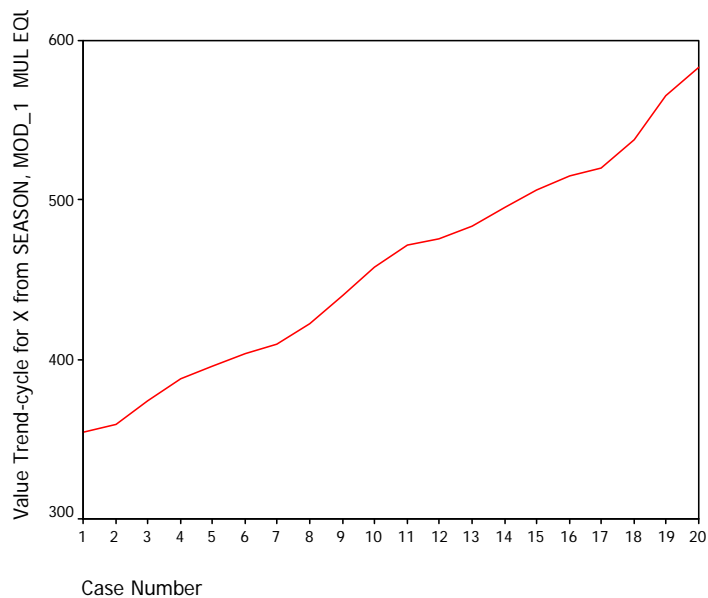


图8 趋势循环stc成分图

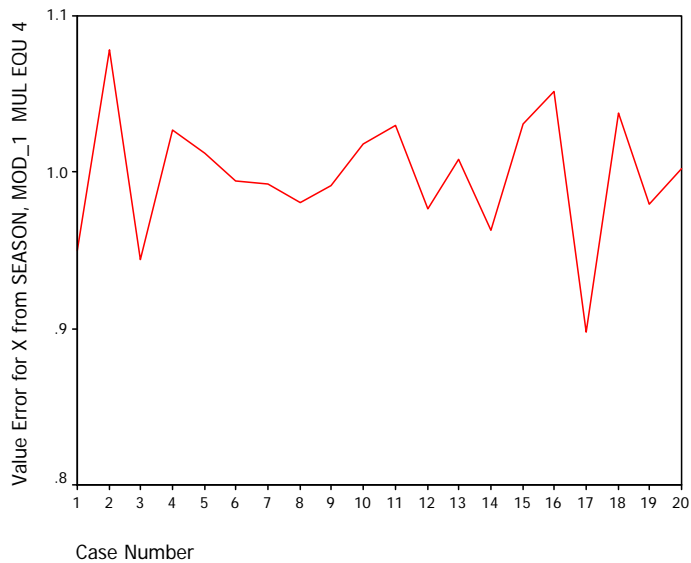


图9 不规则因子err图

1.4 如何应用这些数据进行预测

从长期趋势数据stc的图形可以看出，随着时间的变化，呈现出直线的趋势，可以利用趋势数据stc和t，建立线性回归模型 $stc = 337.329 + 11.499t$ ，预测出趋势的预测值。

表2 预测结果

时间	stc	t	趋势的预测值
1984.1	354.09695	1	348.82856
1984.2	359.04997	2	360.32792

1984. 3	374. 44139	3	371. 82728
1984. 4	388. 22155	4	383. 32664
1985. 1	395. 97579	5	394. 82600
1985. 2	403. 62496	6	406. 32536
1985. 3	409. 90770	7	417. 82472
1985. 4	421. 92421	8	429. 32408
1986. 1	440. 51240	9	440. 82344
1986. 2	458. 24312	10.	452. 32280

1986. 3	471. 68611	11	463. 82216
1986. 4	476. 11337	12	475. 32152
1987. 1	483. 85628	13	486. 82088
1984. 2	495. 29080	14	498. 32024
1987. 3	506. 60126	15	509. 81960
1987. 4	515. 48815	16	521. 31896
1988. 1	519. 66574	17	532. 81832
1988. 2	538. 08290	18	544. 31768

1988. 3	565. 61976	19	555. 81704
1988. 4	583. 04718	20	567. 31640
1989. 1	.	21	578. 81576
1989. 2	.	22	590. 31512
1989. 3	.	23	601. 81448
1989. 4	.	24	613. 31384

578. 81576，590. 31512，601. 81448和613. 31384分别是1989年一季度到四季度的趋势预测值，再根据季节因子一季

度到四季度的季节因子0.94519, 0.98155, 1.01254和1.06072, 便可以比较快捷地估计出1989一季度到四季度的过安全检查门的人数。

1989第一季度的预测值 $=578.81576 \times 0.94519$

1989第二季度的预测值 $=590.31512 \times 0.98155$

1989第三季度的预测值 $=601.81448 \times 1.01254$

1989四季度的预测值 $=613.31384 \times 1.06072$

1.5 练习和理解

下表是北京市 1997 年 1 月到 2003 年 8 月接待海外旅游人数（单位：万人）。

表 3 北京市接待海外旅游人数（单位：万人）

年	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1997	9.4	11.3	16.8	19.8	20.3	18.8	20.9	24.9	24.7	24.3	19.4	18.6
1998	9.6	11.7	15.8	19.9	19.5	17.8	17.8	23.3	21.4	24.5	20.1	15.9
1999	10.1	12.9	17.7	21.0	21.0	20.4	21.9	25.8	29.3	29.8	23.6	16.5
2000	11.4	26.0	19.6	25.9	27.6	24.3	23.0	27.8	27.3	28.5	32.8	18.5
2001	11.5	26.4	20.4	26.1	28.9	28.0	25.2	30.8	28.7	28.1	22.2	20.7

2002	13.7	29.7	23.1	28.9	29.0	27.4	26.0	32.2	31.4	32.6	29.2	22.9
2003	15.4	17.1	23.5	11.6	1.78	2.61	8.8	16.2				

完成下面的工作：

- 1、 练习输入数据和定义时间变量。
- 2、 作图，并观测数据在何处出现了非常值。
- 3、 去掉2003年的SPSS数据，做季节变动分析，请思考为何要去掉2003年的数据。
- 4、 观测输出的结果，给出12个月的季节因子。
- 5、 如果其趋势是直线，预测2003年1月到8月的可能人数，

并计算由于非典的原因，2003年1月到8月，到北京的海外旅游人数可能共计少了多少人。

2 平稳时间序列的 ARMA 模型

2.1 平稳性

一类描述时间序列的重要随机模型受到了人们的广泛关注，这就是所谓的平稳模型。这类模型假设随机过程在一个不变的均值附近保持平衡。其统计规律不会随着时间的推移发生变化。平稳的定义分为严平稳和宽平稳。

定义 1 (严平稳)

设 $X(t)$ ($t \in T$) 是一个随机过程, $X(t)$ 是在不同的时刻 t 的

n 个变量 x_1, x_2, \dots, x_n 组成的, 在不同的时刻 t 是不同的随机变量, 任取 n 个值 t_1, t_2, \dots, t_n 和任意的实数 h , 随机过程 $X(t) (t \in T)$ n 维分布函数满足关系式

$$F_n(x_1, \dots, x_n; t_1, \dots, t_n) = F_n(x_1, \dots, x_n; t_1 + h, \dots, t_n + h)$$

则称 $X(t) (t \in T)$ 为严平稳过程。

这个定义是说将他们的观测时刻 t_1, t_2, \dots, t_n 同时往前或往后 h , 所得到的时刻 $t_1+h, t_2+h, \dots, t_n+h$ 两组时刻观测的联合分布相同。

我们知道分布函数完整描述了随机变量的统计特征。但是严平稳过程的要求过分的严格，在实际中，要想知道 $X(t)(t \in T)$ 所有可能的联合分布函数，这几乎是不可能的。由此我们考虑到是否可以把条件放宽，仅仅要求其数字特征(数学期望和协方差)相等。

定义 2 (宽平稳)

若随机变量 $X(t)(t \in T)$ 的均值（一阶矩）和协方差（二阶矩）存在，且满足：

(1) 任取 $t \in T$, 有 $E(X(t)) = \text{常数}$;

(2) 任取 $t \in T$, $(t + \tau) \in T$, 有

$$E[(X(t) - a)(X(t + \tau) - a)] = R(\tau)$$

协方差是时间间隔的函数。则称 $X(t)(t \in T)$ 为宽平稳过程，其中 $R(\tau)$ 为协方差函数。

注 1: 严平稳和宽平稳的关系。

宽平稳 \Leftarrow 严平稳

宽平稳 \Rightarrow 严平稳

2.2 各种随机时间序列的表现形式

白噪声过程（white noise，如图 10）。属于平稳过程。 $y_t =$

$$u_t, \quad u_t \sim \text{IID}(0, \sigma^2)$$

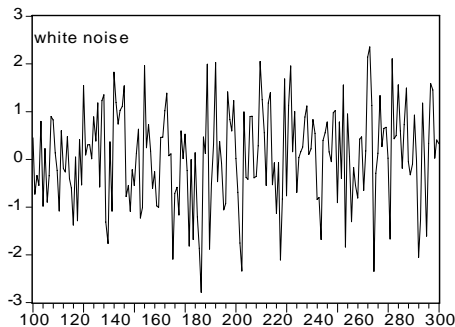


图 10 白噪声序列（ $\sigma^2=1$ ）

随机游走过程（random walk，如图 11）。属于非平稳过程。

$$y_t = y_{t-1} + u_t, \quad u_t \sim \text{IID}(0, \sigma^2)$$

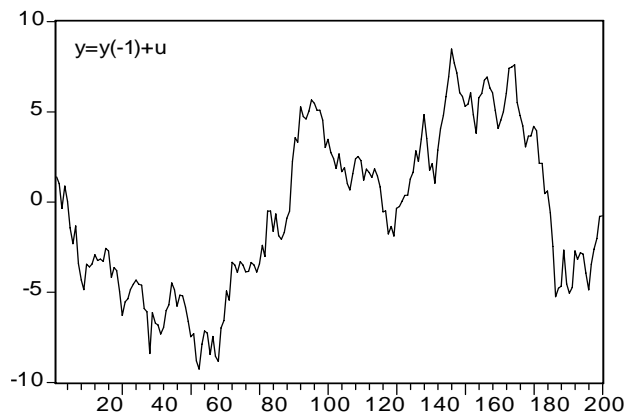


图 11 随机游走序列（ $\sigma^2=1$ ）

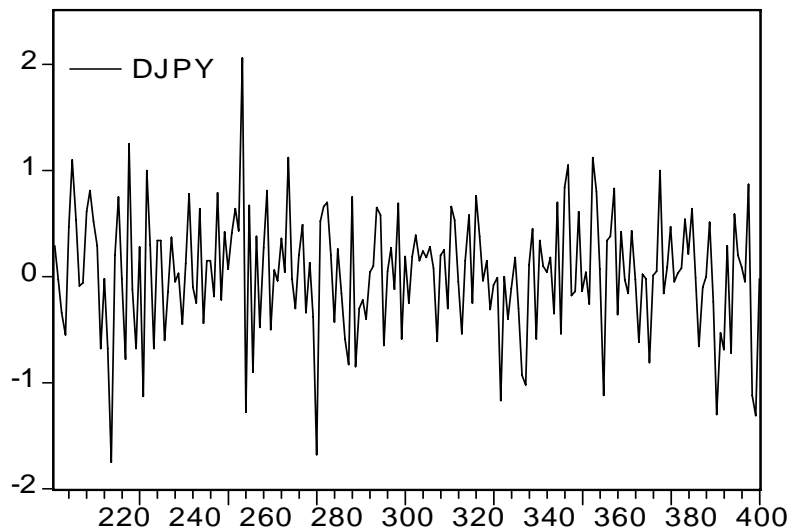


图 12 日元兑美元差分序列

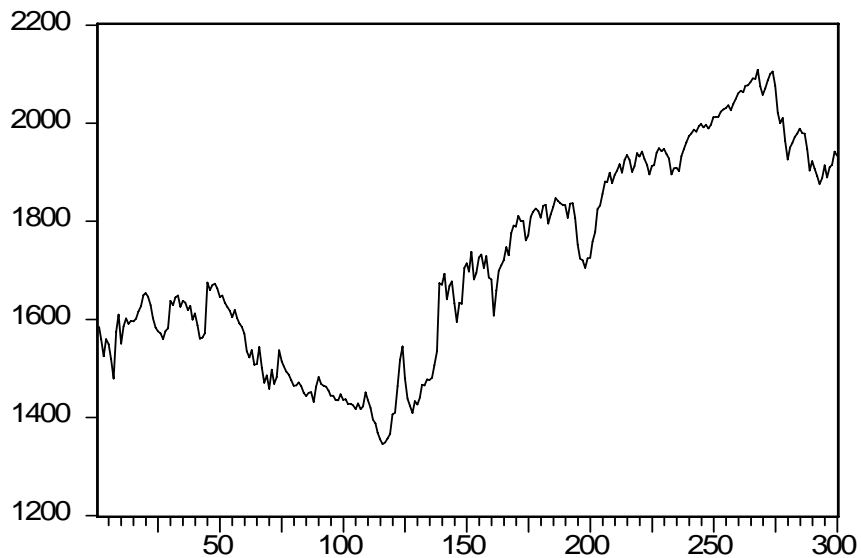


图 13 深圳股票综合指数

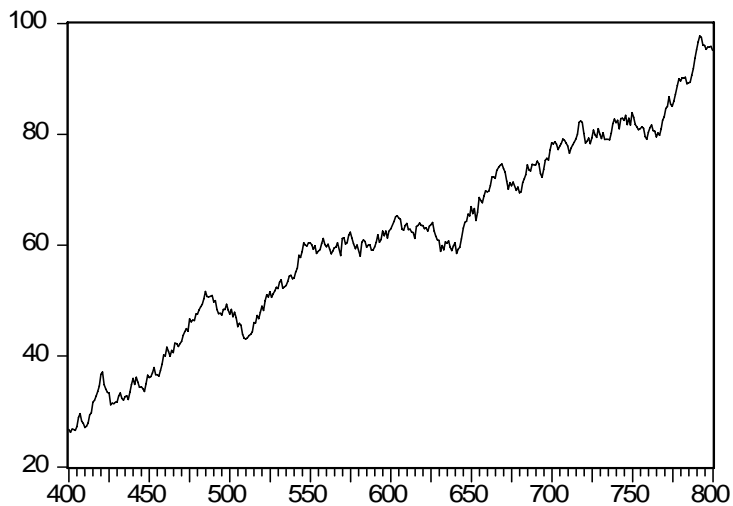


图 14 随机趋势非平稳序列 ($\mu = 0.1$)



图 15 随机趋势非平稳序列 ($\mu = -0.1$)

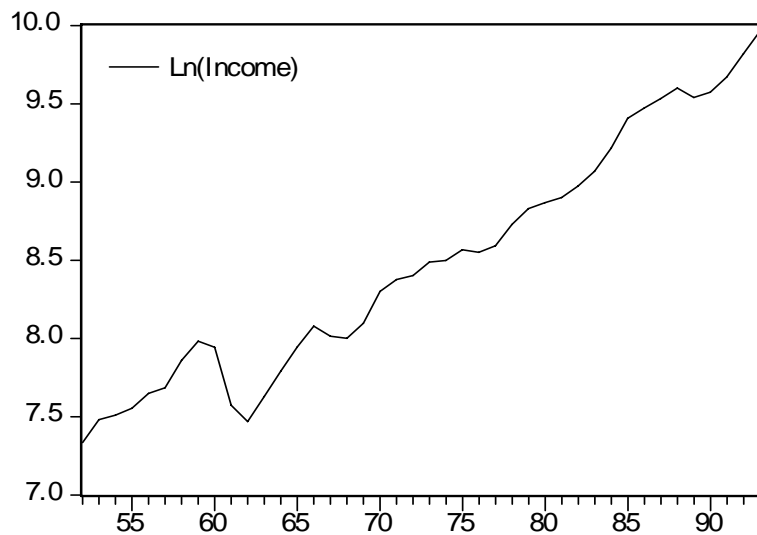


图 16 对数的中国国民收入序列

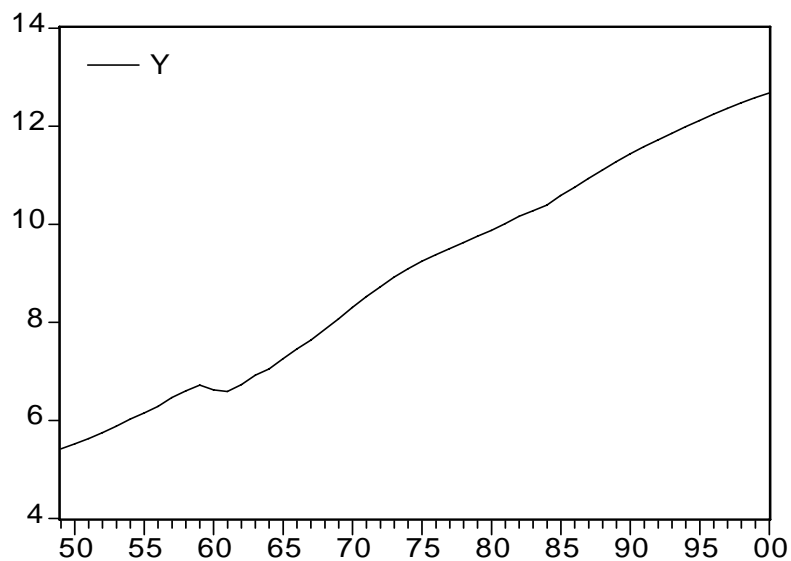


图 17 中国人口序列

注 2: 如果一个时间序列可以表示为白噪声的加权和，且权系数绝对收敛，则该系统是一个平稳的时间序列。(证从略)

2.2 延迟算子

延迟算子类似于一个时间指针，当前序列值乘以一个延迟算子，就相当于把当前序列值的时间向过去拨了一个时刻，记 B 为延迟算子，有 $x_{t-p} = B^p x_t, \forall p \geq 1$ 。

特别 $(1-B)$ 是差分算子。

2.3 ARMA(p,q)模型及其平稳性可逆性

2.3.1 模型类型及其表示

在平稳时间序列的分析中，应用最广泛的是有限参数模型，这种模型是在七十年代初发展活跃起来的，Box-Jenkins的理论提供了三种模型。

p 阶自回归模型：（用自己的过去和现在的随机干扰表 x_t

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t \quad a_t \text{ 是白噪声。}$$

q 阶移动平均模型：（用现在和过去的随机干扰表 x_t

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

p 阶自回归和 q 阶移动平均模型：自己的过去及过去和现在的随机干扰表 x_t

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

其中 ε_t 是白噪声序列。

2.3.2 平稳性

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$ 是平稳时间序列的反映吗？如果它是平稳时间序列的模型，回归系数应该满足何种

条件呢？

设 $\{x_t\}$ 是一个 p 阶自回归模型

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + a_t$$

或 $\Phi(B)X_t = a_t$

其中： $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p$ 。 $\{x_t\}$ 平

稳的充分必要条件是：

$1 - \phi_1 \gamma - \phi_2 \gamma^2 - \phi_3 \gamma^3 - \cdots - \phi_p \gamma^p = 0$ 的根在单位圆外；

$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \phi_3 \lambda^{p-3} - \dots - \phi_p = 0$ 的根在单位圆内³。

2.3.3 可逆性

我们可以考虑到一个时间序列 X_t 是否可以用它的现在值和过去值来表示现在时刻的随机干扰 a_t 呢？即

$$a_t = \mathbf{I}(B)X_t$$

这种表达式称为“逆转形式”。如果一个时间序列具有逆转形式，也就是说逆转形式存在且平稳，通常称该过程 x_t 具

³证明请参看附录 1。

有可逆性。

例 设 x_t 是一阶滑动平均模型，即 $X_t = a_t - \theta_1 a_{t-1}$ 或 $X_t = \Theta(B)a_t$ ，其中 $\Theta(B) = 1 - \theta_1 B$

$$\begin{aligned} \text{则 } a_t &= \frac{1}{(1 - \theta_1 B)} X_t \quad (\text{利用等比级数的通项和公式}) \\ &= \sum_{j=0}^{\infty} \theta_1^j B^j X_t \\ &= \sum_{j=0}^{\infty} \theta_1^j X_{t-j} \end{aligned}$$

对于一阶滑动平均模型 $X_t = a_t - \theta_1 a_{t-1}$ ，无论 θ_1 取何值，

$X_t = a_t - \theta_1 a_{t-1}$ 是一个名副其实的平稳序列，但是对于 $X_t = a_t - \theta_1 a_{t-1}$ 的“逆转形式”是否存在，则取决于 $|\theta_1|$ 是否小于 1。如果 $|\theta_1| > 1$,

$$X_t = -\sum_{j=1}^{\infty} \theta_1^j X_{t-j} + a_t$$

X_{t-j} 的系数随着 j 的增加而趋于无穷大，这显然违背了“远小近大”的原则，由此可见， $X_t = a_t - \theta_1 a_{t-1}$ 的逆转形式存在的充分必要条件为 $|\theta_1| < 1$ ， $|\theta_1| < 1$ 的充分必要条件特征方程 $1 - \theta_1 z = 0$ 的根在单位圆外。

$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} = \Theta(B)a_t$ 可逆的充分必要条件为，方程 $\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q = 0$ 的根在单位圆外。 $\gamma^q - \theta_1 \gamma^{q-1} - \theta_2 \gamma^{q-2} - \cdots - \theta_q = 0$ 的根在单位圆内⁴。

由于自回归模型

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + a_t$ 稍微变形，就是用系统的现在和过去值表示随机干扰项，所以自回归模型自然可逆。

⁴证明参看附录 2。

2.3.4 ARMA (p, q) 的平稳性和可逆性

设时间序列 $\{X_t\}$ 是 ARMA (p,q) 模型

$$\begin{aligned} & X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} \\ &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \end{aligned}$$

$$\text{令 } \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

则模型记为 $\Phi(B)X_t = \Theta(B)a_t$

如果 1. $\phi_p \neq 0, \theta_q \neq 0$;

2. $\Phi(B)$ 和 $\Theta(B)$ 无公共因子;

3. $\Phi(z)=0$ 和 $\Theta(B)=0$ 的根在单位圆外。

则 $\{X_t\}$ 是自回归移动平均模型, 平稳且可逆。它有传递形式

$X_t = \frac{\Theta(B)}{\Phi(B)}a_t$, 由此我们可以认为, 任何一个自回归滑动平均

模型都可以用一个足够高阶的滑动平均模型逼近。逆转形式

$a_t = \frac{\Phi(B)}{\Theta(B)}X_t$, 可见任何一个自回归滑动平均模型都可以用一个

足够高阶的自回归模型逼近。

2.4 平稳时间序列的统计特征

前面的分析讨论说明了三种模型的平稳性和可逆性。进一步我们想知道的是它们的统计特征。

2.4.1 总体的自相关函数和样本的自相关函数

一、AR(p)模型的自相关函数

自相关函数呈负指数衰减。收敛于零。但不等于零。AR(p)模型，自相关函数快速收敛于零，但不等于零，“拖尾”。又因为 ARMA (p , q) 模型 $\Phi(B)x_t = \Theta(B)\varepsilon_t$ 的可逆性，即

$\varepsilon_t = \frac{\Phi(B)}{\Theta(B)} x_t$ ，所以任何一个 ARMA (p, q) 模型都可以表示为一个足够高阶的 AR (p) 模型，所以 ARMA (p, q) 模型与 AR (p) 模型有相同的统计特性。下面从可以从图 18 到图 25 观察时间序列图与其自相关函数图的特点。

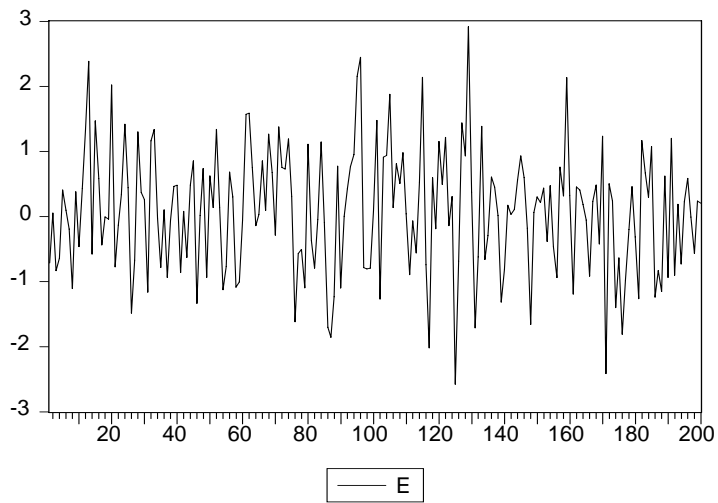


图 18 白噪声序列的自相关函数

Date: 03/15/04 Time: 23:23

Sample: 1 200

Included observations: 200

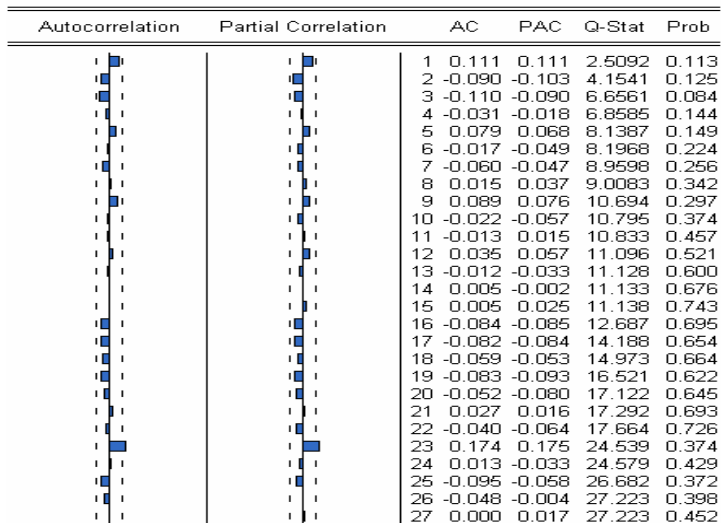


图 19 白噪声序列的自相关函数图

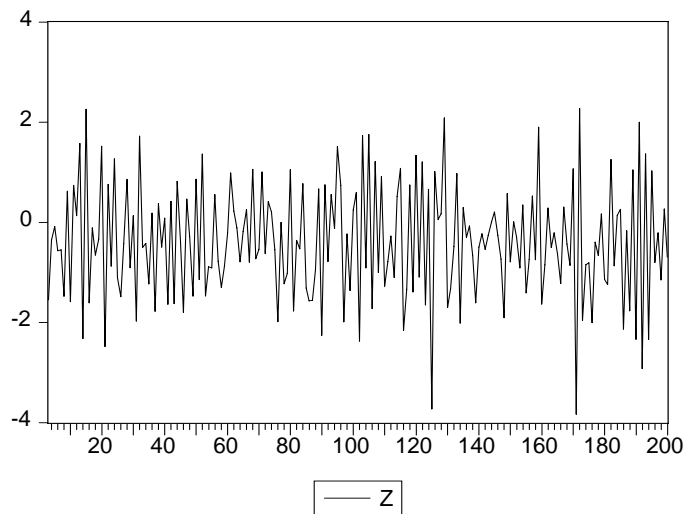


图 20 人工模拟序列 $X_t = 0.65X_{t-1} + 0.36X_{t-2} + a_t$ 图

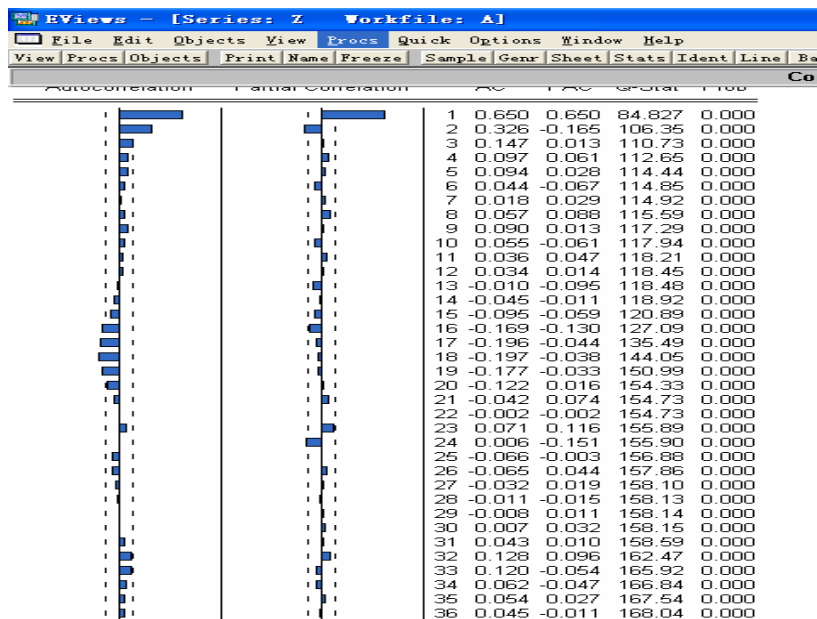


图 21 人工模拟序列 $X_t = 0.65X_{t-1} + 0.36X_{t-2} + a_t$ 的自相关函数图

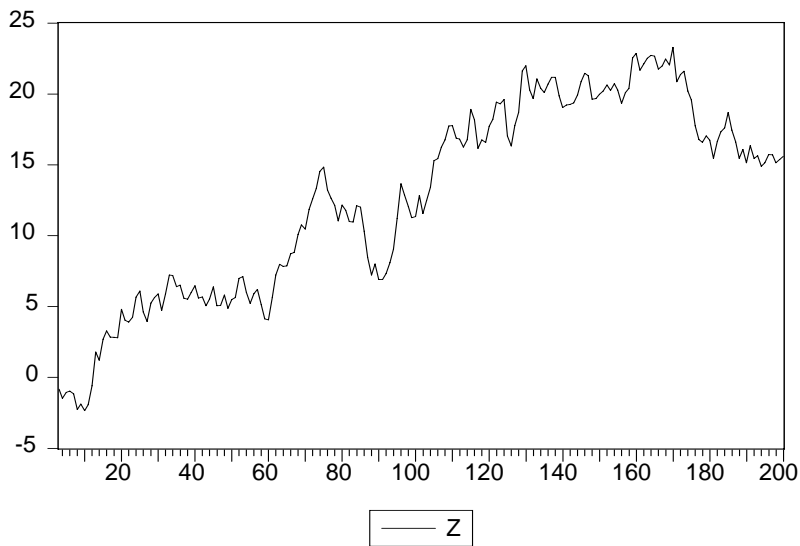


图 22 模拟随机游走序列 $x_t = x_{t-1} + a_t$ 图

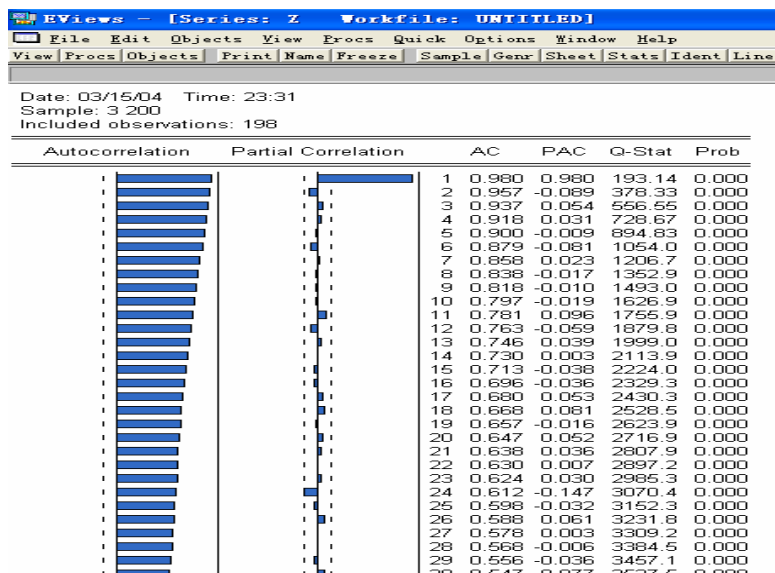


图 23 模拟随机游走序列 $x_t = x_{t-1} + a_t$ 的自相关函数图

二、MA (q) 的自相关函数

结论：MA (q) 模型的自相关函数 q 阶截尾，即在 q+1 及以后为零。图 2-7 是模拟一阶移动平均模型 $X_t = a_t + 0.8a_{t-1}$ 趋势图，图 2-8 是 $X_t = a_t + 0.8a_{t-1}$ 自相关函数图

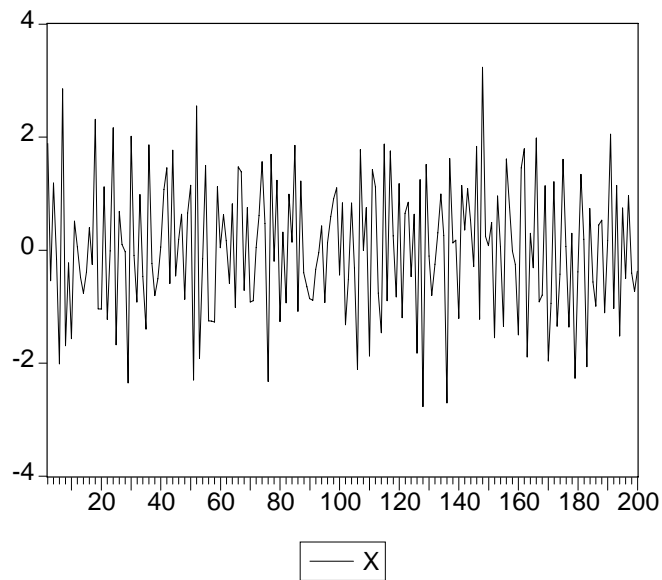


图 24 $X_t = a_t + 0.8a_{t-1}$ 趋势图

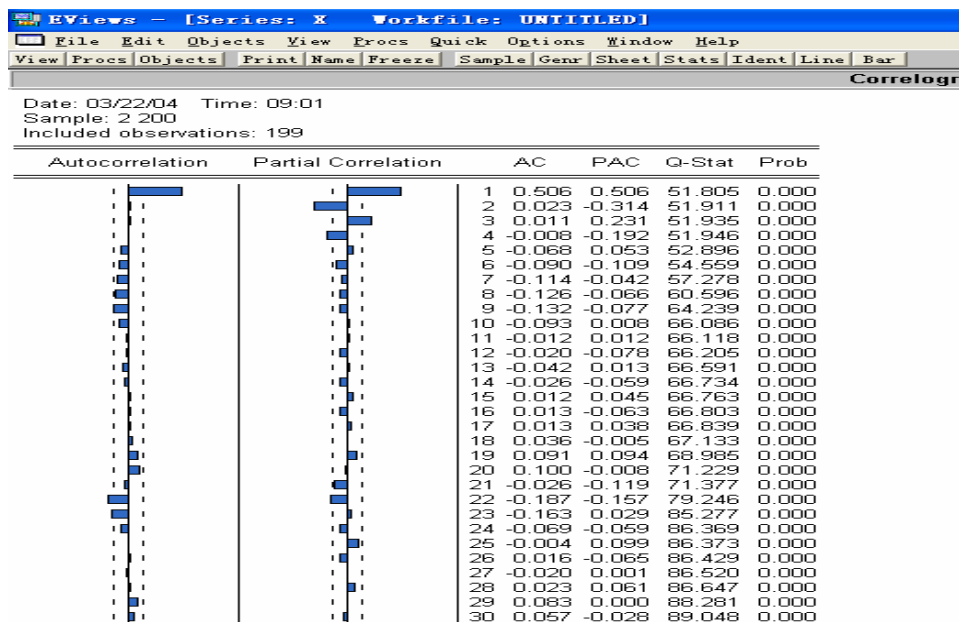


图 25 $X_t = a_t + 0.8a_{t-1}$ 自相关函数图

由此，我们已经有了识别 MA (q) 模型的工具，自相关函数 q 阶截尾。但是对于 AR (p) 和 ARMA (p, q) 模型，则无法区别了。

2.4.2 偏自相关函数 φ_{kk}

由 AR (p) 模型本身看，只涉及到 n 步相关性，但序列的自相关函数 ρ_k 确是拖尾的。例如 AR (1) 自回归模型只涉及到一步相关，结果其自相关函数 ρ_k 当 $k \rightarrow \infty$ 时才趋于零。之所以出现这种情况时因为这种相关关系是连带的，比如 x_t 与

x_{t-2} 的关系是由于 x_{t-1} 。于是我们可以设想，如果我们排除“局外”变量后，两个变量本身可能并不存在相关关系。这种现象说明：用简单的相关系数来描述两个变量之间的相关关系是不够的，需要有另一种表达两个变量之间相互关系的方式，它就是偏自相关系数。

AR (P) 模型的偏自相关函数 p 阶截尾。 $\varphi_{kk} = 0, k > p$ 。

注：偏自相关函数的概率意义是在给定 $X_{t-1}, \dots, X_{t-k+1}$ 的条件下， x_t 和 x_{t-k} 的相关系数。

ARMA(p,q)模型自相关和偏自相关均拖尾，但是快速收敛到零。

前面我们讨论了 ARMA 模型的统计特性。为了便于运用，将零均值平稳时间序列的自相关函数和偏自相关函数的统计特征归纳如下：

模 型	AR (p)	MA (q)	ARMA (P, q)
自相关函数 (ACF)	拖 尾	截 尾	拖 尾

偏自相关函数 (PACF)	截 尾	拖 尾	拖 尾
------------------	-----	-----	-----

掌握了 ARMA 模型的统计特性，我们就可以依据这些统计特征，初步识别模型的类型。对一个实际时间序列，我们能掌握的是一段样本数据，所以首先要利用样本数据估计模型的自相关函数和偏自相关函数。

【例】利用 1997 年 1 月—2002 年 12 月到北京海外旅游人数资料绘制自相关和偏自相关图，在这里去掉了 2003 年的

数据是由于非典的流行使 2003 年到北京旅游的人数锐减，出现奇异值，不具有一般性。如图 26 所示。

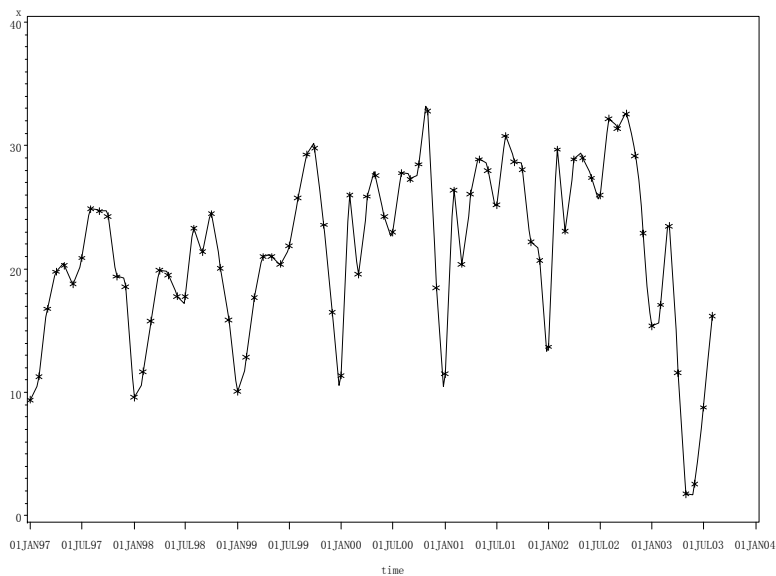


图 26 1997 年 1 月—2003 年 8 月到北京海外旅游人数曲线图

Autocorrelation

[illegible]

图 27 97 年 1 月到 03 年 8 月到北京海外旅游人数自相关图

图 27 显示滞后一期和滞后两期的自相关函数分别为 0.5874 和 0.35818，超过了两倍标准差，显著不为零，以后的自相关函数均显著为零，直到滞后期为周期的长度 12 时，自相关函数出现了峰值，为 0.66015，这是季节性时间序列的十分典型的特征，该序列从自相关函数看长期趋势并不十分显著。

		Partial Autocorrelations																				
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	0.58740								.			*****										
2	0.02008								.			.										
3	-0.07996								.	**		.										
4	0.00272								.			.										
5	0.06444								.			*	.									
6	0.19652								.			****.										
7	-0.25550								*****		.											
8	0.03626								.			*	.									
9	0.22313								.			****.										
10	0.24783								.			*****										
11	0.23918								.			*****										
12	0.39145								.			*****										
13	-0.30474								*****		.											
14	-0.16485								.	***		.										
15	-0.04449								.	*		.										
16	-0.02916								.	*		.										

图 28 97 年 1 月到 03 年 8 月到北京海外旅游人数偏自相关图

偏自相关函数图 28 显示滞后期为 1, 7, 12 和 13 的偏自相关函数分别为 0.5874、-0.2555、0.39145 和-0.30474, 显著不为零, 该时间序列的偏自相关函数显示该时间序列可能适

应 $(1 - \phi_1 B)(1 - \phi_{12} B^{12})X_t = a_t$ 或

$(1 - \phi_1 B - \phi_7 B^7 - \phi_{12} B^{12} - \phi_{13} B^{13})X_t = a_t$ 的模型。

从图 26、图 27 和图 28, 我们已经对 1997 年 1 月—2003 年 8 月到北京海外旅游人数的时间序列有了初步的了解, 在

滞后期为 1、7、12 和 13 的偏自相关系数显著不为零，自相关函数图趋势不明显。初步设定模型为

$$(1 - \phi_1 B - \phi_7 B^7 - \phi_{12} B^{12} - \phi_{13} B^{13}) X_t = a_t$$

利用极大似然估计得参数估计，如表 4 所示。

表 4 模型 $(1 - \phi_1 B - \phi_7 B^7 - \phi_{12} B^{12} - \phi_{13} B^{13}) X_t = a_t$ 的参数估计表

参数	参数估计	标准差	t 值	P 值
ϕ_1	0.52143	0.10861	4.80	<.0001
ϕ_7	0.05217	0.04886	1.07	0.2857

ϕ_{12}	0.75205	0.07390	10.18	<.0001
ϕ_{13}	-0.32729	0.12087	-2.71	0.0068

该模型为

$$X_t = \frac{1}{(1 - 0.521B + 0.0521B^7 - 0.75205B^{12} + 0.32729B^{13})} a_t$$
 , 从检验的 t 统计量看, 滞后期为 7 的自回归参数不显著。改进模型为 $(1 - \phi_1 B)(1 - \phi_{12} B^{12}) X_t = a_t$, 仔细观察改进的模型, 你会发现这是一个自回归参数为疏系数的自回归模型。仍用极大似然估计的方法, 可以得模型的估计。

表 5 型 $(1-\phi_1 B)(1-\phi_{12} B^{12})X_t = a_t$ 的参数估计表

参数	参数估计	标准差	t 值	P 值
ϕ_1	0.88874	0.05294	16.79	<.0001
ϕ_{12}	0.78445	0.07719	10.16	<.0001

表 5 显示，该模型为 $(1-0.889B)(1-0.784B^{12})X_t = a_t$ ，进一步对模型的适应性进行检验，回归系数均显著外，残差的自相关函数均落在两倍标准差内，可以认为残差序列是白噪声序列，如图 29 所示。

Autocorrelation Check of Residuals																								
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	19.699544	1.00000											*****											0
1	-3.262416	-.16561										.***	.											0.111803
2	0.345450	0.01754									.		.											0.114829
3	-0.177902	-.00903									.		.											0.114862
4	1.795260	0.09113									.	**	.											0.114871
5	-0.660013	-.03350									.	*	.											0.115771
6	1.952285	0.09910									.	**	.											0.115893
7	0.170288	0.00864									.		.											0.116947
8	-1.294234	-.06570									.	*	.											0.116955
9	1.049438	0.05327									.	*	.											0.117415
10	-0.169403	-.00860									.		.											0.117717
11	0.745819	0.03786									.	*	.											0.117725
12	-0.162516	-.00825									.		.											0.117877
13	2.614005	0.13269									.	***	.											0.117884

图 29 最终模型残差的自相关函数图

表 29 是最终模型残差的白噪声检验结果，结果表明残差序列可以视为白噪声序列，模型是适应的。

表 6 最终模型残差的白噪声检验结果表

滞后期	χ^2 统计量	自由度	P 值
1—6	4.00	4	0.4066
1—12	4.81	10	0.9036
1—18	11.05	16	0.8065

1—24	15.52	22	0.8391
------	-------	----	--------

当模型通过了检验，我们可以用该模型进行结构分析和预测分析了。

3 时间序列建模的方法

为了对时间序列建模有一个较全面的了解，下面从样本观测数据出发，介绍建立时间序列模型的基本步骤。

Box-Jenkins 方法是以序列的自相关函数和偏自相关函数的统计特性为依据，找出序列可能适应的模型，然后对模型

进行估计。通常可以考虑的模型 ARMA、ARIMA 和乘积型季节模型。

（一）模型的识别

对于一组长度为 N 的样本观测数据 x_1, x_2, \dots, x_N ，首先要对数据进行预处理，预处理的目的是实现平稳化，处理的手段包括差分和季节差分等。经过预处理的新序列能较好满足平稳性条件。

模型的识别包括差分阶数 d 、季节差分阶数 D 、模型阶数、

q 、 k 和 m 的识别。识别的工具是自相关函数和偏自相关函数。如果样本的自相关函数 $\hat{\rho}(s)$ 当 $s > q$ 时显著为零，则序列适应的模型是 $MA(q)$ 。如果样本的偏自相关函数 $\hat{\phi}_{ss}$ 当 $s > p$ 时显著为零，则序列适应的模型是 $AR(p)$ 。若样本的自相关函数和偏自相关函数均拖尾，并且按负指数衰减，则序列是 $ARMA$ 序列，这时应该从高阶到低阶拟合模型，从中选择最佳的。

当自相关函数缓慢下降，或是具有季节变化，那么观测的序列是具有趋势变动或季节变动的非平稳序列，则需要做差

分或季节差分，如果差分后的序列的样本的自相关函数和偏自相关函数既不截尾又不拖尾，而在周期 s 的整倍数时出现峰值，则序列遵从乘积型季节模型，否则遵从 **ARIMA** 模型。

（二）模型的估计

当模型的阶数确定之后，利用有效的拟合方法。如最小二乘估计，极大似然估计等方法，估计模型各部分的参数。

（三）诊断性检验模型选择

检验所许则的模型是否能较好地拟合数据。它包括模型过

拟合和欠拟合检验。通过检验的结果，修改模型。时间序列建模应该基于简约的原则，即用尽可能少的模型参数，对模型做出尽可能精确估计。所以在选择模型时应该反复试探，这是一个识别，建模，再识别，再建模的过程。

附录 1

AR 模型平稳的充分必要条件。

$$\text{由于 } \Phi(B)X_t = a_t$$

$$\text{有 } X_t = \frac{1}{\Phi(B)} a_t$$

设 $\Phi(B) = 0$ 有 $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_p}$ p 个根, 则 $\Phi(B)$ 可表示为

$$\Phi(B) = c(1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_p B), \text{ } c \text{ 为常数, 不妨假设}$$

$$\text{为 1。则 } X_t = \frac{1}{\Phi(B)} a_t = \frac{1}{(1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_p B)} a_t$$

用待定系数法，有

$$X_t = \frac{1}{\Phi(B)} a_t = \sum_{k=1}^p \frac{A_k}{(1 - \lambda_k B)} a_t \quad (\text{其中 } A_k \text{ 是有限实数})$$

再用等比级数通项和公式，有

$$\begin{aligned} X_t &= \frac{1}{\Phi(B)} a_t \\ &= \sum_{k=1}^p \frac{A_k}{(1 - \lambda_k B)} a_t \\ &= \sum_{k=1}^p A_k \left(\sum_{j=0}^{\infty} \lambda_k^j B^j \right) a_t \end{aligned}$$

$$= \sum_{j=0}^{\infty} \left(\sum_{k=1}^p A_k \lambda_k^j \right) a_{t-j}$$

$\sum_{k=1}^p A_k \lambda_k^j$ 是把 X_t 表示为白噪声的加权和的系数，根据前面

的结论，如果 $X_t = \frac{1}{\Phi(B)} a_t$ 平稳，其充分必要条件为权系数绝

对收敛，权系数绝对收敛的充分必要条件为所有 λ_k 的模小于

1，所以其根 $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_p}$ 的模大于 1，即在单位圆外。

$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_p}$ 的模大于 1，则 $\lambda_1, \lambda_2, \dots, \lambda_p$ 的模小于 1。

可见自回归模型的自回归多项式如果有在单位圆上的根，则可以称为时间序列是非平稳的，或存在趋势。

附录 2

MA 模型可逆的充分必要条件

设时间序列 $\{X_t\}$ 是 m 阶滑动平均模型，有

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} = \Theta(B)a_t$$

其中： $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

$\{X_t\}$ 可逆的充分必要条件是：

特征方程 $\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$ 的根在单位圆外。

证： 假 设 $\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$ 有 m 个 根

$\frac{1}{v_1}, \dots, \frac{1}{v_2}$, 则

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q =$$

$$(1 - v_1 B)(1 - v_2 B) \cdots (1 - v_q B)$$

$$\text{故 } a_t = \frac{1}{\Theta(B)} X_t = \frac{1}{(1 - v_1 B)(1 - v_2 B) \cdots (1 - v_q B)} X_t$$

用待定系数法，有上式为：

$$\begin{aligned}
a_t &= \frac{1}{\Theta(B)} X_t = \sum_{i=1}^q \frac{C_i}{(1-v_i B)} X_t \quad (\text{用等比级数通项和公式}) \\
&= \sum_{i=1}^m C_i \sum_{j=0}^{\infty} v_i^j B^j X_t \\
&= \sum_{j=0}^{\infty} \left(\sum_{i=1}^q C_i v_i^j \right) X_{t-j}
\end{aligned}$$

($\sum_{i=1}^q C_i v_i^j$ 是加权和的权数)

可见, $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} = \Theta(B) a_t$ 可逆的充分必要条件为 $\sum_{j=0}^{\infty} \left(\sum_{i=1}^q C_i v_i^j \right)$ 绝对收敛, $\sum_{j=0}^{\infty} \left(\sum_{i=1}^q C_i v_i^j \right)$ 绝对收敛

的充分必要条件为诸 $|v_i|$ 小于 1，故特征方程 $\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q = 0$ 的根在单位圆外。