

## 第二十四章 时间序列模型

时间序列是按时间顺序排列的、随时间变化且相互关联的数据序列。分析时间序列的方法构成数据分析的一个重要领域，即时间序列分析。

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。
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。
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。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  $\tau$  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。我们以后所研究的时间序列主要是宽平稳时间序列。

4. 按时间序列的分布规律来分，有 Gaussian 型时间序列和非 Gaussian 型时间序列。

### §1 时间序列分析方法概述

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。一个时间序列往往是以下几类变化形式的叠加或耦合。

（1）长期趋势变动。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。

（3）循环变动。通常是指周期为一年以上，由非季节因素引起的涨落起伏波形相似的波动。

（4）不规则变动。通常它分为突然变动和随机变动。

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

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

（2）乘法模型

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

（3）混合模型

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

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

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  $\sigma^2$  较小，并且有理由认为过去和现在的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。

### §2 移动平均法

移动平均法是根据时间序列资料逐渐推移，依次计算包含一定项数的时序平均数，以反映长期趋势的方法。当时间序列的数值由于受周期变动和不规则变动的影响，起伏较大，不易显示出发展趋势时，可用移动平均法，消除这些因素的影响，分析、预测序

## 列的长期趋势。

移动平均法有简单移动平均法，加权移动平均法，趋势移动平均法等。

### 2.1 简单移动平均法

设观测序列为  $y_1, \dots, y_T$ ，取移动平均的项数  $N < T$ 。一次简单移动平均值计算公式为：

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

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

$$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}(y_t + \dots + y_{t-N+1}), \quad t = N, N+1, \dots, T, \quad (2)$$

其预测标准误差为：

$$S = \sqrt{\frac{\sum_{t=N+1}^T (\hat{y}_t - y_t)^2}{T - N}}, \quad (3)$$

最近  $N$  期序列值的平均值作为未来各期的预测结果。一般  $N$  取值范围： $5 \leq N \leq 200$ 。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。预测标准误差最小者为好。

例 1 某企业 1 月~11 月份的销售收入时间序列如表 1 示。试用一次简单滑动平均法预测第 12 月份的销售收入。

表 1 企业销售收入

月份 $t$	1	2	3	4	5	6
销售收入 $y_t$	533.8	574.6	606.9	649.8	705.1	772.0
月份 $t$	7	8	9	10	11	
销售收入 $y_t$	816.4	892.7	963.9	1015.1	1102.7	

解： 分别取  $N = 4, N = 5$  的预测公式

$$\hat{y}_{t+1}^{(1)} = \frac{y_t + y_{t-1} + y_{t-2} + y_{t-3}}{4}, \quad t = 4, 5, \dots, 11$$

$$\hat{y}_{t+1}^{(2)} = \frac{y_t + y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}}{5}, \quad t = 5, \dots, 11$$

当  $N = 4$  时，预测值  $\hat{y}_{12}^{(1)} = 993.6$ ，预测的标准误差为

$$S_1 = \sqrt{\frac{\sum_{t=5}^{11} (\hat{y}_t^{(1)} - y_t)^2}{11 - 4}} = 150.5$$

当  $N = 5$  时，预测值  $\hat{y}_{12}^{(2)} = 182.4$ ，预测的标准误差为

$$S_2 = \sqrt{\frac{\sum_{t=6}^{11} (\hat{y}_t^{(2)} - y_t)^2}{11-5}} = 958.2$$

计算结果表明， $N = 4$  时，预测的标准误差较小，所以选取  $N = 4$ 。预测第 12 月份的销售收入为 993.6。

计算的 Matlab 程序如下：

```
clc,clear
y=[533.8 574.6 606.9 649.8 705.1 772.0 816.4 892.7 963.9 1015.1
1102.7];
m=length(y);
n=[4,5]; %n 为移动平均的项数
for i=1:length(n)
    %由于 n 的取值不同，yhat 的长度不一致，下面使用了细胞数组
    for j=1:m-n(i)+1
        yhat{i}(j)=sum(y(j:j+n(i)-1))/n(i);
    end
    y12(i)=yhat{i}(end);
    s(i)=sqrt(mean((y(n(i)+1:m)-yhat{i}(1:end-1)).^2));
end
y12,s
```

简单移动平均法只适合做近期预测，而且是预测目标的发展趋势变化不大的情况。

如果目标的发展趋势存在其它的变化，采用简单移动平均法就会产生较大的预测偏差和滞后。

## 2.2 加权移动平均法

在简单移动平均公式中，每期数据在求平均时的作用是等同的。但是，每期数据所包含的信息量不一样，近期数据包含着更多关于未来情况的信息。因此，把各期数据等同看待是不尽合理的，应考虑各期数据的重要性，对近期数据给予较大的权重，这就是加权移动平均法的基本思想。

设时间序列为  $y_1, y_2, \dots, y_t, \dots$ ；加权移动平均公式为

$$M_{nw} = \frac{w_1 y_t + w_2 y_{t-1} + \dots + w_N y_{t-N+1}}{w_1 + w_2 + \dots + w_N}, \quad t \geq N \quad (4)$$

式中  $M_{nw}$  为  $t$  期加权移动平均数； $w_i$  为  $y_{t-i+1}$  的权数，它体现了相应的  $y_t$  在加权平均数中的重要性。

利用加权移动平均数来做预测，其预测公式为

$$\hat{y}_{t+1} = M_{nw} \quad (5)$$

即以第  $t$  期加权移动平均数作为第  $t+1$  期的预测值。

例 2 我国 1979~1988 年原煤产量如表 2 所示，试用加权移动平均法预测 1989 年的产量。

表 2 我国原煤产量统计数据及加权移动平均预测值表

年份	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
原煤产量 $y_t$	6.35	6.20	6.22	6.66	7.15	7.89	8.72	8.94	9.28	9.8
三年加权移动平均预测值				6.235	6.4367	6.8317	7.4383	8.1817	8.6917	9.0733

相对误差 (%)				6.38	9.98	13.41	14.7	8.48	6.34	7.41
----------	--	--	--	------	------	-------	------	------	------	------

解 取  $w_1 = 3, w_2 = 2, w_3 = 1$ , 按预测公式

$$\hat{y}_{t+1} = \frac{3y_t + 2y_{t-1} + y_{t-2}}{3 + 2 + 1}$$

计算三年加权移动平均预测值, 其结果列于表 2 中。1989 年我国原煤产量的预测值为 (亿吨)

$$\hat{y}_{1989} = \frac{3 \times 9.8 + 2 \times 9.28 + 8.94}{6} = 9.48$$

这个预测值偏低, 可以修正。其方法是: 先计算各年预测值与实际值的相对误差, 例如 1982 年为

$$\frac{6.66 - 6.235}{6.66} = 6.38\%$$

将相对误差列于表 2 中, 再计算总的平均相对误差。

$$\left(1 - \frac{\sum \hat{y}_t}{\sum y_t}\right) \times 100\% = \left(1 - \frac{52.89}{58.44}\right) \times 100\% = 9.5\%$$

由于总预测值的平均值比实际值低 9.5%, 所以可将 1989 年的预测值修正为

$$\frac{9.48}{1 - 9.5\%} = 10.4788$$

计算的 MATLAB 程序如下:

```
y=[6.35 6.20 6.22 6.66 7.15 7.89 8.72 8.94 9.28
9.8];
w=[1/6;2/6;3/6];
m=length(y);n=3;
for i=1:m-n+1
    yhat(i)=y(i:i+n-1)*w;
end
yhat
err=abs(y(n+1:m)-yhat(1:end-1))./y(n+1:m)
T_err=1-sum(yhat(1:end-1))/sum(y(n+1:m))
y1989=yhat(end)/(1-T_err)
```

在加权移动平均法中,  $w_t$  的选择, 同样具有一定的经验性。一般的原则是: 近期数据的权数大, 远期数据的权数小。至于大到什么程度和小到什么程度, 则需要按照预测者对序列的了解和分析来确定。

### 2.3 趋势移动平均法

简单移动平均法和加权移动平均法, 在时间序列没有明显的趋势变动时, 能够准确反映实际情况。但当时间序列出现直线增加或减少的变动趋势时, 用简单移动平均法和加权移动平均法来预测就会出现滞后偏差。因此, 需要进行修正, 修正的方法是作二次移动平均, 利用移动平均滞后偏差的规律来建立直线趋势的预测模型。这就是趋势移动平均法。

一次移动的平均数为

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1})$$

在一次移动平均的基础上再进行一次移动平均就是二次移动平均，其计算公式为

$$M_t^{(2)} = \frac{1}{N}(M_t^{(1)} + \cdots + M_{t-N+1}^{(1)}) = M_{t-1}^{(2)} + \frac{1}{N}(M_t^{(1)} - M_{t-N}^{(1)}) \quad (6)$$

下面讨论如何利用移动平均的滞后偏差建立直线趋势预测模型。

设时间序列  $\{y_t\}$  从某时期开始具有直线趋势，且认为未来时期也按此直线趋势变化，则可设此直线趋势预测模型为

$$\hat{y}_{t+m} = a_t + b_t m, \quad m = 1, 2, \dots \quad (7)$$

其中  $t$  为当前时期数； $m$  为由  $t$  至预测期的时期数； $a_t$  为截距； $b_t$  为斜率。两者又称为平滑系数。

现在，我们根据移动平均值来确定平滑系数。由模型 (7) 可知

$$\begin{aligned} a_t &= y_t \\ y_{t-1} &= y_t - b_t \\ y_{t-2} &= y_t - 2b_t \\ &\dots \\ y_{t-N+1} &= y_t - (N-1)b_t \end{aligned}$$

所以

$$\begin{aligned} M_t^{(1)} &= \frac{y_t + y_{t-1} + \cdots + y_{t-N+1}}{N} = \frac{y_t + (y_t - b_t) + \cdots + [y_t - (N-1)b_t]}{N} \\ &= \frac{Ny_t - [1 + 2 + \cdots + (N-1)]b_t}{N} = y_t - \frac{N-1}{2}b_t \end{aligned}$$

因此

$$y_t - M_t^{(1)} = \frac{N-1}{2}b_t \quad (8)$$

由式 (7)，类似式 (8) 的推导，可得

$$y_{t-1} - M_{t-1}^{(1)} = \frac{N-1}{2}b_t \quad (9)$$

所以

$$y_t - y_{t-1} = M_t^{(1)} - M_{t-1}^{(1)} = b_t \quad (10)$$

类似式 (8) 的推导，可得

$$M_t^{(1)} - M_t^{(2)} = \frac{N-1}{2}b_t \quad (11)$$

于是，由式 (8) 和式 (11) 可得平滑系数的计算公式为

$$\begin{cases} a_t = 2M_t^{(1)} - M_t^{(2)} \\ b_t = \frac{2}{N-1}(M_t^{(1)} - M_t^{(2)}) \end{cases} \quad (12)$$

例 3 我国 1965~1985 年的发电总量如表 3 所示，试预测 1986 年和 1987 年的发电总量。

表3 我国发电量及一、二次移动平均值计算表

年份	$t$	发电总量 $y_t$	一次移动平均, $N=6$	二次移动平均, $N=6$
1965	1	676		
1966	2	825		
1967	3	774		
1968	4	716		
1969	5	940		
1970	6	1159	848.3	
1971	7	1384	966.3	
1972	8	1524	1082.8	
1973	9	1668	1231.8	
1974	10	1688	1393.8	
1975	11	1958	1563.5	1181.1
1976	12	2031	1708.8	1324.5
1977	13	2234	1850.5	1471.9
1978	14	2566	2024.2	1628.8
1979	15	2820	2216.2	1792.8
1980	16	3006	2435.8	1966.5
1981	17	3093	2625	2143.4
1982	18	3277	2832.7	2330.7
1983	19	3514	3046	2530
1984	20	3770	3246.7	2733.7
1985	21	4107	3461.2	2941.2

解 由散点图 1 可以看出, 发电总量基本呈直线上升趋势, 可用趋势移动平均法来预测。

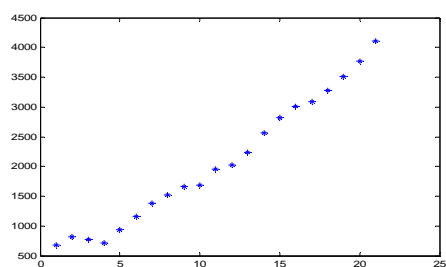


图1 原始数据散点图

取  $N = 6$ , 分别计算一次和二次移动平均值并列于表 3 中。

$$M_{21}^{(1)} = 3461.2, \quad M_{21}^{(2)} = 2941.2$$

再由公式 (12), 得

$$a_{21} = 2M_{21}^{(1)} - M_{21}^{(2)} = 3981.1$$

$$b_{21} = \frac{2}{6-1}(M_{21}^{(1)} - M_{21}^{(2)}) = 208$$

于是, 得  $t = 21$  时直线趋势预测模型为

$$\hat{y}_{21+m} = 3981.1 + 208m$$

预测 1986 年和 1987 年的发电总量为

$$\hat{y}_{1986} = \hat{y}_{22} = \hat{y}_{21+1} = 4192.1$$

$$\hat{y}_{1987} = \hat{y}_{23} = \hat{y}_{21+2} = 4397.1$$

计算的 MATLAB 程序如下：

```
clc,clear
load y.txt    %把原始数据保存在纯文本文件 y.txt 中
m1=length(y);
n=6;    %n 为移动平均的项数
for i=1:m1-n+1
    yhat1(i)=sum(y(i:i+n-1))/n;
end
yhat1
m2=length(yhat1);
for i=1:m2-n+1
    yhat2(i)=sum(yhat1(i:i+n-1))/n;
end
yhat2
plot(1:21,y,'*')
a21=2*yhat1(end)-yhat2(end)
b21=2*(yhat1(end)-yhat2(end))/(n-1)
y1986=a21+b21
y1987=a21+2*b21
```

趋势移动平均法对于同时存在直线趋势与周期波动的序列，是一种既能反映趋势变化，又可以有效地分离出来周期变动的方法。

### § 3 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。一般说来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。

指数平滑法根据平滑次数的不同，又分为一次指数平滑法、二次指数平滑法和三次指数平滑法等，分别介绍如下。

#### 3.1 一次指数平滑法

##### 1. 预测模型

设时间序列为  $y_1, y_2, \dots, y_t, \dots$ ， $\alpha$  为加权系数， $0 < \alpha < 1$ ，一次指数平滑公式为：

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha(y_t - S_{t-1}^{(1)}) \quad (13)$$

式 (13) 是由移动平均公式改进而来的。由式 (1) 知，移动平均数的递推公式为

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

以  $M_{t-1}^{(1)}$  作为  $y_{t-N}$  的最佳估计，则有

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + \left(1 - \frac{1}{N}\right) M_{t-1}^{(1)}$$

令  $\alpha = \frac{1}{N}$ ，以  $S_t$  代替  $M_t^{(1)}$ ，即得式 (13)

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

为进一步理解指数平滑的实质，把式 (13) 依次展开，有

$$S_t^{(1)} = \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha) S_{t-2}^{(1)}] = \cdots = \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j y_{t-j} \quad (14)$$

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

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

由于加权系数符合指数规律，又具有平滑数据的功能，故称为指数平滑。

以这种平滑值进行预测，就是一次指数平滑法。预测模型为

$$\hat{y}_{t+1} = S_t^{(1)}$$

即

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t \quad (15)$$

也就是以第  $t$  期指数平滑值作为  $t + 1$  期预测值。

## 2. 加权系数的选择

在进行指数平滑时，加权系数的选择是很重要的。由式 (15) 可以看出， $\alpha$  的大小规定了在新预测值中新数据和原预测值所占的比重。 $\alpha$  值越大，新数据所占的比重就愈大，原预测值所占的比重就愈小，反之亦然。若把式 (15) 改写为

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \quad (16)$$

则从上式可看出，新预测值是根据预测误差对原预测值进行修正而得到的。 $\alpha$  的大小则体现了修正的幅度， $\alpha$  值愈大，修正幅度愈大； $\alpha$  值愈小，修正幅度也愈小。

若选取  $\alpha = 0$ ，则  $\hat{y}_{t+1} = \hat{y}_t$ ，即下期预测值就等于本期预测值，在预测过程中不考虑任何新信息；若选取  $\alpha = 1$ ，则  $\hat{y}_{t+1} = y_t$ ，即下期预测值就等于本期观测值，完全不相信过去的信息。这两种极端情况很难做出正确的预测。因此， $\alpha$  值应根据时间序列的具体性质在 0~1 之间选择。具体如何选择一般可遵循下列原则：①如果时间序列波动不大，比较平稳，则  $\alpha$  应取小一点，如 (0.1~0.5)。以减少修正幅度，使预测模型能包含较长时间序列的信息；②如果时间序列具有迅速且明显的变动倾向，则  $\alpha$  应取大一点，如 (0.6~0.8)。使预测模型灵敏度高一些，以便迅速跟上数据的变化。

在实用上，类似移动平均法，多取几个  $\alpha$  值进行试算，看哪个预测误差小，就采用哪个。

## 3. 初始值的确定

用一次指数平滑法进行预测，除了选择合适的  $\alpha$  外，还要确定初始值  $s_0^{(1)}$ 。初始值是由预测者估计或指定的。当时间序列的数据较多，比如在 20 个以上时，初始值对以后的预测值影响很少，可选用第一期数据为初始值。如果时间序列的数据较少，在 20 个以下时，初始值对以后的预测值影响很大，这时，就必须认真研究如何正确确定初始



值。一般以最初几期实际值的平均值作为初始值。

例 4 某市 1976~1987 年某种电器销售额如表 4 所示。试预测 1988 年该电器销售额。

解 采用指数平滑法，并分别取  $\alpha = 0.2, 0.5$  和  $0.8$  进行计算，初始值

$$S_0^{(1)} = \frac{y_1 + y_2}{2} = 51$$

即

$$\hat{y}_1 = S_0^{(1)} = 51$$

按预测模型

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

计算各期预测值，列于表 4 中。

表 4 某种电器销售额及指数平滑预测值计算表 （单位：万元）

年份	$t$	实际销售额 $y_t$	预测值 $\hat{y}_t$ $\alpha = 0.2$	预测值 $\hat{y}_t$ $\alpha = 0.5$	预测值 $\hat{y}_t$ $\alpha = 0.8$
1976	1	50	51	51	51
1977	2	52	50.8	50.5	50.2
1978	3	47	51.04	51.25	51.64
1979	4	51	50.23	49.13	47.93
1980	5	49	50.39	50.06	50.39
1981	6	48	50.11	49.53	49.28
1982	7	51	49.69	48.77	48.26
1983	8	40	49.95	49.88	50.45
1984	9	48	47.96	44.94	42.09
1985	10	52	47.97	46.47	46.82
1986	11	51	48.77	49.24	50.96
1987	12	59	49.22	50.12	50.99

从表 4 可以看出， $\alpha = 0.2, 0.5$  和  $0.8$  时，预测值是很不相同的。究竟  $\alpha$  取何值为好，可通过计算它们的预测标准误差  $S$ ，选取使  $S$  较小的那个  $\alpha$  值。预测的标准误差见表 5。

表 5 预测的标准误差

$\alpha$	0.2	0.5	0.8
$S$	4.5029	4.5908	4.8426

计算结果表明： $\alpha = 0.2$  时， $S$  较小，故选取  $\alpha = 0.2$ ，预测 1988 年该电器销售额为  $\hat{y}_{1988} = 51.1754$ 。

计算的 MATLAB 程序如下：

```
clc,clear
load dianqi.txt %原始数据以列向量的方式存放在纯文本文件中
yt=dianqi; n=length(yt);
alpha=[0.2 0.5 0.8];m=length(alpha);
yhat(1,1:m)=(yt(1)+yt(2))/2;
for i=2:n
    yhat(i,:)=alpha*yt(i-1)+(1-alpha).*yhat(i-1,:);
end
```

```

yhat
err=sqrt(mean(( repmat(yt,1,m)-yhat).^2))
xlswrite('dianqi.xls',yhat)
yhat1988=alpha*yt(n)+(1-alpha).*yhat(n,:);

```

### 3.2 二次指数平滑法

一次指数平滑法虽然克服了移动平均法的缺点。但当时间序列的变动出现直线趋势时，用一次指数平滑法进行预测，仍存在明显的滞后偏差。因此，也必须加以修正。修正的方法与趋势移动平均法相同，即再作二次指数平滑，利用滞后偏差的规律建立直线趋势模型。这就是二次指数平滑法。其计算公式为

$$\begin{aligned}
S_t^{(1)} &= \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} \\
S_t^{(2)} &= \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)}
\end{aligned} \quad (17)$$

式中  $S_t^{(1)}$  为一次指数的平滑值； $S_t^{(2)}$  为二次指数的平滑值。当时间序列  $\{y_t\}$ ，从某时期开始具有直线趋势时，类似趋势移动平均法，可用直线趋势模型

$$\hat{y}_{t+m} = a_t + b_t m, \quad m = 1, 2, \dots \quad (18)$$

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \frac{\alpha}{1 - \alpha} (S_t^{(1)} - S_t^{(2)}) \end{cases} \quad (19)$$

进行预测。

例 5 仍以例 3 我国 1965~1985 年的发电总量资料为例，试用二次指数平滑法预测 1986 年和 1987 年的发电总量。

表 6 我国发电总量及一、二次指数平滑值计算表 (单位: 亿度)

年份	$t$	发电总量 $y_t$	一次平滑值	二次平滑值	$y_{t+1}$ 的估计值
1965	1	676	676	676	
1966	2	825	720.7	689.4	676
1967	3	774	736.7	703.6	765.4
1968	4	716	730.5	711.7	784.0
1969	5	940	793.3	736.2	757.4
1970	6	1159	903.0	786.2	875.0
1971	7	1384	1047.3	864.6	1069.9
1972	8	1524	1190.3	962.3	1308.4
1973	9	1668	1333.6	1073.7	1516.1
1974	10	1688	1439.9	1183.6	1705.0
1975	11	1958	1595.4	1307.1	1806.1
1976	12	2031	1726.1	1432.8	2007.2
1977	13	2234	1878.4	1566.5	2145.0
1978	14	2566	2084.7	1722.0	2324.1
1979	15	2820	2305.3	1897.0	2602.9
1980	16	3006	2515.5	2082.5	2888.6
1981	17	3093	2688.8	2264.4	3134.1
1982	18	3277	2865.2	2444.6	3295.0
1983	19	3514	3059.9	2629.2	3466.1
1984	20	3770	3272.9	2822.3	3675.1
1985	21	4107	3523.1	3032.6	3916.6

解 取  $\alpha = 0.3$ ，初始值  $S_0^{(1)}$  和  $S_0^{(2)}$  都取序列的首项数值，即  $S_0^{(1)} = S_0^{(2)} = 676$ 。  
计算  $S_t^{(1)}, S_t^{(2)}$ ，列于表 6。得到

$$S_{21}^{(1)} = 3523.1, S_{21}^{(2)} = 3032.6$$

由公式 (19)，可得  $t = 21$  时

$$a_{21} = 2S_{21}^{(1)} - S_{21}^{(2)} = 4013.7,$$

$$b_{21} = \frac{\alpha}{1-\alpha}(S_{21}^{(1)} - S_{21}^{(2)}) = 210.24$$

于是，得  $t = 21$  时直线趋势方程为

$$\hat{y}_{21+m} = 4013.7 + 210.24m$$

预测 1986 年和 1987 年的发电总量为（单位：亿度）

$$\hat{y}_{1986} = \hat{y}_{22} = \hat{y}_{21+1} = 4223.95$$

$$\hat{y}_{1987} = \hat{y}_{23} = \hat{y}_{21+2} = 4434.19$$

为了求各期的模拟值。可将式 (19) 代入直线趋势模型 (18)，并令  $m = 1$ ，则得

$$\hat{y}_{t+1} = (2S_t^{(1)} - S_t^{(2)}) + \frac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)})$$

即

$$\hat{y}_{t+1} = \left(1 + \frac{1}{1-\alpha}\right) S_t^{(1)} - \frac{1}{1-\alpha} S_t^{(2)} \quad (20)$$

令  $t = 1, 2, \dots, 20$ ，由公式 (20) 可求出各期的模拟值。计算结果见表 6。

计算的 MATLAB 程序如下：

```
clc,clear
load fadian.txt %原始数据以列向量的方式存放在纯文本文件中
yt=fadian; n=length(yt);
alpha=0.3; st1(1)=yt(1); st2(1)=yt(1);
for i=2:n
    st1(i)=alpha*yt(i)+(1-alpha)*st1(i-1);
    st2(i)=alpha*st1(i)+(1-alpha)*st2(i-1);
end
xlswrite('fadian.xls',[st1,st2])
a=2*st1-st2
b=alpha/(1-alpha)*(st1-st2)
yhat=a+b;
xlswrite('fadian.xls',yhat,'Sheet1','C2')
str=char(['C',int2str(n+2)]);
xlswrite('fadian.xls',a(n)+2*b(n),'Sheet1',str)
```

### 3.3 三次指数平滑法

当时间序列的变动表现为二次曲线趋势时，则需要用三次指数平滑法。三次指数平滑是在二次指数平滑的基础上，再进行一次平滑，其计算公式为

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha) S_{t-1}^{(3)} \end{cases} \quad (21)$$

式中  $S_t^{(3)}$  为三次指数平滑值。

三次指数平滑法的预测模型为

$$\hat{y}_{t+m} = a_t + b_t m + C_t m^2, \quad m = 1, 2, \dots \quad (22)$$

其中

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}] \\ c_t = \frac{\alpha^2}{2(1-\alpha)^2} [S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}] \end{cases} \quad (23)$$

例 6 某省 1978~1988 年全民所有制单位固定资产投资总额如表 7 所示，试预测 1989 年和 1990 年固定资产投资总额。

表 7 某省全民所有制单位固定资产投资总额及一、二、三次指数平滑值计算表（单位：亿元）

年份	$t$	投资总额 $y_t$	一次平滑值	二次平滑值	三次平滑值	$y_{t+1}$ 的估计值
1978	1	20.04	21.37	21.77	21.89	21.94
1979	2	20.06	20.98	21.53	21.78	20.23
1980	3	25.72	22.40	21.79	21.78	19.56
1981	4	34.61	26.06	23.07	22.17	24.49
1982	5	51.77	33.78	26.28	23.40	34.59
1983	6	55.92	40.42	30.52	25.54	53.89
1984	7	80.65	52.49	37.11	29.01	64.58
1985	8	131.11	76.07	48.80	34.95	89.30
1986	9	148.58	97.83	63.51	43.52	142.42
1987	10	162.67	117.28	79.64	54.35	176.09
1988	11	232.26	151.77	101.28	68.43	196.26

解 从图 2 可以看出，投资总额呈二次曲线上升，可用三次指数平滑法进行预测。

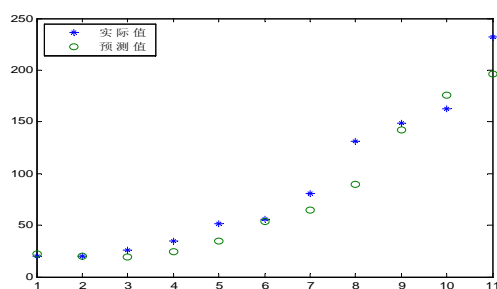


图 2 某省固定资产投资总额趋势图

取  $\alpha = 0.3$ ，初始值  $S_1^{(0)} = S_2^{(0)} = S_3^{(0)} = \frac{y_1 + y_2 + y_3}{3} = 21.94$ 。计算  $S_t^{(1)}, S_t^{(2)}, S_t^{(3)}$

列于表 7 中。得到

$$S_{11}^{(1)} = 151.77, S_{11}^{(2)} = 101.28, S_{11}^{(3)} = 68.43$$

由公式 (23)，可得到当  $t = 11$  时

$$a_{11} = 219.91, b_{11} = 38.38, c_{11} = 1.62$$

于是，得  $t = 11$  时预测模型为

$$\hat{y}_{11+m} = 219.91 + 38.38m + 1.62m^2$$

预测 1989 年和 1990 年的固定资产投资总额为（单位：亿元）

$$\hat{y}_{1989} = \hat{y}_{12} = \hat{y}_{11+1} = a_{11} + b_{11} + c_{11} = 259.91$$

$$\hat{y}_{1990} = \hat{y}_{13} = \hat{y}_{11+2} = a_{11} + 2b_{11} + 2^2c_{11} = 303.16$$

因为国家从 1989 年开始对固定资产投资采取压缩政策，这些预测值显然偏高，应作适当的修正，以消除政策因素的影响。

与二次指数平滑法一样，为了计算各期的模拟值，可将式 (23) 代入预测模型 (22)，并令  $m = 1$ ，则得

$$\hat{y}_{t+1} = \frac{3-3\alpha+\alpha^2}{(1-\alpha)^2} S_t^{(1)} - \frac{3-\alpha}{(1-\alpha)^2} S_t^{(2)} + \frac{1}{(1-\alpha)^2} S_t^{(3)} \quad (24)$$

令  $t = 0, 1, 2, \dots, 10$ ，公式 (24) 可求出各期的模拟值，见表 7。

计算的 MATLAB 程序如下：

```
clc,clear
load touzi.txt %原始数据以列向量的方式存放在纯文本文件中
yt=touzi; n=length(yt);
alpha=0.3; st1_0=mean(yt(1:3)); st2_0=st1_0; st3_0=st1_0;
st1(1)=alpha*yt(1)+(1-alpha)*st1_0;
st2(1)=alpha*st1(1)+(1-alpha)*st2_0;
st3(1)=alpha*st2(1)+(1-alpha)*st3_0;
for i=2:n
    st1(i)=alpha*yt(i)+(1-alpha)*st1(i-1);
    st2(i)=alpha*st1(i)+(1-alpha)*st2(i-1);
    st3(i)=alpha*st2(i)+(1-alpha)*st3(i-1);
end
xlswrite('touzi.xls',[st1',st2',st3'])
st1=[st1_0,st1]; st2=[st2_0,st2]; st3=[st3_0,st3];
a=3*st1-3*st2+st3;
b=0.5*alpha/(1-alpha)^2*((6-5*alpha)*st1-2*(5-4*alpha)*st2+(4-3*alpha)*st3);
c=0.5*alpha^2/(1-alpha)^2*(st1-2*st2+st3);
yhat=a+b+c;
xlswrite('touzi.xls',yhat,'Sheet1','D1')
plot(1:n,yt,'*',1:n,yhat(1:n),'O')
legend('实际值','预测值',2)
xishu=[c(n+1),b(n+1),a(n+1)];
yhat1990=polyval(xishu,2)
```

指数平滑预测模型是以时刻  $t$  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  $\alpha$  是提高预测精度的关键环节。根据实践经验,  $\alpha$  的取值范围一般以  $0.1 \sim 0.3$  为宜。 $\alpha$  值愈大, 加权系数序列衰减速度愈快, 所以实际上  $\alpha$  取值大小起着控制参加平均的历史数据的个数的作用。 $\alpha$  值愈大意味着采用的数据愈少。因此, 可以得到选择  $\alpha$  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  $\alpha$  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的变化, 则  $\alpha$  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。

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  $S_0^{(1)}, S_0^{(2)}, S_0^{(3)}$ 。可以取前  $3 \sim 5$  个数据的算术平均值作为初始值。

#### §4 差分指数平滑法

在上节我们已经讲过, 当时间序列的变动具有直线趋势时, 用一次指数平滑法会出现滞后偏差, 其原因在于数据不满足模型要求。因此, 我们也可以从数据变换的角度来考虑改进措施, 即在运用指数平滑法以前先对数据作一些技术上的处理, 使之能适合于一次指数平滑模型, 以后再对输出结果作技术上的返回处理, 使之恢复为原变量的形态。差分方法是改变数据变动趋势的简易方法。下面我们讨论如何用差分方法来改进指数平滑法。

##### 4.1 一阶差分指数平滑法

当时间序列呈直线增加时, 可运用一阶差分指数平滑模型来预测。其公式如下:

$$\nabla y_t = y_t - y_{t-1} \quad (25)$$

$$\nabla \hat{y}_{t+1} = \alpha \nabla y_t + (1 - \alpha) \nabla \hat{y}_t \quad (26)$$

$$\hat{y}_{t+1} = \nabla \hat{y}_{t+1} + y_t \quad (27)$$

其中的  $\nabla$  为差分记号。式 (25) 表示对呈现直线增加的序列作一阶差分, 构成一个平稳的新序列; 式 (27) 表示把经过一阶差分后的新序列的指数平滑预测值与变量当前的实际值迭加, 作为变量下一期的预测值。对于这个公式的数学意义可作如下的解释。

因为

$$y_{t+1} = y_{t+1} - y_t + y_t = \nabla y_{t+1} + y_t \quad (28)$$

当我们采用按式 (26) 计算的预测值去估计式 (28) 中的  $y_{t+1}$ , 从而式 (28) 等号左边的  $y_{t+1}$  也要改为预测值, 亦即成为式 (27)。

在前面我们已分析过, 指数平滑值实际上是一种加权平均数。因此把序列中逐期增量的加权平均数 (指数平滑值) 加上当前值的实际数进行预测, 比一次指数平滑法只用变量以往取值的加权平均数作为下一期的预测更合理。从而使预测值始终围绕实际值上下波动, 从根本上解决了在有直线增长趋势的情况下, 用一次指数平滑法所得出的结果始终落后于实际值的问题。

例 7 某工业企业 1977~1986 年锅炉燃料消耗量资料如表 8 所示, 试预测 1987 年的燃料消耗量。

表 8 某企业锅炉燃料消耗量的差分指数平滑法计算表 ( $\alpha = 0.4$ ) (单位: 百吨)

年份	$t$	燃料消耗量 $y_t$	差分	差分指数平滑值	预测值
1977	1	24			

1978	2	26	2		
1979	3	27	1	2	28
1980	4	30	3	1.6	28.6
1981	5	32	2	2.16	32.16
1982	6	33	1	2.10	34.10
1983	7	36	3	1.66	34.66
1984	8	40	4	2.19	38.19
1985	9	41	1	2.92	42.92
1986	10	44	3	2.15	43.15
1987	11			2.49	46.49

解 由资料可以看出，燃料消耗量，除个别年份外，逐期增长量大体在 200 吨左右，即呈直线增长，因此可用一阶差分指数平滑模型来预测。我们取  $\alpha = 0.4$ ，初始值为差分序列首项值，计算结果列于表 8 中。预测 1987 年燃料消耗量为

$$\hat{y}_{1987} = 2.49 + 44 = 46.49 \text{ (百吨)}。$$

#### 4.2 二阶差分指数平滑模型

当时间序列呈现二次曲线增长时，可用二阶差分指数平滑模型来预测，计算公式如下：

$$\nabla y_t = y_t - y_{t-1} \quad (29)$$

$$\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} \quad (30)$$

$$\nabla^2 \hat{y}_{t+1} = \alpha \nabla^2 y_t + (1 - \alpha) \nabla^2 \hat{y}_t \quad (31)$$

$$\hat{y}_{t+1} = \nabla^2 \hat{y}_{t+1} + \nabla y_t + y_t \quad (32)$$

其中  $\nabla^2$  表示二阶差分。

因为

$$y_{t+1} = y_{t+1} - y_t + y_t = \nabla y_{t+1} + y_t = (\nabla y_{t+1} - \nabla y_t) + \nabla y_t + y_t = \nabla^2 y_{t+1} + \nabla y_t + y_t$$

同样，用  $\nabla^2 y_{t+1}$  的估计值代替  $\nabla^2 y_{t+1}$  得到式 (32)。

差分方法和指数平滑法的联合运用，除了能克服一次指数平滑法的滞后偏差之外，对初始值的问题也有显著的改进。因为数据经过差分处理后，所产生的新序列基本上是平稳的。这时，初始值取新序列的第一期数据对于未来预测值不会有多大影响。其次，它拓展了指数平滑法的适用范围，使一些原来需要运用配合直线趋势模型处理的情况可用这种组合模型来取代。但是，对于指数平滑法存在的加权系数  $\alpha$  的选择问题，以及只能逐期预测问题，差分指数平滑模型也没有改进。

### § 5 自适应滤波法

#### 5.1 自适应滤波法的基本过程

自适应滤波法与移动平均法、指数平滑法一样，也是以时间序列的历史观测值进行某种加权平均来预测的，它要寻找一组“最佳”的权数，其办法是先用一组给定的权数来计算一个预测值，然后计算预测误差，再根据预测误差调整权数以减少误差。这样反复进行，直至找出一组“最佳”权数，使误差减少到最低限度。由于这种调整权数的过程与通讯工程中的传输噪声过滤过程极为接近，故称为自适应滤波法。

自适应滤波法的基本预测公式为

$$\hat{y}_{t+1} = w_1 y_t + w_2 y_{t-1} + \cdots + w_N y_{t-N+1} = \sum_{i=1}^N w_i y_{t-i+1} \quad (33)$$

式(33)中,  $\hat{y}_{t+1}$  为第  $t+1$  期的预测值,  $w_i$  为第  $t-i+1$  期的观测值权数,  $y_{t-i+1}$  为第  $t-i+1$  期的观测值,  $N$  为权数的个数。其调整权数的公式为

$$w'_i = w_i + 2k \cdot e_{i+1} y_{t-i+1} \quad (34)$$

式中,  $i=1, 2, \dots, N$ ,  $t=N, N+1, \dots, n$ ,  $n$  为序列数据的个数,  $w_i$  为调整前的第  $i$  个权数,  $w'_i$  为调整后的第  $i$  个权数,  $k$  为学习常数,  $e_{i+1}$  为第  $t+1$  期的预测误差。式(34)表明: 调整后的一组权数应等于旧的一组权数加上误差调整项, 这个调整项包括预测误差、原观测值和学习常数等三个因素。学习常数  $k$  的大小决定权数调整的速度。

下面举一个简单的例子来说明此法的全过程。设有一个时间序列包括 10 个观测值, 如表 9 所示。试用自适应滤波法, 以两个权数来求第 11 期的预测值。

表 9 某时间序列数据表

时期 $t$	1	2	3	4	5	6	7	8	9	10
观测值 $y_t$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

本例中  $N=2$ 。取初始权数  $w_1=0.5$ ,  $w_2=0.5$ , 并设  $k=0.9$ 。 $t$  的取值由  $N=2$  开始, 当  $t=2$  时:

(1) 按预测公式(33), 求第  $t+1=3$  期的预测值。

$$\hat{y}_{t+1} = \hat{y}_3 = w_1 y_2 + w_2 y_1 = 0.15$$

(2) 计算预测误差。

$$e_{t+1} = e_3 = y_3 - \hat{y}_3 = 0.3 - 0.15 = 0.15$$

(3) 根据式(34),

$$w'_i = w_i + 2k \cdot e_{i+1} y_{t-i+1}$$

调整权数为

$$w'_1 = w_1 + 2k e_3 y_2 = 0.554$$

$$w'_2 = w_2 + 2k e_3 y_1 = 0.527$$

(1) ~ (3) 结束, 即完成了一次权数调整, 然后  $t$  进 1 再重复以前步骤。当  $t=3$  时:

(1) 利用所得到的权数, 计算第  $t+1=4$  期的预测值。方法是, 舍去最前面的一个观测值  $y_1$ , 增加一个新的观测值  $y_3$ 。即

$$\hat{y}_{t+1} = \hat{y}_4 = w'_1 y_3 + w'_2 y_2 = 0.2716$$

(2) 计算预测误差

$$e_{t+1} = e_4 = y_4 - \hat{y}_4 = 0.13$$

(3) 调整权数

$$w'_1 = 0.554 + 2 \times 0.9 \times 0.13 \times 0.3 = 0.624$$

$$w'_2 = 0.527 + 2 \times 0.9 \times 0.13 \times 0.2 = 0.564$$

这样进行到  $t=10$  时

$$\hat{y}_{t+1} = \hat{y}_{11} = w'_1 y_{10} + w'_2 y_9$$



但由于没有  $t=11$  的观测值  $y_{11}$ ，因此

$$e_{t+1} = e_{11} = y_{11} - \hat{y}_{11}$$

无法计算。这时，第一轮的调整就此结束。把现有的新权数作为初始权数，重新开始  $t=2$  的过程。这样反复进行下去，到预测误差（指新一轮的预测总误差）没有明显改进时，就认为获得了一组“最佳”权数，能实际用来预测第 11 期的数值。本例在调整过程中，可使得误差降为零，而权数达到稳定不变，最后得到的“最佳”权数为

$$w_1' = 2.0, \quad w_2' = -1.0$$

用“最佳”权数预测第 11 期的取值

$$\hat{y}_{11} = w_1' y_{10} + w_2' y_9 = 1.1$$

在实际应用中，权数调整计算工作量可能很大，必须借助于计算机才能实现。

计算的 MATLAB 程序如下：

```
clc,clear
yt=0.1:0.1:1;
m=length(yt); k=0.9;
N=2; Terr=10000;
w=ones(1,N)/N;
while abs(Terr)>0.00001
    Terr=[];
    for j=N+1:m-1
        yhat(j)=w*yt(j-1:-1:j-N)';
        err=yt(j)-yhat(j);
        Terr=[Terr,abs(err)];
        w=w+2*k*err*yt(j-1:-1:j-N);
    end
    Terr=max(Terr);
end
w, yhat
```

## 5.2 $N, k$ 值和初始权数的确定

在开始调整权数时，首先要确定权数个数  $N$  和学习常数  $k$ 。一般说来，当时间序列的观测值呈季节变动时， $N$  应取季节性长度值。如序列以一年为周期进行季节变动时，若数据是月度的，则取  $N=12$ ，若季节是季度的，则取  $N=4$ 。如果时间序列无明显的周期变动，则可用自相关系数法来确定，即取  $N$  为最高自相关系数的滞后时期。

$k$  的取值一般可定为  $1/N$ ，也可以用不同的  $k$  值来进行计算，以确定一个能使  $S$  最小的  $k$  值。

初始权数的确定也很重要，如无其它依据，也可用  $1/N$  作为初始权系数用，即

$$w_i = \frac{1}{N} (i=1, 2, \dots, N)$$

自适应滤波法有两个明显的优点：一是技术比较简单，可根据预测意图来选择权数的个数和学习常数，以控制预测。也可以由计算机自动选定。二是它使用了全部历史数据来寻求最佳权系数，并随数据轨迹的变化而不断更新权数，从而不断改进预测。

由于自适应滤波法的预测模型简单，又可以在计算机上对数据进行处理，所以这种预测方法应用较为广泛。

## §6 趋势外推预测方法

趋势外推法是根据事物的历史和现时资料，寻求事物发展规律，从而推测出事物未来状况的一种比较常用的预测方法。利用趋势外推法进行预测，主要包括六个阶段：(a) 选择应预测的参数；(b) 收集必要的资料；(c) 利用数据拟合曲线；(d) 趋势外推；(e) 预测说明；(f) 研究预测结果在进行决策中应用的可能性。

趋势外推法常用的典型数学模型有：指数曲线、修正指数曲线、生长曲线、包络曲线等。

### 6.1 指数曲线法

一般来说，技术的进步和生产的增长，在其未达饱和之前的新生时期是遵循指数曲线增长规律的，因此可以用指数曲线对发展中的事物进行预测。

指数曲线的数学模型为

$$y = y_0 e^{Kt} \quad (35)$$

其中系数  $y_0$  和  $K$  值由历史数据利用回归方法求得。对式 (35) 取对数可得

$$\ln y = \ln y_0 + Kt \quad (36)$$

令

$$Y = \ln y, \quad A = \ln y_0$$

则

$$Y = A + Kt$$

其中  $A, K$  可以用最小二乘法求得。

### 6.2 修正指数曲线法

利用指数曲线外推来进行预测时，存在着预测值随着时间的推移会无限增大的情况。这是不符合客观规律的。因为任何事物的发展都是有一定限度的。例如某种畅销产品，在其占有市场的初期是呈指数曲线增长的，但随着产品销售量的增加，产品总量接近于社会饱和量时。这时的预测模型应改用修正指数曲线。

$$\hat{y}_t = K + ab^t \quad (37)$$

在此数学模型中有三个参数  $K, a$  和  $b$  要用历史数据来确定。

修正指数曲线用于描述这样一类现象。

(1) 初期增长迅速，随后增长率逐渐降低。

(2) 当  $K > 0, a < 0, 0 < b < 1$  时， $t \rightarrow \infty, ab^t \rightarrow 0$ ，即  $\hat{y}_t \rightarrow K$ 。

当  $K$  值可预先确定时，采用最小二乘法确定模型中的参数。而当  $K$  值不能预先确定时，应采用三和法。

把时间序列的  $n$  个观察值等分为三部分，每部分有  $m$  期，即  $n = 3m$ 。

第一部分：  $y_1, y_2, \dots, y_m$ ；

第二部分：  $y_{m+1}, y_{m+2}, \dots, y_{2m}$ ；

第三部分：  $y_{2m+1}, y_{2m+2}, \dots, y_{3m}$

令每部分的趋势值之和等于相应的观察值之和，由此给出参数估计值。三和法步骤如下：

记观察值的各部分之和

$$S_1 = \sum_{t=1}^m y_t, S_2 = \sum_{t=m+1}^{2m} y_t, S_3 = \sum_{t=2m+1}^{3m} y_t \quad (38)$$

且

$$\begin{cases} S_1 = \sum_{t=1}^m \hat{y}_t = \sum_{t=1}^m (K + ab^t) = mK + ab(1 + b + b^2 + \cdots b^{m-1}) \\ S_2 = \sum_{t=m+1}^{2m} \hat{y}_t = \sum_{t=m+1}^{2m} (K + ab^t) = mK + ab^{m+1}(1 + b + b^2 + \cdots b^{m-1}) \\ S_3 = \sum_{t=2m+1}^{3m} \hat{y}_t = \sum_{t=2m+1}^{3m} (K + ab^t) = mK + ab^{2m+1}(1 + b + b^2 + \cdots b^{m-1}) \end{cases} \quad (39)$$

由于

$$(1 + b + b^2 + \cdots b^{m-1})(b-1) = b^m - 1 \quad (40)$$

则根据(39)式, 得

$$\begin{cases} S_1 = mK + ab \frac{b^m - 1}{b - 1} \\ S_2 = mK + ab^{m+1} \frac{b^m - 1}{b - 1} \\ S_3 = mK + ab^{2m+1} \frac{b^m - 1}{b - 1} \end{cases} \quad (41)$$

由(41)式, 解得

$$\begin{cases} b = \left( \frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ a = (S_2 - S_1) \frac{b-1}{b(b^m - 1)^2} \\ K = \frac{1}{m} \left[ S_1 - \frac{ab(b^m - 1)}{b-1} \right] \end{cases} \quad (42)$$

至此三个参数全部确定了, 于是就可以用式(37)进行预测。

值得注意的是, 并不是任何一组数据都可以用修正指数曲线拟合。采用前应对数据进行检验, 检验方法是看给定数据的逐期增长量的比率是否接近某一常数  $b$ 。即

$$\frac{y_{t+1} - y_t}{y_t - y_{t-1}} \approx b \quad (43)$$

例 8 根据统计资料, 某厂收音机连续 15 年的销售量如表 10。

表 11 某厂收音机销售量

时间(年)	1969	1970	1971	1972	1973	1974	1975	1976
销售量(万部)	42.1	47.5	52.7	57.7	62.5	67.1	71.5	75.7
时间(年)	1977	1978	1979	1980	1981	1982	1983	
销售量(万部)	79.8	83.7	87.5	91.1	94.6	97.9	101.1	

试用修正指数曲线预测 1986 年的销售量。

解 经计算可知

$$\frac{y_{t+1} - y_t}{y_t - y_{t-1}} \in [0.9429, 0.9762]$$

可以认定这组数据可以采用修正指数曲线拟合。现将以上 15 个数据分为三部分，每部分 5 个数据，即  $n=15$ ， $m=5$ ，并以 1969 年作为开始年份  $t=1$ 。

根据式 (38)，得

$$S_1 = 262.5, S_2 = 377.8, S_3 = 472.2$$

再由 (42) 式，得

$$b = 0.9608, a = -143.2063, K = 179.7162$$

故修正指数曲线的数学模型为

$$y = 179.7162 - 143.2063 \times 0.9608^t \quad (44)$$

预测 1986 年的产量时， $t = 1986 - 1969 + 1 = 18$ 。所以

$$y_{1986} = 179.7162 - 143.2063 \times 0.9608^{18} = 110 \text{ (万部)}$$

计算的 MATLAB 程序如下 (所有代码放在一个文件中)：

```
function chanliang
clc,clear
global a b k
load xsh.txt %原始数据存放在纯文本文件 xsh.txt 中
yt=xsh; n=length(yt);m=n/3
cf=diff(yt);
for i=1:n-2
    bzh(i)=cf(i+1)/cf(i);
end
range=minmax(bzh)
s1=sum(yt(1:m)), s2=sum(yt(m+1:2*m)), s3=sum(yt(2*m+1:end))
b=((s3-s2)/(s2-s1))^(1/m)
a=(s2-s1)*(b-1)/(b*(b^m-1)^2)
k=(s1-a*b*(b^m-1)/(b-1))/m
y=yuce(1:18)
%*****
%定义预测函数
%*****
function y=yuce(t)
global a b k
y=k+a*b.^t;
```

### 6.3 Compertz 曲线

曲线的一般形式

$$\hat{y}_t = Ka^{b^t}, K > 0, 0 < a \neq 1, 0 < b \neq 1 \quad (45)$$

采用 Compertz 曲线前应对数据进行检验，检验方法是看给定数据的对数逐期增长量的比率是否接近某一常数  $b$ 。即

$$\frac{\ln y_{t+1} - \ln y_t}{\ln y_t - \ln y_{t-1}} \approx b \quad (46)$$

Compertz 曲线用于描述这样一类现象：初期增长缓慢，以后逐渐加快。当达到一定程度后，增长率又逐渐下降。

参数估计方法如下：

式 (46) 两边取对数, 得

$$\ln \hat{y}_t = \log K + (\ln a)b^t \quad (47)$$

记

$$\hat{y}'_t = \ln \hat{y}_t, K' = \ln K, a' = \ln a$$

得

$$\hat{y}'_t = K' + a'b^t$$

仿照修正指数曲线的三和法估计参数, 令

$$S_1 = \sum_{t=1}^m y'_t, S_2 = \sum_{t=m+1}^{2m} y'_t, S_3 = \sum_{t=2m+1}^{3m} y'_t \quad (48)$$

其中  $y'_t = \ln y_t$ 。则类似式 (42), 得

$$\begin{cases} b = \left( \frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ a' = (S_2 - S_1) \frac{b-1}{b(b^m-1)^2} \\ K' = \frac{1}{m} \left[ S_1 - \frac{a'b(b^m-1)}{b-1} \right] \end{cases} \quad (49)$$

例 9 (续例 8) 根据表 10 的数据, 试确定收音机销售量的 Gompertz 曲线方程, 求出各年收音机销售量的趋势值, 并预测 1986 年的销售量。

解: 已知  $n=15$ ,  $m=5$ , 根据式 (48), 得

$$S_1 = 19.7558, S_2 = 21.6094, S_3 = 22.7333$$

再由式 (49), 得

$$b = 0.9048$$

$$a' = -1.2588, a = 0.284$$

$$K' = 4.8929, K = 133.3341$$

从而收音机销售量的 Compertz 曲线方程为

$$\hat{y}_t = 133.3341 \times 0.284^{0.9048^t}$$

将  $t=18$  代入方程, 得 1986 年收音机销售量的预测值为

$$\hat{y}_{1986} = 108.3143 \text{ (万部)}$$

计算的 MATLAB 程序如下:

```
clc,clear
yuce=@(t,a,b,k) k*a.^(b.^t); %定义预测的匿名函数
load xsh.txt %原始数据存放在纯文本文件 xsh.txt 中
yt=log(xsh); n=length(yt); m=n/3;
s1=sum(yt(1:m)), s2=sum(yt(m+1:2*m)), s3=sum(yt(2*m+1:end))
b=((s3-s2)/(s2-s1))^(1/m)
a=(s2-s1)*(b-1)/(b*(b^m-1)^2)
k=(s1-a*b*(b^m-1)/(b-1))/m
a=exp(a)
k=exp(k)
```

$y=yuce(1:18,a,b,k)$

#### 6.4 Logistic 曲线（生长曲线）

生物的生长过程经历发生、发展到成熟三个阶段，在三个阶段生物的生长速度是不一样的，例如南瓜的重量增长速度，在第一阶段增长的较慢，在发展时期则突然加快，而到了成熟期又趋减慢，形成一条 S 形曲线，这就是有名的 Logistic 曲线（生长曲线），很多事物，如技术和产品发展进程都有类似的发展过程，因此 Logistic 曲线在预测中有相当广泛的应用。

Logistic 曲线的一般数学模型是

$$\frac{dy}{dt} = ry(1 - \frac{y}{L}) \quad (50)$$

式中  $y$  为预测值， $L$  为  $y$  的极限值， $r$  为增长率常数， $r > 0$ 。解此微分方程得

$$y = \frac{L}{1 + ce^{-rt}} \quad (51)$$

式中  $c$  为常数。

下面我们记 Logistic 曲线的一般形式为

$$y_t = \frac{1}{K + ab^t}, \quad K > 0, \quad a > 0, \quad 0 < b \neq 1 \quad (52)$$

检验能否使用 Logistic 曲线的方法，是看给定数据倒数的逐期增长量的比率是否接近某一常数  $b$ 。即

$$\frac{1/y_{t+1} - 1/y_t}{1/y_t - 1/y_{t-1}} \approx b \quad (53)$$

Logistic 曲线中参数估计方法如下：  
作变换

$$y'_t = \frac{1}{y_t}$$

得

$$y'_t = K + ab^t$$

仿照修正指数曲线的三和法估计参数，令

$$S_1 = \sum_{t=1}^m y'_t, S_2 = \sum_{t=m+1}^{2m} y'_t, S_3 = \sum_{t=2m+1}^{3m} y'_t \quad (54)$$

则类似式 (42)，得

$$\begin{cases} b = \left( \frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ a = (S_2 - S_1) \frac{b-1}{b(b^m-1)^2} \\ K = \frac{1}{m} \left[ S_1 - \frac{ab(b^m-1)}{b-1} \right] \end{cases} \quad (55)$$

例 10（续例 8） 根据表 10 的数据，试确定收音机销售量的 Logistic 曲线方程，

求出各年收音机销售量的趋势值，并预测 1986 年的销售量。

解：已知  $n = 15$ ， $m = 5$ ，根据式 (54)，得

$$S_1 = 0.0971, S_2 = 0.0666, S_3 = 0.0531$$

再由式 (55)，得

$$b = 0.8493, a = 0.0174, K = 0.0085$$

从而收音机销售量的 Logistic 曲线方程为

$$\hat{y}_t = \frac{1}{0.0085 + 0.0174 \times 0.8493^t}$$

将  $t = 18$  代入方程，得 1986 年收音机销售量的预测值为

$$\hat{y}_{1986} = 106.3981$$

计算的 MATLAB 程序如下：

```
clc,clear
yuce=@(t,a,b,k) 1./(k+a*b.^t); %定义预测的匿名函数
load xsh.txt %原始数据存放在纯文本文件 xsh.txt 中
yt=1./xsh; n=length(yt);m=n/3;
s1=sum(yt(1:m)), s2=sum(yt(m+1:2*m)), s3=sum(yt(2*m+1:end))
b=((s3-s2)/(s2-s1))^(1/m)
a=(s2-s1)*(b-1)/(b*(b^m-1)^2)
k=(s1-a*b*(b^m-1)/(b-1))/m
y=yuce(1:18,a,b,k)
```

## 6.5 趋势线的选择

趋势线的选择有以下几种方式。

1. 由散点图选择趋势线。
2. 由数据本身的取值规律选择趋势线。
3. 比较预测标准误差大小

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (56)$$

当有几种趋势线可供选择时，应选择  $S$  最小的趋势线。

## 7 平稳时间序列

这里的平稳是指宽平稳，其特性是序列的统计特性不随时间的平移而变化，即均值和协方差不随时间的平移而变化。

### 7.1 平稳随机序列（平稳时间序列）

#### 7.1.1 有关概念

定义 1 给定随机过程  $\{X_t, t \in T\}$ 。固定  $t$ ， $X_t$  是一个随机变量，设其均值为  $\mu_t$ ，当  $t$  变动时，此均值是  $t$  的函数，记为

$$\mu_t = E(X_t) \quad (57)$$

称为随机过程的均值函数。

固定  $t$ ，设  $X_t$  的方差为  $\sigma_t^2$ 。当  $t$  变动时，这个方差也是  $t$  的函数，记为

$$\sigma_t^2 = \text{Var}(X_t) = E[(X_t - \mu_t)^2] \quad (58)$$

称为随机过程的方差函数。方差函数的平方根  $\sigma_t$  称为随机过程的标准差函数，它表示

随机过程  $X_t$  对于均值函数  $\mu_t$  的偏离程度。

定义 2 对随机过程  $\{X_t, t \in T\}$ ，取定  $t, s \in T$ ，定义其自协方差函数为

$$\gamma_{t,s} = \text{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] \quad (59)$$

为刻画  $\{X_t, t \in T\}$  在时刻  $t$  与  $s$  之间的相关性，还可将  $\gamma_{t,s}$  标准化，即定义自相关函数

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}}\sqrt{\gamma_{s,s}}} = \frac{\gamma_{t,s}}{\sigma_t\sigma_s} \quad (60)$$

因此，自相关函数  $\rho_{t,s}$  是标准化自协方差函数。

定义 3 设随机序列  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  满足

- 1)  $E(X_t) = \mu = \text{常数}$ ;
- 2)  $\gamma_{t+k,t} = \gamma_k$  ( $k = 0, \pm 1, \pm 2, \dots$ ) 与  $t$  无关。

则称  $X_t$  为平稳随机序列（平稳时间序列），简称平稳序列。

定义 4 设平稳序列  $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \dots\}$  的自协方差函数  $\gamma_k$  是

$$\gamma_k = \sigma^2 \delta_{k,0} = \begin{cases} 0, & k \neq 0 \\ \sigma^2, & k = 0 \end{cases}$$

其中  $\delta_{k,0}$  当  $k = 0$  时为 1，当  $k \neq 0$  时为 0，则称该序列为平稳白噪声序列。

平稳白噪声序列的方差是常数  $\sigma^2$ ，因为  $\gamma_k = 0 (k \neq 0)$ ，则  $\varepsilon_t$  的任意两个不同点之间是不相关的。平稳白噪声序列是一种最基本的平稳序列。

#### 7.1.2 平稳性检验及自协方差函数、自相关函数的估计

##### 1. 平稳性 Daniel 检验

检验序列平稳性的方法很多，在此介绍其中一种，即 Daniel 检验。Daniel 检验方法建立在 Spearman 相关系数的基础上。

Spearman 相关系数是一种秩相关系数。先讲样本的秩的概念。设  $x_1, x_2, \dots, x_n$  是从一元总体抽取的容量为  $n$  的样本，其顺序统计量是  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ 。若  $x_i = x_{(k)}$ ，则称  $k$  是  $x_i$  在样本中的秩，记作  $R_i$ ，对每一个  $i = 1, 2, \dots, n$ ，称  $R_i$  是第  $i$  个秩统计量。

$R_1, R_2, \dots, R_n$  总称为秩统计量。例如，对样本数据

$$-0.8, \quad -3.1, \quad 1.1, \quad -5.2, \quad 4.2$$

顺序统计量是

$$-5.2, \quad -3.1, \quad -0.8, \quad 1.1, \quad 4.2$$

而秩统计量是

$$3, \quad 2, \quad 4, \quad 1, \quad 5$$

对于二维总体  $(X, Y)$  的样本观测数据  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，可得各分量  $X, Y$  的一元样本数据  $x_1, x_2, \dots, x_n$  与  $y_1, y_2, \dots, y_n$ 。设  $x_1, x_2, \dots, x_n$  的秩统计量是

$$R_1, R_2, \dots, R_n,$$

$y_1, y_2, \dots, y_n$  的秩统计量是

$$S_1, S_2, \dots, S_n,$$



当  $X, Y$  联系比较紧密时, 这两组秩统计量联系也是紧密的。Spearman 相关系数定义为这两组秩统计量的相关系数, 即 Spearman 相关系数是

$$q_{xy} = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (61)$$

其中  $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$ ,  $\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$ 。经过运算, 可证

$$q_{xy} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2 \quad (62)$$

其中  $d_i = R_i - S_i$ ,  $i = 1, 2, \dots, n$ 。

对于 Spearman 相关系数, 作假设检验

$$H_0: \rho_{XY} = 0, \quad H_1: \rho_{XY} \neq 0$$

可以证明, 当  $(X, Y)$  是二元正态总体, 且  $H_0$  成立时, 统计量

$$t = \frac{q_{xy} \sqrt{n-2}}{\sqrt{1-q_{xy}^2}}$$

服从自由度为  $n-2$  的  $t$  分布  $t(n-2)$ 。

对于给定的显著水平  $\alpha$ , 通过  $t$  分布表可查到统计量  $t$  的临界值  $t_{\alpha/2}(n-2)$ , 当  $|t| \leq t_{\alpha/2}(n-2)$  时, 接受  $H_0$ ; 当  $|t| > t_{\alpha/2}(n-2)$  时, 拒绝  $H_0$ 。

对于时间序列的样本  $X_1, X_2, \dots, X_n$ , 记  $X_t$  的秩为  $R_t = R(X_t)$ , 考虑变量对  $(t, R_t)$ ,  $t = 1, 2, \dots, n$  的 Spearman 相关系数  $q_s$ , 有

$$q_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (t - R_t)^2 \quad (63)$$

构造统计量

$$T = \frac{q_s \sqrt{n-2}}{\sqrt{1-q_s^2}}$$

作下列假设检验:

$$H_0: \text{序列 } X_t \text{ 平稳}, \quad H_1: \text{序列 } X_t \text{ 非平稳 (存在上升或下降趋势)}$$

Daniel 检验方法: 对于显著水平  $\alpha$ , 由时间序列  $X_t$  计算  $(t, R_t), t = 1, 2, \dots, n$  的 Spearman 秩相关系数  $q_s$ , 若  $|T| > t_{\alpha/2}(n-2)$ , 则拒绝  $H_0$ , 认为序列非平稳。且当  $q_s > 0$  时, 认为序列有上升趋势;  $q_s < 0$  时, 认为序列有下降趋势。又当  $|T| \leq t_{\alpha/2}(n-2)$  时, 接受  $H_0$ , 可以认为  $X_t$  是平稳序列。

例 11 美国某城镇 1964~1983 年间每年发生的偷窃率 (每万人平均发生的偷盗次数) 的数据见表 12 (时间  $t: 1 \sim 20$  表示从 1964 年~1983 年)。问偷盗率数据是否可看成是平稳的?

表 12 美国某城镇年偷盗率数据

时间	1	2	3	4	5	6	7	8	9	10
偷盗率	1.37	2.96	1.91	3.10	2.08	2.54	4.07	3.62	2.91	1.94
时间	11	12	13	14	15	16	17	18	19	20
偷盗率	3.96	4.19	2.71	3.42	3.02	3.54	2.66	4.11	4.25	3.76

解 对显著水平  $\alpha = 0.05$ ，算得  $q_s = 0.5624$ ，计算得  $t$  统计量的值为  $T = 2.8857$ ，上  $\alpha/2$  分位点的值  $t_{\alpha/2}(n-2) = 2.1009$ ，所以  $|T| > t_{\alpha/2}(n-1)$ ，故拒绝  $H_0$ ，认为序列是非平稳的；因  $q_s > 0$ ，故序列有上升趋势，即认为该城镇偷盗率有上升趋势。

计算的 Matlab 程序如下：

```
clc,clear
x0=[1.37 2.96 1.91 3.10 2.08 2.54 4.07 3.62 2.91 1.94
3.96 4.19 2.71 3.42 3.02 3.54 2.66 4.11 4.25 3.76]; %x0 为 2×10 矩阵
x0=x0'; x0=x0(:)'; n=length(x0);
[xs,ind]=sort(x0) %按从小到大的次序排列数据
Rt(ind)=1:n %计算秩
t=1:n;
Qs=1-6/(n*(n^2-1))*sum((t-Rt).^2) %计算 Qs 的值
t=Qs*sqrt(n-2)/sqrt(1-Qs^2) %计算 t 统计量的值
t_0=tinv(0.975,n-2) %计算上 alpha/2 分位点
```

## 2. 平稳序列自协方差函数与自相关函数的估计

对平稳序列  $X_t$  的样本  $X_1, X_2, \dots, X_n$ ，可以用样本均值估计随机序列的均值

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (64)$$

$\gamma_k$  通常有如下两种估计

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_{t+k} - \bar{X})(X_t - \bar{X}), \quad 0 \leq k \leq K \quad (65)$$

且  $\hat{\gamma}_{-k} = \hat{\gamma}_k$ 。或者

$$\hat{\gamma}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_{t+k} - \bar{X})(X_t - \bar{X}), \quad 0 \leq k \leq K \quad (66)$$

$\rho_k$  的估计为

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \quad 0 \leq k \leq K \quad (67)$$

## 7.2 ARMA 时间序列及其特性

下面介绍一种重要的平稳时间序列—ARMA 时间序列。ARMA 时间序列分为三种类型：

- 1) AR 模型，即自回归序列 (Auto Regressive Model)；
- 2) MA 序列，即滑动平均序列 (Moving Average Model)；
- 3) ARMA 序列，即自回归滑动平均序列 (Auto Regressive Moving Average Model)。

可证 ARMA 时间序列具有遍历性，因此可以通过它的一个样本估计自协方差函数及自相关函数。

### 7.2.1 ARMA 时间序列的定义

#### 1. AR( $p$ ) 序列

设  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  是零均值平稳序列，满足下列模型：

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t \quad (68)$$

其中  $\varepsilon_t$  是零均值、方差是  $\sigma_\varepsilon^2$  的平稳白噪声。则称  $X_t$  是阶数为  $p$  的自回归序列，简记为 AR( $p$ ) 序列，而

$$\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$$

称为自回归参数向量，其分量  $\varphi_j, j=1, 2, \dots, p$  称为自回归系数。

引进后移算子对描述式 (68) 比较方便。算子  $B$  定义如下：

$$BX_t \equiv X_{t-1}, \quad B^k X_t \equiv X_{t-k}$$

记算子多项式

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p,$$

则式 (68) 可以改写为

$$\varphi(B)X_t = \varepsilon_t$$

#### 2. MA( $q$ ) 序列

设  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  是零均值平稳序列，满足下列模型：

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (69)$$

其中  $\varepsilon_t$  是零均值、方差是  $\sigma_\varepsilon^2$  的平稳白噪声，则称  $X_t$  是阶数为  $q$  的滑动平均序列，简记为 MA( $q$ ) 序列，而

$$\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$$

称为滑动平均参数向量，其分量  $\theta_j, j=1, 2, \dots, q$  称为滑动平均系数。

在工程上，一个平稳白噪声发生器通过一个线性系统，如果其输出是白噪声的线性叠加，那么这一输出服从 MA 模型。

对于线性后移算子  $B$ ，有

$$B\varepsilon_t \equiv \varepsilon_{t-1}, \quad B^k \varepsilon_t \equiv \varepsilon_{t-k};$$

再引进算子多项式

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

则式 (69) 可以改写为

$$X_t = \theta(B)\varepsilon_t$$

#### 3. ARMA( $p, q$ ) 序列

设  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  是零均值平稳序列，满足下列模型：

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (70)$$

其中  $\varepsilon_t$  是零均值、方差是  $\sigma_\varepsilon^2$  的平稳白噪声，则称  $X_t$  是阶数为  $p, q$  的自回归滑动平均序列，简记为 ARMA( $p, q$ ) 序列。当  $q=0$  时，它是 AR( $p$ ) 序列；当  $p=0$  时，它为

MA( $q$ )序列。

应用算子多项式 $\varphi(B), \theta(B)$ ，式(70)可以写为

$$\varphi(B)X_t = \theta(B)\varepsilon_t$$

对于一般的平稳序列 $\{X_t, t=0, \pm 1, \pm 2, \dots\}$ ，设其均值 $E(X_t) = \mu$ ，满足下列模型：

$$(X_t - \mu) - \varphi_1(X_{t-1} - \mu) - \dots - \varphi_p(X_{t-p} - \mu) = \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q} \quad (71)$$

其中 $\varepsilon_t$ 是零均值、方差是 $\sigma_\varepsilon^2$ 的平稳白噪声，利用后移算子 $\varphi(B), \theta(B)$ ，式(71)可表为

$$\varphi(B)(X_t - \mu) = \theta(B)\varepsilon_t,$$

关于算子多项式 $\varphi(B), \theta(B)$ ，通常还要作下列假定：

- 1)  $\varphi(B)$ 和 $\theta(B)$ 无公共因子，又 $\varphi_p \neq 0$ ， $\theta_q \neq 0$ ；
- 2)  $\varphi(B)=0$ 的根全在单位圆外，这一条件称为模型的平稳性条件；
- 3)  $\theta(B)=0$ 的根全在单位圆外，这一条件称为模型的可逆性条件。

### 7.2.2 ARMA 序列的平稳性与可逆性

在理论与实际问题的探讨中，ARMA 序列的平稳性与可逆性是非常重要的。

#### 1. ARMA 序列的平稳性

设 $\{\varepsilon_t, t=0, \pm 1, \pm 2, \dots\}$ 是零均值平稳白噪声， $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ 。若 $\{G_k, k=1, 2, \dots\}$ 是一数列，满足

$$\sum_{k=0}^{\infty} |G_k| < +\infty, \quad G_0 = 1 \quad (72)$$

定义随机序列

$$X_t = \sum_{k=0}^{\infty} G_k \varepsilon_{t-k}, \quad (73)$$

则 $X_t$ 称为随机线性序列。在条件(72)下，可证式(73)中的 $X_t$ 是平稳序列，故它为平稳线性序列。 $X_t$ 有明显的物理意义，它表明， $X_t$ 可以通过以往白噪声 $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-j}, \dots$ ，加权叠加而生成，其中 $\varepsilon_{t-j}$ 所加的权是 $G_j$ 。可以设计一个线性系统实现式(73)的运算，其输入是平稳白噪声 $\varepsilon_t$ ，其输出是平稳线性序列 $X_t$ 。

$X_t$ 的平稳性证明如下。显然

$$E(X_t) = \sum_{k=0}^{\infty} G_k E(\varepsilon_{t-k}) = 0$$

$$\gamma_{t+k,t} = E(X_{t+k} X_t) = E\left[\left(\sum_{j=0}^{\infty} G_j \varepsilon_{t+k-j}\right)\left(\sum_{i=0}^{\infty} G_i \varepsilon_{t-i}\right)\right]$$

式中，只有当 $t+k-j=t-i$ ，即 $j=i+k$ 时

$$E(\varepsilon_{t+k-j} \varepsilon_{t-i}) = \sigma_\varepsilon^2$$

而当 $j \neq i+k$ 时，上式为0，故

$$\gamma_{t+k,t} = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} G_{i+k} G_i \quad (k \geq 0)$$

从而  $X_t$  是平稳的。严格地说, 在条件 (72) 下, 级数  $\sum_{i=0}^{\infty} G_{i+k} G_i$  是收敛的。

先来看个简单例子, 说明某些简单的 ARMA 序列在平稳性条件下可以表示成上述形式的平稳线性序列。

引进算子  $G(B)$ :

$$G(B) = \sum_{k=0}^{\infty} G_k B^k, \quad G_0 = 1,$$

则

$$X_t = \sum_{k=0}^{\infty} G_k \varepsilon_{t-k} = G(B) \varepsilon_t \quad (74)$$

若零均值平稳序列  $X_t$  能表示为式 (74) 的形式, 这种形式称为传递形式,  $\{G_k, k = 0, 1, 2, \dots\}$  称为 Green 函数。

例 12 求下列时间序列的 Green 函数与传递形式。

$$\text{AR}(1): X_t - \varphi_1 X_{t-1} = \varepsilon_t$$

$$\text{解 } \varphi(B) = 1 - \varphi_1 B, \quad \varphi(B) X_t = \varepsilon_t, \quad X_t = \varphi^{-1}(B) \varepsilon_t = \frac{1}{1 - \varphi_1 B} \varepsilon_t.$$

当  $|\varphi_1| < 1$  时, 有

$$\frac{1}{1 - \varphi_1 B} = 1 + \varphi_1 B + \varphi_1^2 B^2 + \dots = \sum_{k=0}^{\infty} \varphi_1^k B^k$$

从而

$$X_t = \left( \sum_{k=0}^{\infty} \varphi_1^k B^k \right) \varepsilon_t = \sum_{k=0}^{\infty} \varphi_1^k \varepsilon_{t-k} \quad (75)$$

即

$$G_k = \varphi_1^k \quad (k = 0, 1, 2, \dots)$$

注意到  $\varphi(B) = 1 - \varphi_1 B = 0$  的根是  $\frac{1}{\varphi_1}$ 。当  $|\varphi_1| < 1$  时,  $\left| \frac{1}{\varphi_1} \right| > 1$ , 即  $\varphi(B)$  的根在单位圆外, 满足平稳性条件。说明在满足平稳性条件时, 序列存在传递形式 (75)。

对于 ARMA( $p, q$ ) 序列

$$\varphi(B) X_t = \theta(B) \varepsilon_t$$

当  $\varphi(B)$  满足平稳性条件 ( $\varphi(B)$  的根全在单位圆外) 时, 可证  $G_k$  是负指数下降的, 即存在  $g_1, g_2 > 0$ , 使得

$$|G_k| < g_2 e^{-g_1 k} \quad (k \geq 0) \quad (76)$$

当  $G_k$  满足式 (76) 时, 必满足式 (72), 从而线性序列式 (73) 是平稳的。

2. ARMA 序列的可逆性

设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是 ARMA( $p, q$ ) 序列

$$\varphi(B) X_t = \theta(B) \varepsilon_t$$

若  $X_t$  能够表示为

$$X_t - I_1 X_{t-1} - I_2 X_{t-2} - \cdots = X_t - \sum_{k=1}^{\infty} I_k X_{t-k} = I(B)X_t = \varepsilon_t, \quad (77)$$

其中

$$I(B) = 1 - \sum_{k=1}^{\infty} I_k B^k = -\sum_{k=0}^{\infty} I_k B^k, \quad (I_0 = -1)$$

则称 ARMA( $p, q$ ) 序列  $X_t$  具有逆转形式, 而  $\{I_k, k = 0, 1, 2, \dots\}$  称为逆函数。

当  $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$  的根全在单位圆外, 即  $\theta(B)$  满足可逆条件时, 可证存在  $h_1, h_2 > 0$ , 使得

$$|I_k| < h_2 e^{-h_1 k},$$

即  $I_k$  是被负指数控制的, 此时, 逆转形式 (77) 必然存在。

例 13 求下列随机序列的逆函数与逆转形式。

$$\text{MA}(1): X_t = (1 - \theta_1 B)\varepsilon_t$$

$$\text{解 } \varepsilon_t = \frac{1}{1 - \theta_1 B} X_t = \sum_{k=0}^{\infty} \theta_1^k X_{t-k},$$

从而

$$I_k = -\theta_1^k, \quad k = 0, 1, 2, \dots$$

### 7.2.3 ARMA 序列的相关特性

我们要阐述 AR, MA, ARMA 序列的相关特性, 这些特性对于序列的建模起着重要的作用。相关特性包括相关函数, 以及下面将要阐述的偏相关函数。

#### 1. MA 序列的自相关函数

设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是 MA( $q$ ) 序列

$$X_t = \theta(B)\varepsilon_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q},$$

其中  $\varepsilon_t$  是零均值, 方差是  $\sigma_\varepsilon^2$  的平稳白噪声, 则  $X_t$  是零均值平稳序列, 且其自协方差函数为

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2), & k = 0 \\ \sigma_\varepsilon^2 (-\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q), & 1 \leq k \leq q \\ 0, & k > q \end{cases} \quad (78)$$

自相关函数为

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & k = 0 \\ \frac{-\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, & 1 \leq k \leq q \\ 0, & k > q \end{cases} \quad (79)$$

注意到, 当  $k > q$  时

$$\gamma_k = 0, \quad \rho_k = 0$$

## 2. AR( $p$ )序列的自相关函数

$$E(X_{t-k}\varepsilon_t)=0, \ k \geq 1 \quad (80)$$
$$X_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j}, \quad X_{t-k} = \sum_{j=0}^{\infty} G_j \varepsilon_{t-k-j}$$
$$E(X_{t-k}\varepsilon_t) = E(\sum_{j=0}^{\infty} G_j \varepsilon_{t-k-j} \varepsilon_t) = \sum_{j=0}^{\infty} G_j E(\varepsilon_{t-k-j} \varepsilon_t) = 0, \quad (k \geq 1)$$
$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t,$$
$$E(X_t X_{t-k}) = \varphi_1 E(X_{t-1} X_{t-k}) + \varphi_2 E(X_{t-2} X_{t-k}) + \cdots + \varphi_p E(X_{t-p} X_{t-k}) + E(\varepsilon_t X_{t-k})$$
$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \cdots + \varphi_p \gamma_{k-p}, \quad k > 0,$$
$$\varphi(B)\gamma_k = 0, \quad k > 0.$$
$$\varphi(B)\rho_k=0, \, k>0 \tag{81}$$
[illegible]
$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix} \quad (82)$$

还可证明, 对于  $\text{AR}(p)$  序列

$$\sigma_\varepsilon^2 = \gamma_0 - \sum_{j=1}^p \varphi_j \gamma_j \quad (83)$$

事实上

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = E[(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p})^2] = \gamma_0 - 2 \sum_{j=1}^p \varphi_j \gamma_j + \sum_{j=1}^p \sum_{i=1}^p \varphi_j \varphi_i \gamma_{j-i}$$

由于  $\gamma_j = \sum_{i=1}^p \varphi_i \gamma_{j-i}$ , 从而

$$\sigma_\varepsilon^2 = \gamma_0 - 2 \sum_{j=1}^p \varphi_j \gamma_j + \sum_{j=1}^p \varphi_j \left( \sum_{i=1}^p \varphi_i \gamma_{j-i} \right) = \gamma_0 - 2 \sum_{j=1}^p \varphi_j \gamma_j + \sum_{j=1}^p \varphi_j \gamma_j = \gamma_0 - \sum_{j=1}^p \varphi_j \gamma_j$$

式(83)获证。

可以证明,  $\text{AR}(p)$  序列的自协方差函数  $\gamma_j$  是被负指数控制的

$$|\gamma_k| \leq c_2 e^{-c_1 k}, \quad (c_1, c_2 > 0)$$

对于自相关函数  $\rho_k$  也有同样性质。这说明  $\gamma_k$  或  $\rho_k$  不能在有限步之后截尾, 而是按负指数率衰减。这种性质称为拖尾性。

### 3. $\text{ARMA}(p, q)$ 序列的自相关函数

设  $\{X_t, t=0, \pm 1, \pm 2, \cdots\}$  是  $\text{ARMA}(p, q)$  序列

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

则其自协方差函数满足

$$\varphi(B)\gamma_k = \begin{cases} 0, & k > q \\ -\theta_q \sigma_\varepsilon^2, & k = q \end{cases} \quad (84)$$

由(84)式, 取  $k = q+1, q+2, \cdots, q+p$ , 得到方程组

$$\begin{cases} \gamma_{q+1} = \varphi_1 \gamma_q + \varphi_2 \gamma_{q-1} + \cdots + \varphi_p \gamma_{q+1-p} \\ \gamma_{q+2} = \varphi_1 \gamma_{q+1} + \varphi_2 \gamma_q + \cdots + \varphi_p \gamma_{q+2-p} \\ \cdots \cdots \cdots \\ \gamma_{q+p} = \varphi_1 \gamma_{q+p-1} + \varphi_2 \gamma_{q+p-2} + \cdots + \varphi_p \gamma_q \end{cases}$$

将其写成矩阵形式有

$$\begin{bmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} = \begin{bmatrix} \gamma_q & \gamma_{q-1} & \cdots & \gamma_{q+1-p} \\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q+2-p} \\ \vdots & \vdots & & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix} \quad (85)$$

解此方程组, 可得自回归系数向量  $(\varphi_1, \varphi_2, \cdots, \varphi_p)^T$ 。

其次, 令  $Y_t = \varphi(B)X_t = X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}$ , 则  $\{Y_t, t=0, \pm 1, \pm 2, \cdots\}$  是参数为  $(\theta_1, \theta_2, \cdots, \theta_q)^T$  的  $\text{MA}(q)$  序列。 $\{Y_t, t=0, \pm 1, \pm 2, \cdots\}$  的自协方差函数



$$\gamma_k^{(Y)} = E(Y_{t+k} Y_t) = E\left[-\sum_{i=0}^p \varphi_i X_{t+k-i}\right] \left[-\sum_{j=0}^p \varphi_j X_{t-j}\right] = \sum_{i,j=0}^p \varphi_i \varphi_j \gamma_{k+j-i}, \quad (\varphi_0 = -1)$$

将  $\gamma_k^{(Y)}$  代入 (78) 式, 即

$$\sum_{i,j=0}^p \varphi_i \varphi_j \gamma_{k+j-i} = \begin{cases} \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2), & k = 0 \\ \sigma_\varepsilon^2 (-\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q), & 1 \leq k \leq q \end{cases} \quad (86)$$

解此方程组, 得滑动平均系数向量  $(\theta_1, \theta_2, \dots, \theta_q)^T$  及  $\sigma_\varepsilon^2$ 。

可以证明, ARMA( $p, q$ ) 序列的自协方差函数 (自相关函数) 为拖尾的。

对于重要的 ARMA(1,1) 序列

$$X_t - \varphi_1 X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

它平稳且可逆的条件是

$$|\varphi_1| \leq 1, \quad |\theta_1| \leq 1.$$

可求得它的自协方差函数为

$$\gamma_k = \begin{cases} \frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{1 - \varphi_1^2} \sigma_\varepsilon^2, & k = 0 \\ \frac{(1 - \varphi_1\theta_1)(\varphi_1 - \theta_1)}{1 - \varphi_1^2} \sigma_\varepsilon^2 \varphi_1^{k-1}, & k \geq 1 \end{cases} \quad (87)$$

自相关函数为

$$\rho_k = \frac{(1 - \varphi_1\theta_1)(\varphi_1 - \theta_1)}{1 + \theta_1^2 - 2\varphi_1\theta_1} \varphi_1^{k-1}, \quad k \geq 1 \quad (88)$$

例 14 随机模拟下列序列, 样本容量 10000。

$$X_t - 0.8X_{t-1} = \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \text{其中 } \varepsilon_t \sim N(0,1)$$

利用模型数据研究上述序列自相关特性。

解 利用如下的 Matlab 程序:

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=0.8*x(j-1)+elps(j)-0.4*elps(j-1); %产生样本点
end
y=(x-mean(x)); %把数据中心化
gama0=var(x); %求样本方差
for j=1:10
    gama(j)=y(j+1:end)*y(1:end-j)'/10000; %求自协方差函数
end
rho=gama/gama0 %求自相关函数
rho2=autocorr(x) %直接利用 Matlab 工具箱求自相关函数
由样本算得样本自相关函数如下 (其中的一次运行结果):
```

$$\hat{\rho}_1 = 0.5284, \quad \hat{\rho}_2 = 0.4287, \quad \hat{\rho}_3 = 0.3442, \quad \hat{\rho}_4 = 0.2832$$

由式 (88), 理论自相关函数是

$$\rho_1 = 0.5231, \quad \rho_2 = 0.4185, \quad \rho_3 = 0.3348, \quad \rho_4 = 0.2678$$

计算理论自相关函数的 Matlab 程序为

```
clc,clear
rho=@(k,p1,p2) (1-p1*p2)*(p1-p2)/(1+p2^2-2*p1*p2)*p1.^(k-1);
a=rho([1:10],0.8,0.4)
```

由此可见, 模拟计算与理论值基本一致。

#### 4. 偏相关函数及 $AR(p)$ 序列偏相关函数的截尾性

设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是零均值平稳序列, 从时间序列预报的角度引出偏相关函数的定义。如果已知  $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$  的值, 要求对  $X_t$  作出预报。此时, 可以考虑由  $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$  对  $X_t$  的线性最小均方估计, 即选择系数  $\varphi_{k,1}, \varphi_{k,2}, \dots, \varphi_{k,k}$ , 使得

$$\delta = E[(X_t - \sum_{j=1}^k \varphi_{k,j} X_{t-j})^2] = \min$$

将  $\delta$  展开, 得

$$\delta = \gamma_0 - 2 \sum_{j=1}^k \varphi_{k,j} \gamma_j + \sum_{j=1}^k \sum_{i=1}^k \varphi_{k,j} \varphi_{k,i} \gamma_{j-i}$$

令  $\frac{\partial \delta}{\partial \varphi_{k,j}} = 0, \quad j = 1, 2, \dots, k$ , 得

$$-\gamma_j + \sum_{i=1}^k \varphi_{k,i} \gamma_{j-i} = 0, \quad j = 1, 2, \dots, k$$

两端同除  $\gamma_0$  并写成矩阵形式, 可知  $\varphi_{k,j}$  应满足下列线性方程组

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k,1} \\ \varphi_{k,2} \\ \vdots \\ \varphi_{k,k} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (89)$$

式 (89) 与式 (82) 形式上相同, 也称为 Yule-Walker 方程。

我们称  $\{\varphi_{k,k}, k = 1, 2, \dots\}$  为  $X_t$  的偏相关函数。

式 (89) 的系数矩阵是 Toeplitz 矩阵。这一线性方程组的解  $\varphi_{k,1}, \dots, \varphi_{k,k}$  有下列递推公式

$$\begin{cases} \varphi_{1,1} = \rho_1 \\ \varphi_{k+1,k+1} = (\rho_{k+1} - \sum_{j=1}^k \rho_{k+1-j} \varphi_{k,j}) (1 - \sum_{j=1}^k \rho_j \varphi_{k,j})^{-1} \\ \varphi_{k+1,j} = \varphi_{k,j} - \varphi_{k+1,k+1} \varphi_{k,k+1-j}, \quad j = 1, 2, \dots, k \end{cases} \quad (90)$$

$AR(p)$  序列的偏相关函数  $\{\varphi_{k,k}, k = 1, 2, \dots\}$  是  $p$  步截尾的, 且有

$$\begin{cases} \varphi_{k,k} = \varphi_k, & 1 \leq k \leq p \\ \varphi_{k,k} = 0, & k > p \end{cases}$$

对于  $\text{MA}(q)$  和  $\text{ARMA}(p, q)$  序列，其偏相关函数  $\{\varphi_{k,k}, k = 1, 2, \dots\}$  具有拖尾性。

例 15 讨论例 14 所述时间序列的偏相关特性。

解 在式 (90) 中，以  $\hat{\rho}_k$  代替  $\rho_k$ ， $\hat{\varphi}_{k,j}$  代替  $\varphi_{k,j}$ ，计算得到的  $\{\hat{\varphi}_{k,k}, k = 1, 2, \dots\}$  称为样本偏相关函数。计算得到样本偏相关函数（其中的一次运行结果）是

$$\hat{\varphi}_{1,1} = 0.5351, \quad \hat{\varphi}_{2,2} = 0.2075, \quad \hat{\varphi}_{3,3} = 0.0683, \quad \hat{\varphi}_{4,4} = 0.0256$$

由此可见，偏相关函数是拖尾的。

计算的 Matlab 程序如下：

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=0.8*x(j-1)+elps(j)-0.4*elps(j-1); %产生样本点
end
y=(x-mean(x)); %把数据中心化
gama0=var(x); %求样本方差
L=10;
for j=1:L
    gama(j)=y(j+1:end)*y(1:end-j)'/10000; %求自协方差
end
rho=gama/gama0 %求自相关系数
f(1,1)=rho(1);
for k=2:L
    s1=rho(k);s2=1; %式 (90) 中第二式计算的初始值
    for j=1:k-1
        s1=s1-rho(k-j)*f(k-1,j);
        s2=s2-rho(j)*f(k-1,j);
    end
    f(k,k)=s1/s2; %式 (90) 中第二式的计算值
    for j=1:k-1
        f(k,j)=f(k-1,j)-f(k,k)*f(k-1,k-j); %式 (90) 中第三式的计算值
    end
end
end
f
pcorr=diag(f)' %提取偏相关函数
pcorr2=parcorr(x) %直接利用Matlab工具箱计算偏相关函数
```

## ~~§8 ARMA 时间序列的建模与预报~~

在实际问题中，若考察的时间序列是 ARMA 序列

$$\varphi(B)X_t = \theta(B)\varepsilon_t,$$

首先要进行模型的识别与定阶，即要判断是  $\text{AR}(p)$ ， $\text{MA}(q)$ ， $\text{ARMA}(p, q)$  模型的类别，并估计阶数  $p, q$ 。其实，这都归结到模型的定阶问题。当模型定阶后，就要对

模型参数  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$  及  $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$  进行估计。定阶与参数估计完成后，还要对模型进行检验，即要检验  $\varepsilon_t$  是否为平稳白噪声。若检验获得通过，则 ARMA 时间序列的建模完成。作为时间序列建模之后的一个重要应用，我们还要讨论 ARMA 时间序列的预报。

### 8.1 ARMA 序列参数的矩估计

#### 1. AR( $p$ ) 序列参数的矩估计

设  $X_t$  是 AR( $p$ ) 序列

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t$$

其中  $p$  已知， $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$  未知，需要估计。在式 (82) 中，以  $\hat{\rho}_j$  代替  $\rho_j$ ， $j = 1, 2, \dots, p$ ，得  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$  的矩估计  $\hat{\varphi} = (\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p)^T$  所满足的方程组

$$\begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} \quad (91)$$

在式 (83) 中，以  $\hat{\gamma}_j$  代替  $\gamma_j$ ， $\hat{\varphi}_j$  代替  $\varphi_j$ ，得  $\sigma_\varepsilon^2$  的矩估计为

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0 - \sum_{j=1}^p \hat{\varphi}_j \hat{\gamma}_j \quad (92)$$

$\hat{\varphi}$  还可以由递推计算得到，式 (90) 中，以  $\hat{\rho}_j$  代替  $\rho_j$ ，以  $\hat{\varphi}_{k,j}$  代替  $\varphi_{k,j}$ ，得

$$\begin{cases} \hat{\varphi}_{1,1} = \hat{\rho}_1 \\ \hat{\varphi}_{k+1,k+1} = (\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\rho}_{k+1-j} \hat{\varphi}_{k,j})(1 - \sum_{j=1}^k \hat{\rho}_j \hat{\varphi}_{k,j})^{-1} \\ \hat{\varphi}_{k+1,j} = \hat{\varphi}_{k,j} - \hat{\varphi}_{k+1,k+1} \hat{\varphi}_{k,k+1-j}, \quad j = 1, 2, \dots, k \end{cases}$$

当递推计算到第  $p$  步，得

$$\hat{\varphi}_j = \hat{\varphi}_{p,j}, \quad j = 1, 2, \dots, p$$

$\hat{\varphi}$  称为  $\varphi$  的 Yule-Walker 估计。当样本容量  $n$  较大时，Yule-Walker 估计  $\hat{\varphi}$  的效果较好。

#### 2. MA( $q$ ) 序列参数的矩估计

设  $X_t$  为 MA( $q$ ) 序列，其中  $q$  已知， $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$  未知。在式 (78) 中，以  $\hat{\gamma}_k$  代替  $\gamma_k$ ，可得  $(\theta_1, \theta_2, \dots, \theta_q)^T$  和  $\sigma_\varepsilon^2$  的矩估计  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)^T$ 、 $\hat{\sigma}_\varepsilon^2$  所满足的方程组为

$$\begin{cases} \hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0 (1 + \hat{\theta}_1^2 + \dots + \hat{\theta}_q^2)^{-1} \\ \hat{\theta}_k = \hat{\theta}_1 \hat{\theta}_{k+1} + \dots + \hat{\theta}_{q-k} \hat{\theta}_q - \frac{\hat{\gamma}_k}{\hat{\sigma}_\varepsilon^2}, \quad k = 1, 2, \dots, q \end{cases} \quad (93)$$

下面介绍 MA( $q$ ) 序列参数估计的线性迭代算法。给定  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)^T$  与  $\hat{\sigma}_\varepsilon^2$  的

一组初值 (例如  $\hat{\theta} = 0, \hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0$ ), 代入式 (93) 右边, 由此得到等式左边的值  $\hat{\sigma}_\varepsilon^2(1)$  与  $\hat{\theta}(1) = (\hat{\theta}_1(1), \hat{\theta}_2(1), \dots, \hat{\theta}_q(1))^T$ , 称为  $\hat{\sigma}_\varepsilon^2$  与  $\hat{\theta}$  的一步迭代值, 再将它们代入式 (93) 的右边得到  $\hat{\sigma}_\varepsilon^2(2)$  和  $\hat{\theta}(2)$ 。以此方式迭代下去, 直到某步的  $\hat{\sigma}_\varepsilon^2(m)$  和  $\hat{\theta}(m)$  与前一步的  $\hat{\sigma}_\varepsilon^2(m-1)$  和  $\hat{\theta}(m-1)$  相差不大时停止迭代, 并取

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(m), \quad \hat{\theta} = \hat{\theta}(m)$$

需注意, 算得的  $\hat{\theta}$  必须验证是否满足可逆性条件。如不满足, 可改动初值反复试算, 直到满足要求为止。

### 3. ARMA(p, q) 序列参数的矩估计

设  $X_t$  为 ARMA(p, q) 序列, 为求该模型参数的矩估计, 采用先求  $(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)^T$ , 再求  $\hat{\sigma}_\varepsilon^2$  和  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)^T$  的方法, 分以下三步进行。

(1) 首先求回归参数  $(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)^T$  的估计

由式 (91) 得, 自回归系数  $(\phi_1, \phi_2, \dots, \phi_p)^T$  的矩估计  $(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)^T$  为

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}$$

(2) 令  $Y_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p}$ , 则  $Y_t$  的自协方差函数

$$\gamma_k^{(Y)} = E(Y_{t+k} Y_t) = \sum_{i,j=0}^p \hat{\phi}_i \hat{\phi}_j \gamma_{k+j-i}, \quad (\phi_0 = -1)$$

以  $\hat{\gamma}_k$  代替  $\gamma_k$ , 得  $\gamma_k^{(Y)}$  的估计

$$\hat{\gamma}_k^{(Y)} = \sum_{i,j=0}^p \hat{\phi}_i \hat{\phi}_j \hat{\gamma}_{k+j-i}, \quad \hat{\phi}_0 = -1$$

(3) 把  $Y_t$  近似看作 MA(q) 序列, 则

$$Y_t \approx \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

再根据前面介绍的 MA(q) 序列的参数估计方法求出  $\sigma_\varepsilon^2$  和  $(\theta_1, \theta_2, \dots, \theta_q)^T$  的矩估计  $\hat{\sigma}_\varepsilon^2$  和  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)^T$ 。

计算实践表明, 当  $n$  充分大时, AR(p) 序列的矩估计效果较好, 但 MA(q), ARMA(p, q) 序列的矩估计则不太理想, ARMA 序列参数的矩估计可以作为初估计。

### 8.2 ARMA(p, q) 序列参数的逆函数估计法

因为 MA(q) 序列和 ARMA(p, q) 序列参数的矩估计涉及到非线性运算, 所以只能采取迭代算法得到近似解。而逆函数估计法的优点是它仅仅涉及到线性方程组的求解, 因而计算简便, 快速, 易于实现, 便于对高阶 ARMA 序列参数进行初估计。

用逆函数法求模型参数估计的步骤如下:

第一步, 令  $p^* = \max(p, q) + q$ , 对 AR(p\*) 模型的自回归系数用 Yule-Walker

方程组求解或采用偏相关递推公式求出，将它们近似地当作相应  $\text{ARMA}(p, q)$  模型的前  $p^*$  个逆函数估计值  $\hat{I}_1, \hat{I}_2, \dots, \hat{I}_{p^*}$ 。

第二步，解下列方程组

$$\begin{cases} (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \hat{I}_j = 0 \\ j = \max(p, q) + 1, \max(p, q) + 2, \dots, p^* \end{cases}$$

得到滑动平均系数向量  $(\theta_1, \theta_2, \dots, \theta_q)^T$  的初估计  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)^T$ 。

第三步，将第一、二步得到的  $\hat{I}_1, \hat{I}_2, \dots, \hat{I}_{p^*}$  和  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$  代入下式

$$\varphi_j = I_j - \theta_1 I_{j-1} - \theta_2 I_{j-2} - \dots - \theta_q I_{j-q} + \theta_j$$

其中当  $j > q$  时， $\theta_j = 0$ 。得到  $\varphi_1, \varphi_2, \dots, \varphi_p$  的初估计  $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p$ 。

值得注意的是，由上述三步所得的  $\text{ARMA}(p, q)$  模型参数的初估计，并不能保证它们一定满足可逆性条件。不过经过适当调整后，既能保证模型的可逆性又能保持动态统计性不变，只需不断地将  $\hat{\theta}(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_q B^q = 0$  在单位圆内的根，逐步用相应根的倒数代替。那么换根后所产生的  $\theta^*(B)$  既满足可逆性又能保持模型的统计特性不变。

参数逆函数估计法在计算机上容易实现，这种方法无论对低阶模型还是高阶模型都是很有用的。由该方法得到的模型参数的估计，可以作为最小二乘估计等精估计进行迭代计算的初值。

### 8.3 $\text{ARMA}(p, q)$ 序列参数的最小二乘估计

#### 8.3.1 $\text{AR}(p)$ 序列参数的最小二乘估计

设  $X_t$  是  $\text{AR}(p)$  序列，则

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t$$

当获得观测值  $X_1, X_2, \dots, X_n$  后，由上式可得如下  $n - p$  个方程

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad t = p+1, p+2, \dots, n$$

对待估计参数  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$ ，定义函数

$$S(\varphi) = \sum_{t=p+1}^n (X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p})^2$$

称使函数  $S(\varphi)$  达到极小的  $\hat{\varphi}_L$  为参数  $\varphi$  的最小二乘估计，记作  $\hat{\varphi}_L = (\hat{\varphi}_1^L, \dots, \hat{\varphi}_p^L)^T$ 。令

$$\frac{\partial S(\varphi)}{\partial \varphi} = 0, \quad \text{得 } \varphi \text{ 满足正规方程}$$

$$(A^T A) \varphi = A^T \alpha$$

其中

$$A = \begin{bmatrix} X_p & X_{p-1} & \cdots & X_1 \\ X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & & \vdots \\ X_{n-1} & X_{n-2} & \cdots & X_{n-p} \end{bmatrix}, \quad \alpha = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{bmatrix}$$

令  $\hat{\gamma}_{ij}^L = \frac{1}{n} \sum_{i=p+1}^n X_{t-i} X_{t-j}$ ,  $i=0,1,\dots,p$ ,  $j=1,2,\dots,p$ , 则

$$\hat{\phi}_L = \begin{bmatrix} \hat{\phi}_1^L \\ \hat{\phi}_2^L \\ \vdots \\ \hat{\phi}_p^L \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_{11}^L & \hat{\gamma}_{12}^L & \cdots & \hat{\gamma}_{1p}^L \\ \hat{\gamma}_{21}^L & \hat{\gamma}_{22}^L & \cdots & \hat{\gamma}_{2p}^L \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{p1}^L & \hat{\gamma}_{p2}^L & \cdots & \hat{\gamma}_{pp}^L \end{bmatrix}^{-1} \begin{bmatrix} \hat{\gamma}_{01}^L \\ \hat{\gamma}_{02}^L \\ \vdots \\ \hat{\gamma}_{0p}^L \end{bmatrix}$$

若用  $\hat{\phi}_L$  代替  $\phi$ , 则得  $\varepsilon_t$  的估计

$$\hat{\varepsilon}_t^L = X_t - \hat{\phi}_1^L X_{t-1} - \cdots - \hat{\phi}_p^L X_{t-p}, \quad t = p+1, p+2, \dots, n$$

从而方差  $\sigma_\varepsilon^2$  的最小二乘估计为

$$\hat{\sigma}_L^2 = \frac{1}{n-p} \sum_{t=p+1}^n (X_t - \hat{\phi}_1^L X_{t-1} - \cdots - \hat{\phi}_p^L X_{t-p})^2 = \frac{1}{n-p} S(\hat{\phi}_L)$$

当  $n$  充分大时

$$\hat{\gamma}_{ij}^L = \hat{\gamma}_{ji}^L \approx \hat{\gamma}_{|i-j|}$$

此时, 最小二乘估计  $\hat{\phi}_L$  与矩估计  $\hat{\phi}$  是非常接近的。

### 8.3.2 MA( $q$ ) 与 ARMA( $p, q$ ) 序列参数的最小二乘估计

ARMA( $p, q$ ) 与 MA( $q$ ) 这两种序列参数的最小二乘估计方法基本相同, 以下只叙述 ARMA( $p, q$ ) 序列参数的估计方法。设  $X_t$  是 ARMA( $p, q$ ) 序列,  $X_t$  的一个观测样本是  $X_1, X_2, \dots, X_n$ 。如 7.2.3 小节讨论偏相关函数时, 由  $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$  预报  $X_t$ , 现取  $k = t-1$ , 即由  $\{X_{t-1}, X_{t-2}, \dots, X_1\}$  的线性组合预报  $X_t$ ,  $t = 2, 3, \dots, n$ , 估计量是

$$\sum_{j=1}^{t-1} \varphi_{t-1,j} X_{t-j}, \quad j = 2, 3, \dots, n$$

由式 (89), 知  $\varphi_{t-1,1}, \varphi_{t-1,2}, \varphi_{t-1,t-1}$  满足下列方程组

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{t-2} \\ \rho_1 & 1 & \cdots & \rho_{t-3} \\ \vdots & \vdots & & \vdots \\ \rho_{t-2} & \rho_{t-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{t-1,1} \\ \varphi_{t-1,2} \\ \vdots \\ \varphi_{t-1,t-1} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{t-1} \end{bmatrix}, \quad t = 2, 3, \dots, n$$

因为在理论上,  $\rho_j$  是  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$ ,  $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$  的函数, 所以  $\varphi_{t-1,j}$  是  $\varphi, \theta$  的函数, 从而预报的残差平方和也是  $\varphi, \theta$  的函数

$$S(\varphi, \theta) = \sum_{i=2}^n (X_i - \sum_{j=1}^{t-1} \varphi_{t-1,j} X_{i-j})^2 \quad (94)$$

在  $X_t$  的平稳可逆域中寻求  $\hat{\phi}_l, \hat{\theta}_l$ , 使得

$$S(\hat{\phi}_l, \hat{\theta}_l) = \min \quad (95)$$

$\hat{\varphi}_L, \hat{\theta}_L$  称为  $\varphi, \theta$  的 (无条件) 最小二乘估计 (ULS 估计)。

## 8.4 ARMA 序列参数的精估计

ARMA 序列参数的精估计有许多种方法，在此重点介绍 2 种方法：条件最小二乘估计（CLS）和最大似然估计（ML）。

### 1. 条件最小二乘估计 (CLS 估计)

设  $X_t$  是  $\text{ARMA}(p, q)$  序列

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (96)$$

其中  $\varepsilon_t$  是零均值, 方差为  $\sigma_\varepsilon^2$  的平稳白噪声。设  $X_t$  具有逆转形式

$$X_t - \sum_{j=1}^{\infty} I_j X_{t-j} = \varepsilon_t \quad (97)$$

式 (96)、式 (97) 用算子形式可写为

$$(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p) X_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \varepsilon_t \quad (98)$$

$$\varepsilon_t = (1 - I_1 B - I_2 B^2 - \dots) X_t \quad (99)$$

将式 (99) 代入式 (98), 得到算子恒等式

$$1 - \varphi_1 B - \cdots - \varphi_n B^n = (1 - \theta_1 B - \cdots - \theta_q B^q)(1 - I_1 B - I_2 B^2 - \cdots)$$

比较等式两边  $B$  的相同幂次, 得

[illegible]

其中, 当  $j > q$  时, 令  $\theta_j = 0$ ;  $i > p$  时, 令  $\varphi_i = 0$ 。将式 (100) 改写为

[illegible]

给定  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$ ,  $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ , 由式(101)可以递推算得逆函数  $\{I_j, j = 1, 2, \dots\}$ , 从而得到逆转形式



$$I(B)X_t = X_t - \sum_{j=1}^{\infty} I_j X_{t-j} = \varepsilon_t$$

需要注意， $\varphi, \theta$  的取值必须使得 ARMA( $p, q$ ) 序列  $X_t$  满足平稳可逆条件；换句话说， $\varphi, \theta$  必须在  $X_t$  的平稳可逆域中取值。

条件最小二乘估计方法使得下列残差平方和

$$\sum_{j=1}^n \hat{\varepsilon}_t^2 = \sum_{j=1}^n (X_t - \sum_{j=1}^{\infty} I_j X_{t-j})^2 = \min$$

其中约定： $X_t = 0$ ，当  $t \leq 0$ （这是由于观测样本是  $X_1, X_2, \dots, X_n$ ）。而  $I_j$  由式 (101) 递推算得，因为  $I_j$  是  $\varphi, \theta$  的函数，故残差平方和也是  $\varphi, \theta$  的函数，即

$$S(\varphi, \theta) = \sum_{j=1}^n (X_t - \sum_{j=1}^{\infty} I_j X_{t-j})^2 \quad (102)$$

注意到上述平方和是  $\varphi, \theta$  的非线性函数，最小化式 (102) 需用非线性最小二乘法。即在  $X_t$  的平稳可逆域中寻求  $\hat{\varphi}_L^*, \hat{\theta}_L^*$ ，使得

$$S(\hat{\varphi}_L^*, \hat{\theta}_L^*) = \min \quad (103)$$

$\hat{\varphi}_L^*, \hat{\theta}_L^*$  称为  $\varphi, \theta$  的条件最小二乘估计 (CLS 估计)。它之所以称为条件最小二乘估计，是因为它假定过去未观测值等于零。又  $\sigma_\varepsilon^2$  的估计是

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-r} S(\hat{\varphi}_L^*, \hat{\theta}_L^*)$$

其中  $r = p + q$ 。

## 2. 最大似然估计 (ML 估计)

假设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是零均值正态 ARMA( $p, q$ ) 序列，已取得样本  $X = (X_1, X_2, \dots, X_n)^T$ ，它的概率密度是

$$f(x | \beta, \sigma_\varepsilon^2) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} x^T \Sigma^{-1} x)$$

其中  $\beta = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)^T$ ， $x = (x_1, x_2, \dots, x_n)^T$ ， $\Sigma = E(XX^T)$ 。令  $M = \sigma_\varepsilon^2 \Sigma^{-1}$ ，则有

$$f(x | \beta, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} |M|^{\frac{1}{2}} \exp(-\frac{x^T M x}{2\sigma_\varepsilon^2})$$

取对数似然函数

$$L(\beta, \sigma_\varepsilon^2 | x) = \ln f(x | \beta, \sigma_\varepsilon^2) = \frac{1}{2} \ln |M| - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_\varepsilon^2 - \frac{x^T M x}{2\sigma_\varepsilon^2}$$

这里， $x = (x_1, x_2, \dots, x_n)^T$  理解为样本观测值。

我们下面说明  $x^T M x$  仅与  $\beta$  有关，而与  $\sigma_\varepsilon^2$  无关。事实上  $\Sigma$  的第  $i$  行第  $j$  列的元素是

$$r_{ij} = E(X_i X_j) = r_{|i-j|}$$

设  $X_t$  的传递形式是  $X_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j}$ ，则

$$r_{|i-j|} = (\sum_{k=0}^{\infty} G_k G_{k+|i-j|}) \sigma_{\varepsilon}^2$$

由此即知  $M = \sigma_{\varepsilon}^2 \Sigma^{-1}$  与  $\sigma_{\varepsilon}^2$  无关，从而  $x^T M x$  仅与  $\beta$  有关，与  $\sigma_{\varepsilon}^2$  无关。

既然  $x^T M x$  仅与  $\beta$  有关，记

$$S(\beta) = x^T M x \quad (104)$$

则称  $S(\beta)$  为平方和函数。进一步记

$$\hat{\varepsilon}_i = E(\varepsilon_i | x_1, x_2, \dots, x_n)$$

即  $\hat{\varepsilon}_i$  是  $\varepsilon_i$  关于  $x_1, x_2, \dots, x_n$  的条件期望，也就是已知  $x_1, x_2, \dots, x_n$  时关于  $\varepsilon_i$  的估计，则可证

$$S(\beta) = \sum_{i=-\infty}^n \hat{\varepsilon}_i^2 \quad (105)$$

可见  $S(\beta)$  是一残差平方和的形式。

若  $\hat{\beta}_s$  使得式 (104) 达到最小，即

$$S(\beta) = \min$$

则  $\hat{\beta}_s$  称为最小平方和估计。

现在讨论最大似然估计。对数似然函数是

$$L(\beta, \sigma_{\varepsilon}^2 | x) = \frac{1}{2} \ln |M| - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_{\varepsilon}^2 - \frac{S(\beta)}{2\sigma_{\varepsilon}^2} \quad (106)$$

注意到  $M, S(\beta)$  均与  $\sigma_{\varepsilon}^2$  无关，令

$$\frac{\partial L}{\partial (\sigma_{\varepsilon}^2)} = -\frac{n}{2\sigma_{\varepsilon}^2} + \frac{S(\beta)}{2(\sigma_{\varepsilon}^2)^2} = 0 \quad (107)$$

解得

$$\hat{\sigma}_{\varepsilon}^2 = \frac{S(\beta)}{n} = \frac{x^T M x}{n} \quad (108)$$

将其代入式 (106) 中，得

$$L(\beta, \hat{\sigma}_{\varepsilon}^2 | x) = \frac{1}{2} \ln |M| - \frac{n}{2} \ln(x^T M x) + c \quad (109)$$

其中  $c = -\frac{n}{2} [\ln(2\pi) + 1 - \ln n]$ 。略去常数  $c$ ，得

$$L(\beta) = \frac{1}{2} \ln |M| - \frac{n}{2} \ln(x^T M x) \quad (110)$$

$\beta$  的最大似然估计应使式 (110) 达到最大。若  $\hat{\beta}_m$  使得

$$L(\hat{\beta}_m) = \max$$

则  $\hat{\beta}_m$  是  $\beta$  的最大似然估计 (ML 估计)，代入式 (108) 即得  $\sigma_{\varepsilon}^2$  的最大似然估计。

对于  $AR(p)$  序列

$$\varphi(B)X_t = X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = \varepsilon_t$$

当  $n \geq 2p$  时, 其平方和函数为

$$S(\varphi) = x^T M x = \sum_{i=0}^n \sum_{j=0}^n d_{ij} \varphi_i \varphi_j \quad (111)$$

其中  $\varphi_0 = -1$ ,  $d_{ij} = \sum_{t=1+j}^{n-i} x_{t+i-j} x_t = d_{ji}$ ,  $i, j = 0, 1, \dots, p$ 。记

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & d_{22} & \cdots & d_{2p} \\ \vdots & \vdots & & \vdots \\ d_{p1} & d_{p2} & \cdots & d_{pp} \end{bmatrix}, \quad d = \begin{bmatrix} d_{01} \\ d_{02} \\ \vdots \\ d_{0p} \end{bmatrix}$$

式 (111) 又可表为

$$S(\varphi) = \varphi^T D \varphi - 2\varphi^T d + d_{00}$$

使  $S(\varphi)$  达到最小的解是

$$\hat{\varphi}_s = D^{-1}d \quad (112)$$

它是  $\varphi$  的最小平方差估计, 近似于 ULS 估计。

### 3. Matlab 工具箱相关命令

Matlab 中 GARCH 工具箱可以实现时间序列建模的功能。工具箱中的模型 ARMAX(R,M,Nx) 的标准形为

$$y_t = C + \sum_{i=1}^R \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^M \theta_j \varepsilon_{t-j} + \sum_{k=1}^{N_x} \beta_k X(t, k) \quad (113)$$

其中  $\{\phi_i\}$  是自回归系数,  $\{\theta_j\}$  是滑动平均系数;  $X$  是解释回归矩阵, 它的每一个列是一个时间序列,  $X(t, k)$  表示第  $t$  行第  $k$  列的数据。

在工具箱中可以通过命令 `garchset` 指定模型的结构, `garchset` 的语法为

`Spec=garchset('属性 1', 属性 1 的值, '属性 2', 属性 2 的值, ...)`

常需指定如下属性:

**R, M** 表示 ARMA(R,M) 模型的阶次;

**C** 表示模型中的常数;

**AR** 表示自回归中的  $R$  个系数向量;

**MA** 表示滑动平均中的  $M$  个系数向量;

**Regress** 表示回归系数  $\beta_k$ 。

在工具箱中可以通过命令 `garchfit` 对模型中的参数进行估计, `garchfit` 的语法为

`[Coeff, Errors, LLF, Innovations, Sigmas] = garchfit(Spec, Series)`

其中输入参数 **Spec** 指定模型的结构, **Series** 为时间序列的样本观测值。输出参数 **Coeff** 是模型的参数估计值, **Errors** 是模型参数的标准差, **LLF** 是最大似然估计法中的对数目标函数值, **Innovations** 是残差向量, **Sigmas** 是对应于 **Innovations** 的标准差。

例 16 对于下列时间序列

$$X_t - 0.8X_{t-1} = \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0,1)$$

利用模拟产生的样本容量为 10000 的数据，对其参数进行估计。

解 编写如下程序

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=0.8*x(j-1)+elps(j)-0.4*elps(j-1); %产生样本点
end
spec = garchset('R',1,'M',1,'Display','off'); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec,x) %拟合参数
求得最大似然估计的参数值（其中的一次运行结果） $\hat{\phi}_1 = 0.7894$ ,  $\hat{\theta}_1 = 0.3984$ 。
```

例 17 对于下列时间序列

$$X_t + 0.5X_{t-1} - 0.3X_{t-2} = \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

利用模拟产生的样本容量为 10000 的数据，对其参数进行估计。

解 编写如下程序

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; x(2)=0; %赋初始值
for j=3:10000
    x(j)=-0.5*x(j-1)+0.3*x(j-2)+elps(j); %产生样本点
end
spec = garchset('R',2,'M',0,'Display','off'); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec,x) %拟合参数
求得最大似然估计的参数值（其中的一次运行结果） $\hat{\phi}_1 = -0.4968$ ,  $\hat{\phi}_2 = 0.2971$ 。
```

## 8.5 ARMA 模型的定阶与检验

现在讨论时间序列建模的另外两个问题，即模型的定阶与检验问题。对于一个时间序列，若已知服从  $ARMA(p, q)$  模型，关键是定阶，即估计  $p, q$  的值。实际上在建模过程中，首先要解决定阶问题，才能进行参数估计。在定阶与参数估计后，对建立的模型要进行检验。其基本做法是检验模型误差  $\varepsilon_t$  是否为白噪声。若检验认为  $\varepsilon_t$  是白噪声，则建模获得通过，否则要重新进行定阶与参数估计。

### 1. 模型定阶的 AIC 准则

AIC 准则又称 Akaike 信息准则，是由日本统计学家 Akaike 于 1974 年提出的。AIC 准则是信息论与统计学的重要研究成果，具有重要的意义。

AIC 准则起源于 Kullback—Leibler 信息量。设  $X$  是随机变量，它的概率密度是  $f(x)$ ，其中含有  $k$  个未知参数，设未知参数向量为  $\beta^0 = (\beta_1^0, \beta_2^0, \dots, \beta_k^0)^T$ 。 $f(x)$  属于分布族  $g(x|\beta)$ ，其中  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ 。显然

$$f(x) = g(x|\beta^0)$$

K—L 信息量可以用来刻画  $g(x|\beta)$  与  $f(x)$  的接近程度，其定义为

$$I(f(\cdot), g(\cdot|\beta)) = \int f(x) \ln \frac{f(x)}{g(x|\beta)} dx \quad (114)$$

则有  $I(f(\cdot), g(\cdot|\beta)) \geq 0$ ，且有  $I(f(\cdot), g(\cdot|\beta))|_{\beta=\beta_0} = 0$ 。

K-L 信息量是寻求最接近于  $f(x)$  的参数概率密度  $g(x|\beta)$ ，使得

$$I(f(\cdot), g(\cdot|\beta)) = \min \quad (115)$$

经过理论分析，当给定样本观测值  $x = (x_1, x_2, \dots, x_n)$ （它是容量为  $n$  的样本），设  ${}_k\hat{\beta}^{(m)}$  是模型参数  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ （未知参数个数是  $k$ ， $k$  未知）的最大似然估计，这里标出左足标“ $k$ ”是为了强调未知参数个数  $k$  是未知的，是需要估计的。设  $\ln(L(\beta))$  是其对数似然函数，AIC 信息准则是：使得式 (115) 中的  $k$ （ $k$  确定后， ${}_k\hat{\beta}^{(m)}$  就确定）满足

$$\text{AIC}(k) = -2\ln(L({}_k\hat{\beta}^{(m)})) + 2k = \min \quad (116)$$

设  $X_t$  是 ARMA( $p, q$ ) 序列，其中未知参数的个数是  $k = p + q + 1$  个，包括自回归参数  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^T$ ，滑动平均参数  $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$  及  $\sigma_\varepsilon^2$ 。在式 (106) 中，略去  $\frac{1}{2}\ln|M| - \frac{n}{2}\ln(2\pi)$ ，得最小平方和估计对应的对数似然函数

$$L'(\beta, \sigma_\varepsilon^2 | x) = -\frac{n}{2}\ln\sigma_\varepsilon^2 - \frac{S(\beta)}{2\sigma_\varepsilon^2} \quad (117)$$

由式 (108) 可得， $\sigma_\varepsilon^2$  的最大似然估计

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}S(\hat{\beta})$$

将上式代入式 (117)，得

$$L'(\beta, \sigma_\varepsilon^2 | x) = -\frac{n}{2}\ln\sigma_\varepsilon^2 - \frac{n}{2}$$

因此，ARMA( $p, q$ ) 序列 AIC 定阶准则为：选  $p, q$ ，使得

$$\text{AIC} = n\ln\hat{\sigma}_\varepsilon^2 + 2(p + q + 1) = \min \quad (118)$$

其中， $n$  是样本容量， $\hat{\sigma}_\varepsilon^2$  与  $p$  和  $q$  有关。若当  $p = \hat{p}$ ， $q = \hat{q}$  时，式 (118) 达到最小值，则认为序列是 ARMA( $\hat{p}, \hat{q}$ )。

当 ARMA( $\hat{p}, \hat{q}$ ) 序列含有未知均值参数  $\mu$  时，模型为

$$\varphi(B)(X_t - \mu) = \theta(B)\varepsilon_t \quad (119)$$

这时，未知参数个数为  $k = p + q + 2$ ，AIC 准则为：选取  $p, q$ ，使得

$$\text{AIC} = n\ln\hat{\sigma}_\varepsilon^2 + 2(p + q + 2) = \min \quad (120)$$

实际上，式 (118) 与式 (120) 有相同的最小值点  $\hat{p}, \hat{q}$ 。

## 2. 模型检验的 $\chi^2$ 检验

若拟合模型的残差记为  $\hat{\varepsilon}_t$ ，它是  $\varepsilon_t$  的估计。例如，对 AR( $p$ ) 序列，设未知参数的估计是  $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p$ ，则残差

$$\hat{\varepsilon}_t = X_t - \hat{\varphi}_1 X_{t-1} - \dots - \hat{\varphi}_p X_{t-p}, \quad t = 1, 2, \dots, p, \quad (\text{设 } X_0 = X_{-1} = \dots = X_{1-p} = 0)$$

记

$$\eta_k = \frac{\sum_{t=1}^{n-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad k=1,2,\dots,m$$

Ljung-Box 的  $\chi^2$  检验统计量是

$$\chi^2 = n(n+2) \sum_{k=1}^m \frac{\eta_k^2}{n-k} \quad (121)$$

检验的假设是

$$H_0: \rho_k = 0, \text{ 当 } k \leq m; \quad H_1: \text{ 对某些 } k \leq m, \quad \rho_k \neq 0$$

在  $H_0$  成立时, 若  $n$  充分大,  $\chi^2$  近似于  $\chi^2(m-r)$  分布, 其中  $r$  是估计的模型参数个数。

**$\chi^2$  检验法:** 给定显著性水平  $\alpha$ , 查表得上  $\alpha$  分位数  $\chi_\alpha^2(m-r)$ , 则当  $\chi^2 > \chi_\alpha^2(m-r)$  时拒绝  $H_0$ , 即认为  $\varepsilon_t$  非白噪声, 模型检验未通过; 而当  $\chi^2 \leq \chi_\alpha^2(m-r)$  时, 接受  $H_0$ , 认为  $\varepsilon_t$  是白噪声, 模型通过检验。

例 18 模拟产生下列时间序列的 10000 个观测数据:

$$X_t - 0.8X_{t-1} = \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0,1)$$

用实际计算验证用 AIC 准则定阶的正确性。

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=0.8*x(j-1)+elps(j)-0.4*elps(j-1); %产生样本点
end
for i=0:3
    for j=0:3
        spec= garchset('R',i,'M',j,'Display','off'); %指定模型的结构
        [coeffX,errorsX,LLFX] = garchfit(spec,x); %拟合参数
        num=garchcount(coeffX); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic,bic]=aicbic(LLFX,num,10000);
        fprintf('R=%d,M=%d,AIC=%f,BIC=%f\n',i,j,aic,bic); %显示计算结果
    end
end
```

其中的一次计算结果如下:

```
R=0,M=0,AIC=32484.707852,BIC=32499.128533
R=0,M=1,AIC=30332.919112,BIC=30354.550133
R=0,M=2,AIC=29494.702222,BIC=29523.543583
R=0,M=3,AIC=29123.131961,BIC=29159.183663
R=1,M=0,AIC=29124.394412,BIC=29146.025433
R=1,M=1,AIC=28620.943324,BIC=28649.784685
R=1,M=2,AIC=28622.928551,BIC=28658.980253
R=1,M=3,AIC=28624.851198,BIC=28668.113240
R=2,M=0,AIC=28696.403650,BIC=28725.245012
```

R=2,M=1,AIC=28622.928290,BIC=28658.979992  
R=2,M=2,AIC=28624.372827,BIC=28667.634869  
R=2,M=3,AIC=28626.264289,BIC=28676.736671  
R=3,M=0,AIC=28639.063004,BIC=28675.114706  
R=3,M=1,AIC=28624.860895,BIC=28668.122937  
R=3,M=2,AIC=28626.271477,BIC=28676.743860  
R=3,M=3,AIC=28617.468215,BIC=28675.150938

计算结果显示应为 ARMA(1,1) 序列。

若实际序列是 ARMA( $p, q$ ) 序列，应对各种可能的  $p, q$  值计算其 AIC。因此，计算量较大。只有认真反复计算，才能求得真阶  $p$  和  $q$ 。

例 19 模拟产生下列时间序列的 10000 个观测数据：

$$X_t - 0.8X_{t-1} = \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0,1)$$

建立 ARMA(1,1) 模型，并用  $\chi^2$  检验法进行模型检验。

解 编写程序如下：

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=0.8*x(j-1)+elps(j)-0.4*elps(j-1); %产生样本点
end
spec= garchset('R',1,'M',1); %指定模型的结构
[coeffX,errorsX,LLFX]= garchfit(spec,x) %拟合参数
e(1)=0; %残差赋初值
for j=2:10000
    e(j)=x(j)-coeffX.AR*x(j-1) -coeffX.MA*e(j-1); %计算残差
end
%下面进行 chi2 检验，是否服从 0 均值的正态分布，nparam 表示估计参数的个数
[h,p,stats]= chi2gof(e,'cdf',@(z)normcdf(z,0,std(e)),'nparams',1)
```

## 8.6 平稳线性最小均方预报

时间序列的  $m$  步预报，是根据  $\{X_k, X_{k-1}, \dots\}$  的取值对未来  $k+m$  时刻的随机变量  $X_{k+m}$  ( $m > 0$ ) 做出估计。估计量记作  $\hat{X}_k(m)$ ，它是  $X_k, X_{k-1}, \dots$  的线性组合。引进估计空间

$$\Omega_k \equiv \{x \mid x = \sum_{j=0}^{\infty} c_j X_{k-j}, \sum_{j=0}^{\infty} c_j^2 < \infty\}$$

现设  $X_t$  是零均值正态 ARMA( $p, q$ ) 序列。所谓平稳线性最小均方预报  $\hat{X}_k(m)$  是指：

$\hat{X}_k(m) \in \Omega_k$ ，且使得

$$E[(X_{k+m} - \hat{X}_k(m))^2] = \min \quad (122)$$

在正态性的条件下  $\hat{X}_k(m)$  是  $X_{k+m}$  对于  $\{X_k, X_{k-1}, \dots\}$  的条件期望，即

$$\hat{X}_k(m) = E(X_{k+m} \mid X_k, X_{k-1}, \dots) = E(X_{k+m} \mid \Omega_k) \quad (123)$$

因为  $\{X_k, X_{k-1}, \dots\}$  一般是相关的, 因此, 将  $\hat{X}_k(m)$  用  $\{X_k, X_{k-1}, \dots\}$  的元素的线性表示不甚方便。考虑到对于  $\text{ARMA}(p, q)$  序列, 总存在着传递形式, 令

$$\Xi_k = \{\varepsilon \mid \varepsilon = \sum_{j=0}^{\infty} d_j \varepsilon_{k-j}, \sum_{j=0}^{\infty} d_j^2 < +\infty\}, \quad (124)$$

则可以证明  $\Omega_k = \Xi_k$ 。这就是说,  $\hat{X}_k(m)$  可以用  $\Xi_k$  中的元素来表示。

平稳线性最小均方预报的性质:

1) 设零均值  $\text{ARMA}(p, q)$  序列  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  具有传递形式

$$X_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j},$$

则

$$\hat{X}_k(m) = \sum_{j=0}^{\infty} G_{j+m} \varepsilon_{k-j} = G_m \varepsilon_k + G_{m+1} \varepsilon_{k-1} + G_{m+2} \varepsilon_{k-2} + \dots \quad (125)$$

若令  $e_k(m) = X_{k+m} - \hat{X}_k(m)$ , 则

$$e_k(m) = \varepsilon_{k+m} + G_1 \varepsilon_{k+m-1} + \dots + G_{m-1} \varepsilon_{k+1} \quad (126)$$

有

$$\text{Var}(e_k(m)) = \sigma_\varepsilon^2 (1 + G_1^2 + G_2^2 + \dots + G_{m-1}^2) \quad (127)$$

当  $m=1$  时, 由式 (126),

$$e_k(1) = X_{k+1} - \hat{X}_k(1) = \varepsilon_{k+1}$$

说明  $k$  时刻的一步预报误差是  $\varepsilon_{k+1}$ 。

平稳线性最小均方预报即条件期望

$$\hat{X}_k(m) = E(X_{k+m} \mid \Omega_k) = E(X_{k+m} \mid \Xi_k)$$

有简单性质

$$E\left(\sum_{j=0}^{\infty} c_j X_{k+j} \mid \Omega_k\right) = \sum_{j=0}^{\infty} c_j E(X_{k+j} \mid \Omega_k)$$

$$E(X_{k+m} \mid \Omega_k) = E(X_{k+m} \mid \Xi_k) = \begin{cases} \hat{X}_k(m), & m > 0 \\ X_{k+m}, & m \leq 0 \end{cases}$$

$$E(\varepsilon_{k+m} \mid \Omega_k) = E(\varepsilon_{k+m} \mid \Xi_k) = \begin{cases} 0, & m > 0 \\ \varepsilon_{k+m}, & m \leq 0 \end{cases}$$

2) 如  $\{X_t, t=0, \pm 1, \pm 2, \dots\}$  是零均值  $\text{ARMA}(p, q)$ , 则下列预报差分方程成立

$$\hat{X}_k(m) = \varphi_1 \hat{X}_k(m-1) + \varphi_2 \hat{X}_k(m-2) + \dots + \varphi_p \hat{X}_k(m-p), \quad m > p \quad (128)$$

事实上, 因

$$X_{k+m} = \varphi_1 X_{k+m-1} + \varphi_2 X_{k+m-2} + \dots + \varphi_p X_{k+m-p} - \theta_1 \varepsilon_{k+m-1} - \dots - \theta_q \varepsilon_{k+m-q} + \varepsilon_{k+m}$$

当  $m > p$  时

$$\hat{X}_k(m) = E(X_{k+m} \mid \Omega_k) = \varphi_1 E(X_{k+m-1} \mid \Omega_k) + \dots + \varphi_p E(X_{k+m-p} \mid \Omega_k)$$



$$\begin{aligned}
& -\theta_1 E(\varepsilon_{k+m-1} | \Omega_k) - \cdots - \theta_q E(\varepsilon_{k+m-q} | \Omega_k) + E(\varepsilon_{k+m} | \Omega_k) \\
& = \varphi_1 \hat{X}_k(m-1) + \varphi_2 \hat{X}_k(m-2) + \cdots + \varphi_p \hat{X}_k(m-p)
\end{aligned}$$

故式 (128) 成立。

当  $X_t$  是  $\text{MA}(q)$  序列时, 有

$$\hat{X}_k(m) = 0, \quad m > q \quad (129)$$

3) 下列递推预报公式成立:

$$\hat{X}_{k+1}(m) = \hat{X}_k(m+1) + G_m[X_{k+1} - \hat{X}_k(1)] \quad (130)$$

这是因为由式 (125) 及  $\varepsilon_{k+1} = X_{k+1} - \hat{X}_k(1)$ , 得

$$\begin{aligned}
\hat{X}_{k+1}(m) &= \sum_{j=0}^{\infty} G_{j+m} \varepsilon_{k+1-j} = G_m \varepsilon_{k+1} + \sum_{j=1}^{\infty} G_{m+j} \varepsilon_{k+1-j} \\
&= G_m (X_{k+1} - \hat{X}_k(1)) + \sum_{i=0}^{\infty} G_{m+1+i} \varepsilon_{k-i} = \hat{X}_k(m+1) + G_m (X_{k+1} - \hat{X}_k(1))
\end{aligned}$$

从而式 (130) 成立。

式 (130) 可以这样理解:  $X_{k+1} - \hat{X}_k(1)$  表示取得观测值  $X_{k+1}$  后带来的“新息”,  $k+1$  时刻的  $m$  步预报值等于  $k$  时刻的  $m+1$  步预报值与“新息”的加权和, 这里取“权”为  $G_m$ 。

4) 设  $\text{ARMA}(p, q)$  序列  $X_t$  的逆转形式是

$$\varepsilon_t = X_t - \sum_{j=1}^{\infty} I_j X_{t-j}$$

则

$$\hat{X}_k(m) = \sum_{j=1}^{\infty} I_j^{(m)} X_{k+1-j} \quad (131)$$

其中

$$\begin{cases} I_j^{(1)} = I_j \\ I_j^{(m)} = I_{j+m-1} + \sum_{i=1}^{m-1} I_i I_j^{(m-i)} \end{cases} \quad (132)$$

式 (131) 表示  $\hat{X}_k(m)$  是  $X_k, X_{k-1}, \dots$  的线性组合, 其系数是  $I_j^{(m)}$ 。又式 (132)  $I_j^{(m)}$  的递推计算式由逆函数  $\{I_j, j=1, 2, \dots\}$  决定。

## 8.7 $\text{ARMA}(p, q)$ 序列的预报

### 1. $\text{AR}(p)$ 序列的预报

由预报差分方程 (128)

$$\hat{X}_k(m) = \varphi_1 \hat{X}_k(m-1) + \varphi_2 \hat{X}_k(m-2) + \cdots + \varphi_p \hat{X}_k(m-p), \quad m > 0$$

又

$$\hat{X}_k(m) = X_{k-m}, \quad m \leq 0$$

这就给出  $\text{AR}(p)$  序列的预报递推公式:

$$\begin{cases} \hat{X}_k(1) = \varphi_1 X_k + \varphi_2 X_{k-1} + \cdots + \varphi_p X_{k-p+1} \\ \hat{X}_k(2) = \varphi_1 \hat{X}_k(1) + \varphi_2 X_k + \cdots + \varphi_p X_{k-p+2} \\ \dots\dots\dots \\ \hat{X}_k(p) = \varphi_1 \hat{X}_k(p-1) + \varphi_2 \hat{X}_k(p-2) + \cdots + \varphi_{p-1} \hat{X}_k(1) + \varphi_p X_k \\ \hat{X}_k(m) = \varphi_1 \hat{X}_k(m-1) + \varphi_2 \hat{X}_k(m-2) + \cdots + \varphi_p \hat{X}_k(m-p), m > p \end{cases} \quad (133)$$

由此可见,  $\hat{X}_k(m) (m \geq 1)$  仅仅依赖于  $X_t$  的  $k$  时刻以前的  $p$  个时刻的值  $X_k, X_{k-1}, \dots, X_{k-p+1}$ 。这是 AR( $p$ ) 序列预报的特点。

## 2. MA( $q$ ) 与 ARMA( $p, q$ ) 序列的预报

关于 MA( $q$ ) 序列  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  的预报, 有

$$\hat{X}_k(m) = 0, \quad m > q$$

因此, 只需要讨论  $\hat{X}_k(m), m = 1, 2, \dots, q$ 。为此, 定义预报向量

$$\hat{X}_k(q) = (\hat{X}_k(1), \hat{X}_k(2), \dots, \hat{X}_k(q))^T \quad (134)$$

所谓递推预报是求  $\hat{X}_k^{(q)}$  与  $\hat{X}_{k+1}^{(q)}$  的递推关系, 由递推预报公式 (130), 对 MA( $q$ ) 序列,  $G_m = -\theta_m, m = 1, 2, \dots, q$ , 得

$$\begin{aligned} \hat{X}_{k+1}(1) &= \theta_1 \hat{X}_k(1) + \hat{X}_k(2) - \theta_1 X_{k+1} \\ \hat{X}_{k+1}(2) &= \theta_2 \hat{X}_k(1) + \hat{X}_k(3) - \theta_2 X_{k+1} \\ &\dots\dots\dots \\ \hat{X}_{k+1}(q-1) &= \theta_{q-1} \hat{X}_k(1) + \hat{X}_k(q) - \theta_{q-1} X_{k+1} \\ \hat{X}_{k+1}(q) &= \theta_q \hat{X}_k(1) - \theta_q X_{k+1} \end{aligned}$$

从而得

$$\hat{X}_{k+1}^{(q)} = \begin{bmatrix} \theta_1 & 1 & 0 & \cdots & 0 \\ \theta_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \theta_{q-1} & 0 & 0 & \cdots & 1 \\ \theta_q & 0 & 0 & 0 & 0 \end{bmatrix} \hat{X}_k^{(q)} - \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{bmatrix} X_{k+1} \quad (135)$$

递推初值可取  $\hat{X}_{k_0}^{(q)} = 0$  ( $k_0$  较小)。因为模型的可逆性保证了递推式渐近稳定, 即当  $n$  充分大后, 初始误差的影响可以逐渐消失。

对于 ARMA( $p, q$ ) 序列, 由式 (128)

$$\hat{X}_k(m) = \varphi_1 \hat{X}_k(m-1) + \varphi_2 \hat{X}_k(m-2) + \cdots + \varphi_p \hat{X}_k(m-p), \quad m > p$$

因此, 只需要知道  $\hat{X}_k(1), \hat{X}_k(2), \dots, \hat{X}_k(p)$ , 就可以递推算得  $\hat{X}_k(m), m > p$ 。仍定义预报向量 (134)。令

$$\varphi_j^* = \begin{cases} \varphi_j, & j=1,2,\dots,p \\ 0, & j>p \end{cases}$$

可证下列递推预报公式

$$\hat{X}_{k+1}^{(q)} = \begin{bmatrix} -G_1 & 1 & 0 & \cdots & 0 \\ -G_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -G_{q-1} & 0 & 0 & \cdots & 1 \\ -G_q + \varphi_q^* & \varphi_{q-1}^* & \varphi_{q-2}^* & \cdots & \varphi_1^* \end{bmatrix} \hat{X}_k^{(q)} + \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{q-1} \\ G_q \end{bmatrix} X_{k+1} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sum_{j=q+1}^p \varphi_j^* X_{k+q+1-j} \end{bmatrix} \quad (136)$$

式(136)中第三项当  $p \leq q$  时为零。由可逆性条件保证, 当  $k_0$  较小时, 可令初值  $\hat{X}_k^{(q)} = 0$ 。

在实际中, 模型参数是未知的。若已建立了时间序列的模型, 则理论模型中的未知参数用其估计替代, 再用上面介绍的方法进行预报。

下面, 用随机模拟的方法, 对所讨论的时间序列进行预报, 并从理论上予以验证。

例 20 随机产生下列时间序列的 10000 个观测值

$$X_t + 0.65X_{t-1} = \varepsilon_t$$

估计其模型参数, 计算 1, 2, 3 步预测, 并对计算结果从理论上验证其正确性。

解 利用模拟数据计算得到其 1, 2, 3 步预测 (其中的一次运行结果) 为

$$\begin{matrix} -0.9901 & 0.6329 & -0.4281 \end{matrix}$$

以下进行理论验证。模拟数据中  $X_{10000} = 1.4925$ , 故理论预报值为 (由式(133))

$$\hat{X}_{10000}(1) = -0.65 \times 1.4925 = -0.9701$$

$$\hat{X}_{10000}(2) = -0.65 \times (-0.9701) = 0.6306$$

$$\hat{X}_{10000}(3) = -0.65 \times 0.6306 = -0.4099$$

又 3 步预报理论标准差计算如下。因  $G_1 = \varphi_1 = -0.65$ ,  $G_2 = \varphi_1^2 = 0.4225$ ,

$G_3 = \varphi_1^3 = -0.2746$ 。标准差以 Std 表示, 由式(127)

$$\text{Std}[e_{10000}(1)] = \sqrt{\sigma_\varepsilon^2} = 1$$

$$\text{Std}[e_{10000}(2)] = \sqrt{\sigma_\varepsilon^2(1 + G_1^2)} = 1.1927$$

$$\text{Std}[e_{10000}(3)] = \sqrt{\sigma_\varepsilon^2(1 + G_1^2 + G_2^2)} = 1.2653$$

计算结果与理论值十分接近。

计算的 Matlab 程序如下:

```
clc,clear
randn('state',sum(clock)); %初始化随机数发生器
elps=randn(1,10000); %产生 10000 个服从标准正态分布的随机数
x(1)=0; %赋初始值
for j=2:10000
    x(j)=-0.65*x(j-1)+elps(j); %产生样本点
end
```

```

spec= garchset('R',1); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec,x) %拟合参数
%下面计算 3 步预测值, 其中第一个参数返回值是标准差, 第二个参数返回值为预
报值
[sigmaForecast,x_Forecast] = garchpred(coeffX,x,3)
x_theory(1)=-0.65*x(10000); %计算一步理论预测值
x_theory(2)=-0.65*x_theory(1); %计算二步理论预测值
x_theory(3)=-0.65*x_theory(2) %计算三步理论预测值

```

## §9 ~~ARIMA 序列与季节性序列~~

在前面两节中介绍了平稳时间序列, 主要是 ARMA 序列的建模与预报。在实际中遇到的时间序列往往有三个特性: 趋势性、季节性与非平稳性。本节主要采用 Box—Jenkins 方法, 即差分方法, 有时还要用时间序列的变换方法, 消除其趋势性、季节性, 使得变换后的序列是平稳序列, 并假设为 ARMA 序列, 再用上面介绍的方法去研究。

### 9.1 ARIMA 序列及其预报

我们先看一个例子。考虑研究时间序列  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ , 满足

$$X_t - 1.5X_{t-1} + 0.5X_{t-2} = \varepsilon_t \quad (137)$$

改写为  $\varphi(B)X_t = \varepsilon_t$ , 其中  $\varphi(B) = 1 - 1.5B + 0.5B^2$ 。 $\varphi(B) = 0$  的根是  $B_1 = 1$ ,  $B_2 = 2$ , 其中  $B_1 = 1$  在单位圆周上, 并非在单位圆外, 即原序列非平稳, 因而不是 AR(2) 序列。但若改写式 (137) 为

$$(X_t - X_{t-1}) - 0.5(X_{t-1} - X_{t-2}) = \varepsilon_t$$

令

$$\nabla X_t = X_t - X_{t-1} = W_t \quad (138)$$

则有

$$W_t - 0.5W_{t-1} = \varepsilon_t$$

$W_t$  是 AR(1) 序列。式 (138) 定义的运算  $\nabla$  称为 1 阶向后差分运算。经过这样的 1 阶差分运算, 原来的非平稳序列  $X_t$  转化为平稳序列  $W_t$ 。

由上例可见, 差分运算可以使一类非平稳序列 (即带有趋势性的序列) 平稳化。如果 1 阶差分还不能使时间序列平稳化, 还可以进行 2 阶差分, 3 阶差分, 直至第  $d$  阶差分, 最后将序列化为平稳序列。

1 阶差分:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

2 阶差分:

$$\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2} = (1 - B)^2 X_t$$

一般地,  $d$  阶差分:

$$\nabla^d X_t = (1 - B)^d X_t$$

其中  $\nabla^d$  称为  $d$  阶差分算子,

$$\nabla^d \equiv (1 - B)^d = 1 - \binom{d}{1}B + \binom{d}{2}B^2 - \dots + (-1)^{d-1} \binom{d}{d-1}B^{d-1} + (-1)^d B^d。$$

设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是非平稳序列。若存在正整数  $d$ , 使得

$$\nabla^d X_t = W_t$$

而  $\{W_t, t = 0, \pm 1, \pm 2, \dots\}$  是  $\text{ARMA}(p, q)$  序列, 则称  $X_t$  是  $\text{ARIMA}(p, d, q)$  序列。这时,  $X_t$  满足

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t \quad (139)$$

若  $\nabla^d X_t$  为平稳序列, 但均值  $\mu \neq 0$ , 则  $\nabla^d X_t - \mu$  为平稳零均值序列, 满足

$$\varphi(B)(\nabla^d X_t - \mu) = \theta(B)\varepsilon_t, \quad t > d \quad (140)$$

此时, 称  $X_t$  为一般  $\text{ARIMA}(p, d, q)$  序列, 若  $\mu$  未知, 可用  $\nabla^d X_t$  的平均值  $\bar{x}$  估计。

若  $X_t$  的观测样本是  $X_1, X_2, \dots, X_n$ , 经过 1 阶差分后, 数据减少为  $n-1$  个; 2 阶差分以后, 数据为  $n-2$  个; 一般地,  $d$  阶差分以后, 数据为  $n-d$  个。由  $d$  阶差分  $\nabla^d X_t$  复原数据, 需要给定初值  $X_1, X_2, \dots, X_d$ 。

在确定模型时, 往往采用下面的方法。先对  $X_t$  的样本  $X_1, X_2, \dots, X_n$ , 计算样本自相关函数与样本偏相关函数。如果是截尾的或者是拖尾的 (即被负指数控制的), 说明已服从  $\text{ARMA}$  模型。若自相关函数与偏相关函数至少有 1 个不是截尾的或拖尾的, 说明  $X_t$  不是平稳的, 可以作 1 阶差分  $\nabla X_t, t = 2, 3, \dots, n$ , 并求其样本自相关函数与样本偏相关函数, 再用上述方法讨论。这样, 直至判断  $\nabla^d X_t$  是平稳序列为止。在实际计算中, 若遇到样本自相关函数或样本偏相关函数的图形虽然下降, 但下降很慢, 应认为是非平稳序列, 需作差分运算。

若初值  $X_1, X_2, \dots, X_d$  已知, 由

$$W_t = \nabla^d Z_t, \quad t = d+1, \dots, n$$

可以复原  $X_t$ 。给出  $d=1, 2$  时的复原公式, 这是不难证明的。

(1)  $d=1$

$$X_t = X_1 + \sum_{j=1}^{t-1} W_{j+1} = X_k + \sum_{j=1}^{t-k} W_{j+k}, \quad t > k \geq 1 \quad (141)$$

(2)  $d=2$

$$\begin{aligned} X_t &= X_2 + (t-2)(X_2 - X_1) + \sum_{j=1}^{t-2} (t-j-1)W_{j+2} \\ &= X_k + (t-k)(X_k - X_{k-1}) + \sum_{j=1}^{t-k} (t-j-k+1)W_{j+k}, \quad t > k \geq 2 \end{aligned} \quad (142)$$

设  $X_t$  是  $\text{ARIMA}(p, d, q)$  序列, 则当  $p=0$  时, 称为  $\text{IMA}(d, q)$  序列; 当  $q=0$  时, 称为  $\text{ARI}(p, d)$  序列。

下面简单介绍  $\text{ARIMA}$  序列的预报。

设  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  是  $\text{ARIMA}(p, d, q)$  序列, 仅讨论  $d=1$  与  $d=2$  的情形, 这在实用中是最常见的。

(1) 当  $d=1$  时,  $\nabla X_t = W_t$ , 有

$$\nabla \hat{X}_k(m) = \hat{W}_k(m),$$

即

$$\hat{X}_k(m) - \hat{X}_k(m-1) = \hat{W}_k(m)$$

由此得

$$\hat{X}_k(m) = \hat{X}_k(m-1) + \hat{W}_k(m) = X_k + \sum_{j=1}^m \hat{W}_k(j) \quad (143)$$

(2) 当  $d=2$  时,  $\nabla^2 \hat{X}_k(m) = \hat{W}_k(m)$ , 即

$$\hat{X}_k(m) - 2\hat{X}_k(m-1) + \hat{X}_k(m-2) = \hat{W}_k(m)$$

复原  $\hat{X}_k(m)$ , 可得

$$\hat{X}_k(m) = X_k + m(X_k - X_{k-1}) + \sum_{j=1}^m (m+1-j)\hat{W}_k(j) \quad (144)$$

例 21 考虑 IMA(1,1) 序列:  $\nabla X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ , 即

$$X_t - X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (145)$$

(1) 求  $X_t$  的逆转形式;

(2) 求  $X_t$  的预报。

解 (1) 序列 (145) 是实用中相当重要的时间序列, 称为求和滑动平均序列。  $X_t$  经 1 阶差分以后成为 MA(1,1) 序列。式 (145) 是可逆的, 因为

$$\theta(B) = 1 - \theta_1 B$$

当  $|\theta_1| < 1$  时,  $\theta(B) = 0$  的根  $1/\theta_1$  在单位圆外, 故可求逆转形式。设

$$I(B)X_t = \varepsilon_t \quad (146)$$

即

$$X_t - I_1 X_{t-1} - I_2 X_{t-2} \cdots = \varepsilon_t$$

由式 (145), 得

$$(1-B)X_t = (1-\theta_1 B)\varepsilon_t \quad (147)$$

将式 (146) 代入式 (147), 得

$$(1-B)X_t = (1-\theta_1 B)I(B)X_t$$

因此

$$1-B = (1-\theta_1 B)I(B)$$

即

$$I(B) = \frac{1-B}{1-\theta_1 B} = \frac{1-\theta_1 B - (1-\theta_1)B}{1-\theta_1 B} = 1 - (1-\theta_1)(B + \theta_1 B^2 + \theta_1^2 B^3 + \cdots)$$

得

$$I_j = (1-\theta_1)\theta_1^{j-1}, \quad j=1, 2, \cdots$$

故  $X_t$  的逆转形式是

$$\varepsilon_t = X_t - (1-\theta_1) \sum_{j=1}^{\infty} \theta_1^{j-1} X_{t-j} \quad (148)$$

(2) 记  $W_t = X_t - X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ , 则  $W_t$  是 MA(1) 序列。由式 (143), 得

$$\begin{aligned}\hat{X}_{k+1}(1) &= X_{k+1} + \hat{W}_{k+1}(1) = X_{k+1} + \theta_1 \hat{W}_k(1) - \theta_1 W_{k+1} \\ &= X_{k+1} + \theta_1 [\hat{W}_k(1) - X_{k+1} + X_k] = (1 - \theta_1) X_{k+1} + \theta_1 \hat{X}_k(1)\end{aligned}\quad (149)$$

若当  $k = k_0$ , 取  $\hat{X}_{k_0}(1) = X_{k_0}$ , 按上式递推, 得  $\hat{X}_k(1)$ 。又当  $m \geq 2$ ,  $\hat{W}_k(m) = 0$ , 故由式 (143), 得

$$\hat{X}_k(m) = \hat{X}_k(1), \quad m \geq 2$$

式 (149) 的含义是:  $k+1$  时刻  $X_t$  的 1 步预报是  $X_{k+1}, \hat{X}_k(1)$  的加权线性组合, 其权分别是  $1 - \theta_1, \theta_1$ 。

式 (148) 取  $t = k+1$ , 并两边对  $\Omega_k$  取条件期望, 得

$$\hat{X}_k(1) = (1 - \theta_1) \sum_{j=1}^{\infty} \theta_1^{j-1} X_{k+1-j} \quad (150)$$

表明  $X_t$  在  $k$  时刻的 1 步预报是  $X_k, X_{k-1}, \dots$  的加权线性组合, 其权是  $1 - \theta_1, (1 - \theta_1)\theta_1, (1 - \theta_1)\theta_1^2, \dots$ , 即是  $X_k, X_{k-1}, \dots$  的指数平滑。这是时间序列预报的指数平滑法的由来。实际上, 当模型是 IMA(1,1) (或 MA(1), ARMA(1,1)) 时, 采用指数平滑法是合适的。式 (149), 式 (150) 是时间序列 IMA(1,1) 预报的特点。

下面, 通过实例说明如何实现时间序列的建模与预报。

例 22 某化工生产过程每 2 小时的浓度读数如表 13 所示。

表 13 化工生产过程浓度数据

17.0	16.6	16.3	16.1	17.1	16.9	16.8	17.4	17.1	17.0
16.7	17.4	17.2	17.4	17.4	17.0	17.3	17.2	17.4	16.8
17.1	17.4	17.4	17.5	17.4	17.6	17.4	17.3	17.0	17.8
17.5	18.1	17.5	17.4	17.4	17.1	17.6	17.7	17.4	17.8
17.6	17.5	16.5	17.8	17.3	17.3	17.1	17.4	16.9	17.3
17.6	16.9	16.7	16.8	16.8	17.2	16.8	17.6	17.2	16.6
17.1	16.9	16.6	18.0	17.2	17.3	17.0	16.9	17.3	16.8
17.3	17.4	17.7	16.8	16.9	17.0	16.9	17.0	16.6	16.7
16.8	16.7	16.4	16.5	16.4	16.6	16.5	16.7	16.4	16.4
16.2	16.4	16.3	16.4	17.0	16.9	17.1	17.1	16.7	16.9
16.5	17.2	16.4	17.0	17.0	16.7	16.2	16.6	16.9	16.5
16.6	16.6	17.0	17.1	17.1	16.7	16.8	16.3	16.6	16.8
16.9	17.1	16.8	17.0	17.2	17.3	17.2	17.3	17.2	17.2
17.5	16.9	16.9	16.9	17.0	16.5	16.7	16.8	16.7	16.7
16.6	16.5	17.0	16.7	16.7	16.9	17.4	17.1	17.0	16.8
17.2	17.2	17.4	17.2	16.9	16.8	17.0	17.4	17.2	17.2
17.1	17.1	17.1	17.4	17.2	16.9	16.9	17.0	16.7	16.9
17.3	17.8	17.8	17.6	17.5	17.0	16.9	17.1	17.2	17.4
17.5	17.9	17.0	17.0	17.0	17.2	17.3	17.4	17.4	17.0
18.0	18.2	17.6	17.8	17.7	17.2	17.4			

(1) 对这一生产过程建模;

(2) 对这一生产过程进行 10 步预测。

解 通过计算自相关函数和偏相关函数, 确定取  $d = 1$ 。利用 AIC 准则定阶, 取

ARIMA(3,1,3) 模型。计算得

$$\hat{\phi}_1 = 1.1762, \hat{\phi}_2 = -1.1158, \hat{\phi}_3 = 0.1706$$

$$\hat{\theta}_1 = 1.8129, \hat{\theta}_2 = -1.8174, \hat{\theta}_3 = 0.8412$$

得到模型为

$$(1 - 1.1762B + 1.1158B^2 - 0.1706B^3)(1 - B)X_t$$

$$= (1 - 1.8129B + 1.8174B^2 - 0.8412B^3)\varepsilon_t$$

(2) 经计算, 其 10 步预报值见表 14。

表 14 10 步预报值

步数	1	2	3	4	5
预报值	17.3611	17.4446	17.5811	17.6441	17.5821
步数	6	7	8	9	10
预报值	17.4644	17.4079	17.4644	17.5759	17.6364

计算的 Matlab 程序如下:

```

clc,clear
a=textread('hua.txt'); %把原始数据按照原来的排列格式存放在纯文本文件 hua.txt
a=nonzeros(a); %按照原来数据的顺序去掉零元素
r11=autocorr(a) %计算自相关函数
r12=parcorr(a) %计算偏相关函数
da=diff(a); %计算 1 阶差分
r21=autocorr(da) %计算自相关函数
r22=parcorr(da) %计算偏相关函数
n=length(da); %计算差分后的数据个数
for i=0:3
    for j=0:3
        spec= garchset('R',i,'M',j,'Display','off'); %指定模型的结构
        [coeffX,errorsX,LLFX] = garchfit(spec,da); %拟合参数
        num=garchcount(coeffX); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic,bic]=aicbic(LLFX,num,n);
        fprintf('R=%d,M=%d,AIC=%f,BIC=%f\n',i,j,aic,bic); %显示计算结果
    end
end
r=input('输入阶数 R=');m=input('输入阶数 M=');
spec2= garchset('R',r,'M',m,'Display','off'); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec2,da) %拟合参数
[sigmaForecast,w_Forecast] = garchpred(coeffX,da,10) %计算 10 步预报值
x_pred=a(end)+cumsum(w_Forecast) %计算原始数据的 10 步预测值

```

## 9.2 季节性序列及其预报

在不少实际问题中, 时间序列有很明显的周期规律性, 例如气温、雨量、用电量等。由季节性因素或其它周期因素引起的周期性变化的时间序列, 我们称为季节性时间序列, 相应的模型为季节性模型。例如, 以电力负荷为例, 如果  $X_t$  表示时刻  $t$  的用电



负荷量，以 1 小时为采样间隔，显然  $X_t$  将包含 24 小时的周期性变化。一般，上午有一个用电高峰，晚上又有一个用电高峰；中午有一个用电较少的低谷，深夜则出现一天用电量最少的低谷。

一般地，对周期  $s$  的序列，可进行差分运算：

$$\nabla_s X_t = (1 - B^s) X_t$$

$$\nabla_s^D = (1 - B^s)^D X_t$$

对于季节性模型，可以用一类疏系数 ARIMA 模型去描述。设  $s$  为一正整数。一个时间序列  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  如果满足下列模型：

$$\Phi(B^s) \nabla_s^D X_t = \Theta(B^s) E_t \quad (151)$$

则称  $X_t$  是周期为  $s$  的季节性序列。其中

$$\begin{cases} \Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps} \\ \Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{qs} \end{cases}$$

式 (151) 中， $E_t$  一般不必是白噪声，而可设它是另一个 ARIMA( $p, d, q$ ) 序列

$$\varphi(B) \nabla^d E_t = \theta(B) \varepsilon_t$$

由式 (151)

$$\Phi(B^s) \nabla^d \nabla_s^D X_t = \Theta(B^s) \nabla^d E_t$$

$$\varphi(B) \Phi(B^s) \nabla^d \nabla_s^D X_t = \Theta(B^s) \varphi(B) \nabla^d E_t = \Theta(B^s) \theta(B) \varepsilon_t$$

令  $W_t = \nabla^d \nabla_s^D X_t$ ，得

$$\varphi(B) \Phi(B^s) W_t = \theta(B) \Theta(B^s) \varepsilon_t \quad (152)$$

式 (152) 称为乘积型季节性模型，其阶数常用  $(p, d, q) \times (P, D, Q)$  表示。式 (152)

中的  $W_t$  是由原来序列  $X_t$  经差分  $\nabla^d \nabla_s^D$  得到的，经过这样的运算，可以消除序列  $X_t$  的非平稳性（即趋势性）与季节性（周期性）。式 (152) 是一种疏系数模型，它反映了序列相邻时刻与相隔为周期  $s$  的时刻之间复杂变化的规律。例如，考察

$$W_t = (1 - \theta_1 B)(1 - \Theta_1 B^s) \varepsilon_t$$

展开成

$$W_t = (1 - \theta_1 B - \Theta_1 B^s + \theta_1 \Theta_1 B^{s+1}) \varepsilon_t$$

它是 MA( $s+1$ ) 模型，但是其系数  $\theta_1, \theta_2, \dots, \theta_{s+1}$  中有很多为 0，即

$$\theta_2 = \dots = \theta_{s-1} = 0, \quad \theta_s = \Theta_1, \quad \theta_{s+1} = -\theta_1 \Theta_1$$

对于这样的疏系数模型，对它进行相关分析，较易识别它的特性。

下面是几种重要的乘积型季节性模型：

1)  $(0, d, 1) \times (0, D, 1)_s$  型

$$W_t = (1 - \theta_1 B)(1 - \Theta_1 B^s) \varepsilon_t \quad (153)$$

2)  $(0, d, 1) \times (1, D, 1)_s$  型

$$(1 - \Phi_1 B^s) W_t = (1 - \theta_1 B)(1 - \Theta_1 B^s) \varepsilon_t \quad (154)$$

3)  $(0, d, 2) \times (0, D, 2)_s$  型

$$W_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^s - \Theta_2 B^{2s})\varepsilon_t \quad (155)$$

例 23 测得某地区一口井 7 年的地下水埋深数据如表 15 所示。试预报第 8 年全年的地下水埋深。

表 15 井水埋深数据

月份 年份	1	2	3	4	5	6	7	8	9	10	11	12
1	9.40	8.81	8.65	10.01	11.07	11.54	12.73	12.43	11.64	11.39	11.1	10.85
2	10.71	10.24	8.48	9.88	10.31	10.53	9.55	6.51	7.75	7.8	5.96	5.21
3	6.39	6.38	6.51	7.14	7.26	8.49	9.39	9.71	9.65	9.26	8.84	8.29
4	7.21	6.93	7.21	7.82	8.57	9.59	8.77	8.61	8.94	8.4	8.35	7.95
5	7.66	7.68	7.85	8.53	9.38	10.09	10.59	10.83	10.49	9.21	8.66	8.39
6	8.27	8.14	8.71	10.43	11.47	11.73	11.61	11.93	11.55	11.35	11.11	10.49
7	10.16	9.96	10.47	11.70	10.1	10.37	12.47	11.91	10.83	10.64	10.29	10.34

解 (1) 首先进行时间序列模型定阶。因为数据有下降趋势，又有 12 个月的季节性，故对数据作下列差分运算

$$W_t = \nabla \nabla_{12} X_t$$

对  $W_t$  进行稠密系数 ARMA 模型拟合。用选取的  $p, q$  的各种阶数形式进行试算，用 AIC 准则寻求最优模型。求得当  $p = 1, q = 13$  时模型最优。

计算的 Matlab 程序如下：

```

clc,clear
load water.txt; %把原始数据按照表中的格式存放在纯文本文件 water.txt
water=water';x=water(:)';
s=12; %周期 s=12
n=12; %预报数据的个数
m1=length(x); %原始数据的个数
for i=s+1:m1
    y(i-s)=x(i)-x(i-s);
end
m2=length(y); %周期差分后数据的个数
w=diff(y); %消除趋势性的差分运算
m3=length(w); %计算最终差分后数据的个数
for i=0:3
    for j=0:s+1
        spec= garchset('R',i,'M',j,'Display','off'); %指定模型的结构
        [coeffX,errorsX,LLFX] = garchfit(spec,w); %拟合参数
        num=garchcount(coeffX); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic,bic]=aicbic(LLFX,num,m3);
        fprintf('R=%d,M=%d,AIC=%f,BIC=%f\n',i,j,aic,bic); %显示计算结果
    end
end
save bdata x y w n m1 m2 s

```

(2) 建立模型并进行预测，算得第 8 年全年预报值如表 16 所示。

表 16 12 个月地下水埋深预报值

步数	1	2	3	4	5	6
预报值	9.57	9.0978	10.7288	11.4250	10.7821	11.9707
步数	7	8	9	10	11	12
预报值	12.8529	12.5094	12.2442	11.3361	11.0203	10.991

计算的 Matlab 程序如下:

```

clc,clear
load bdata
spec2= garchset('R',1,'M',13,'Display','off'); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec2,w); %拟合参数
[sigmaForecast,w_Forecast] = garchpred(coeffX,w,n) %求 w 的预报值
yhat=y(m2)+cumsum(w_Forecast) %求 y 的预报值
for j=1:n
    x(m1+j)=yhat(j)+x(m1+j-s);
end
x_hat=x(m1+1:end) %复原到原始数据的预报值

```

例 24 国际航线月度旅客总数 (1949.1~1960.12, 单位: 千人) 见表 17。

表 17 国际航线月度旅客总数

月份 年份	1	2	3	4	5	6	7	8	9	10	11	12
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	408	461	390	432

对所给时间序列建模, 并对时间序列进行 2 年 (24 个月) 的预报。

解 (1) 这一时间序列的图形见图 3。

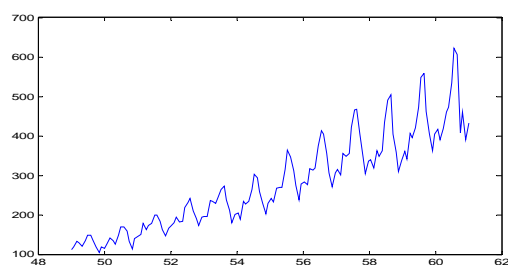


图 3 数据变化趋势

从直观上看,这一时间序列有上升的趋势,且有明显的 12 个月的季节性( $s=12$ ),而且波动的幅度越来越大。这个例子是 Box—Jenkins 建模的范例。该范例采用了数据变换的方法使数据平稳化。对  $X_t$  作对数变换

$$Y_t = \ln X_t$$

设  $Y_t$  为一个乘积型季节性序列。对  $Y_t$  再作差分

$$W_t = \nabla \nabla_{12} Y_t$$

用 AIC 准则对模型定阶,当  $p=0, q=13$  时 AIC 达到最小值-431.5504。由此可见,取模型

$$W_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_{13} B^{13}) \varepsilon_t$$

计算的 Matlab 程序如下:

```
clc,clear
load hang.txt; %把原始数据按照表中的格式存放在纯文本文件 hang.txt
hang=hang';x0=hang(:)';
m1=length(x0); %原始数据的个数
plot(linspace(49,61,m1),x0)
x=log(x0); %做对数变换 y=lnx
s=12; %周期 s=12
n=24; %预报数据的个数
for i=s+1:m1
    y(i-s)=x(i)-x(i-s);
end
m2=length(y); %周期差分后数据的个数
w=diff(y); %消除趋势性的差分运算
m3=length(w); %计算最终差分后数据的个数
for i=0:3
    for j=0:s+1
        spec=garchset('R',i,'M',j,'Display','off'); %指定模型的结构
        [coeffX,errorsX,LLFX]=garchfit(spec,w); %拟合参数
        num=garchcount(coeffX); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic,bic]=aicbic(LLFX,num,m3);
        fprintf('R=%d,M=%d,AIC=%f,BIC=%f\n',i,j,aic,bic); %显示计算结果
    end
end
end
save bdata x y w n m1 m2 s
```

(2) 建立模型并进行预测。根据上面确定的模型阶数,建立  $W_t$  的 MA(13) 模型,并预报  $Y_t$  的 24 个月的价值。

然而,我们要求的是  $X_t$  的 24 个月的预报值。因为  $Y_t = \ln X_t$ , 并假设  $Y_t$  是正态序列。在概率论中有这一事实: 设  $X$  是随机变量, 又  $\ln X$  服从正态分布  $N(\mu, \sigma^2)$ , 则  $X$  服从对数正态分布,  $X$  的均值是

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

这个例子中， $X_t$  是对数正态序列， $\hat{Y}_k(m)$  代表  $Y_t$  在时刻  $k(k=144)$  的  $m$  步预报， $\hat{\sigma}_k(m)$  代表预报标准差， $\hat{X}_k(m)$  代表  $X_t$  在时刻  $k(k=144)$  的  $m$  步预报值，则有

$$\hat{X}_k(m) = \exp(\hat{Y}_k(m) + \frac{1}{2} \hat{\sigma}_k^2(m))$$

由此算得  $X_t$  的 24 个月预报值如下：

466.884	410.313	458.429	495.865	508.518	580.256	685.96	636.911	495.594
504.109	425.486	476.18	513.984	451.762	504.6	546.089	560.091	639.183
755.713	701.763	546.123	555.574	468.981	524.921			

计算的 Matlab 程序如下：

```
clc,clear
load bdata
spec2= garchset('R',0,'M',13,'Display','off'); %指定模型的结构
[coeffX,errorsX,LLFX] = garchfit(spec2,w); %拟合参数
[sigmaForecast,w_Forecast] = garchpred(coeffX,w,n); %求 w 的预报值
yhat=y(m2)+cumsum(w_Forecast); %求 y 的预报值
for j=1:n
    x(m1+j)=yhat(j)+x(m1+j-s);
end
x_hat=x(m1+1:end); %复原到对数数据的预报值
x0_hat=exp(x_hat+sigmaForecast.^2/2) %恢复到原始数据的预报值
x0_hat_check=exp(x_hat)
```

## § 10 异方差性

### 10.1 简介

类似于线性回归分析，许多时间序列表现出异方差（非常数方差）结构。在一个线性回归模型中，假如响应变量  $Y$  具有如下形式的非常数方差结构

$$Y = X\beta + e, \quad \text{var}(e) = \sigma^2 \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

那么，考虑到  $e$  的非齐性，我们不是使用普通最小二乘法而是使用广义最小二乘（GLS）法。在时间序列领域，常常可以观察到在一段连续的时间周期内时间序列的变化相当小，而后的一段时间其变化则大得多，然后又莫名其妙地变小。假如这些变化（波动）能够合并到模型中，则是最惬意的事。

一个相关的例子是资产价格序列。如果一个对数资产时间序列（例如，对数股价、对数债券价格、对数汇率等等）被差分，则差分序列看起来像白噪声。换言之，如果  $\{Y_t\}$  表示在  $t$  期的一份主要资产的对数价格序列，那么我们通常最后会用 ARIMA(0,1,0) 模型来表示  $\{Y_t\}$ ：

$$\Delta Y_t = X_t = \varepsilon_t, \quad \varepsilon_t \sim N(0,1) \text{ i.i.d.} \quad (156)$$

（这里 i.i.d. 表示独立同分布）这个方程只是表述了主要资产的收益  $X_t = Y_t - Y_{t-1}$  的行为像高斯（Gaussian）白噪声序列。这和著名的随机游动假设是一致的。虽然作为一阶

逼近这似乎合情合理，然而，进一步的分析揭示出不能由方程（156）充分解释的别的类型的结构。这些特征的典型标志是：

- （1） $\{X_t\}$  是厚尾的，且比高斯白噪声厚得多。
- （2）虽然  $\{X_t\}$  的相关函数没能揭示出多少结构性的东西，但是序列  $\{X_t^2\}$  却是高度相关的。有时，这些相关总是非负的。
- （3） $\{X_t\}$  的变化显示出聚类的特点。那就是， $\{X_t\}$  的大变化往往跟随着大变化，而  $\{X_t\}$  的小变化则往往由小变化跟随。

这最后一点值得注意。对于金融市场来说，尽力理解方差或波动  $\{\sigma_t\}$  的变化是重要的。投资者要求较高的期望收益作为拥有较高风险的资产的一种补偿。进一步，方差随时间变化的时间序列当然牵涉到有关参数的统计推断的合理性和有效性。例如，著名的 Black-Scholes 期权定价公式就需要波动过程  $\{\sigma_t\}$  的知识。

正是部分地由于这些原因，带有异方差误差的时间序列模型得到了发展。在本节中，我们将简要地讨论受欢迎的 ARCH 和 GARCH 模型。也有别的选择可用来捕捉异方差效应，例如，随机波动模型。

## 10.2 ARCH

最早的异方差时间序列模型之一是 ARCH 模型。在它的最简单形式里，一个 ARCH 模型将收益序列  $X_t$  表示为

$$X_t = \sigma_t \varepsilon_t$$

其中通常假设  $\varepsilon_t \sim N(0,1)$  i.i.d. 并且  $\sigma_t$  满足

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad (157)$$

设  $\Omega_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$  表示由直到时刻  $t-1$  的过去信息的线性组合生成的空间。则

$$E(X_t^2 | \Omega_{t-1}) = E(\sigma_t^2 \varepsilon_t^2 | \Omega_{t-1}) = \sigma_t^2 E(\varepsilon_t^2 | \Omega_{t-1}) = \sigma_t^2$$

这个等式意味着  $X_t$  的条件方差，像  $AR(p)$  模型一样，根据  $X_t^2$  的过去值演变。因此，我们给予它  $p$ -阶自回归条件异方差 ARCH( $p$ ) 模型之名。显然，为了使方程（157）是一个明确定义的过程，我们需要把一些条件加在系数上。例如，为了确保  $\sigma_t^2 \geq 0$  以及  $X_t$  是明确定义的，一个充分条件是

$$\alpha_i \geq 0, \quad i = 0, \dots, p$$

并且

$$\alpha_1 + \dots + \alpha_p < 1$$

这里我们不去寻求此论断的证明了。感兴趣的可以从有关 ARCH 的标准参考书中找到有关的讨论。

例 25 设  $X_t = \sigma_t \varepsilon_t$  而  $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$ 。反复迭代这个递推式，我们有

$$\begin{aligned} X_t^2 &= \sigma_t^2 \varepsilon_t^2 = \varepsilon_t^2 (\alpha_0 + \alpha_1 X_{t-1}^2) = \alpha_0 \varepsilon_t^2 + \alpha_1 X_{t-1}^2 \varepsilon_t^2 = \alpha_0 \varepsilon_t^2 + \alpha_1 \varepsilon_t^2 (\sigma_{t-1}^2 \varepsilon_{t-1}^2) \\ &= \alpha_0 \varepsilon_t^2 + \alpha_1 \varepsilon_t^2 \varepsilon_{t-1}^2 (\alpha_0 + \alpha_1 X_{t-2}^2) \\ &= \alpha_0 \sum_{j=0}^n \alpha_1^j \varepsilon_t^2 \cdots \varepsilon_{t-j}^2 + \alpha_1^{n+1} \varepsilon_t^2 \varepsilon_{t-1}^2 \cdots \varepsilon_{t-n}^2 X_{t-n-1}^2 \end{aligned}$$

假如  $\alpha_1 < 1$ ，则当  $n$  趋近于无穷大时，以上表达式中的最后一项趋近于零。此时，

$$X_t^2 = \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_t^2 \cdots \varepsilon_{t-j}^2 \quad (158)$$

特别地， $EX_{t-1}^2 = \alpha_0 / (1 - \alpha_1)$ 。

另一方面，我们可以将 ARCH(1) 模型表示为

$$X_t^2 = \sigma_t^2 + X_t^2 - \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \sigma_t^2 (\varepsilon_t^2 - 1) = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t$$

其中， $v_t = \sigma_t^2 (\varepsilon_t^2 - 1)$ 。我们可以将一个 ARCH(1) 看成过程  $\{X_t^2\}$  的由新噪音  $\{v_t\}$  驱动的 AR(1) 模型。

### 10.3 GARCH

类似于从 AR 模型到 ARMA 模型的推广，人们可以将 ARCH 模型的概念推广到广义 ARCH (GARCH) 模型。确切地，一个 GARCH 模型可以用如下形式来表示

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j X_{t-j}^2 \quad (159)$$

为了使方程 (159) 是明确定义的，关于诸  $\alpha$  和  $\beta$  的条件必须提出来。然而，由于对一般的 GARCH 模型去寻求确切的条件很棘手，人们不得不求助于逐个地研究。考虑 GARCH 扩展的主要原因之一是借助于在方程 (159) 中允许过去的波动影响现在的波动，我们可以得到一个更简约的模型。我们不去探求完全一般形式的 GARCH 模型。相反，让我们来考虑如下起说明作用的 GARCH(1,1) 模型。

例 26 设  $X_t$  是 GARCH(1,1) 模型，满足

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

由于  $\sigma_t^2$  是不可观测的，一种方法是利用最初一段的数据，比如说前 50 个，去估计它。这种方法常常不能令人满意。而一种更好的方法是将它参数化，并且把它看成一个 ARMA 过程。确切地，考虑

$$\begin{aligned} X_t^2 &= \sigma_t^2 + (X_t^2 - \sigma_t^2) = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + X_t^2 - \sigma_t^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 - \beta_1 (X_{t-1}^2 - \sigma_{t-1}^2) + X_t^2 - \sigma_t^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 + v_t - \beta_1 v_{t-1} \end{aligned}$$

其中  $v_t = X_t^2 - \sigma_t^2 = \sigma_t^2 (\varepsilon_t^2 - 1)$ 。利用这个表达式，过程  $\{X_t^2\}$  可以看成由噪音  $v_t$  驱动的 ARMA(1,1) 过程。

在一般情况下，这一结论仍然正确。人们可以容易地推导出以下定理。

定理 1 设  $X_t$  是一个 GARCH( $p, q$ ) 过程，则就  $v_t = \sigma_t^2 (\varepsilon_t^2 - 1)$  而言， $X_t^2$  是一个 ARMA( $m, p$ ) 过程，其中  $m = \max\{p, q\}$ ， $\alpha_i = 0 (i > q)$  以及  $\beta_j = 0 (j > p)$ 。

### 10.4 ARCH 的估计与检验

根据 ARCH 的定义，显然  $X_t | \Omega_{t-1} \sim N(0, \sigma_t^2)$  具有条件概率密度函数

$$f(x_t | \Omega_{t-1}) = (2\pi\sigma_t^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_t^2} x_t^2\right)$$

其中  $\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2$ 。迭代这个条件密度，我们得

$$\begin{aligned} f(x_n, \dots, x_1 | x_0) &= f(x_n | x_{n-1}, \dots, x_0) \cdots f(x_2 | x_1, x_0) f(x_1 | x_0) \\ \log f(x_n, \dots, x_1 | x_0) &= \sum_{t=1}^n \log f(x_t | \Omega_{t-1}) \\ &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \left( \log \sigma_t^2 + \frac{x_t^2}{\sigma_t^2} \right) \end{aligned} \quad (160)$$

因此，对不同的  $(\alpha_0, \dots, \alpha_p)$  的值，将

$$\sigma_t^2(\alpha_0, \alpha_1, \dots, \alpha_p) = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2$$

代入以上方程中并最大化这个对数似然函数，我们可以得到最大似然估计的数值解。如果  $\varepsilon_t$  不是正态的，人们仍然可以使用这个方法去获得拟（准）最大似然估计（PMLE 或 QMLE）。

注意这个方法也适用于 GARCH 模型。回想一下在线性时间序列的情况下，由于 MA 部分的存在，ARMA 模型的估计通常比单纯的 AR 模型的估计更棘手。在估计 GARCH 模型的时候，会遭遇到类似的困难。为了说明这一点，考虑一个简单的 GARCH(1,1) 模型，其中

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (161)$$

依波动  $\sigma_t^2$  重写以上的方程得

$$\sigma_t^2 = (1 - \beta_1 B)^{-1} (\alpha_0 + \alpha_1 X_{t-1}^2) = (1 + \beta_1 B + \beta_1^2 B^2 + \cdots) (\alpha_0 + \alpha_1 X_{t-1}^2)$$

因此，波动  $\sigma_t^2$  依赖于所有的  $X_t$  的过去值。为了从这个方程估计出  $\sigma_t^2$ ，我们截断  $X_t$  和  $\sigma_t^2$  的值：对于  $t \leq 0$ ，令  $X_t = 0$ ， $\sigma_t^2 = 0$ 。于是，对于  $t = 1, 2, \dots$ ， $\sigma_t^2$  可由  $\tilde{\sigma}_t^2$  递归地逼近，这里

$$\tilde{\sigma}_t^2 = \alpha_0 + \alpha_1 \tilde{X}_{t-1}^2 + \beta_1 \tilde{\sigma}_{t-1}^2$$

其中，对于  $t \leq 0$ ， $\tilde{X}_t = 0$ ， $\tilde{\sigma}_t^2 = 0$ ；对于  $t > 0$ ， $\tilde{X}_t = X_t$ 。迭代这个方程，我们得

$$\begin{aligned} \tilde{\sigma}_1^2 &= \alpha_0 \\ \tilde{\sigma}_2^2 &= \alpha_0 + \alpha_1 \tilde{X}_1^2 + \beta_1 \tilde{\sigma}_1^2 = \alpha_0 + \alpha_1 X_1^2 + \beta_1 \alpha_0 \\ &\vdots \end{aligned}$$

对于各个  $\theta = (\alpha_0, \alpha_1, \beta_1)$  的值，将  $\tilde{\sigma}_t^2$  的这个表达式代入对数似然函数（160）中并将它最大化，我们可以求得最大似然估计的数值解。同样的思路适用于一般的 GARCH(p, q) 模型。

最后值得一提的是另一个受欢迎的估计方法：广义矩法或有效矩法。

就检验而言，有许多方法可用。下面，我们讨论三个简单的方法。

（1）时间序列检验。既然 ARCH(p) 意味着  $\{X_t^2\}$  遵从 AR(p)，为了识别 AR 的性质，我们可以利用 Box-Jenkins 法去研究  $X_t^2$  的相关性结构。

（2）残差的 Portmanteau 检验。设 Portmanteau 统计量定义为



$$Q = n(n+2) \sum_{j=1}^h r^2(j)/(n-j)$$

其中,  $r(j)$  表示  $X_t^2$  的拟合残差的 ACF (自相关函数)。如果模型是被正确设定的, 那么, 对于大的样本容量,  $Q \xrightarrow{L} \chi_{h-m}^2$ , 其中,  $m$  是模型中独立参数的个数, 而  $\xrightarrow{L}$  表示当样本容量  $n \rightarrow \infty$  时, 依分布收敛。

(3) Lagrange 乘数检验。

定理 2 设  $X_t$  是 ARCH( $p$ ) 模型,  $X_t^2$  关于  $X_{t-1}^2, \dots, X_{t-p}^2$  的拟合回归方程是

$$\hat{X}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^p \hat{\alpha}_i X_{t-i}^2 \quad (162)$$

令  $R^2$  表示来自 (162) 的判定系数。那么, 在零假设  $H_0: \alpha_1 = \dots = \alpha_p = 0$  之下,  $nR^2 \xrightarrow{L} \chi_p^2$ 。

#### 习题二十四

1. 我国 1974~1981 年布的产量如表 18 所示。

表 18 1974~1981 年布的产量

年份	1974	1975	1976	1977	1978	1979	1980	1981
产量(亿米)	80.8	94.0	88.4	101.5	110.3	121.5	134.7	142.7

(1) 试用趋势移动平均法 (取  $N = 3$ ), 建立布的年产量预测模型。

(2) 分别取  $\alpha = 0.3$ ,  $\alpha = 0.6$ ,  $S_0^{(1)} = S_0^{(2)} = \frac{y_1 + y_2 + y_3}{3} = 87.7$ , 建立布的直线指数平滑预测模型。

线指数平滑预测模型。

(3) 计算模型拟合误差, 比较 3 个模型的优劣。

(4) 用最优的模型预测 1982 年和 1985 年布的产量。

2. 1960 年~1982 年全国社会商品零售额如表 19 所示 (单位: 亿元)。

表 19 全国社会商品零售额数据

年份	1960	1961	1962	1963	1964	1965	1966	1967
零售总额	696.9	607.7	604	604.5	638.2	670.3	732.8	770.5
年份	1968	1969	1970	1971	1972	1973	1974	1975
零售总额	737.3	801.5	858	929.2	1023.3	1106.7	1163.6	1271.1
年份	1976	1977	1978	1979	1980	1981	1982	
零售总额	1339.4	1432.8	1558.6	1800	2140	2350	2570	

试用三次指数平滑法预测 1983 年和 1985 年全国社会商品零售额。

3. 某地区粮食产量 (亿千克), 从 1969~1983 年依次为: 3.78, 4.19, 4.83, 5.46, 6.71, 7.99, 8.60, 9.24, 9.67, 9.87, 10.49, 10.92, 10.93, 12.39, 12.59, 试选用 2~3 种适当的曲线预测模型, 预测 1985 年和 1990 年的粮食产量。

4. 随机模拟下列序列, 样本容量 10000。

$$X_t + 0.5X_{t-1} - 0.3X_{t-2} = \varepsilon_t, \text{ 其中 } \varepsilon_t \sim N(0,1)$$

利用模拟数据研究上述序列自相关特性。

5. 1952 年~1997 年我国人均国内生产总值（单位：元）数据如表 20 所示。

表 20 1952 年~1997 年我国人均国内生产总值

年代	人均生产总值	年代	人均生产总值	年代	人均生产总值
1952	119	1968	222	1984	682
1953	142	1969	243	1985	853
1954	144	1970	275	1986	956
1955	150	1971	288	1987	1104
1956	165	1972	292	1988	1355
1957	168	1973	309	1989	1512
1958	200	1974	310	1990	1634
1959	216	1975	327	1991	1879
1960	218	1976	316	1992	2287
1961	185	1977	339	1993	2939
1962	173	1978	379	1994	3923
1963	181	1979	417	1995	4854
1964	208	1980	460	1996	5576
1965	240	1981	489	1997	6079
1966	254	1982	525		
1967	235	1983	580		

(1) 用 ARIMA(2,1,1) 模型拟合，求模型参数的估计值。

(2) 求数据的 10 步预报值。

6. 某地区山猫的数量在前连续 114 年的统计数据如表 21 所示。分析该数据，得出山猫的生长规律，并预测以后两个年度山猫的数量。

表 21 山猫数据

269	321	585	871	1475	2821	3928	5943	4950	2577	523	98
184	279	409	2285	2685	3409	1824	409	151	45	68	213
546	1033	2129	2536	957	361	377	225	360	731	1638	2725
2871	2119	684	299	236	245	552	1623	3311	6721	4254	687
255	473	358	784	1594	1676	2251	1426	756	299	201	229
469	736	2042	2811	4431	2511	389	73	39	49	59	188
377	1292	4031	3495	537	105	153	387	758	1307	3465	6991
6313	3794	1836	345	382	808	1388	2713	3800	309	2985	3790
674	71	80	108	229	399	1132	2432	3575	2935	1537	529
485	662	1000	1520	2657	3396						

7. 1946 年~1970 年美国各季耐用品支出资料如表 22 所示。

(1) 对所给时间序列建模；

(2) 对时间序列进行两年（8 个季度）的预报。

表 22 1946 年~1970 年美国各季耐用品支出资料

年度	一季	二季	三季	四季
1946	7.5	8.9	11.1	13.4
1947	15.5	15.7	15.6	16.7
1948	18.0	17.4	17.9	18.8
1949	17.6	17.0	16.1	15.7
1950	15.9	17.9	20.3	20.4
1951	20.2	20.5	20.9	20.9
1952	21.1	21.4	18.2	20.1
1953	21.4	21.3	21.9	21.3

1954	20.4	20.4	20.7	20.7
1955	20.9	23.0	24.9	26.5
1956	25.6	26.1	27.0	27.2
1957	28.1	28.0	29.1	28.3
1958	25.7	24.5	24.4	25.5
1959	27.0	28.7	29.1	29.0
1960	29.6	31.2	30.6	29.8
1961	27.6	27.7	29.0	30.3
1962	31.0	32.1	33.5	33.2
1963	33.2	33.8	35.5	36.8
1964	37.9	39.0	41.0	41.6
1965	43.7	44.4	46.6	48.3
1966	50.2	52.1	54.0	56.0
1967	53.9	55.6	55.4	56.2
1968	57.9	57.3	58.8	60.4
1969	63.1	83.5	64.8	65.7
1970	64.8	65.6	67.2	62.1