

第 14 讲 时间序列分析

数据分为横剖面数据和纵剖面数据两类（或者叫做静态数据和动态数据）。横剖面数据是由若干相关现象在某一时点上所处的状态组成的，它反映一定时间、地点等客观条件下诸相关现象之间存在的内在数值联系。研究这种数据结构的统计方法是多元统计分析。纵剖面数据是由某一现象或若干现象在不同时刻上的状态所形成的数据，它反映的是现象以及现象之间关系的发展变化规律性。研究这种数据的统计方法就是时间序列分析。

14.1 时间序列分析简介

14.1.1 时间序列的定义

从统计意义上讲，所谓时间序列就是将某一个指标在不同时间上的不同数值，按照时间的先后顺序排列而成的数列。这种数列由于受到各种偶然因素的影响，往往表现出某种随机性，彼此之间存在着统计上的依赖关系。

从数学意义上讲，如果我们对某一过程中的某一个变量或一组变量 $X(t)$ 进行观察测量，在一系列时刻 t_1, t_2, \dots, t_N （ t 为自变量，且 $t_1 < t_2 < \dots < t_N$ ）得到的离散有序数集合

$$X_{t_1}, X_{t_2}, \dots, X_{t_i}, \dots, X_{t_N}$$

称为离散数字时间序列。

值得注意的是，时间顺序中的“时间”也可以具有不同的物理意义，例如长度、温度、速度等等。

时间序列的主要分类有如下几种：

(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。

我们主要研究离散时间序列，并用 X_t 表示。

(3) 按序列的统计特性分，有平稳时间序列和非平稳时间序列两类。如果一个时间序列的概率分布与时间 t 无关，则称该序列为严格的（狭义的）平稳时间序列。如果序列的一、二阶矩存在，而且对任意时刻 t 满足：

i) 均值为常数

ii) 协方差为时间间隔 τ 的函数

则称该序列为宽平稳时间序列，也叫广义平稳时间序列。我们以后所研究的时间序列主要是宽平稳时间序列。如果不明确提出严平稳，所谓的平稳即指宽平稳。

(4) 按序列的分布规律来分，有高斯型(Gaussian)时间序列和非高斯型(non-Gaussian)时间序列。

14.1.2 时间序列基本样式

时间序列的数据有各种样式，根据数据的特征，可将其分为两大类：一类是平稳时间序列，基本样式是白噪声时间序列；另一类为非平稳时间序列，根据其具体的数据特征，其基本样式包括趋势性时间序列、季节性、条件异方差和异常观测值等。对具有不同数据特征的时间序列我们需要选用不同的分析方法。

下面采用图形法和回归分析两种分析工具来说明时间序列的基本样式。图形法，是将数据转化为时序图，用图形说明时间序列基本样式的数据特征。用横轴表示时间，纵轴表示时间序列的观测值，在 $t-X_t$ 平面直角坐标系中画出序列图。回归分析，是用回归模型说明数据特征。

1. 平稳时间序列

当一个时间序列满足下面 2 个条件

$$E(X_t) = \mu, \quad t = 1, 2, \dots, T, \quad (1)$$

$$E[(X_t - \mu)(X_{t-k} - \mu)] = \gamma_k, \quad t = 1, 2, \dots, T \text{ 且 } k = \dots, -2, -1, 0, 1, 2, \dots, \quad (2)$$

其中 μ, γ_k 为常数，我们称该序列是平稳的。

白噪声时间序列是一种典型的平稳时间序列。如果时刻 t 以前的信息不能对时刻 t 的值 X_t 提供任何信息，并且如果对 X_{n+h} 的最好预测值或期望都等于 0，那么这样的时间序列就是白噪声时间序列。白噪声的正式定义为

$$E(a_t) = 0, \quad t = 1, 2, \dots, n, \quad (3)$$

$$E(a_t^2) = \sigma^2, \quad t = 1, 2, \dots, n, \quad (4)$$

$$E(a_t a_s) = 0, \quad t, s = 1, 2, \dots, n \text{ 且 } t \neq s. \quad (5)$$

在本书的讨论中，一般考虑的是均值为 0 的 Gaussian 白噪声过程。图 14.1 就是一个均值为 0，方差为 1 的 Gaussian 白噪声序列（样本容量为 200）。

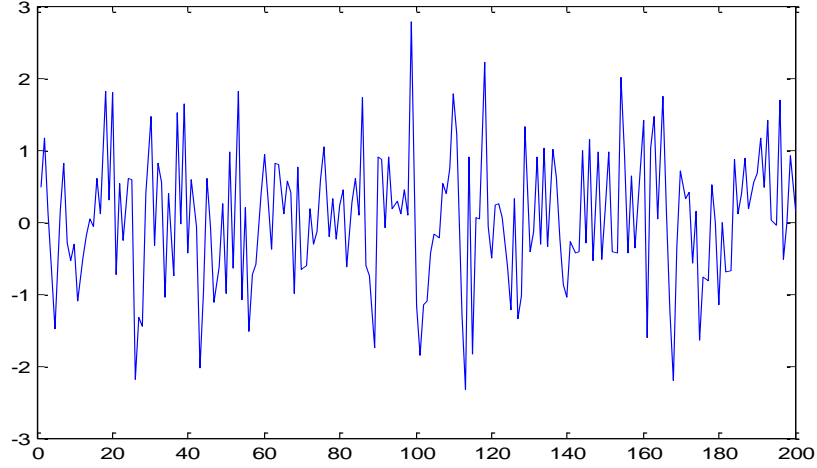


图 14.1 白噪声序列 ($E(a_t) = 0$, $E(a_t^2) = 1$)

画图的 MATLAB 程序如下：

```
clc, clear
a=normrnd(0,1,200,1); t=[1:200]';
plot(t,a)
```

2. 趋势性时间序列

在经济领域我们碰到的大部分时间序列都是非平稳的，这些非平稳序列常具有某种趋势，表现为沿时间有向上、向下的趋势，趋势有可能比较陡峭，也有可能比较平缓，有时候呈指数趋势，有时候呈线性趋势。

非平稳的时间序列有多种样式，分析序列的协方差过程，常常会发现它们具有时变的均值 μ_t 或时变的二阶矩如时变方差 σ_t^2 ，或者是同时具有时变均值和时变方差两种数据特征。因此，可根据序列的一阶矩和二阶矩将非平稳时间序列分为均值非平稳和方差非平稳两大类。

根据非平稳过程的均值函数是否可由一个特定的时间趋势来表示，又将均值非平稳序列分为确定性趋势和随机趋势。

(1) 确定性趋势时间序列

当一个时间序列为均值非平稳，且均值可由一个特定的时间趋势表示时，称该序列为确定性趋势时间序列。可以用一个一般的回归模型来描述

$$y_t = f(t) + a_t, \quad t = 1, 2, \dots, \quad (6)$$

其中， $f(t)$ 表示 y_t 中随时间变化的均值， a_t 是 y_t 剔除趋势性或周期性后的部分，通常为平稳的过程，可用平稳序列模型来描述。根据 $f(t)$ 的形式，确定性趋势时间序列又可进一步分为线性趋势序列、指数趋势序列、多项式函数序列、正弦函数序列等多种样式。下面介绍两种常用的确定性趋势序列：线性趋势和指数趋势。

i) 线性趋势时间序列

对具有线性趋势性的时间序列，通常的办法是通过将序列 y_t 写成

$$y_t = \alpha_0 + \alpha_1 t + a_t, \quad t = 1, 2, \dots. \quad (7)$$

取 $\alpha_0 = 20$ ， $\alpha_1 = 0.95$ 即

$$y_t = 20 + 0.95t + a_t, \quad t = 1, 2, \dots, 200, \quad (8)$$

其中， a_t 为均值为 0，方差为 1 的正态白噪声序列。得到的一个典型的线性趋势如图 14.2 所

示。

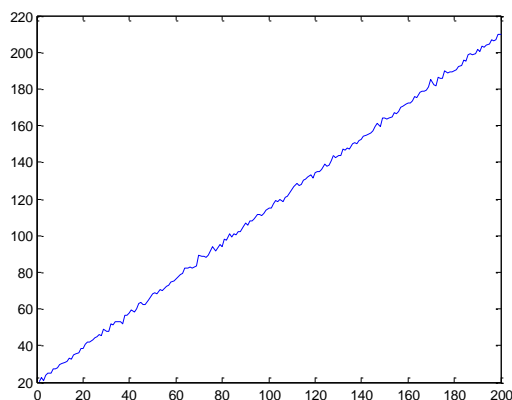


图 14.2 线性趋势图

画图的 MATLAB 程序为

```
clc, clear
at=normrnd(0,1,200,1); t=[1:200]';
yt=20+0.95*t+at;
plot(t,yt)
```

澳大利亚 1948—2005 年人口数见表 14.1。图 14.3 画出了澳大利亚 1948—2005 年的人口数。这个序列呈明显的线性趋势，反映出人口随时间不断增加的事实。

我们用线性时间趋势模型拟合该模型，结果如下

$$y_t = 7.5274 + 0.2173t, \quad t = 1, 2, \dots, 58, \quad (9)$$

这里 $t = 1, 2, \dots, 58$ 分别对应于 1948 年, 1949 年, \dots , 2005 年。

表 14.1 澳大利亚 1948—2005 年人口数（单位：百万）

1948	7.7	1968	12.278	1988	16.394
1949	7.9	1969	12.512	1989	16.638
1950	8.219	1970	12.728	1990	16.873
1951	8.414	1971	12.925	1991	17.097
1952	8.61	1972	13.105	1992	17.31
1953	8.805	1973	13.275	1993	17.518
1954	9.002	1974	13.445	1994	17.726
1955	9.201	1975	13.625	1995	17.941
1956	9.403	1976	13.816	1996	18.162
1957	9.611	1977	14.017	1997	18.388
1958	9.825	1978	14.224	1998	18.618
1959	10.046	1979	14.433	1999	18.846
1960	10.277	1980	14.638	2000	19.072
1961	10.516	1981	14.84	2001	19.293
1962	10.763	1982	15.039	2002	19.512
1963	11.016	1983	15.241	2003	19.728
1964	11.271	1984	15.449	2004	19.942
1965	11.525	1985	15.669	2005	20.14
1966	11.779	1986	15.902		
1967	12.032	1987	16.145		

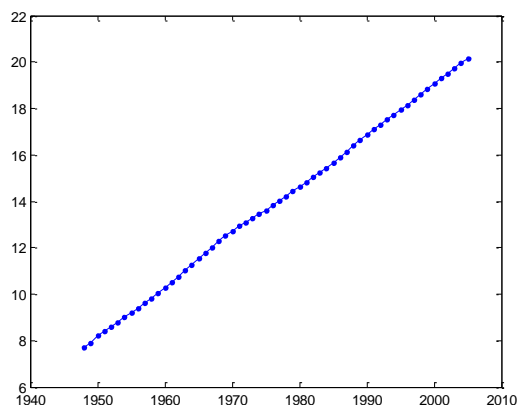


图 14.3 1984—2004 年间澳大利亚每年的人口数（单位：百万人）

画图和计算的 MATLAB 程序如下

```
clc, clear
a=textread('data13.txt');
a=a(:,[2:2:end]); a=nonzeros(a); %提取人口数据并转化为列向量
n=length(a); t1=[1948:2005]'; t2=[1:n]';
plot(t1,a,'-')
xishu=[ones(n,1),t2];
cs=xishu\ a %拟合参数
```

有时趋势的方向不时发生变化，表 14.2 是 1983—2004 年澳大利亚失业率年度数据，图 14.4 描述的是 1983—2004 年澳大利亚失业率年度数据。从 1984 年到 1989 年，失业率存在一个下降的趋势，从 1989 年到 1994 年，趋势由下降趋势变为上升趋势，1994 年以后，又变为下降趋势。通过对该时间序列的观测，我们可看到其趋势是在不断变化的。

表 14.2 1983—2004 年澳大利亚失业率年度数据

1983	9.958	1991	9.292	1999	6.967
1984	8.983	1992	10.492	2000	6.317
1985	8.275	1993	10.6	2001	6.733
1986	7.933	1994	9.458	2002	6.317
1987	7.9	1995	8.217	2003	5.933
1988	7.008	1996	8.233	2004	5.533
1989	5.958	1997	8.258		
1990	6.675	1998	7.758		

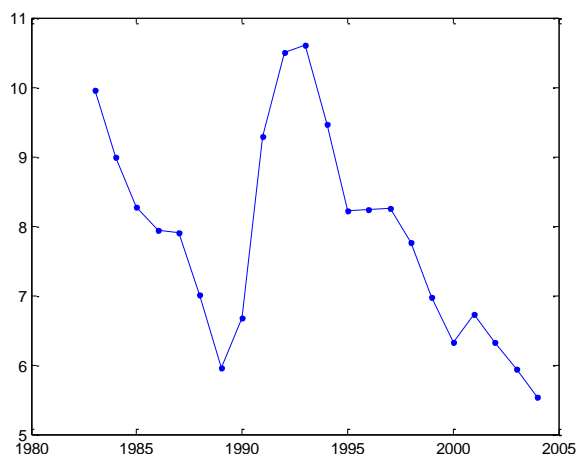


图 14.4 1983—2004 年澳大利亚失业率年度数据

ii) 指数趋势时间序列

当一个时间序列从一个时期到另一个时期的平均增长率为恒定时，它就服从指数趋势。很多经济方面的时间序列表现出非线性的趋势性，对这类时间序列数据用指数趋势能更好地逼近。

在实践中，时间序列中的指数趋势可以通过建立有线性趋势的自然对数模型得到

$$\log(y_t) = \beta_0 + \beta_1 t + a_t, \quad t = 1, 2, \dots \quad (10)$$

将两边指数化后， y_t 便表现出指数趋势： $y_t = \exp(\beta_0 + \beta_1 t + a_t)$ 。这是在线性回归模型中表示指数趋势时间序列的最方便的办法。

表 14.3 给出了中国 1978—2005 年每年的 GDP 指数的数据，图 14.5 给出了中国 1978—2005 年每年的 GDP 指数的曲线，从时序图我们可以初步得出该序列具有明显的指数趋势。

对该序列拟合指数趋势模型，得到下面的回归方程

$$y_t = \exp(4.478 + 0.0939t), \quad t = 1, 2, \dots, 28, \quad (11)$$

其中 $t = 1, 2, \dots, 28$ 分别对应于 1978 年，1979 年， \dots ，2005 年。

表 14.3 中国 1978—2005 年 GDP 指数

1978	100	1988	260.7	1998	651.2
1979	107.6	1989	271.3	1999	700.9
1980	116	1990	281.7	2000	759.9
1981	122.1	1991	307.6	2001	823
1982	133.1	1992	351.4	2002	897.8
1983	147.6	1993	400.4	2003	987.8
1984	170	1994	452.8	2004	1087.4
1985	192.9	1995	502.3	2005	1198.7
1986	210	1996	552.6		
1987	234.3	1997	603.9		

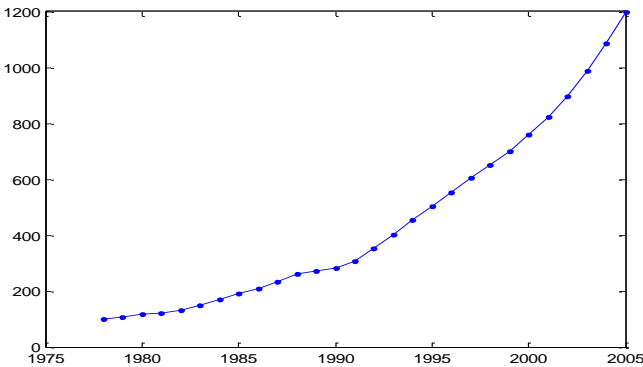


图 14.5 中国 1978—2005 年 GDP 指数

(2) 随机趋势时间序列

当我们可以用差分的方式去掉时间序列的趋势性时，就称序列具有随机趋势。随机趋势序列可以用下面的模型来表示，该模型又称带漂移的随机游走模型

$$y_t = \mu + y_{t-1} + a_t, \quad t = 1, 2, \dots, T, \quad (12)$$

我们构造一个随机趋势模型

$$y_t = 0.3 + y_{t-1} + a_t, \quad t = 1, 2, \dots, 200, \quad (13)$$

其中， a_t 为均值为 0，方差为 1 的正态白噪声序列。得到的一个典型的随机趋势序列，图形见图 14.6 所示。

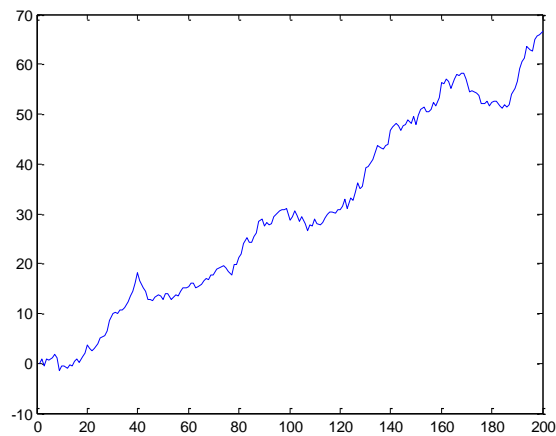


图 14.6 随机趋势模型 ($y_1 = 0$)

画图的 MATLAB 程序如下

```
clc, clear
y(1)=0; at=normrnd(0,1,1,200);
for i=2:200
    y(i)=0.3+y(i-1)+at(i);
end
t=1:200; plot(t,y)
```

我们通过构造一个随机趋势与一个确定性趋势来说明这二者的区别。序列由下面的公式生成

$$\begin{aligned} y_t &= 0.3 + y_{t-1} + a_t, \quad t = 1, 2, \dots, 200, \quad y_1 = 0, \\ x_t &= 0.3t + a_t, \quad t = 1, 2, \dots, 200, \end{aligned} \quad (14)$$

其中, a_t 为均值为 0, 方差为 1 的正态白噪声序列。

显然, 由图 14.7 我们可以明显观测到, 这两个序列都有明显的上升趋势, 但具有随机趋势特征的序列在较长的时间里表现出对这一趋势的偏离。

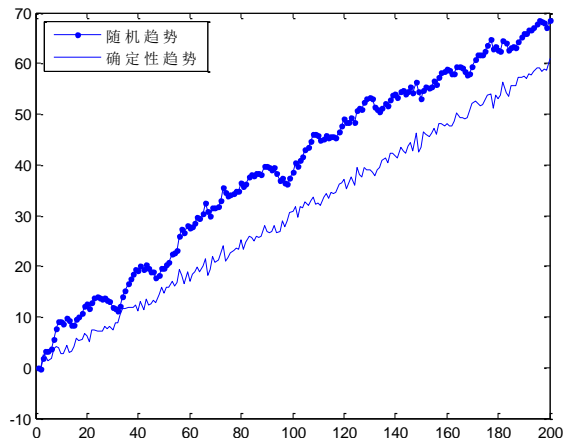


图 14.7 确定性趋势与随机趋势

画图的 MATLAB 程序如下

```
clc, clear
y(1)=0; at=normrnd(0,1,1,200);
for i=2:200
    y(i)=0.3+y(i-1)+at(i);
end
t=1:200; plot(t,y,'-')
hold on
xt=0.3*t+at;
```

```
plot(t,xt)
legend('随机趋势','确定性趋势',2)
```

表 14.4 给出了美国（1991.1—2006.4）工业总产出的季度数据。时序图 14.8 表明该序列存在一个明显的趋势。首先，我们考虑一个线性趋势模型（见图 14.9），模型的拟合结果表明拟合值与实际值偏差较大。

表 14.4 美国（1991.1—2006.4）工业总产出的季度数据（逐列排列）

5.111	5.404	15.929	29.342	52.82	70.74
5.15	5.757	15.763	29.538	54.909	71.389
5.718	6.06	15.254	30.037	55.394	71.583
5.581	5.463	14.774	30.449	56.424	72.016
6.1	5.62	14.803	31.163	56.701	73.138
5.816	6.413	14.735	31.663	56.633	74.14
5.708	6.344	15.508	32.142	56.432	75.449
4.954	6.618	16.81	33.356	56.636	76.388
4.249	7.147	18.161	34.13	56.868	77.93
4.21	7.049	18.475	34.844	54.466	79.066
4.308	7.637	18.847	35.549	53.602	79.266
4.592	8.058	18.906	36.568	55.655	80.006
4.954	8.557	18.455	37.331	55.786	80.698
5.287	8.929	18.582	37.909	55.996	81.099
5.571	9.086	19.013	38.271	56.59	82.757
6.286	8.939	18.739	38.06	55.345	83.808
6.472	7.333	19.287	37.905	54.26	85.062
6.824	6.452	20.668	38.166	53.586	86.769
6.628	6.237	21.079	39.187	52.779	87.948
6.413	6.883	21.373	39.747	51.771	89.856
6.501	7.539	21.392	40.282	52.332	92.26
5.992	7.744	20.384	40.595	53.54	93.318
5.845	7.793	19.738	41.211	55.423	94.074
6.334	8.381	19.659	42.013	56.888	94.955
6.697	9.35	19.757	42.196	58.584	96.194
6.716	9.046	20.306	42.703	59.511	97.351
6.736	9.262	21.402	42.435	59.936	98.317
7.03	9.771	22.332	41.407	60.018	99.374
7.049	10.349	22.675	41.179	60.198	101.349
7.098	11.259	23.213	41.032	60.297	102.66
7.265	12.013	23.301	40.145	60.226	104.139
7.343	12.493	23.204	40.919	60.625	104.01
7.353	12.767	23.076	41.29	60.982	103.584
7.284	13.393	23.918	41.417	60.62	102.076
7.157	13.217	24.143	42.377	60.871	100.656
6.942	13.952	23.84	44.149	61.56	99.217
7.167	15.126	23.928	45.003	62.382	97.923
7.274	15.939	22.92	45.614	63.474	98.582
7.549	16.409	21.627	47.2	64.598	100.119
7.93	17.27	21.265	48.529	66.144	100.698
8.185	18.024	22.391	48.939	66.696	100.602
8.459	18.24	23.399	49.384	67.258	101.174
8.586	18.171	24.427	50.104	67.603	100.345
7.979	18.181	25.612	49.656	68.162	100.957
7.549	18.122	24.564	49.708	68.415	101.771
7.255	17.78	24.682	49.499	68.108	102.588
6.618	16.918	26.17	47.552	67.692	103.226

6.168	14.578	25.602	44.398	67.997	103.69
6.051	12.963	25.181	43.8	68.494	104.79
6.06	12.424	24.574	44.919	68.967	105.976
5.62	12.747	24.202	45.896	69.205	106.698
5.12	13.961	25.162	47.281	68.165	106.911
4.837	14.656	25.964	47.887	66.835	108.15
4.288	15.009	26.875	48.511	67.286	109.485
4.19	15.019	27.276	49.435	68.224	111.226
4.523	15.048	27.541	50.442	68.398	112.333
4.308	15.459	27.834	51.98	68.339	112.178
5.189	15.616	28.069	52.609	69.519	
6.227	15.782	28.588	52.986	70.046	

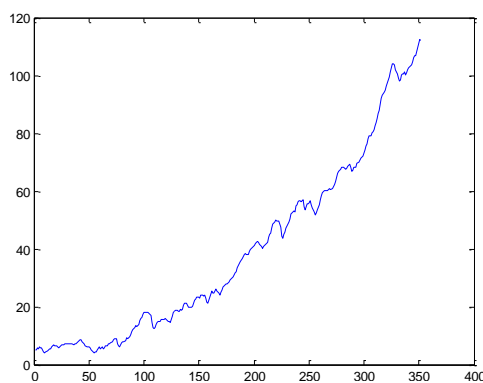


图 14.8 美国（1991.1—2006.4）工业总产出的季度数据

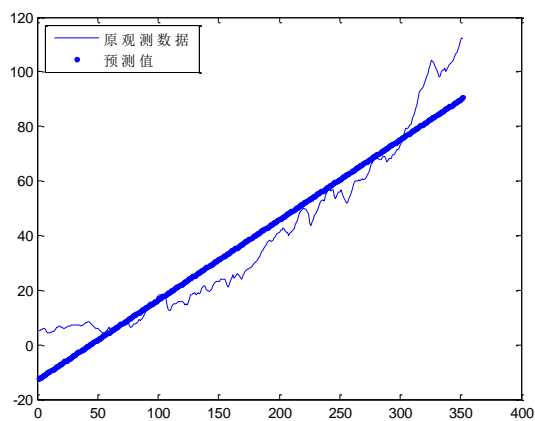


图 14.9 线性趋势模型

线性趋势模型为

$$y_t = -12.7826 + 0.2936t. \quad (15)$$

我们再尝试用指数趋势模型建模（见图 14.10），求得

$$\ln y_t = 1.5872 + 0.0095t.$$

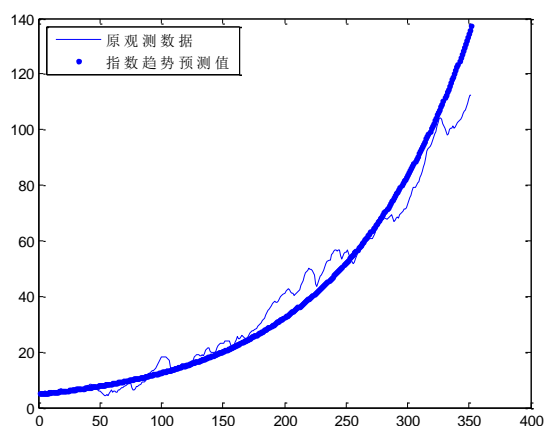


图 14.10 指数趋势模型

用随机趋势模型对该序列拟合，得到模型

$$y_t = 0.0731 + 1.006y_{t-1}.$$

模型很好地提取了序列的随机趋势（见图 14.11）。

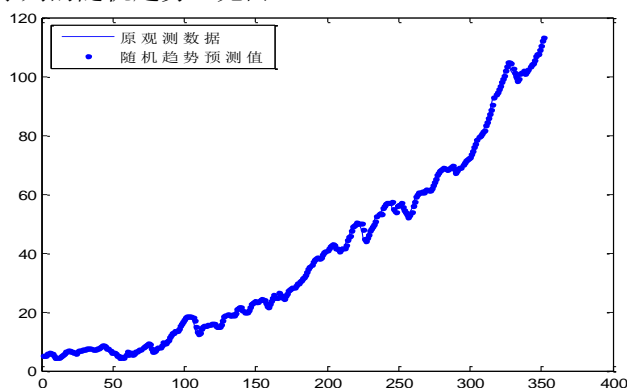


图 14.11 随机趋势模型

计算和画图的 MATLAB 程序如下

```
clc, clear
a=textread('data16.txt');
a=nonzeros(a); n=length(a);
plot(a)
t=[1:n]'; %构造时间的取值
xishu=[ones(n,1),t]'; %构造系数矩阵
cs1=xishu\log(a) %拟合线性趋势模型的参数
figure, plot(a)
hold on
yhat=xishu*cs1; plot(yhat,'r');
legend('原观测数据','预测值',2)
cs2=xishu\log(a) %拟合指数趋势模型的参数
figure, plot(a), hold on
yhat2=exp(xishu*cs2); plot(yhat2,'r')
legend('原观测数据','指数趋势预测值',2)
xishu2=[ones(n-1,1),a(1:end-1)]; %构造系数矩阵
cs3=xishu2\log(a(2:end));
yhat3(1)=a(1);
for i=2:n
    yhat3(i)=cs3(1)+cs3(2)*a(i-1);
end
figure, plot(a), hold on
plot(yhat3,'r')
```

legend('原观测数据','随机趋势预测值',2)

通过上面对非平稳时间序列趋势性的数据特征的讨论我们可以看出,时间序列的趋势性是非常复杂的,具有多种形式。我们需要针对趋势的不同数据特征来选择不同的类型。因此时间序列趋势性的判断和检验是建立时序模型中非常重要的一步,直接影响到模型的适应与否和预测结果。

3. 季节性时间序列

对一个给定的时间序列 y_t , 当观测值呈现出周期性变化规律时, 称序列为季节性时间序列。在实际问题中, 许多时间序列数据是按月或季(甚至每周或每天)观测的, 这样的序列就有可能呈现出显著的季节性。季节性的表现形式一般有两种: 一种是, 在某个季节的观测值具有与其它季节的观测值显著不同的特征, 如芬兰工业总产出月度指数(见图 14.12); 另一种是时间序列的方差呈现周期性的波动, 如股市的收益率数据。

表 14.5 芬兰 1975. 1—2005. 12 工业总产出月度指数(逐列排列)

38.352	40.145	30.432	51.965	57.248	61.325
39.547	40.645	44.82	52.749	55.827	38.469
39.596	38.312	47.564	52.426	54.445	56.199
39.822	38.538	48.349	30.883	54.396	62.07
37.763	40.096	49.309	48.3	52.7	63.393
34.049	40.008	48.3	51.367	55.876	64.658
20.67	39.91	46.829	51.505	54.445	61.325
34.598	40.557	48.3	53.524	54.817	59.688
35.833	37.577	48.113	50.044	36.617	61.482
36.842	19.661	49.809	50.867	52.494	61.482
39.224	37.44	51.328	50.632	56.297	61.649
37.028	41.194	46.379	51.328	58.365	64.079
36.793	42.017	30.06	53.71	59.267	63.236
37.489	42.987	46.878	54.161	57.15	39.635
38.077	42.066	49.221	53.798	54.925	59.051
37.901	42.439	50.319	33.686	56.405	62.913
39.224	43.673	49.721	49.221	57.777	63.393
35.744	44.585	48.799	52.612	58.15	64.765
19.151	44.908	48.986	54.259	59.581	61.012
35.382	44.997	49.035	56.278	58.042	61.433
38.998	42.664	49.358	55.817	38.518	58.277
39.498	25.071	50.181	53.024	55.503	63.305
39.596	41.468	50.407	53.494	59.002	59.336
39.596	44.271	49.77	54.876	59.796	64.53
38.724	45.232	29.697	55.827	59.953	59.355
40.047	47.016	46.143	56.356	58.787	40.312
35.558	45.095	48.986	55.239	58.306	60.423
36.019	45.918	48.799	34.019	59.532	62.07
38.401	46.143	50.132	50.956	58.042	64.834
37.166	47.427	49.123	54.553	60.885	63.217
19.151	48.3	47.937	55.827	61.482	53.308
35.744	47.339	49.035			
39.91	47.564	49.221			

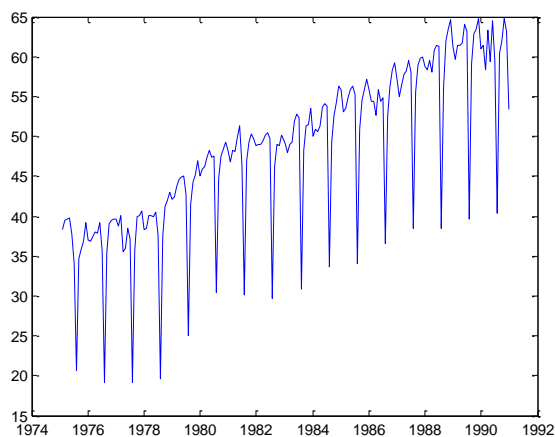


图 14.12 芬兰工业总产出月度指数（1975.1—1990.12）

画图的 MATLAB 程序如下

```
clc, clear
a=textread('data17.txt'); a=nonzeros(a);
t=[1975+1/12:1/12:1991]';
plot(t,a)
```

我们以 S 来表示季节的数目。如果以年作为考虑的基准，对于季度数据 S 就等于 4，月度数据 S 为 12。如果有 N 年的观测 y_t ，则总的观测数就等于 SN 。

一般情况下，数据的季节性可从时序图观测到。但在某些情况下，仅通过简单的图形还不能获得足够的信息来捕获数据的季节性特征，这时可借助于回归方程来帮助探索数据特征。一般来讲，我们可以在模型中包括一组季节性虚拟变量来解释因变量或自变量中的季节性。办法很简单，假设我们有月度数据，而且认为每一年中的季节变化大致相同。

$$y_t - y_{t-1} = \mu_1 D_{1t} + \mu_2 D_{2t} + \cdots + \mu_S D_{St} + \varepsilon_t, \quad (16)$$

这里 D_{st} 是季节哑变量，且有以下定义

$$D_{st} = 1, \quad s = 1, 2, \dots, S, \quad t = (T-1)S + s, \quad T = 1, 2, \dots, N,$$

$$D_{st} = 0, \quad \text{其它.}$$

许多时间序列的季节性随着时间在改变。有时这些变化发生的很慢，有时这些改变发生的很突然，就像图 14.13 中国 1993 年 1 月到 2007 年 5 月的社会商品零售总额数据。该时间序列数据不仅有一个向上的移动趋势，而且季节性发生了改变。

像中国社会消费品零售总额这样既有季节性又有趋势性的时间序列，要用既包含了时间趋势虚拟变量又包括了季节性虚拟变量的回归模型进行估计。

表 14.6 中国社会消费品零售总额（逐列排列）

977.5	1639.7	2454.9	2572	3422.4	5488.3
892.5	1623.6	2881.7	2637	3661.9	5105.4
942.3	1637.1	2549.5	2645	3733.1	5085.3
941.3	1756	2306.4	2597	4404.4	5407.7
962.2	1818	2279.7	2636	4087.1	5318.5
1005.7	1935.2	2252.7	2854	3856.6	5286.8
963.8	2389.5	2265.2	3029	3645.2	5340.4
959.8	1909.1	2326	3108	3535.2	5753.6
1023.3	1911.2	2286.1	3680	3536.6	6119.6
1051.1	1860.1	2314.6	3332.8	3732	5978.7
1102	1854.8	2443.1	3047.1	3718.4	6542
1415.5	1898.3	2536	2876.1	3769	6641.6
1192.2	1966	2652.2	2820.9	4139	6001.9
1162.7	1888.7	3131.4	2929.6	4365.4	5796.7

1167.5	1916.4	2662.1	2908.7	4358.6	5774.6
1170.4	2083.5	2538.4	2851.4	4859	6175.6
1213.7	2148.3	2403.1	2889.4	4569.4	6057.8
1281.1	2290.1	2356.8	3136.9	4211.4	6012.2
1251.5	2848.6	2364	3347.3	4049.8	6077.4
1286	2288.5	2428.8	3421.7	4001.8	6553.6
1396.2	2213.5	2380.3	4033.3	4166.1	6997.7
1444.1	2130.9	2410.9	3596.1	4250.7	6821.7
1553.8	2100.5	2604.3	3324.4	4209.2	7499.2
1932.2	2108.2	2743.9	3114.8	4262.7	7488.3
1602.2	2164.7	2781.5	3052.2	4717.7	7013.7
1491.5	2102.5	3405.7	3202.1	4983.2	6685.8
1533.3	2104.4	2774.7	3158.8	4965.6	6672.5
1548.7	2239.6	2805	3096.6	5562.5	7157.5
1585.4	2348	2627	3143.7	5750.3	

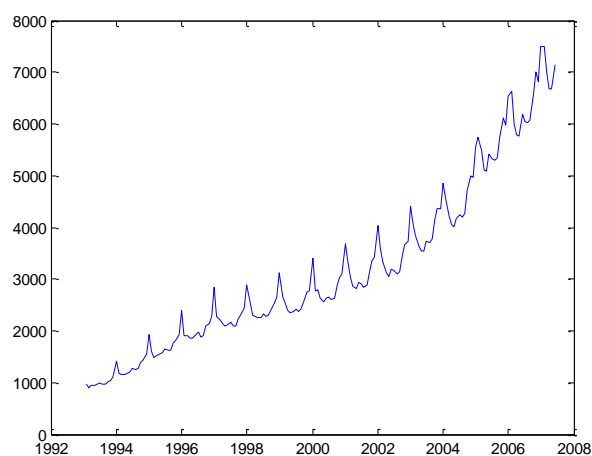


图 14.13 中国社会消费品零售额（1993.1—2007.5）

4. 条件异方差

条件异方差是指时间序列的方差不再是常数，表现为时序图上异常观测值倾向于成群出现。条件异方差更多地出现于经济时间序列，尤其是金融时间序列，在金融时间序列中常常观测到大的正的或负的观测值倾向于成群出现。图 14.14 为 2003 年 1 月 3 日至 2006 年 12 月 29 日（共 209 个交易周）香港恒生指数的周对数超额收益序列，可以看出，收益围绕某个水平处于不断波动之中，在几个特定的时段，收益的波动明显较大，而在其余时段的波动则较小。也就是说，收益的大波动后面倾向于跟着一个大的波动，而收益的小波动后面倾向于一个小的波动，出现了所谓的波动聚集性效应。这意味着前后期的波动之间存在着某种相关关系。

表 14.7 香港恒生指数周超额收益（逐列排列）

9583.85	11170.6	11910.8	13574.9	15165.8	16901.9
9721.5	10883.5	11276.9	13494.8	14983.2	16313.4
9614.59	10968.4	11576	13481	15144	15895.1
9460.6	11290.2	12116.9	13650.1	15428.5	15912.7
9258.95	11608.7	12022.6	13585.2	14847.8	15628.7
9150.95	11935.8	12396.4	13845.6	14485.9	15842.7
9201.76	12044.5	11855.6	14087.9	14487.9	15808.8
9250.86	11736.4	12185.5	14157.1	14215.8	16267.6
9122.66	12190.1	12220.1	13730.8	14585.8	16459.8
8907.1	12215.2	12202.3	13890.9	14740.6	16135.7
8956.17	12203.5	12059.2	13828.4	14883.3	16464.2
9179.19	11839.8	12353	13597.1	15081.5	16955

8863.36	12317.5	12238	13491.4	15200.4	16887.8
8822.45	12314.7	12478.7	13666.7	14910.5	17250
8645.65	12594.4	12359.8	13638.8	15029.8	17330.7
8579.14	12371.8	12376.9	13693.6	15183.6	16955.5
8409.01	12456.7	12818.4	13909	14876.4	17423.7
8808.18	12801.5	12948.1	14034	15344.4	17145.8
9084.16	13385.8	13004	13866.8	15788	17237.7
9093.18	13167.8	13224.9	13717.4	15662.1	17600.7
9303.73	13750.6	13066.8	13714.8	15753.1	17543.1
9487.38	13289.4	13120	13818.5	15429.7	17903.4
9694.63	13309.6	13241.5	13934.8	15426	17988.9
9855.64	13739.8	13059.4	13912	15475.7	18113.6
9930.31	13868.4	13015.2	14230.3	15856.1	18297.6
9657.21	13907	13054.7	14201.1	15802	18749.7
9636.81	13454.8	13495	13964.5	15445.1	18891.1
9911.5	12932.2	13784.5	14504.3	15801.7	19182.7
10140.8	12790.6	13787.7	14786.5	15716.5	19260.3
9939.2	12483.2	13895	14881	15805	18690.8
10248.6	12731.8	14211.8	15051.3	16471.8	18740
9945.22	12909.4	13901.8	15451	16429.5	19110.7
10424.6	12458.4	13992.4	15038.6	16912.2	19320.5
10760.7	12383.9	14194.9	14982.9	16661.3	19964.7
10909	11943	14230.1	15221.9	17013.9	

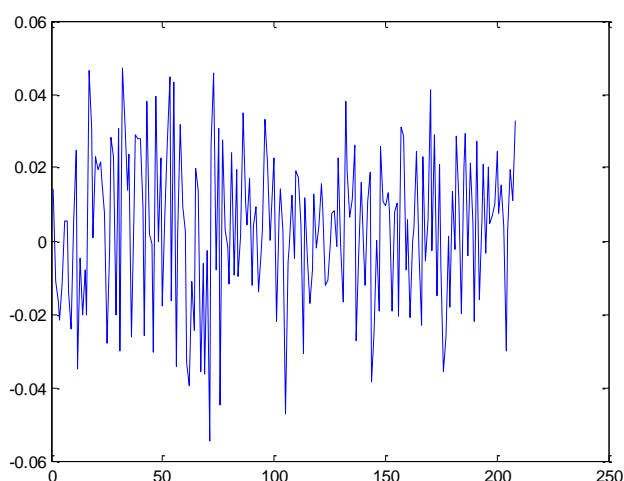


图 14.14 香港恒生指数的周对数超额收益序列

记原始时间序列为 x_t ，先作对数变换 $y_t = \ln x_t$ ，再作一个简单的回归模型

$$(y_t - y_{t-1})^2 = \alpha + \rho(y_{t-1} - y_{t-2})^2 + a_t, \quad t = 3, 4, \dots, 209, \quad (17)$$

$(y_t - y_{t-1})^2$ 表示收益的方差。估计出 $\alpha = 0.0004$ ， $\rho = 0.0222$ 。

计算的 MATLAB 程序如下

```
clc, clear
a=textread('gushi.txt');
a=nonzeros(a); b=log(a);
c=diff(b); n=length(c);
xishu=[ones(n-1,1),c(2:end).^2];
cs=xishu\c(1:end-1).^2 %线性最小二乘拟合
```

14.1.3 异常观测值

对一给定时间序列 y_t ，如果一个或多个观测值与其它时刻的观测值有显著的不同，就

称该序列出现了异常的观测值，又叫离群点（Outlier）。根据异常值对时间序列的影响效果的差异，常见的异常值可分为五类：可加的异常观测值，新生的异常观测值，永久的水平迁移，临时的水平迁移，趋势改变。

下面我们分别介绍这五种异常值，并用生成数据，说明各类异常值的数据特征。为了叙述上的方便，我们对所研究的序列做几个假定：

- （1）令 x_t 为不受异常值影响的基本时间序列过程，在实际分析中是不可观测的真实时间序列，并假定序列 x_t 为平稳的时间序列；
- （2）令 y_t 为观察到的时间序列；
- （3） a_t 一个均值为 0，方差为 1 的 Gaussian 白噪声序列。下面分析中使用的是样本容量为 200 的相同的 a_t ；
- （4） $I_t^{(T)}$ 为虚拟变量。

1. 可加性异常观测值

造成这种异常值的干扰，只影响该干扰发生的那一时刻 T 上的序列值 y_T ，而不影响该时刻以后的序列值 y_{T+1}, y_{T+2}, \dots 。因此我们常将这类异常值看作是一个真实的观测值加上或减去一个误差。在实际的时间序列中，发生干扰的时刻 T 是很难确定的，而且干扰值的大小常是随机的。显然，对这类异常值的预测几乎是不可能的。我们用一个模拟的序列来说明可加性异常观测值的数据特征。

假定数据由下式生成

$$y_t = x_t + 10I_t[t=100], \tag{18}$$

$$x_t = 0.7x_{t-1} + a_t, \quad x_1 = 0, \quad t = 2, 3, \dots, 200. \tag{19}$$

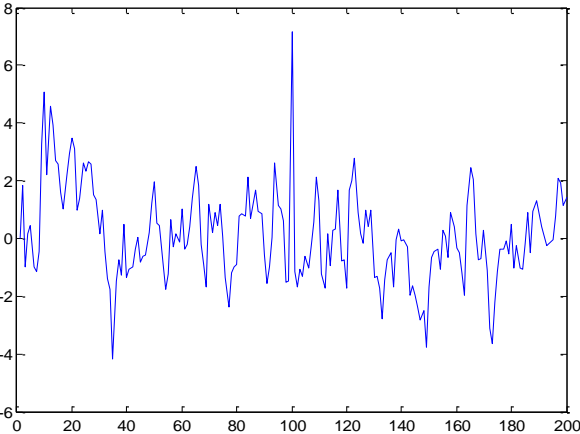


图 14.15 有一个异常观测值的模型

表 14.8 国际原油价格（逐列排列）

18.63	15.75	20.50	19.65	18.48	18.48
18.63	14.40	19.20	20.53	18.83	16.03
18.63	11.93	17.95	21.43	18.70	16.73
18.63	12.60	16.35	22.00	18.85	17.18
18.58	14.35	15.30	20.08	18.48	16.95
19.08	16.23	15.73	17.75	17.43	16.23
20.03	16.38	19.23	18.15	16.85	16.80
18.63	17.23	27.80	17.45	16.73	17.23
18.48	20.45	41.00	19.09	17.23	17.98
18.80	20.15	34.30	19.65	15.43	18.98
17.70	18.25	31.20	20.70	14.23	17.58
17.60	18.28	28.35	20.60	13.18	16.58
18.28	16.30	20.70	20.38	14.83	15.78
14.73	17.20	19.50	19.65	13.59	16.20
15.65	18.23	18.00	20.18	13.25	16.58

16.60	18.93	19.70	19.13	15.35	16.58
16.20	18.48	19.00	19.25	16.18	17.13
14.18	21.05	18.48	17.85	17.45	18.65

图 14.16 为国际原油价格 1987 年 1 月—1995 年 12 月的月度数据,至少有一个明显的异常值 (41.00)。

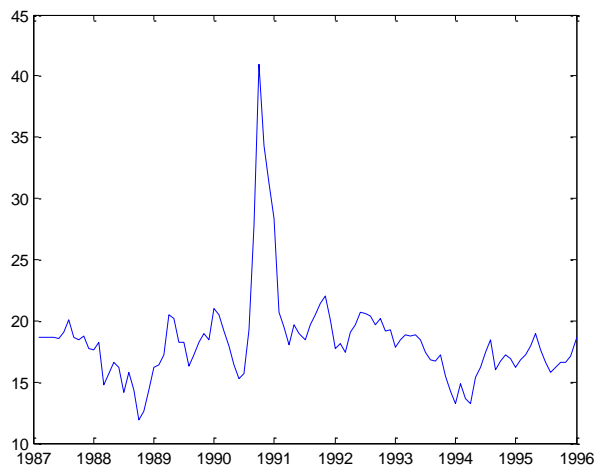


图 14.16 1987.1—1995.12 国际原油价格

2. 新生的异常观测值

造成这种异常值的干扰,不仅作用于 y_t , 而且影响 T 时刻以后序列的所有观测值 y_{T+1}, y_{T+2}, \dots , 它的出现意味着外部干扰作用于系统的开始, 并且其作用方式与系统的动态模型有关。

为对比 IO 和 AO 异常值对时间序列的干扰的差异, 我们用生成 AO 的基本数据来生成一个 IO 的时间序列, a_t 为均值为 0, 方差为 1 的正态白噪声序列, 数据生成过程为

$$\begin{aligned} y_t &= 0.7y_{t-1} + 10I_t[t=100] + a_t, \\ y_1 &= 0, \quad t = 2, 3, \dots, 200. \end{aligned} \quad (20)$$

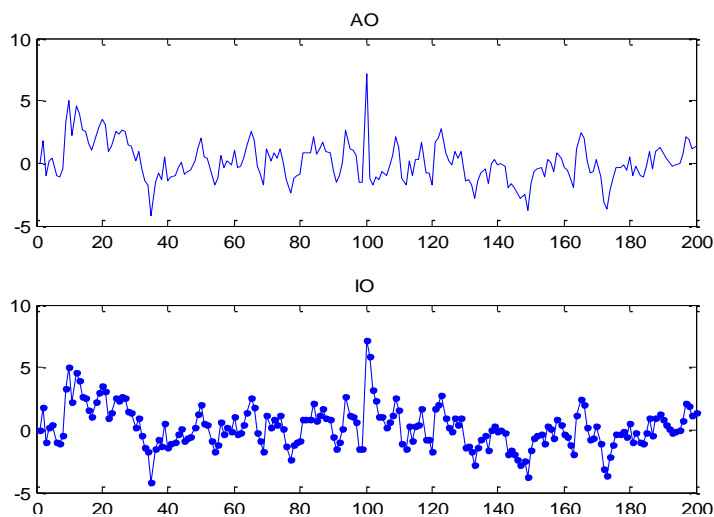


图 14.17 有一个异常观测值的 IO 模型

图 14.17 清楚地表明一个 IO 不仅影响了时刻 $t=100$ 的值, 而且对时刻 $t > 100$ 的观测值也产生了影响, 影响的大小随着时间的增加而减少。

MATLAB 程序如下

```
clc, clear
```

```

at=normrnd(0,1,1,200);
xt(1)=0;
for i=2:200
    xt(i)=0.7*xt(i-1)+at(i);
end
t=1:200;
yt1=xt+10*(t==100);
subplot(211), plot(yt1), title('AO')
yt2(1)=0;
for t=2:200
    yt2(t)=0.7*yt2(t-1)+10*(t==100)+at(t);
end
subplot(212), plot(yt2,'-'), title('IO')

```

3. 永久水平迁移异常值

造成这种异常值的干扰是在某一时刻 T ，系统的结构发生变化，持续影响 T 时刻以后的所有行为，且影响的效应相同，在序列上往往表现出 T 时刻前后序列均值发生水平位移。

图 14.18 为模拟的具有不同水平迁移效应的时间序列， a_t 为均值为 0，方差为 1 的正态白噪声序列，数据生成过程为

$$\begin{aligned}
 y_t &= 0.7y_{t-1} + \omega I_t[t \geq 100] + a_t, \\
 y_1 &= 0, \quad t = 2, 3, \dots, 200.
 \end{aligned}
 \tag{21}$$

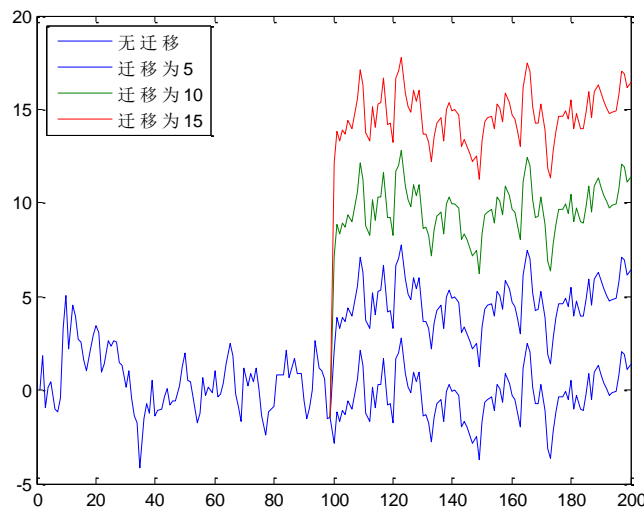


图 14.18 AR (1) 过程中 ω 不同值的水平迁移效应

画图的 MATLAB 程序如下

```

clc, clear
at=normrnd(0,1,1,200);
yt1(1)=0; yt2(1)=0; yt3(1)=0; yt4(1)=0;
for i=2:200
    yt1(i)=0.7*yt1(i-1)+at(i);
    yt2(i)=0.7*yt2(i-1)+5*(i>=100)+at(i);
    yt3(i)=0.7*yt3(i-1)+10*(i>=100)+at(i);
    yt4(i)=0.7*yt4(i-1)+15*(i>=100)+at(i);
end
t2=99:200;
y2=yt2(t2); y3=yt3(t2); y4=yt4(t2);
plot(yt1), hold on
plot(t2,[y2',y3',y4'])
legend('无迁移','迁移为 5','迁移为 10','迁移为 15',2)

```


随着 ω 的增加, 时间序列呈向上的趋势, 佩罗恩 (Perron, 1989) 证明, 如果一个序列存在水平迁移, 随着 ω 的增大, y_{t-1} 的系数估计值越接近于 1, 这个可以通过序列的自相关函数 (ACF) 值来说明, 自相关函数 $\hat{\rho}_i$ 的取值随 ω 值的增加而增加, 说明具有水平迁移的系统有较长的记忆性。

计算自相关函数的 MATLAB 程序如下

```
clc, clear
at=normrnd(0,1,1,200);
yt1(1)=0; yt2(1)=0; yt3(1)=0; yt4(1)=0;
for i=2:200
    yt1(i)=0.7*yt1(i-1)+at(i);
    yt2(i)=0.7*yt2(i-1)+2*(i>=100)+at(i);
    yt3(i)=0.7*yt3(i-1)+4*(i>=100)+at(i);
    yt4(i)=0.7*yt4(i-1)+6*(i>=100)+at(i);
end
r1=autocorr(yt1); % 计算自相关函数
r2=autocorr(yt2);
r3=autocorr(yt3);
r4=autocorr(yt4);
biao=[r1(2:6); r2(2:6); r3(2:6); r4(2:6)]
```

4. 暂时水平迁移异常值

造成这种离群点的干扰是在时刻 T 干扰发生时具有一定的初始效应 ω , 以后随着时间根据衰减因子 δ 的大小呈指数衰减的一类干扰事件。我们用一个白噪声序列来说明暂时变更性异常值对时间序列的影响效应 (见图 14.19)。 y_t 由一个简单的白噪声过程生成, a_t 为均值为 0, 方差为 1 的正态白噪声序列, 数据生成过程为

$$y_t = \omega I_t(t=51) + \omega\delta I_t(t=52) + \omega\delta^2 I_t(t=53) + \cdots + a_t, \quad (22)$$

$$y_1 = 0, t = 2, 3, \dots, 200.$$

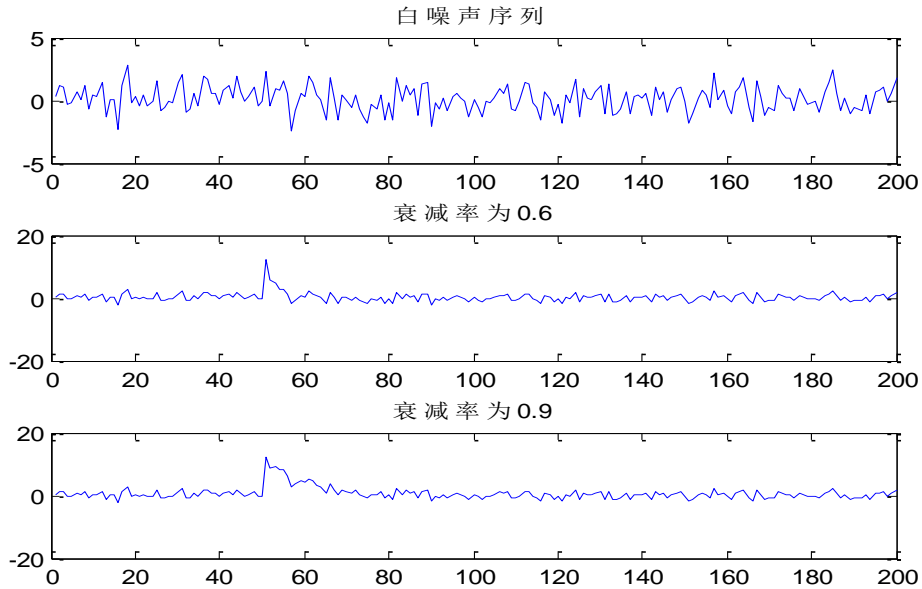


图 14.19 临时的水平迁移效应

MATLAB 程序如下

```
clc, clear
at=normrnd(0,1,1,200);
omega=10; delta1=0.6; delta2=0.9;
y1(1:50)=at(1:50); y2(1:50)=at(1:50);
for i=51:200
```

```

rd1=omiga*delta1^(i-51); rd2=omiga*delta2^(i-51);
y1(i)=rd1+at(i); y2(i)=rd2+at(i);
end
subplot(311), plot(at), title('白噪声序列')
subplot(312), plot(y1), title('衰减率为 0.6')
subplot(313), plot(y2), title('衰减率为 0.9')

```

5. 趋势改变

在经济领域的时间序列，还有一类重要的异常值，这类异常值的存在改变了时间序列的趋势，我们用一个回归模型来描述这类异常值的数据特征，如下式

$$y_t = vt + \omega I_t(t \geq T)t + a_t, \quad (23)$$

(23) 式说明在时刻 T 以后，序列的趋势由 vt 变为 $(v + \omega)t$ 。

图 14.20 为模型的具有改变的时间序列， a_t 为均值为 0，方差为 1 的正态白噪声序列，数据生成过程为

$$y_t = 0.95t - 0.45I_t[t \geq 71]t + a_t, \quad t = 1, 2, \dots, 200. \quad (24)$$

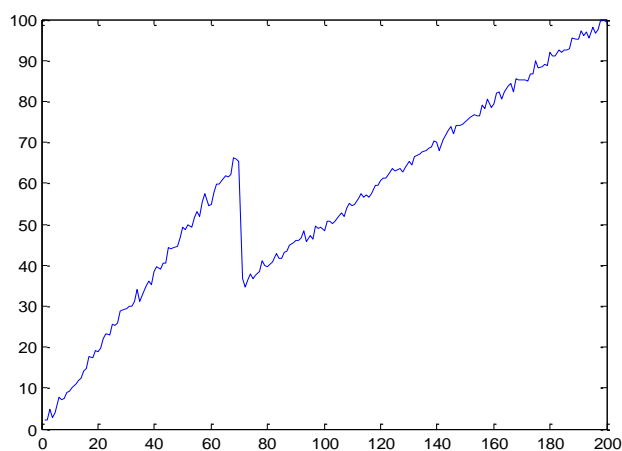


图 14.20 趋势改变模型

画图的 MATLAB 程序如下

```

clc, clear
at=normrnd(0,1,1,200);
t=1:200;
yt=0.95*t-0.45*(t>=71).*t+at;
plot(yt)

```

14.2 时间序列的预处理

14.2.1 非平稳时间序列平稳化的处理

时间序列的预处理，一方面能够使序列的特征体现的更加明显，利于分析模型的选择；另一方面也使数据满足模型的要求。例如通过对序列取对数以及对序列进行标准化、中心化、归一化处理等方法进行数据变换，使偏态分布的序列变成对称的分布，消除序列中的异方差性，使变量间的非线性关系转换成线性关系，在时间序列数据数量很大的时候会起到显著改善计算精度的作用。

时间序列往往具有明显的长期趋势和不规则变动叠加于随机波动之上，因此，大部分时间序列都是非平稳的时间序列，只有通过各种数据处理方法将数据的非平稳特性从序列中分离出来，才能将其转换为平稳的时间序列。因此，时间序列的平稳化处理是时间序列分析的重要步骤。

时间序列非平稳性表现出多样性和复杂性。一个均值平稳过程其方差和自协方差可能是非平稳的；而一个均值非平稳过程也可能是方差和自协方差非平稳过程。因而时间序列平稳化的方法也是多种多样的。对于不同形式的非平稳性，应采取不同的平稳化方法。

方差非平稳序列的平稳化处理可采用 BOX—COX 变换, 该方法最早由 BOX 和 COX 于 1964 年提出。如果序列的方差同序列的发展水平成比例, 可对原序列做对数变换。但对数变换有时候可能会对某些序列产生过度的修正数据, 这时候常采用 BOX—COX 变换。变换方法如下式

$$x_t^{(\lambda)} = \begin{cases} \ln x_t, & \lambda = 0, \\ \frac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \end{cases} \quad (25)$$

其中 λ 为变换参数。

均值非平稳序列的平稳化处理, 往往采用差分的方法对原序列进行预处理。下面我们介绍三种常用的方法。

1. 差分

差分是通过逐项相减消除前后期数据相关性的方法, 可剔除序列中的趋势性, 是非平稳序列的均值平稳化的预处理。差分运算可用后移算法 B 或者差分算子 ∇ 以及相应的阶数 d 表示。

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 ∇^d 称为 d 阶差分算子,

$$\nabla^d \equiv (1 - B)^d = 1 - C_d^1 B + C_d^2 B^2 + \cdots + (-1)^{d-1} C_d^{d-1} B^{d-1} + (-1)^d B^d,$$

其中 $C_d^k = \frac{d!}{k!(d-k)!}$ 。

下面从实例出发, 对差分这一方法做进一步的分析。比如中国 1998 年—2006 年房地产销售价格指数序列 X_t (季度数据) 见表 14.9, 做一次差分后记为 ∇X_t , 则

$$\nabla X_t = X_t - X_{t-1}. \quad (26)$$

表 14.9 中国 1998—2006 年房地产价格指数 (逐列排列)

101.3	99.9	101.9	104	107.7	106.1
102.1	100.7	102.5	103.5	109.1	106.5
101.3	100.7	102.7	104.8	109.9	105.5
101	101.1	101.8	105	110.8	105.7
99.7	101.5	104.3	104.1	109.8	105.5
99.6	101.2	102.8	105.1	108	

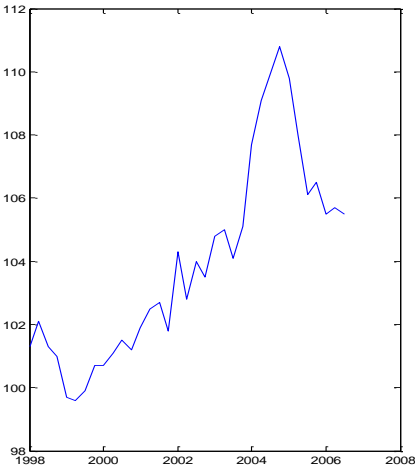


图 14.21 中国房地产销售价格指数

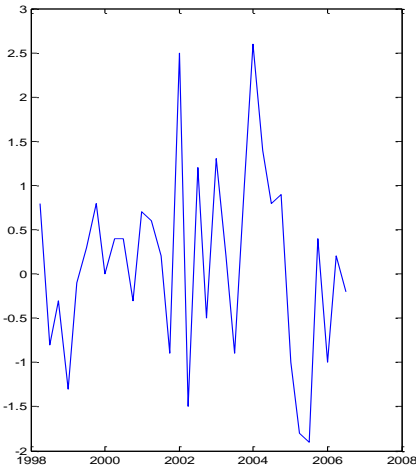


图 14.22 中国房地产销售价格指数一阶差分

X_t 和 ∇X_t 的趋势图如图 14.21 和图 14.22 所示。可以看出 X_t 具有线性上升趋势，而其差分后的序列 ∇X_t 则可以认为是稳定在零点附近。一般而言，若某序列具有线性的趋势，则可以通过对其进行一次差分而将线性趋势剔除掉（这一点从线性函数的性质是不难理解的），然后对差分后的序列拟合 ARMA 模型进行分析与预测，最后再通过差分的反运算得到 X_t 的有关结果。由(26)式可得

$$X_t = \nabla X_t + X_{t-1}, \quad (27)$$

因而

$$\hat{X}_t = \nabla \hat{X}_t + X_{t-1}. \quad (28)$$

若序列 X_t 具有 d 阶多项式趋势，则通过 d 阶差分后可变为平稳序列，当差分阶数为 d 时，最后的数据就会损失 d 个。下面讨论 X_t 与 $\nabla^d X_t = W_t$ 的关系。

当 $d = 2$ 时，有

$$W_t = \nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}, \quad (29)$$

因而

$$X_t = W_t + 2X_{t-1} - X_{t-2}. \quad (30)$$

一般有

$$W_t = \nabla^d X_t = (1-B)^d X_t = [1 + \sum_{k=1}^d (-1)^k C_d^k B^k] X_t, \quad (31)$$

因而

$$X_t = W_t - \sum_{k=1}^d (-1)^k C_d^k X_{t-k}.$$

2. 季节差分

反映经济现象的序列，不少都具有周期性。设 X_t 为一含有周期为 S 的周期性波动序列，则 $X_t, X_{t+S}, X_{t+2S}, \dots$ 为各相应周期点的数值，它们则表现出非常相近或呈现某一趋势的特征，如果把每一观察值同下一周期相应时刻的观察值相减，这就叫季节差分，它可以消除周期性的影响。季节差分常用 ∇_S 表示， $\nabla_S X_t = X_t - X_{t-S}$ ，其中 S 为周期。

例如，某市 1985 年—1994 年各月工业生产总产值见表 14.10，如图 14.23 所示，可以看出对应的时间序列 Z_t 具有明显的周期性，做一次差分， $Y_t = Z_t - Z_{t-12}$ ， Y_t 如图 14.24 所示，周期性明显被消除掉。

表 14.10 某市 1985—1993 年各月工业生产总产值（单位：万元）

obs	GY					
1985.01	10.93	9.34	11.00	10.98	11.29	11.84
1985.07	10.62	10.90	12.77	12.15	12.24	12.30
1986.01	9.91	10.24	10.41	10.47	11.51	12.45
1986.07	11.32	11.73	12.61	13.04	13.14	14.15
1987.01	10.85	10.30	12.74	12.73	13.08	14.27
1987.07	13.18	13.75	14.42	13.95	14.53	14.91
1988.01	12.94	11.43	14.36	14.57	14.25	15.86
1988.07	15.18	15.94	16.54	16.90	16.88	18.10
1989.01	13.70	10.88	15.79	16.36	17.22	17.75
1989.07	16.62	16.96	17.69	16.40	17.51	19.73
1990.01	13.73	12.85	15.68	16.79	17.59	18.51
1990.07	16.80	17.27	20.83	19.18	21.40	23.76
1991.01	15.73	13.14	17.24	17.93	18.82	19.12
1991.07	17.70	19.87	21.17	21.44	22.14	22.45
1992.01	17.88	16.00	20.29	21.03	21.78	22.51
1992.07	21.55	22.01	22.68	23.02	24.55	24.67
1993.01	19.61	17.15	22.46	23.19	23.40	26.26
1993.07	22.91	24.03	23.94	24.12	25.87	28.25
1994.01	20.99	17.04	23.16	24.96	26.2	27.89

1994.07	24.77	26.32	26.75	26.51	28.57	31.14
---------	-------	-------	-------	-------	-------	-------

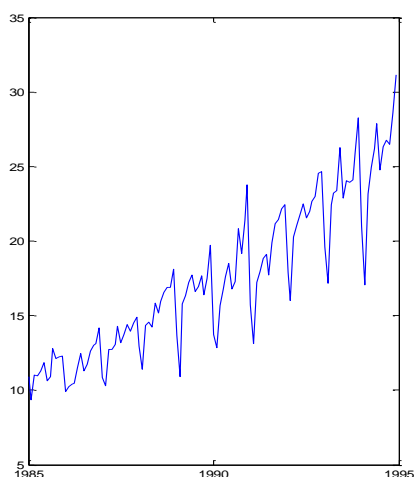


图 14.23 某市各月工业生产总产值数据图

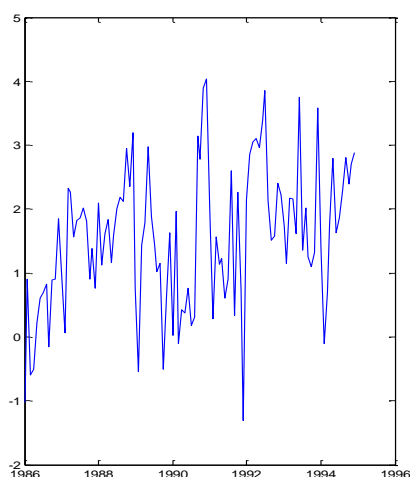


图 14.24 某市工业生产总产值数据季节差分图

3. 对数变换与差分运算的结合运用

如果序列 X_t 含有指数趋势，则可以通过取对数将指数趋势转化为线性趋势，然后再进行差分以消除线性趋势。图 14.25 是我国 1952 年—2005 年社会消费品零售总额（见表 14.11）数据图。可以看出，有一定指数趋势。取对数后如图 14.26（a）所示，具有明显的非线性增长趋势，前期和后期增长速度较快，而中期稍慢。一阶差分后如图 14.26（b）所示，前期和后期数据偏大，而中期较小。具有一定的非平稳性。二阶差分后如图 14.26（c）所示， $\nabla^2 \ln X_t$ 基本平稳。

表 14.11 我国社会消费品零售总额

obs	SHXZ				
1951		262.7	328.8	356.1	364.0
1956	424.0	441.6	481.2	556.5	595.4
1961	537.7	543.7	544.8	572.7	590.1
1966	632.8	679.1	649.2	698.2	728.8
1971	776.9	853.5	917.7	967.4	1046.4
1976	1099.0	1174.3	1264.9	1476.0	1794.0
1981	2002.5	2181.5	2426.1	2899.2	3801.4
1986	4374.0	5115.0	6534.6	7074.2	8300.1
1991	9415.6	10993.7	14270.4	18622.9	23613.8
1996	28360.2	31252.9	33378.1	35647.9	39105.7
2001	43055.4	48135.9	52516.3	59501	67176.6

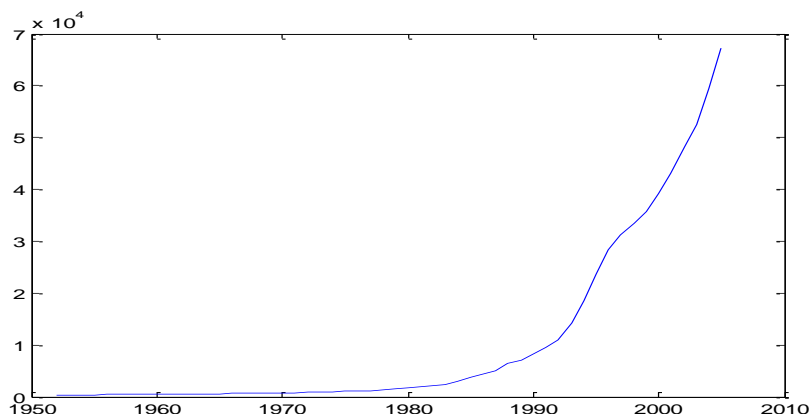


图 14.25 我国 1952—2005 年社会消费品零售总额

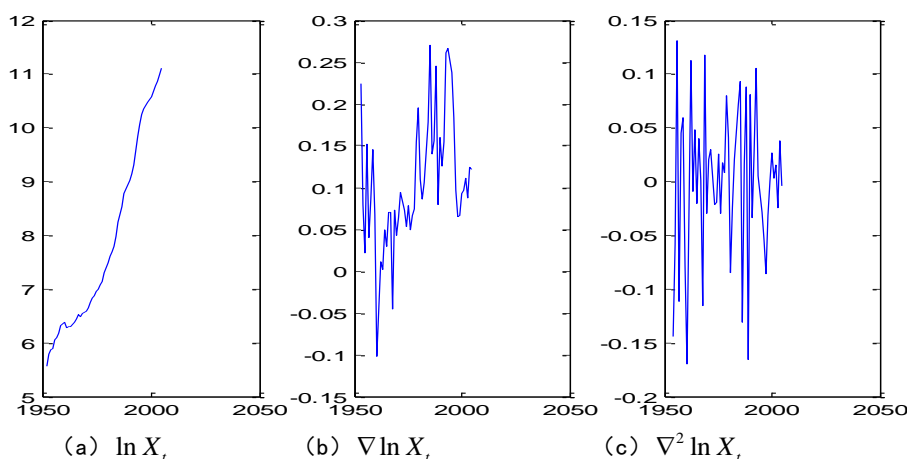


图 14.26 变换后的趋势图

画图的 MATLAB 程序如下

```
clc, clear
xt=textread('data23.txt');
xt=xt'; xt=nonzeros(xt);
yt=log(xt); zt1=diff(yt); zt2=diff(zt1);
t=[1952:2005]';
plot(t,xt)
figure, subplot(131), plot(t,yt)
subplot(132), plot(t(2:end),zt1)
subplot(133), plot(t(3:end),zt2)
```

14.2.2 异常值的处理

离群点 (Outlier) 也称之为奇异值或野值, 是指一个时间序列中, 远离一般水平的极端大值和极端小值。离群点是由于系统受外界的干扰而产生的。首先可能是采样中的误差, 如记录仪出现偏误、工作人员出现笔误、计算误差等; 其次可能是研究现象本身受各种偶然的非正常的因素影响而造成的。

离群点的作用

(1) 离群点会直接影响模型的拟合精度, 甚至会产生一些虚假信息, 所以往往被分析人员看成是一个“坏值”。

(2) 离群点不仅提示我们认真检查采样中是否有差错, 在进行时间序列分析之前认真确认序列, 而且当确认离群点是由于系统受外部突发因素刺激而引起时, 它会提供系统稳定性、灵敏性等重要信息。

1. 离群点的检验

方法 1: 将序列值与平滑值进行比较, 检测其是否显著地大 (或小)。

用 \bar{X}_t^2 表示先对 X_t 取均值再平方, $\overline{X_t^2}$ 表示先对 X_t 平方再取均值, $S_t^2 = \overline{X_t^2} - \bar{X}_t^2$ 表示样本方差, S_t 表示标准差。如果

$$\bar{X}_t - kS_t < X_{t+1} < \bar{X}_t + kS_t, \quad (32)$$

则认为 X_{t+1} 是正常的, 否则认为 X_{t+1} 是一个离群点。其中, k 一般取 3—9 的整数, 开始时可取 $k=6$ 。

如果 X_{t+1} 是一个离群点, 则用

$$\hat{X}_{t+1} = 2X_t - X_{t-1} \quad (33)$$

来代替。这实际上是线性外推。这种方法需要事先规定连续外推的次数, 以免出现无休止的外推。因为连续检测到一些离群点后, 最终的外推结果可能偏差很远, 以致会排除本来是正常的数据点。

方法 2: 检测序列值与其相应的曲线平滑估计值的绝对离差是否大于某一预先设定值 k 。

首先根据序列 X_t 从首项开始取 5 项移动中位数生成一个新序列 X_t' ，即取 X_1, \dots, X_5 的中位数作为 X_3' ，然后舍去 X_1 加入 X_6 取中位数得 X_4' ，依次类推，直到加入序列最后一个数 X_N 。其次用同样的方法根据序列 X_t' ，从首项开始取 3 项移动中位数构成序列 X_t'' 。再次由序列 X_t'' 构造

$$X_t''' = \frac{1}{4}X_{t-1}'' + \frac{1}{2}X_t'' + \frac{1}{4}X_{t+1}'' \quad (34)$$

最后，检测是否存在 $|X_t - X_t'''| > k$ （其中 k 为预先设定值），若有，则该点是一个离群点。

14.2.3 缺损值的补足

分析人员可根据具体序列，选择一定方法，如增长量推算法、发展速度推算法、比例推算法、平滑法、插值估算法等来补足缺损值。

表 14.12 我国 1949—2005 年人口自然增长率（逐列排列）

16	10.19	26.08	11.61	15.04	8.18
19	-4.57	25.83	11.87	14.39	7.58
20	3.78	23.33	16.61	12.98	6.95
20	26.99	22.16	15.73	11.6	6.45
23	33.33	20.89	15.04	11.45	6.01
24	27.64	17.48	14.39	11.21	5.87
20.32	28.38	15.69	14.26	10.55	5.89
20.5	26.22	12.66	15.57	10.42	
23.23	25.53	12.06	16.61	10.06	
17.24	27.38	12	15.73	9.14	

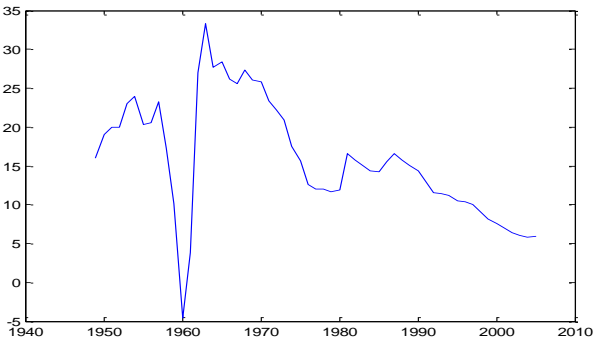


图 14.27 我国 1949—2005 年人口自然增长率数据图

根据图 14.27 可以看出，1960 年的人口自然增长率极端异常，从数值上看为 -4.57%，下面我们数值检验的第一种方法来检测该序列是否存在离群点。

设观测序列为 X_1, \dots, X_T ，一次指数平滑公式为

$$\bar{X}_t = \alpha X_t + (1 - \alpha) \bar{X}_{t-1} = \bar{X}_{t-1} + \alpha(X_t - \bar{X}_{t-1}),$$

设加权系数 $\alpha = 0.2$ ，并取 $k = 3$ ，根据(32)式， $\bar{X}_t - kS_t$ ， X_{t+1} 及 $\bar{X}_t + kS_t$ 的数据如图 14.28 所示。

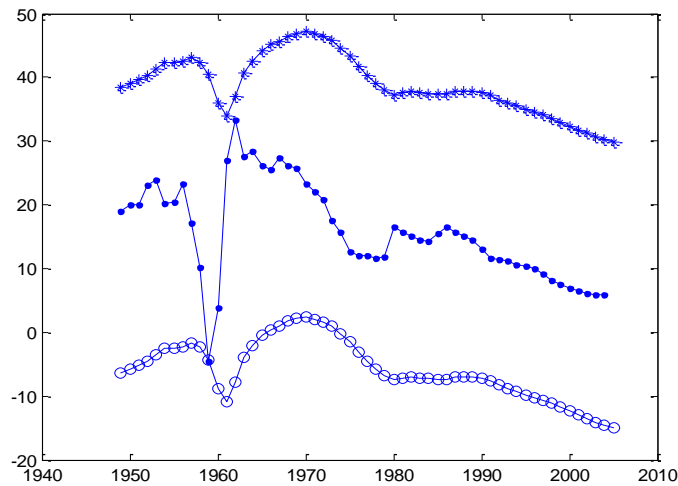


图 14.28 人口自然增长率的离群点检测图

可见，在 X_{12} 处， $X_{12} < \bar{X} - 6S$ ，即 $-4.57\% < -4.3371\%$ ，除此之外，所有的观测值均满足(32)式。所以，我们认为 X_{12} 为一个离群点，根据(33)式，我们可以用

$$2X_{11} - X_{10} = 3.14\%$$

代替原 X_{12} 序列的观测值 -4.57% 。

画图和计算的 MATLAB 程序如下

```
clc, clear
xt=textread('data24.txt');
xt=nonzeros(xt); n=length(xt);
t=[1949:2005]'; sigma=std(xt);
s(1)=xt(1); alpha=0.2; k=3;
for i=2:n
    s(i)=alpha*xt(i)+(1-alpha)*s(i-1);
end
plot(t,s-k*sigma,'o-'), hold on
plot(t(1:end-1),xt(2:end),'-')
plot(t,s+k*sigma,'*-')
xia=s(11)-k*sigma %计算下界
qiyi=xt(12)
```

14.2.4 模型分析

用模型检验异常值是指对数据进行模型分析，然后根据拟合模型后的剩余序列计算特定的统计量，测出显著的异常值及其类型，并用相应的模型进行修正，然后再对修正模型的剩余序列重复上述程序，依次测出各个异常值。

模型分析检验异常值有两个基本假定：首先，时间序列已经过平稳化检验为平稳序列，并根据自相关函数和偏自相关函数初步判断序列服从 $ARIMA(p, d, q)$ ；其次，对时刻 t 的值（或近似值）有一个先验的了解。则异常值的检验步骤如下：

- (1) 对给定序列做平稳性检验，将非平稳序列转换为平稳序列；
- (2) 对 x_t 做自相关函数（ACF）和偏自相关函数（PACF）图；
- (3) 根据 ACF 和 PACF 图对 x_t 拟合 $ARIMA(p, d, q)$ 模型；
- (4) 估计模型残差 a_t ；
- (5) 对残差估计值做正态性检验；
- (6) 如果 JB 检验值较大，拒绝正态性假设，我们可以判断时间序列存在异常值；
- (7) 对检验到的异常值根据所研究问题的需要剔除或通过将哑变量加入到模型的分析中。

考虑芬兰 1949—2004 年的通货膨胀率，数据见表 14.13。从序列的时序图 14.29 来看，有不规则的数据点。

表 14.13 芬兰 1949—2004 年通货膨胀率（逐行排列）

0.02	0.14	0.175	0	0.03	0	-0.029	0.104	0.108	0.073
0.011	0.034	0.022	0.043	0.051	0.097	0.053	0.034	0.057	0.092
0.028	0.027	0.06	0.075	0.105	0.169	0.176	0.146	0.124	0.078
0.075	0.116	0.12	0.097	0.083	0.071	0.058	0.029	0.042	0.05
0.067	0.061	0.041	0.027	0.02	0.011	0.01	0.006	0.012	0.014
0.012	0.034	0.026	0.016	0.009	0.002				

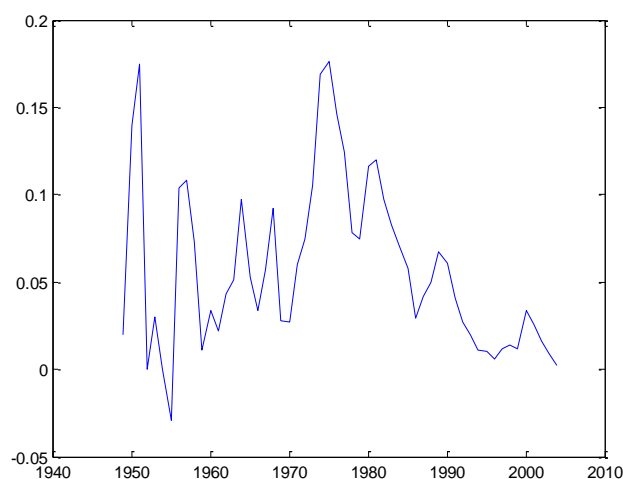


图 14.29 芬兰 1949—2004 年通货膨胀率

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。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。

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

$$q_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (t - R_t)^2, \quad (35)$$

构造统计量

$$T = \frac{q_s \sqrt{n-2}}{\sqrt{1-q_s^2}}. \quad (36)$$

作下列假设检验

H_0 : 序列 x_t 平稳, H_1 : 序列 x_t 非平稳 (存在上升或下降趋势)。

Daniel 检验方法: 对于显著水平 α ，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。

利用上面的 Daniel 检验方法，在 5% 的水平下，认为序列是平稳的。做 x_t 的自相关函数和偏自相关函数（见图 14.30，图 14.31），可对 x_t 拟合 AR(1) 模型

$$x_t = 0.8422x_{t-1} + a_t, \quad t = 1950, \dots, 2004. \quad (37)$$

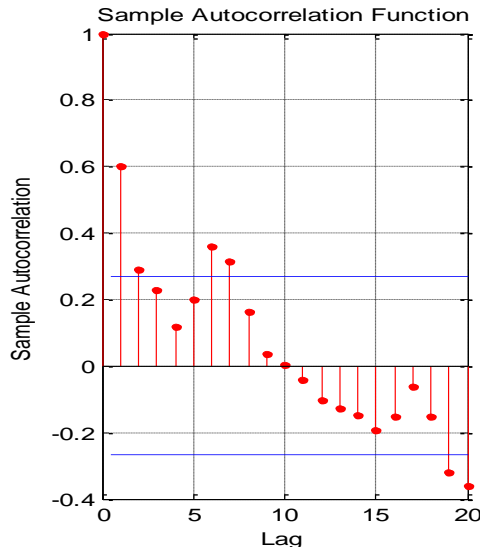


图 14.30 x_t 自相关函数

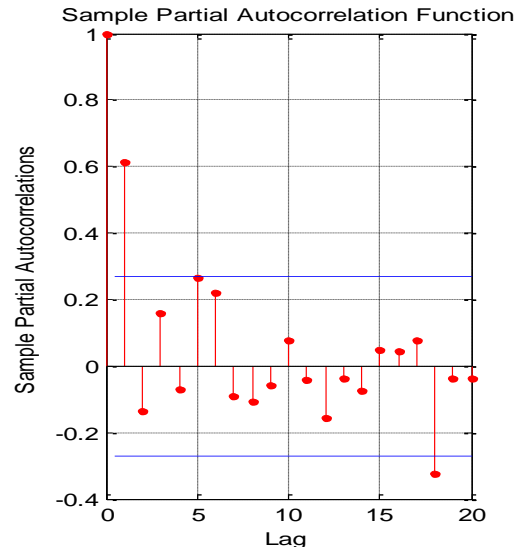


图 14.31 x_t 偏自相关函数

模型残差通过了 JB 检验，说明模型是适宜的。

计算的 MATLAB 程序如下

```
clc, clear
a=textread('data25.txt');
a=a'; a=a(:); a(end-3:end)=[];
Rt=tiedrank(a); %求原始时间序列的秩
n=length(a); t=[1:n]'; t2=[1949:2004]';
plot(t2,a)
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 分位数
figure,subplot(121), autocorr(a) %画自相关函数图
subplot(122), parcorr(a) %画偏相关函数图
cs=ar(a,1) %拟合模型
figure,res=resid(cs,a); %计算残差向量
h=lbqtest(res) %进行 LB (chi2) 检验
```

14.3 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。

设 $\{X_t, t \in T\}$ 是一个随机过程，如果对任意 $t \in T$ ， $E[X_t]$ 存在，则分别称函数

$$\mu_X(t) = E[X_t], \quad t \in T;$$

$$r_X(s, t) = E[(X_s - \mu_X(s))(X_t - \mu_X(t))], \quad s, t \in T;$$

$$D_X(t) = r_X(t, t) = E[X_t - \mu_X(t)]^2, \quad t \in T$$

为 $\{X(t), t \in T\}$ 的均值函数、协方差函数、方差函数。

设随机过程 $\{X_t, t \in T\}$ 的所有二阶矩都存在，且满足

$$(i) \quad E[X_t] = \mu \text{ 为常数}, \quad \forall t \in T;$$

$$(ii) \quad E[(X_{t+\tau} - \mu)(X_t - \mu)] = R(\tau), \quad \forall t, t+\tau \in T;$$

则称 $\{X_t, t \in T\}$ 为宽平稳随机过程，简称平稳过程， $R(\tau)$ 为 X_t 的协方差函数。若 T 是离散集，则称平稳过程为平稳序列。

若时间序列是平稳的，则可考虑建立自回归模型 (Auto Regressive Model)，简称 AR 模型，移动平均模型 (Moving Average Model) 简称 MA 模型和自回归移动平均模型 (Auto Regressive Moving Average Model) 简称 ARMA 模型。

AR (n) 即 n 阶自回归模型为

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_n X_{t-n} + a_t.$$

MA (m) 即 m 阶移动平均模型为

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_m a_{t-m}.$$

ARMA (n, m) 即 n 阶自回归 m 阶移动平均模型为

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_m a_{t-m}.$$

若时间序列是非平稳的, 则可先对序列进行差分运算, 然后再建立 ARMA 模型, 简称 ARIMA 模型。

从模型形式上可以直观看出, AR 模型描述的是系统对过去自身状态的记忆, MA 模型描述的是系统对过去时刻进入系统的噪声的记忆, 而 ARMA 模型则是系统对过去自身状态以及各时刻进入的噪声的记忆。

14.3.1 自回归模型

1. 一阶自回归模型

定义 14.1 如果时间序列 X_t ($t=1, \dots, n$) 的任一时刻的行为主要与其前一时刻的行为有关, 而与其前一时刻以前的行为无直接关系, 即已知 X_{t-1} , X_t 主要与 X_{t-1} 相关。用记忆性来说, 就是一期记忆, 也就是一阶动态性, 那么描述这种关系的数学模型就是一阶自回归模型

$$X_t = \phi_1 X_{t-1} + a_t, \quad (38)$$

记作 AR (1), 其中 X_t 为零均值平稳序列, ϕ_1 为 X_t 对 X_{t-1} 的依赖程度, a_t 为随机扰动。

一阶自回归模型的基本假设

- (1) X_t 对 X_{t-1} 有直接相关关系;
- (2) a_t 为正态独立同分布序列, 记作 $a_t \in NID(0, \sigma_a^2)$;
- (3) a_t 与 X_{t-j} ($j=2, 3, \dots$) 独立。

AR (1) 与普通一元线性回归的关系

普通一元线性回归模型可表示为

$$Y_i = bX_i + \varepsilon_i, \quad i=1, 2, \dots, \quad (39)$$

模型 (38) 与 (39) 的主要区别有:

- (1) 普通线性回归模型需要一组确定性变量值和相应的观测值; AR (1) 模型只需要一组随机变量的观测值。
- (2) 普通一元线性回归表示的是一个随机变量对另一个确定性变量的依存关系; 而 AR (1) 表示的是一个随机变量对其自身过去值的依存关系。
- (3) 普通线性回归是在静态的条件下研究的; AR (1) 是在动态的条件下研究的。
- (4) 二者的假设不同。
- (5) 普通回归模型, 实质上是一种条件回归, 而 AR (1) 是无条件回归。

主要联系表现为:

固定时刻 $t-1$, 且观察值 X_{t-1} 已知时, AR (1) 就是一个普通的一元线性回归模型了。

2. 一般自回归模型

一般自回归模型

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_n X_{t-n} + a_t,$$

$$a_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_n X_{t-n}.$$

AR(n)模型通过把 X_t 中依赖于 $X_{t-1}, X_{t-2}, \dots, X_{t-n}$ 的部分消除之后, 使得具有 n 阶动态性的序列 X_t 转化为独立的白噪声序列 a_t 。因此拟合 AR(n)模型的过程也就是使相关序列独立化的过程。

14.3.2 移动平均模型

AR 系统的特征: 系统在 t 时刻的响应 X_t 仅与其以前时刻的响应 $X_{t-1}, X_{t-2}, \dots, X_{t-n}$ 有关, 而与其以前时刻进入系统的扰动无关。

MA 系统的特征: X_t 与其以前时刻的响应 X_{t-1}, X_{t-2}, \dots 无关, 而与其以前时刻进入系统

的扰动 a_{t-1}, a_{t-2}, \dots 存在着一定的相关关系。

1. 一阶移动平均模型 MA (1)

$$X_t = a_t - \theta_1 a_{t-1}. \quad (40)$$

基本假设:

(1) 系统在 t 时刻的响应 X_t 仅与其前一时刻进入系统的扰动 a_{t-1} 存在着一定的依存关系;

(2) a_t 为白噪声。

2. 一般移动平均模型 MA (m)

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_m a_{t-m}. \quad (41)$$

基本假设

(1) X_t 仅与 $a_{t-1}, a_{t-2}, \dots, a_{t-m}$ 有关, 而与 a_{t-j} ($j = m+1, m+2, \dots$) 无关;

(2) a_t 为白噪声。

一个关于产科医院的例子。

设 a_t 是第 t 天新住院的病员人数, 假设 a_t 是白噪声序列, 即某一天住院人数与第二天住院人数无关。再假设典型的情形是: 10% 的病人住院 1 天, 50% 的病人住院 2 天, 30% 的病人住院 3 天, 10% 的病人住院 4 天, 那么第四天住院的病人数 X_t 满足

$$X_t = a_t + 0.9a_{t-1} + 0.4a_{t-2} + 0.1a_{t-3}.$$

由于上式的各项是随机变化的, 故 $X_t \sim \text{MA}(3)$ 。因此, 我们可以预测一天、两天、三天后的住院人数。四天以后, 各天住院人数的预测值均为序列的均值。

14.3.3 自回归移动平均模型

ARMA 系统的特征: 系统在 t 时刻的响应 X_t 不仅与其以前时刻的响应有关, 而且与其以前时刻进入系统的扰动存在着一定的依存关系。

对于 ARMA 系统来说, 要使响应 X_t 转化为独立序列 a_t , 不仅要消除 X_t 依赖于 t 时刻以前的自身部分, 还必须消除 X_t 依赖于 t 时刻以前进入系统的扰动部分。

ARMA (2, 1) 模型为

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t - \theta_1 a_{t-1}. \\ X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} &= a_t - \theta_1 a_{t-1}. \end{aligned}$$

基本假设:

X_t 除了对 X_{t-1}, X_{t-2} 和 a_{t-1} 存在依存关系之外, 在 X_{t-1} 和 X_{t-2} 已知的条件下对其它的 X_{t-j} ($j = 3, 4, \dots$) 和 a_{t-k} ($k = 2, 3, \dots$) 不存在相关关系。此时, a_t 一定独立于 a_{t-k} ($k = 2, 3, \dots$), 当然也就独立于 X_{t-j} ($j = 3, 4, \dots$)。

ARMA(n, m) 模型为

$$X_t - \phi_1 X_{t-1} - \dots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \dots - \theta_m a_{t-m}.$$

14.4 ARMA 模型的特性

对于静态回归来说, 在时刻 t 进入系统的扰动 ε_t 只影响系统 t 时刻的响应, 而对后继行为无任何影响。

在 ARMA 模型的动态形式下, 一个影响系统的扰动 a_t 被记忆一定时期, 从而影响系统的后继行为。

14.4.1 格林函数和平稳性

任何一个 ARMA 模型都是一个线性差分方程。因此, ARMA 模型的性质往往取决于差分方程的性质。

1. AR (1) 系统的格林函数

格林函数就是描述系统记忆扰动程度的函数。

AR(1)模型为

$$X_t - \phi_1 X_{t-1} = a_t, \quad (42)$$

设 $X_{t-1} = y(t)$, 则有

$$y(t+1) - \phi_1 y(t) = a_t. \quad (43)$$

显然是一个一阶非齐次差分方程。

由于

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + a_t = \phi_1(\phi_1 X_{t-2} + a_{t-1}) + a_t \\ &= \phi_1^2 X_{t-2} + \phi_1 a_{t-1} + a_t = \phi_1^3 X_{t-3} + \phi_1^2 a_{t-2} + \phi_1 a_{t-1} + a_t, \end{aligned}$$

依次递推下去，可得到

$$X_t = \sum_{j=0}^{\infty} \phi_1^j a_{t-j}. \quad (44)$$

显然(44)式是差分方程(42)的解。

方程的解(44)式是驱动函数 a_t 的一个线性组合，或者说系统是怎样记忆扰动 a_t 的。同样也可以解释为，在某一时刻进入系统的 a_t 对后继行为的影响程度，也就是过去扰动的权重函数。如果 $\phi_1 \rightarrow 1$ ，则 ϕ_1^j 随 j 的增大而缓慢减少，表明系统的记忆较强；相反 $\phi_1 \rightarrow 0$ ，则 ϕ_1^j 随 j 的增大而急剧减少，表明系统的记忆较弱。

定义 14.2 方程解的系数函数 ϕ_1^j 客观地描述了该系统的动态性，故这个系统函数就叫做记忆函数，也叫格林函数，记作 $G_j = \phi_1^j$ 。

利用 Green 函数，可将系统的解写成

$$X_t = \sum_{j=0}^{\infty} G_j a_{t-j}. \quad (45)$$

或写成

$$X_t = \sum_{k=-\infty}^t G_{t-k} a_k, \quad (46)$$

这里 $G_0 = \phi_1^0 = 1$ 。

注：注意到

$$X_t = \sum_{j=0}^{\infty} G_j a_{t-j} = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots,$$

所以，AR(1)模型可用一个无穷阶 MA 来逼近。

AR(1)模型的后移算子表达式及格林函数

后移算子用 B 表示， B 的次数表示后移期数，即

$$BX_t = X_{t-1}, B^2 X_t = X_{t-2}, \cdots,$$

这样，AR(1)可写成

$$(1 - \phi_1 B)X_t = a_t.$$

它的解为

$$X_t = \frac{1}{1 - \phi_1 B} a_t = (1 + \phi_1 B + \phi_1^2 B^2 + \cdots) a_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots = \sum_{j=0}^{\infty} G_j a_{t-j}.$$

格林函数的意义

由于格林函数就是差分方程解的系数函数，格林函数的意义可概括如下：

(1) G_j 是前 j 个时间单位以前进入系统的扰动 a_{t-j} 对系统现在行为（响应）影响的权重数。

(2) G_j 客观地刻画了系统动态响应衰减的快慢程度。

(3) G_j 是系统动态的真实描述。系统的动态性就是蕴含在时间序列中的数据依存关系。对于一个平稳系统来说，在某一时刻由于受到进入系统的扰动 a_t 的作用，离开其平衡位置（即平均数—零）， G_j 描述系统回到平衡位置的速度， ϕ_1 的值较小，速度较快； ϕ_1 的值较大，回复的速度就较慢。

AR(1)系统的平稳性

(1) 系统稳定性与非平稳性

渐近稳定性是指系统受扰后达到任意初始状态，由此出发的状态向量都随时间的增长而趋于平衡状态。渐近稳定一定是平稳的。

系统不稳定是指如果系统受扰后达到任意初始状态,由此出发的状态向量将随时间的增长而趋于无穷。不稳定系统一定是非平稳的。

如果系统受扰后达到任意初始状态,由此出发的状态向量随时间的增长既不回到均衡位置,又不趋于无穷,这就是系统的临界稳定性。

一个临界稳定系统其响应可能是等幅振荡,也可能是常数,因而一定存在趋势或季节性。所以它既可能是平稳的,也可能是非平稳的。

以后所讨论的平稳系统就是指渐近稳定系统。

(2) AR(1)系统的平稳性条件

对于平稳的 AR (1) 系统来说,如果系统受扰后,该扰动的作用渐渐减少,直至趋于零,即系统的响应随时间增长回到均衡位置。相对于 Green 函数来说,就是随 $j \rightarrow \infty$, 扰动权数 $G_j = \phi_1^j \rightarrow 0$, 故 AR (1) 系统渐近稳定的条件,也就是平稳性条件为 $|\phi_1| < 1$ 。

当 $|\phi_1| = 1$, 即 $G_j = \begin{cases} 1, & \phi_1 = 1 \\ (-1)^j, & \phi_1 = -1 \end{cases}$ 时,虽然响应不回到均衡位置,但仍是有限界的,此时系统为临界稳定。

2. ARMA(2,1) 系统的格林函数

ARMA(2,1) 系统的格林函数的隐式

用比较系数法来求得 ARMA (2,1) 系统的 Green 函数。

ARMA(2,1) 模型是一个二阶非齐次差分方程

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = a_t - \theta_1 a_{t-1}. \quad (47)$$

设该二阶非齐次差分方程的解为

$$X_t = \sum_{j=0}^{\infty} G_j a_{t-j}. \quad (48)$$

为方便起见,可用 B 算子:

$$(1 - \phi_1 B - \phi_2 B^2) X_t = (1 - \theta_1 B) a_t, \quad (49)$$

$$X_t = \left(\sum_{j=0}^{\infty} G_j B^j \right) a_t. \quad (50)$$

把(49)式代入(50), 比较两边 B 的同次幂的系数得

$$G_0 = 1, \quad G_1 = \phi_1 - \theta_1, \quad G_j - \phi_1 G_{j-1} - \phi_2 G_{j-2} = 0, \quad j = 2, 3, \dots$$

ARMA(2,1) 系统的格林函数的显式

ARMA(2,1) 模型的 Green 函数满足二阶齐次差分方程

$$G_{j+2} - \phi_1 G_{j+1} - \phi_2 G_j = 0, \quad j \geq 0.$$

特征方程为

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0,$$

特征根为

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \quad \lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

通解为

$$G_j = g_1 \lambda_1^j + g_2 \lambda_2^j,$$

其中 c_1 和 c_2 是任意常数,其值由初始条件唯一地确定。这里的初始条件为 $G_0 = 1, G_1 = \phi_1 - \theta_1$,

$$g_1 = \frac{\lambda_1 - \theta_1}{\lambda_1 - \lambda_2}, \quad g_2 = \frac{\lambda_2 - \theta_1}{\lambda_2 - \lambda_1}.$$

则 ARMA(2,1) 系统的格林函数为:

$$G_j = \frac{\lambda_1 - \theta_1}{\lambda_1 - \lambda_2} \lambda_1^j + \frac{\lambda_2 - \theta_1}{\lambda_2 - \lambda_1} \lambda_2^j.$$

ARMA (2, 1) 系统的平稳性

(1) 用特征根表示的平稳性

如果 $G_j \rightarrow 0$ ($j \rightarrow \infty$), 则系统是平稳的。

由于 ARMA (2,1) 系统的 Green 函数为

$$G_j = g_1 \lambda_1^j + g_2 \lambda_2^j,$$

所以 ARMA (2,1) 系统的平稳性条件为特征根的模在单位圆内, 即 $|\lambda_1| < 1$, $|\lambda_2| < 1$ 。

(2) 用自回归系数表示的平稳性条件

由于 ARMA (2,1) 系统的平稳性条件为 $|\lambda_1| < 1$, $|\lambda_2| < 1$, 而 $\lambda_1 \lambda_2 = \phi_2$, 故显然有 $|\phi_2| < 1$ 。

另外, 我们有

$$\begin{aligned} \phi_1 + \phi_2 &= \lambda_1 + \lambda_2 - \lambda_1 \lambda_2 = \lambda_1 + \lambda_2 (1 - \lambda_1) < \lambda_1 + 1 - \lambda_1 = 1, \\ \phi_2 - \phi_1 &= -\lambda_1 \lambda_2 - \lambda_1 - \lambda_2 = -\lambda_1 (1 + \lambda_2) - \lambda_2 < 1 + \lambda_2 - \lambda_2 = 1. \end{aligned}$$

综上所述, 我们得到平稳性条件

$$\begin{cases} \phi_1 + \phi_2 < 1, \\ \phi_2 - \phi_1 < 1, \\ |\phi_2| < 1. \end{cases}$$

14.4.2 逆函数和可逆性

前面的格林函数, 把 X_t 表示为过去 a_t 对 X_t 的影响, 或者说系统对过去 a_t 的记忆性, 也就是用一个 MA 模型来逼近 X_t 的行为。平稳序列 X_t 的这种表达形式称为 X_t 的“传递形式”。同样我们也可以用过去的 X_t 的一个线性组合来逼近系统现在时刻的行为。即

$$X_t = I_1 X_{t-1} + I_2 X_{t-2} + \cdots + a_t = \sum_{j=1}^{\infty} I_j X_{t-j} + a_t,$$

$$\text{或 } a_t = X_t - \sum_{j=1}^{\infty} I_j X_{t-j} = (1 - I_1 B - I_2 B^2 - \cdots) X_t.$$

我们把这种表达形式称为 X_t 的“逆转形式”。其中的系数函数 I_j ($I_0 = 1$) 称为逆函数。可见它是一个无穷阶的自回归模型。一个过程是否具有逆转形式, 也就是说逆函数是否存在的性质, 通常称为过程是否具有可逆性, 如果一个过程可以用一个无限阶的自回归模型逼近, 即逆函数存在, 我们就称该过程具有可逆性, 否则, 就是不可逆的。

AR(1)模型和 MA(1)模型的逆函数

1. AR(1)模型的逆函数

根据 AR(1)模型 $(1 - \phi_1 B) X_t = a_t$, 立即可得 $I_1 = \phi_1$, $I_j = 0$, $j = 2, 3, \cdots$ 。

注意到 AR(1)模型的逆函数 I_j 由算子 $(1 - \phi_1 B)$ 求得, 而 Green 函数 G_j 由算子 $\frac{1}{1 - \phi_1 B}$ 求得, 这是逆函数称呼的由来。

所谓可逆性, 是指移动平均模型可以用 AR 模型表示。

2. MA(1)模型的逆函数

MA(1)模型为 $X_t = (1 - \theta_1 B) a_t$, 于是

$$a_t = \frac{X_t}{1 - \theta_1 B} = (1 + \theta_1 B + \theta_1^2 B^2 + \cdots) X_t = X_t - \sum_{j=1}^{\infty} (-\theta_1^j) X_{t-j},$$

故有 $I_j = -\theta_1^j$, $j = 1, 2, \cdots$ 。

显然, 只有 $|\theta_1| < 1$, 才有 $I_j \rightarrow 0$ ($j \rightarrow \infty$), 故 MA(1)的可逆性条件为 $|\theta_1| < 1$ 。

3. G_j 和 I_j 之间的关系

比较 AR(1)和 MA(1)模型的 Green 函数和逆函数

$$\text{AR(1): } \begin{cases} G_j = \phi_1^j, \\ I_0 = 1, \\ I_1 = \phi_1, \\ I_j = 0, j > 0. \end{cases} \quad \text{MA(1): } \begin{cases} G_0 = 1, \\ G_1 = -\theta_1, \\ G_j = 0, j > 1, \\ I_j = -\theta_1^j. \end{cases}$$

可以看出, AR(1)的 G_j (I_j) 和 MA(1)的 I_j (G_j) 形式一致, 只是符号相反,

参数互换。这种对偶性并不是一阶模型所特有，对任意阶模型都是正确的。

ARMA (2,1) 模型的 Green 函数为

$$G_0 = 1, \quad G_1 = \phi_1 - \theta_1, \quad (1 - \phi_1 B - \phi_2 B^2)G_j = 0.$$

ARMA(1,2)模型的逆函数可推导如下

$$\begin{aligned} X_t - \phi_1 X_{t-1} &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \\ (1 - \phi_1 B)X_t &= (1 - \theta_1 B - \theta_2 B^2)a_t, \end{aligned}$$

将 $a_t = (1 - I_1 B - I_2 B^2 - \dots)X_t$ 代入上式并整理，得

$$1 - \phi_1 B = 1 - (I_1 + \theta_1) - (I_2 - I_1 \theta_1 + \theta_2)B^2 - (I_3 - I_1 \theta_2 - I_2 \theta_1)B^3 - (I_4 - I_3 \theta_1 - I_2 \theta_2)B^4 - \dots.$$

于是

$$\begin{aligned} -(I_1 + \theta_1) &= -\phi_1, & \Rightarrow I_1 &= \phi_1 - \theta_1, \\ I_2 - I_1 \theta_1 + \theta_2 &= 0, & \Rightarrow I_2 &= I_1 \theta_1 - \theta_2, \\ I_3 - I_1 \theta_2 - I_2 \theta_1 &= 0, & \Rightarrow I_3 &= I_1 \theta_2 + I_2 \theta_1, \\ I_j &= I_{j-1} \theta_1 + I_{j-2} \theta_2, & j &\geq 3. \end{aligned}$$

所以，在 ARMA (2,1) 的 Green 函数表达式中，用 $-I_j$ 代替 G_j ，用 ϕ 代替 θ ，用 θ 代替 ϕ ，即可得到 ARMA (1,2) 逆函数 I_j 。

例题 ARMA (2,1) 的逆函数为

$$I_1 = \phi_1 - \theta_1, \quad I_2 = \phi_2 + I_1 \theta_1, \quad I_3 = I_{j-1} \theta_1, \quad j \geq 3;$$

它的格林函数为

$$G_1 = \phi_1 - \theta_1, \quad G_2 = G_1 \phi_1 + \phi_2, \quad G_j = \phi_1 G_{j-1} + \phi_2 G_{j-2}, \quad j \geq 3.$$

当取 $\phi_1 = 1.3$ ， $\phi_2 = -0.4$ ， $\theta_1 = 0.4$ 时，模型的格林函数与逆函数对比图见图 14.32。

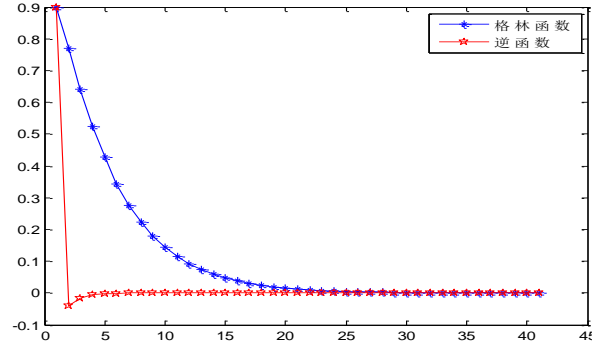


图 14.32 格林函数与逆函数的对比图

计算及画图的 MATLAB 程序如下

```
clc, clear
phi1=1.3; phi2=-0.4; theta1=0.4;
G(1)=phi1-theta1; G(2)=G(1)*phi1+phi2;
I(1)=phi1-theta1; I(2)=phi2+I(1)*theta1;
for i=3:41
    G(i)=phi1*G(i-1)+phi2*G(i-2);
    I(i)=I(i-1)*theta1;
end
t=1:41;
plot(t,G,'-s'), hold on, plot(t,I,'-rP')
legend('格林函数','逆函数')
```

14.4.3 自协方差函数

自协方差函数客观地描述了系统响应的分布特征。

我们知道 X_t 可以用 a_t 的线性组合表示，而 $a_t \sim NID(0, \sigma_a^2)$ ，且有

$$E(a_t) = 0, \quad E(a_t a_{t-k}) = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

a_t 是一个严平稳正态过程，其概率特性完全由自协方差函数来描述。显然 X_t 也是一个正态过程，它的特性也完全取决于自协方差函数。

1. 理论自相关函数和样本自相关函数

(1) 理论自相关函数

设 X_t 是一个零均值化序列，则自协方差函数为 $\gamma_k = E(X_t X_{t-k})$ ，自相关函数为

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{E(X_t X_{t-k})}{E(X_t^2)}.$$

(2) 样本自相关函数

样本自协方差函数一般有下面两种形式

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=k+1}^N X_t X_{t-k}, \quad k=0,1,2,\dots,N-1, \quad (51)$$

$$\hat{\gamma}_k^* = \frac{1}{N-k} \sum_{t=k+1}^N X_t X_{t-k}, \quad k=0,1,2,\dots,N-1. \quad (52)$$

样本自相关函数便有

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{N} \sum_{t=k+1}^N X_t X_{t-k}}{\frac{1}{N} \sum_{t=1}^N X_t^2} = \frac{\sum_{t=k+1}^N X_t X_{t-k}}{\sum_{t=1}^N X_t^2}, \quad (53)$$

或

$$\hat{\rho}_k^* = \frac{\hat{\gamma}_k^*}{\hat{\gamma}_0^*} = \frac{\frac{1}{N-k} \sum_{t=k+1}^N X_t X_{t-k}}{\frac{1}{N} \sum_{t=1}^N X_t^2} = \frac{N}{N-k} \cdot \frac{\sum_{t=k+1}^N X_t X_{t-k}}{\sum_{t=1}^N X_t^2}. \quad (54)$$

说明：

当序列并非零均值且理论均值未知时，用样本均值近似理论均值，样本协方差为

$$\hat{\gamma}_k' = \frac{1}{N} \sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X}), \quad t=0,1,2,\dots,N-1, \quad (55)$$

$$\hat{\gamma}_k^{*'} = \frac{1}{N-k} \sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X}), \quad t=0,1,2,\dots,N-1. \quad (56)$$

2. Green 函数和自协方差函数之间的关系

(1) AR (1) 模型的自协方差函数

假设 X_t 为零均值序列，在 AR(1)模型 $X_t = \phi_1 X_{t-1} + a_t$ 的两端同乘以 X_{t-k} 并取期望，得

$$E(X_t X_{t-k}) = \phi_1 E(X_{t-1} X_{t-k}) + E(a_t X_t).$$

当 $k=0$ 时，有 $E(X_t X_t) = \phi_1 E(X_{t-1} X_t) + E(a_t X_t)$ ，即

$$\gamma_0 = \phi_1 \gamma_1 + \sigma_a^2. \quad (57)$$

当 $k=1$ 时，有 $E(X_t X_{t-1}) = \phi_1 E(X_{t-1} X_{t-1}) + E(a_t X_{t-1})$ ，即

$$\gamma_1 = \phi_1 \gamma_0. \quad (58)$$

当 $k=2$ 时，有 $E(X_t X_{t-2}) = \phi_1 E(X_{t-1} X_{t-2}) + E(a_t X_{t-2})$ ，即

$$\gamma_2 = \phi_1 \gamma_1. \quad (59)$$

依次类推，便有一般式

$$\gamma_k = \phi_1 \gamma_{k-1}, \quad k > 0. \quad (60)$$

将(58)代入(57)，有

$$\gamma_0 = \phi_1 (\phi_1 \gamma_0) + \sigma_a^2 \Rightarrow \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2}. \quad (61)$$

相应的自相关函数为

$$\rho_0 = \gamma_0 / \gamma_0 = 1,$$

$$\rho_k = \gamma_k / \gamma_0 = \phi_1 \gamma_{k-1} / \gamma_0 = \phi_1 \rho_{k-1}.$$

(2) MA (1) 模型的自协方差函数

将 MA (1) 模型 $X_t = a_t - \theta_1 a_{t-1}$ 两端同乘以 X_{t-k} ，取期望，得

$$\begin{aligned}
 E(X_t X_{t-k}) &= E(a_t X_{t-k}) - \theta_1 E(a_{t-1} X_{t-k}) = E(a_t \sum_{j=0}^{\infty} G_j a_{t-k-j}) - \theta_1 E(a_{t-1} \sum_{j=0}^{\infty} G_j a_{t-k-j}) \\
 &= \sum_{j=0}^{\infty} G_j E(a_t a_{t-k-j}) - \theta_1 \sum_{j=0}^{\infty} G_j E(a_{t-1} a_{t-k-j}) \\
 &= G_0 E(a_t a_{t-k}) - \theta_1 [G_0 E(a_{t-1} a_{t-k}) + G_1 E(a_{t-1} a_{t-k-1})] \\
 &= E(a_t a_{t-k}) - \theta_1 E(a_{t-1} a_{t-k}) + \theta_1^2 E(a_{t-1} a_{t-k-1}),
 \end{aligned} \tag{62}$$

当 $k=0$ 时，有

$$\gamma_0 = E(X_t X_t) = E(a_t a_t) - \theta_1 E(a_{t-1} a_t) + \theta_1^2 E(a_{t-1} a_{t-1}) = (1 + \theta_1^2) \sigma_a^2,$$

当 $k=1$ 时，有

$$\gamma_1 = E(X_t X_{t-1}) = E(a_t a_{t-1}) - \theta_1 E(a_{t-1} a_{t-1}) + \theta_1^2 E(a_{t-1} a_{t-2}) = -\theta_1 \sigma_a^2,$$

当 $k=2$ 时，有

$$\gamma_2 = E(X_t X_{t-2}) = E(a_t a_{t-2}) - \theta_1 E(a_{t-1} a_{t-2}) + \theta_1^2 E(a_{t-1} a_{t-3}) = 0.$$

对于 MA (1) 模型来说，有

$$\begin{cases} \gamma_0 = (1 + \theta_1^2) \sigma_a^2, \\ \gamma_1 = -\theta_1 \sigma_a^2, \\ \gamma_k = 0, \quad k \geq 2. \end{cases} \Rightarrow \begin{cases} \rho_0 = 1, \\ \rho_1 = -\frac{\theta_1}{1 + \theta_1^2}, \\ \rho_k = 0, \quad k \geq 2. \end{cases}$$

也就是说，MA (1) 模型的自相关函数在滞后 1 期以后为零，即具有截尾性。

MA (q) 模型的自相关函数在 q 步后同样截尾，可以证明如下。

G_j 和 γ_k 之间有以下关系

$$\begin{aligned}
 \gamma_k &= E(X_t X_{t-k}) = E \left[\left(\sum_{i=0}^{\infty} G_i a_{t-i} \right) \left(\sum_{j=0}^{\infty} G_j a_{t-(j+k)} \right) \right] \\
 &= E \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_i G_j a_{t-i} a_{t-(j+k)} \right] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_i G_j E(a_{t-i} a_{t-(j+k)}) \\
 &= \sum_{j=0}^{\infty} G_{j+k} G_j \sigma_a^2.
 \end{aligned} \tag{63}$$

对于 MA (q) 模型来说

$$G_0 = 1, \quad G_1 = -\theta_1, \quad G_2 = -\theta_2, \quad \dots, \quad G_q = -\theta_q, \quad G_k = 0, \quad k > q,$$

于是根据(63)式，有

$$\gamma_k = \sum_{j=0}^{\infty} G_{j+k} G_j \sigma_a^2 = (G_k G_0 + G_{k+1} G_1 + \dots + G_q G_{q-k}) \sigma_a^2.$$

$$\text{当 } k=0 \text{ 时, } \gamma_0 = (1 + G_1^2 + G_2^2 + \dots + G_q^2) \sigma_a^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2.$$

$$\text{当 } 1 \leq k \leq q \text{ 时, } \gamma_k = (-\theta_k + \theta_{k+1} \theta_1 + \theta_{k+2} \theta_2 + \dots + \theta_q \theta_{q-k}) \sigma_a^2.$$

$$\text{当 } k > q \text{ 时, } \gamma_k = 0.$$

注意：对于 AR 和 ARMA 模型来说，由于永远不会精确地等于零，因而模型表现为拖尾而不是截尾。可见截尾性是 MA 模型所特有。那么，怎样识别 AR 模型呢？这就需要研究偏自相关函数。

14.4.4 偏自相关函数

对于 $k=1, 2, \dots$ ，分别用 $X_{t-k}, X_{t-k+1}, \dots, X_{t-1}$ 对 X_t 作最小方差估计，即选择系数 ϕ_{kj} ($j=1, 2, \dots, k$)，使得

$$\delta = E \left(X_t - \sum_{j=1}^k \phi_{kj} X_{t-j} \right)^2 = \gamma_0 - 2 \sum_{j=1}^k \phi_{kj} \gamma_j + \sum_{i,j=1}^k \phi_{ki} \phi_{kj} \gamma_{j-i}$$

达到极小，“使残差的方差达到极小的 k 阶自回归模型的第 k 项系数”，又称为偏自相关系数。

下面来说明为什么称为偏自相关函数。

令 $\frac{\partial \delta}{\partial \phi_{ki}} = 0$, 得

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{bmatrix},$$

[illegible][illegible]

当用样本自相关函数代替 Yule-Wolker 方程中的理论自相关函数后,得到的偏自相关系数就称为样本偏自相关函数。

$$\delta = E[X_t - (\phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \dots + \phi_{kk}X_{t-k})]^2$$

对于 AR 模型来说, 第 k 个偏自相关系数 ϕ_{kk} 就是 AR 模型中 X_{t-k} 的回归系数, 那么, 对于 $AR(p)$ 模型, 有

$$\phi_{11} = \phi_1, \quad \phi_{22} = \phi_2, \quad \dots, \quad \phi_{nn} = \phi_n, \quad \phi_{kk} = 0, \quad k > p.$$

35

14.5 平稳时间序列模型的建立

本节讨论平稳时间序列模型的拟合问题,也就是从观察到的有限长度的平稳序列样本出发,通过模型的识别、模型的定阶、模型的参数估计等步骤建立起适合序列的模型。

针对序列的均值是未知的问题,有两种处理办法:一种是用样本均值 \bar{X} 作为序列均值 μ 的估计,建模前先用样本数据减去其均值,然后对所得到的零均值序列建模。另一种较为恰当的方法是把序列均值当作一个未知参数来估计,模型可写作

$$(X_t - \mu) - \phi_1(X_{t-1} - \mu) - \cdots - \phi_n(X_{t-n} - \mu) = a_t - \theta_1 a_{t-1} \cdots - \theta_m a_{t-m}, \quad (64)$$

可变形为

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_n X_{t-n} = \theta_0 + a_t - \theta_1 a_{t-1} \cdots - \theta_m a_{t-m}, \quad (65)$$

其中 $\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_n)$ 。

14.5.1 模型识别

所谓的模型识别,就是对一个观察序列,从各种模型族中选择一个与其实过程相吻合的模型结构。本节介绍 Box-Jenkins 模型识别方法,即根据样本自相关、偏自相关函数的截尾、拖尾性初步判断 X_t 所适合的模型类型。

这种方法简单易懂,但精度不够,尤其当样本序列未达到足够长度时,其精度更不理想。

零均值平稳序列的自相关函数和偏自相关函数的统计特性如表 14.14 所列。

表 14.14 零均值平稳序列的统计特性

模型	AR(n)	MA(m)	ARMA(n, m)
自相关函数	拖尾	截尾	拖尾
偏自相关函数	截尾	拖尾	拖尾

1. 若 ρ_k 序列在 q 步截尾(即 $k < q$ 时, ρ_k 不显著地接近于零,而当 $k > q$ 时, ρ_k 显著地等于零)并且 ϕ_{kk} 序列被负指数函数控制收敛到零,即可判断 X_t 为 MA(q) 序列。

具体操作是:若 $k > q$ 时应有 $\rho_k = 0$, 则 $\hat{\rho}_k$ 应渐近于正态分布

$$\hat{\rho}_k \sim N\left(0, \frac{1}{N} (1 + 2 \sum_{l=1}^q \rho_l^2)\right). \quad (66)$$

因此有

$$P\left\{|\hat{\rho}_k| \leq \frac{1}{\sqrt{N}} (1 + 2 \sum_{l=1}^q \rho_l^2)^{\frac{1}{2}}\right\} = 68.3\%, \quad (67)$$

$$P\left\{|\hat{\rho}_k| \leq \frac{2}{\sqrt{N}} (1 + 2 \sum_{l=1}^q \rho_l^2)^{\frac{1}{2}}\right\} = 95.5\%. \quad (68)$$

对每个 $k > 0$, 分别检验 $\hat{\rho}_{k+1}, \hat{\rho}_{k+2}, \dots, \hat{\rho}_{k+M}$ (通常取 $M = \lceil \sqrt{N} \rceil$ 或 $M = \lceil N/10 \rceil$) 中满足

$$|\hat{\rho}_{k+i}| \leq \frac{1}{\sqrt{N}} (1 + 2 \sum_{l=1}^q \hat{\rho}_l^2)^{\frac{1}{2}}, \quad i = 1, 2, \dots, M$$

的比例是否达到 68.3%。若 $k = 1, 2, \dots, q-1$ 都未达到, 而 $k = q$ 达到了, 就说 ρ_k 在 q 步截尾。同理也可以利用 (68) 式进行判断。

2. 若 ϕ_{kk} 序列在 n 步截尾, 并且 ρ_k 序列被负指数函数控制收敛到零, 则可判断 X_t 为 AR(n) 序列。

具体操作步骤是:如果 $k > n$ 时应有 $\phi_{kk} = 0$, 则 $\hat{\phi}_{kk}$ 的分布渐近于 $N(0, 1/N)$, 于是有

$$P\left\{|\hat{\phi}_{kk}| > \frac{1}{N}\right\} = 31.7\%, \quad P\left\{|\hat{\phi}_{kk}| > \frac{2}{N}\right\} = 4.5\%.$$

对每个 $k > 0$, 分别检验 $\hat{\phi}_{k+1, k+1}, \hat{\phi}_{k+2, k+2}, \dots, \hat{\phi}_{k+M, k+M}$ 中满足 $|\hat{\phi}_{k+i, k+i}| > \frac{1}{\sqrt{N}}$, $i = 1, 2, \dots, M$ 的个数是否超过了 31.7%, 若在 $k = 1, 2, \dots, n-1$ 时都超过了, 而 $k = n$ 时没有超过, 就可以认为 ϕ_{kk} 在 n 步截尾。

3. 若 ρ_k 序列和 ϕ_{kk} 序列都不截尾, 但都被负指数函数控制收敛到零, 则 X_t 很可能是 ARMA

模型。

注 14.1 若 ρ_k 序列和 ϕ_{kk} 序列无以上特征，而是出现了缓慢衰减或周期性衰减等情况，说明序列不是平稳的。

例 14.1 表 14.15 是一磨轮剖面资料的数据，共 250 个。试对该序列建立合适的时间序列模型。

表 14.15 磨轮剖面资料（横读）

13.5	4.0	4.0	4.5	3.0	3.0	10.0	10.2	9.0	10.0
8.5	7.0	10.5	7.5	7.0	10.5	9.5	7.0	12.0	13.5
12.5	15.0	13.0	11.0	9.0	10.5	10.5	11.5	10.5	9.0
8.2	8.5	9.2	8.5	10.0	14.5	13.0	2.0	6.0	6.0
11.0	9.5	12.5	13.8	12.0	12.0	12.0	13.0	12.0	14.0
14.5	13.5	12.3	7.0	7.0	7.0	6.5	12.5	15.0	12.5
11.6	11.0	10.0	8.5	3.0	11.5	11.5	11.5	11.0	9.0
2.5	7.0	6.0	6.6	14.0	11.0	9.0	6.5	4.0	6.0
12.0	11.0	12.0	12.5	12.5	13.6	13.0	8.0	6.5	6.8
6.0	7.2	10.2	8.0	7.5	11.0	11.8	11.8	6.5	8.0
9.0	8.0	8.0	9.0	9.5	10.0	9.0	12.0	13.5	13.8
15.0	12.5	11.0	11.5	14.5	11.5	11.8	13.0	15.0	14.5
13.0	9.0	11.0	9.0	10.0	14.0	13.5	3.0	2.2	6.0
8.0	9.0	9.0	9.0	7.0	6.0	6.5	7.0	7.5	8.5
9.0	9.5	10.0	11.5	11.2	12.5	11.6	8.0	7.0	6.0
6.0	6.0	9.0	12.0	13.5	13.0	3.5	1.8	1.6	7.5
8.0	7.9	11.6	12.5	10.5	8.0	9.0	11.6	11.8	12.6
10.2	10.0	5.0	7.0	-1.0	0.0	0.0	3.0	11.0	12.0
12.2	11.0	8.0	7.0	5.5	10.0	11.5	7.0	4.0	7.0
7.0	10.0	9.0	8.0	10.0	13.0	10.0	6.5	11.0	13.0
13.0	14.0	13.0	12.5	12.0	9.0	8.5	7.0	8.5	10.0
8.0	4.0	3.0	10.0	13.0	13.0	13.0	12.0	11.0	11.0
11.0	14.5	14.0	14.0	13.5	10.0	9.5	10.0	12.5	10.0
9.0	9.0	4.0	3.0	6.0	5.0	7.0	6.0	5.0	8.5
10.5	11.1	11.0	10.0	11.2	8.0	2.5	5.0	13.2	14.0

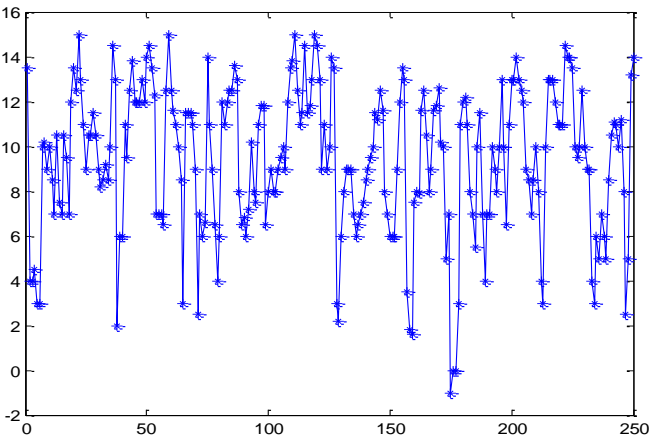


图 14.33 磨轮剖面资料

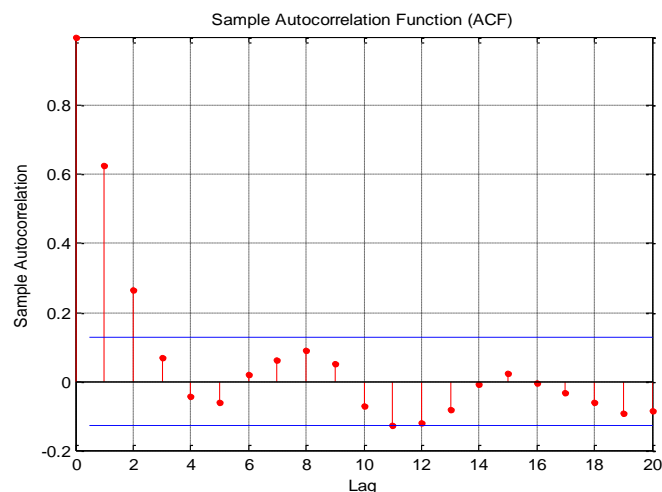


图 14.34 自相关函数图

从数据图可初步断定该序列是平稳的，且样本均值显著非零。用 Y_t 表示原序列， \bar{Y} 表示样本均值， $\bar{Y}=9.4188$ 。接下来考虑序列 $X_t = Y_t - \bar{Y}$ 。用 MATLAB 软件计算 X_t 的自相关函数 (ACF) 和偏相关函数 (PACF) 如图 14.34 和图 14.35。可以发现 $\hat{\phi}_{kk}$ 于 $k=2$ 以后波动于零值附近，且满足 $|\hat{\phi}_{kk}| > \frac{2}{\sqrt{250}}$ 的值未出现，可以认为 $\hat{\phi}_{kk}$ 于 $k=2$ 后截尾。又注意到 $\hat{\rho}_k$ 有收敛到零的趋势，可以认为是拖尾的，故可初步断定序列适合 AR 模型。

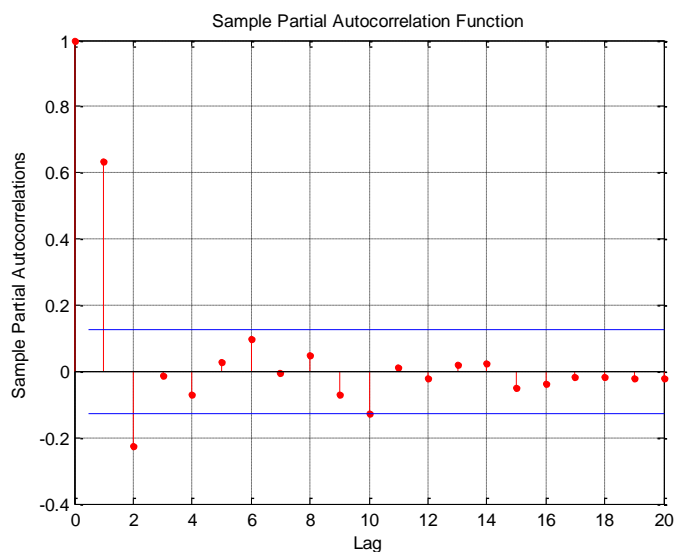


图 14.35 偏相关函数图

计算的 MATLAB 程序如下

```
clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
plot(yt,'*-') %画原始数据散点图
ar=autocorr(xt) %计算自相关函数
figure, autocorr(xt) %画自相关函数
pr=parcorr(xt) % 计算偏相关函数
figure, parcorr(xt) %画偏相关函数
```

14.5.2 模型定阶

1. 残差方法图定阶法

在 AR 模型的情况下，通过假设检验来确定阶数。也就是以 $AR(n)$ 模型作为零假设，而以 $AR(n+1)$ 模型作为对立假设，然后增大 n 的值，直到检验结果不显著未知。

假定模型是有限阶的自回归模型，如果选择的 n 小于真正的阶数，则是一种不足拟合，因而剩余平方和 Q 必然偏大，残差方差 $\hat{\sigma}_a^2$ 将比真正模型的残差方差 σ_a^2 大。这是因为我们把模型中本来应有的一些高阶项略去了，而这些项对减小残差方差是有明显贡献的。另一方面，如果 n 已经达到了真值，那么再进一步增加阶数就是过度拟合，这并不会使 $\hat{\sigma}_a^2$ 有显著减小，甚至还略有增加。

这样用一系列阶数逐渐递增的模型进行拟合，每次都求出 $\hat{\sigma}_a^2$ ，然后画出 n 和 $\hat{\sigma}_a^2$ 的图形，称为残差方差图。开始时， $\hat{\sigma}_a^2$ 会下降，当达到 n 的真值后渐趋平缓。 $\hat{\sigma}_a^2$ 停止下降转为平缓时所对应的阶数就是模型的适合阶数。残差方差的估计式为

$$\hat{\sigma}_a^2 = \frac{\text{模型的剩余平方和}}{\text{实际观察值个数} - \text{模型的参数个数}} \quad (69)$$

式中的“实际观察值个数”是指拟合模型时实际使用的观察值项数。例如，对于一个具有 N 个观察值的序列，若拟合 $AR(n)$ 模型，则实际使用的观察值最多为 $N-n$ 项。式中的“参数个数”是指所建立的模型中实际包含的参数个数。若模型中不含均值项 μ ，则模型的参数个数就等于模型的阶数。具体来说，对于 N 个观察值的序列，相应于 AR, MA, ARMA 三种模型的残差方差估计式分别为

$$\text{AR 模型: } \hat{\sigma}_a^2(n) = \frac{Q(\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_n)}{(N-n) - (n+1)};$$

$$\text{MA 模型: } \hat{\sigma}_a^2(m) = \frac{Q(\hat{\mu}, \hat{\theta}_1, \dots, \hat{\theta}_m)}{N - (m+1)};$$

$$\text{ARMA 模型: } \hat{\sigma}_a^2(n, m) = \frac{Q(\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_n, \hat{\theta}_1, \dots, \hat{\theta}_m)}{(N-n) - (n+m+1)}.$$

这种方法的最大不足，就是没有一个定量的准则，判定的阶数具有主观性。

图 14.36 是对例 14.1 中数据拟合 $AR(n)$ 模型所得的残差方差图。由此可以看出，模型阶数 n 从 1 升至 2，残差方差大幅度减少； n 继续升至 5，残差方差非但不再减少，反而略有增加；继续升高模型阶数，在 $n=6$ 时，残差方差略有回落后继续增加。虽然在 $n=6$ 时残差方差也较小，但从建模的“约简”性原则出发我们初步判断合适的模型为 $AR(2)$ 。

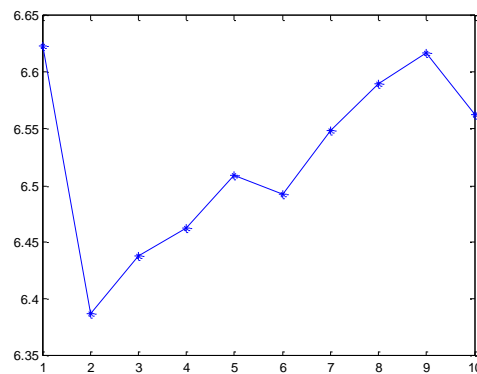


图 14.36 残差方差图

计算的 MATLAB 程序如下

```
clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
N=length(yt);
for i=1:10
    m=ar(xt,i) %拟合模型
```

```

xp=predict(m,xt); %计算预测值
%res=resid(m,xt); %计算残差
Q(i)=sum((xp-xt).^2)/(N-i-1); %计算残差方差
end
plot(Q,'*-')

```

2. 自相关函数和偏自相关函数定阶法

对于单纯的 AR 或 MA 模型，用 ACF 和 PACF 判定模型形式的同时也初步断定了模型的阶数。对于 ARMA 模型来说，需要用其他的方法来判定。需要指出的是，即使是 AR 或 MA 模型，用 ACF 和 PACF 给出的模型的阶数也只是初步的，还需要结合其他方法确定出精确阶数。

3. F 检验定阶法

下面把 F 检验法用于 ARMA(n, m) 模型定阶。

采用过拟合的办法。先对观察数据用 ARMA(n, m) 模型进行拟合，再假定 ϕ_n, θ_m 高阶系数中某些取值为零，用 F 检验法判定降阶后模型与 ARMA(n, m) 模型之间是否存在显著性差异。如果差异显著，说明模型的阶数仍存在升高的可能性；若差异不显著，则模型的阶数可以降低。低阶模型和高阶模型之间的差异用残差平方和来衡量。

例如，假定原假设为 $H_0: \phi_n = 0, \theta_m = 0$,

记 Q_0 为 ARMA(n, m) 模型的残差平方和， Q_1 为 ARMA($n-1, m-1$) 模型的残差平方和，则

$$F = \frac{Q_1 - Q_0}{2} \bigg/ \frac{Q_0}{N - n - (m + n)} \sim F(2, N - 2n - m),$$

其中 N 为样本长度， $\gamma = m + n$ 是模型参数的总个数， $S = 2$ 是被检验的参数的个数。如果 $F > F_\alpha$ ，则 H_0 不成立，模型阶数仍有上升的可能；否则 H_0 成立，即 ARMA($n-1, m-1$) 是适合的模型。

例 14.1 的定阶问题。对序列 X_t 分别拟合 1 至 3 阶 AR 模型，残差平方和 Q_1, Q_2, Q_3 分别为 1619.3805, 1474.0022, 1473.5843。

对于 AR(1) 与 AR(2) 模型，有

$$F = \frac{Q_1 - Q_2}{1} \bigg/ \frac{Q_2}{250 - 4} = 24.2626.$$

给定显著性水平 $\alpha = 0.05$ ，查 F 分布表得 $F_\alpha(1, 250 - 4) = 3.8792$ ，故 $F > F_\alpha$ ，说明 AR(1) 和 AR(2) 之间存在显著的差异，模型阶数还有上升的可能。

对于 AR(2) 与 AR(3) 模型，有

$$F = \frac{Q_2 - Q_3}{1} \bigg/ \frac{Q_3}{250 - 6} = 0.0692.$$

同理查 F 分布表得 $F_\alpha(1, 250 - 6) = 3.8799$ ，故 $F < F_\alpha$ ，说明 AR(3) 与 AR(2) 没有显著差异。

计算的 MATLAB 程序如下：

```

clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
N=length(yt);
for i=1:3
    m=ar(xt,i); %拟合模型
    xp=predict(m,xt); %计算预测值
    %res=resid(m,xt); %计算残差
    Q(i)=sum((xp-xt).^2); %计算残差平方和
end
Q
F1=(Q(1)-Q(2))/(Q(2)/(N-4)) %计算 F 统计量
Falpha1=finv(0.95,1,N-2) %求 F 分布的 alpha 分位数
F2=(Q(2)-Q(3))/(Q(3)/(N-6)) %计算 F 统计量

```


Falpha2=finv(0.95,1,N-6) %求 F 分布的 alpha 分位数

4. 最佳准则函数定阶法

基本思想：既要考虑用某一模型拟合时对原始数据的接近程度，又要考虑模型中所含待估参数的个数，确定出一个准则函数。建模时按照准则函数的取值确定模型的优劣，选择使准则函数达到极小的最佳模型。

准则函数方法首先由日本学者赤池 (Akaike) 提出。1971 年他提出了一种识别 AR 模型阶数的准则，称为最小最终预报误差准则，简称为最小 FPE (Final Prediction Error) 准则。1977 年赤池又将此方法推广到辨识 ARMA 模型阶数，称为最小信息准则或 AIC (A-Information Criterion) 准则。近年来 AIC 准则得到广泛应用，进而又推广为 BIC 准则等等。

(1) FPE 准则

FPE 准则是由模型的预报误差来判明自回归模型的阶数是否恰当，其判据就是最终预报误差最小。

基本原理：若数据所符合的真实模型是 $AR(n)$ ，而我们用 $AR(p)$ 模型进行拟合。则不论是缺参数拟合 ($p < n$)，还是超参数拟合 ($p > n$)，都会使预报误差的方差增大。要找出 AR 模型的合理阶数，就是寻找使一步预报误差方差达到极小的 AR 模型的阶数。

具体做法：设 $\{X_t: 1 \leq t \leq N\}$ 所适合的真实模型为

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_n X_{t-n} + a_t, \quad (70)$$

其中 $Ea_t = 0$, $Ea_t^2 = \sigma_a^2$ 。

设 ϕ_i 的估计值为 $\hat{\phi}_i$ ($1 \leq i \leq n$)，用 $\hat{X}_{t-1}(1)$ 表示 $t-1$ 时对 X_t 的一步预报值，应有

$$\hat{X}_{t-1}(1) = \hat{\phi}_1 X_{t-1} + \hat{\phi}_2 X_{t-2} + \cdots + \hat{\phi}_n X_{t-n}. \quad (71)$$

可以证明，一步预报方差为

$$E[X_t - \hat{X}_{t-1}(1)]^2 \approx (1 + n/N) \sigma_a^2,$$

用 $\hat{\sigma}_a^2$ 表示以 $\hat{\phi}_i$ ($1 \leq i \leq n$) 为参数的拟合模型的残差方差。可以证明，当 N 足够大时有 $E[\hat{\sigma}_a^2] \approx (1 - n/N) \sigma_a^2$ 。换句话说，当 N 足够大时 $\hat{\sigma}_a^2 / (1 - n/N)$ 是 σ_a^2 的无偏估计。于是得到

$$E[X_t - \hat{X}_{t-1}(1)]^2 \approx (1 + n/N) \frac{\hat{\sigma}_a^2}{1 - n/N} = \frac{N + n}{N - n} \hat{\sigma}_a^2, \quad (72)$$

式中第一个因子 $(N + n)/(N - n)$ 随着阶数的增加而增大；第二个因子 $\hat{\sigma}_a^2$ 随着阶数的增加而减少，反映了模型的拟合程度。

对于以 $\hat{\phi}_i$ ($1 \leq i \leq n$) 为参数的 $AR(n)$ 模型应有 $\hat{\sigma}_a^2 = \gamma_0 - \sum_{i=1}^n \hat{\phi}_i \gamma_i$ ，其中 $\gamma_0, \gamma_1, \dots, \gamma_n$ 是数据序列的样本协方差函数在不同滞后时的值。由此，以 $\hat{\phi}_i$ ($1 \leq i \leq n$) 为参数的 $AR(n)$ 模型的最终预报误差定义为

$$FPE(n) = (1 + \frac{n}{N})(1 - \frac{n}{N})^{-1} (\gamma_0 - \sum_{i=1}^n \hat{\phi}_i \gamma_i). \quad (73)$$

在具体应用时，通常是根据观察数据从低阶到高阶建立 AR 模型，并计算出相应的 FPE 值，由此确定出使 FPE 达到极小的 AR 模型。拟合的最高阶数 $M(N)$ 通常取 $\left[\frac{1}{3}N \right] \sim \left[\frac{2}{3}N \right]$ 之间的某个整数。故最佳模型的阶数 n_0 满足

$$FPE(n_0) = \min_{1 \leq n \leq M(N)} FPE(n).$$

(2) AIC 准则

设 $\{X_t, 1 \leq t \leq N\}$ 为一随机序列，用 $AR(n)$ 模型来拟合， $\hat{\sigma}_a^2$ 表示拟合残差方差。定义 AIC 准则函数如下

$$AIC(n) = \ln \hat{\sigma}_a^2 + \frac{2n}{N}. \quad (74)$$

当阶数 n 增高时，第一项减少，第二项增大。可以想象，从 $n=1$ 开始，逐次增加模型阶数对数据进行自回归模型拟合时， $AIC(n)$ 的值有下降的趋势，因为这时起决定作用的是模

型残差方差。当达到某一阶数 n_0 时, $AIC(n_0)$ 达到极小。随后随着模型阶数的继续增高, 残差方差改进甚微, 于是模型的阶数起了关键的作用, $AIC(n)$ 的值随 n 而增长。

对事先给定的最高阶数 $M(N)$, 若 $AIC(n_0) = \min_{1 \leq n \leq M(N)} AIC(n)$, 我们便取 n_0 为最佳自回归模型阶数。

AIC 准则可类似用于 ARMA 模型定阶。假定我们用 ARMA (n, m) 模型对随机序列 $\{X_t: 1 \leq t \leq N\}$ 进行拟合, $\hat{\sigma}_a^2(n, m, \mu)$ 是残差方差。如果序列均值也是待估参数, 那么参数的个数为 $n+m+1$ 。定义 AIC 函数为

$$AIC(n, m, \mu) = \ln \hat{\sigma}_a^2(n, m, \mu) + \frac{2(n+m+1)}{N}.$$

取不同的 n, m 及模型参数对 X_t 进行拟合, 并计算相应的 AIC 值。然后改变模型的阶数及参数, 使 AIC 值达到最小的模型, 认为是最佳模型。

(3) BIC 准则及其他准则

定义与 AIC 函数类似的 BIC 函数

$$BIC(n) = \ln \hat{\sigma}_a^2(n) + \frac{n}{N} \ln N.$$

若某一 n'_0 满足 $BIC(n'_0) = \min_{1 \leq n \leq M(N)} BIC(n)$, 则取 n'_0 为最佳阶数。

说明: BIC 函数的第二项中用 $\ln N$ 代替了 AIC 函数的第二项中的系数 2。一般来说, $\ln N > 2$, 因此 AIC 达到极小时所对应的阶数往往比 BIC 准则所定出的阶数高, 即一般有 $n'_0 \leq n_0$ 。

其他准则: 如 $BIC_1(n) = \ln \hat{\sigma}_a^2(n) + C \frac{n}{N} \ln(\ln N)$, 其中 C 是给定常数。定义不同的准则, 其目的是为了对拟合残差与参数个数之间进行不同的权衡, 以体现使用者对残差与阶数二者重要性的不同侧重。当然, 用不同准则挑出的最优模型, 其渐近性质是不同的。例如当 $N \rightarrow \infty$ 时, 用 AIC 准则挑选模型的阶数往往比真实模型阶数高, 而用 BIC 准则确定的最佳模型的阶数往往与真实模型阶数相一致。

对例 14.1 逐个拟合 AR (n) 模型, 并求出相应的 FPE (n) , AIC (n) , 结果如图 14.37 所示, FPE 准则下最佳模型阶数为 2, AIC 准则下最佳模型阶数也为 2。

本节介绍的几种定阶方法对 AR 和 MA 模型比较有效, 但对 ARMA 模型不太方便。1984 年 Tiao 和 Tsay 提出了一种可同时鉴定平稳的 ARMA 和非平稳 ARIMA 模型阶数的延伸自相关函数方法, 有兴趣的读者可参阅相关文献。

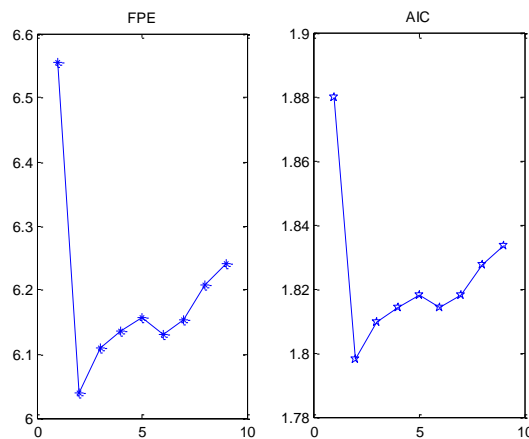


图 14.37 最佳准则函数定阶图

画图 37 的 MATLAB 程序如下

```
clc, clear
yt=load('data52.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
```

```

N=length(yt);
for i=1:9
    m=ar(xt,i) %拟合模型
    fp(i)=fpe(m); %计算 FPE 的值
    am(i)=aic(m); %计算 AIC 的值
end
subplot(1,2,1), plot(fp,'*-'), title('FPE')
subplot(1,2,2), plot(am,'P-'), title('AIC')

```

14.5.3 模型参数估计

模型参数估计方法有矩估计法、最小二乘估计法和极大似然估计法等。下面简要介绍参数估计的基本思路。

1. 模型参数的相关矩估计

(1) AR 模型参数的矩估计

Yule-Walker 方程的矩阵形式

$$\begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix},$$

这里也可以把 ρ 换成 γ 。

$$\begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}. \quad (75)$$

由 $\gamma_k = E(X_t X_{t-k}) = E[X_{t-k}(\phi_1 X_{t-1} + \cdots + \phi_n X_{t-n} + a_t)]$, 当 $k=0$ 时,

$$\gamma_0 = \phi_1 \gamma_1 + \cdots + \phi_n \gamma_n + \sigma_a^2, \quad (76)$$

即

$$\sigma_a^2 = \gamma_0 - \phi_1 \gamma_1 - \cdots - \phi_n \gamma_n. \quad (77)$$

用 $\hat{\phi}_1, \dots, \hat{\phi}_n$ 代替 ϕ_1, \dots, ϕ_n , $\hat{\gamma}_1, \dots, \hat{\gamma}_n$ 代替 $\gamma_1, \dots, \gamma_n$, 代入 (77) 式, 得

$$\hat{\sigma}_a^2 = \hat{\gamma}_0 - \hat{\phi}_1 \hat{\gamma}_1 - \cdots - \hat{\phi}_n \hat{\gamma}_n = \hat{\gamma}_0 (1 - \sum_{i=1}^n \hat{\phi}_i \hat{\rho}_i). \quad (78)$$

例如, 对于 AR(1) 模型来说, Yule-Walker 方程为 $\gamma_1 - \phi_1 \gamma_0 = 0$, $\phi_1 = \gamma_1 / \gamma_0 = \rho_1$, $\hat{\phi}_1 = \hat{\rho}_1$, $\hat{\sigma}_a^2 = \hat{\gamma}_0 (1 - \hat{\rho}_1^2)$ 。

(2) MA 模型参数的矩估计

根据下列方程组估计参数

$$\begin{cases} \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_m^2) \sigma_a^2, \\ \gamma_k = (-\theta_k + \theta_{k+1} \theta_1 + \theta_{k+2} \theta_2 + \cdots + \theta_m \theta_{m-k}) \sigma_a^2, k=1, 2, \dots, m, \end{cases} \quad (79)$$

求解非线性方程组 (79) 常用的解法有三种方法, 直接解法, 线性迭代法和 牛顿-拉普森 (Newton-Raphson) 算法。

(3) ARMA 模型参数的矩估计

对于低阶的模型 (比如 ARMA (1, 1)), 也可以利用模型的自协方差函数和模型参数之间的非线性关系直接求解, 高阶的可以通过以下步骤得到近似解。

第一步, 给出 AR 部分 ϕ_1, \dots, ϕ_n 的矩估计。在 ARMA (n, m) 模型

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \cdots - \theta_m a_{t-m}$$

两端同乘以 X_{t-k} 并取期望, 得

$$\gamma_k - \phi_1 \gamma_{k-1} - \cdots - \phi_n \gamma_{k-n} = E(X_{t-k} a_t) - \theta_1 E(X_{t-k} a_{t-1}) - \cdots - \theta_m E(X_{t-k} a_{t-m}),$$

由于

$$E(X_{t-k}a_{t-l}) = E\left(\sum_{j=1}^{\infty} G_j a_{(t-k)-j} a_{t-l}\right) = \begin{cases} G_{l-k} \sigma_a^2, & k \leq l, \\ 0, & k > l, \end{cases} \quad (0 \leq l \leq m),$$

所以

$$\begin{aligned} k=0, & \quad 1 - \phi_1 \gamma_1 - \cdots - \phi_n \gamma_n = (1 - \theta_1 G_1 - \cdots - \theta_m G_m) \sigma_a^2; \\ k=1, & \quad \gamma_1 - \phi_1 \gamma_0 - \cdots - \phi_n \gamma_{n-1} = (-\theta_1 - \theta_2 G_1 - \cdots - \theta_m G_{m-1}) \sigma_a^2; \\ k=2, & \quad \gamma_2 - \phi_1 \gamma_1 - \cdots - \phi_n \gamma_{n-2} = (-\theta_2 - \theta_3 G_1 - \cdots - \theta_m G_{m-2}) \sigma_a^2; \\ & \dots\dots\dots \\ k=m, & \quad \gamma_m - \phi_1 \gamma_{m-1} - \cdots - \phi_n \gamma_{n-m} = -\theta_m \sigma_a^2; \\ k>m, & \quad \gamma_k - \phi_1 \gamma_{k-1} - \cdots - \phi_n \gamma_{k-n} = 0; \end{aligned}$$

或有 $k > m$, $\rho_k - \phi_1 \rho_{k-1} - \cdots - \phi_n \rho_{k-n} = 0$.

用 $\hat{\rho}_k$ 代替上式中的 ρ_k , k 取 $m+1, m+2, \dots, m+n$ 便可得到

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_n \end{bmatrix} = \begin{bmatrix} \hat{\rho}_m & \hat{\rho}_{m-1} & \cdots & \hat{\rho}_{m-n+1} \\ \hat{\rho}_{m+1} & \hat{\rho}_m & \cdots & \hat{\rho}_{m-n+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{m+n-1} & \hat{\rho}_{m+n-2} & \cdots & \hat{\rho}_m \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_{m+1} \\ \hat{\rho}_{m+2} \\ \vdots \\ \hat{\rho}_{m+n} \end{bmatrix}.$$

第二步: 设 $y_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_n X_{t-n}$, 计算其协方差函数

$$\begin{aligned} \gamma_k(y_t) &= E[y_t y_{t+k}] \\ &= E[(X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_n X_{t-n})(X_{t+k} - \hat{\phi}_1 X_{t+k-1} - \cdots - \hat{\phi}_n X_{t+k-n})] = \sum_{i,j=0}^n \hat{\phi}_i \hat{\phi}_j \gamma_{k+j-i}, \end{aligned}$$

其中 $\phi_0 = -1$ 。再以 $\hat{\gamma}_k$ 代替 γ_k , 便计算出 $\gamma_k(y_t)$ 。

第三步: 把 y_t 近似看出 MA(m) 序列, $y_t \approx a_t - \theta_1 a_{t-1} - \cdots - \theta_m a_{t-m}$, 利用前面介绍的 MA 参数估计方法解下列方程

$$\begin{cases} \gamma_0(y_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_m^2) \sigma_a^2, \\ \gamma_k(y_t) = (-\theta_k + \theta_{k+1} \theta_1 + \theta_{k+2} \theta_2 + \cdots + \theta_m \theta_{m-k}) \sigma_a^2, k=1, 2, \dots, m, \end{cases}$$

其值就是 ARMA 模型移动平均参数和 σ_a^2 的矩估计。

说明: 对 ARMA 模型来说, 用矩估计方法得到的参数估计精度较差。

2. 最小二乘估计

最小二乘估计是一种古老而又广泛使用的估计方法, 是以实际值和模型拟合值之差的平方和达到最小为原则对参数进行估计。

对于 AR 模型, 可以采用与普通多元回归参数最小二乘估计类似的方法得到其参数的线性最小二乘估计, 并且当样本容量充分大时, 矩估计与最小二乘估计结果十分接近。

对于 ARMA(n, m) 模型 ($m \neq 0$), 滑动平均部分的滞后变量 a_{t-j} ($j=1, 2, \dots, m$) 是不可观测的, 利用 ARMA 模型的逆转形式可将其表示为关于观测值 X_t 为线性、而关于模型参数 (ϕ, θ) 为非线性的函数, 因此模型参数只能用非线性最小二乘法进行估计。

3. 极大似然估计

最小二乘估计不需要样本分布的有关信息, 而极大似然估计是以已知样本数据的分布类型为前提的。实际中最普遍采用的是假定样本数据的分布类型为正态分布。极大似然估计的基本思想是: 模型参数应使样本数据出现的可能性最大。可由样本分布的概率密度函数给出样本数据出现的可能性, 即似然函数, 它表现为模型参数的函数, 通过极大化似然函数就可得到模型的参数估计。

对于 ARMA(n, m) 模型, 似然函数公式复杂, 严格的极大化似然函数非常困难, 实际中经常使用近似的极大似然估计方法。

综上所述, 在 ARMA 模型参数的三种估计方法中。矩估计法相对简单, 计算量小, 但精度低, 只宜作为初估计。但对于 AR 模型, 当样本容量充分大时, 三种估计结果十分接近。最小二乘估计和极大似然估计精度高, 一般称之为模型参数的精估计, 但计算量都较大, 计算复杂。

对例 14.1 的 AR (2) 模型参数进行最小二乘估计, 求得的模型为

$$X_t = 0.7852X_{t-1} - 0.2241X_{t-2},$$

残差方差为 5.919。

计算的 MATLAB 程序如下

```
clc, clear
yt=load('data52.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
N=length(yt); %样本点的个数
n=2; %AR 模型的阶数
m=ar(xt,n,'ls') %拟合模型
pa=predict(m,xt) %计算模型的预测值
```

14.5.4 模型的适应性检验

所谓模型的适应性是指一个 ARMA 模型已经完全或基本上解释了系统的动态性 (即数据的相关性), 从而模型中的 a_t 是独立的。所以模型的适应性检验实际上就是 a_t 的独立性检验。

残差序列 $\{a_t\}$ 可由估计出来的模型计算得到。如果估计的模型是 ARMA 形式, 即

$$X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_n X_{t-n} = a_t - \hat{\theta}_1 a_{t-1} - \cdots - \hat{\theta}_m a_{t-m},$$

通常采用递推方法得到残差序列 $\{a_t\}$:

$$\hat{a}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_n X_{t-n} + \hat{\theta}_1 \hat{a}_{t-1} + \cdots + \hat{\theta}_m \hat{a}_{t-m}, \quad t = 0, 1, 2, \dots, \quad (80)$$

递推的初值 $\hat{a}_0, \hat{a}_{-1}, \dots, \hat{a}_{1-m}, X_0, X_{-1}, \dots, X_{1-n}$ 可取为零。计算出 $\{\hat{a}_t\}$ 后就可进行模型的适应性检验。

1. 相关函数法

相关函数是通过计算并考察 $\{\hat{a}_t\}$ 的自相关函数来判断残差序列的独立性。设 $\hat{\rho}_k$ 表示 $\{\hat{a}_t\}$ 序列的自相关函数, 即

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^n \hat{a}_t^2}. \quad (81)$$

如果残差序列 $\{\hat{a}_t\}$ 为白噪声序列, 可以证明, 样本个数充分大时, $\hat{\rho}_k$ 是互不相关的, 且近似于正态分布, 即

$$\hat{\rho}_k \sim N(0, 1/N),$$

因此, 如果

$$|\hat{\rho}_k| \leq 1.96/\sqrt{N},$$

就可以在 0.05 的显著性水平下接受 $\hat{\rho}_k = 0$ 的假设, 认为 $\{\hat{a}_t\}$ 是独立的。

2. χ^2 检验法

上面是对 $\{\hat{a}_t\}$ 的自相关函数 $\hat{\rho}_k$ 单个进行检验。在有些情况下, 由于偶然因素, 可能有个别的 $\hat{\rho}_k$ 不满足 $|\hat{\rho}_k| \leq 1.96/\sqrt{N}$, 此时也可以将自相关函数 $\hat{\rho}_k$ 放在一起整体进行检验, 为此, 构造如下的 Box—Pierce 统计量

$$Q = N \sum_{k=1}^{L(N)} \hat{\rho}_k^2, \quad (82)$$

$L(N)$ 取得充分大, 实际中一般取 $L(N) = [N/10]$ 或 $[\sqrt{N}]$ 。可以证明, 在模型适合的条件下 (82) 近似为 $\chi^2(L(N) - n - m)$ 。在给定的显著性水平 α 下, 如果统计量 $Q \leq \chi_{1-\alpha}^2(L(N) - n - m)$, 就认为模型是适合的, 否则不能认为模型是适合的, 模型还需要进一步修正。

鉴于在模型适合的条件下 χ^2 分布不能对 Box—Pierce 统计量的分布提供足够的近似, Q 统计量的值比在 χ^2 分布下所预期的略为偏小, Ljung 和 Box 提出了一种修正的统计量, 即修正的 Ljung—Box—Pierce 统计量

$$\tilde{Q} = N(N+2) \sum_{k=1}^{L(N)} (N-k)^{-1} \hat{\rho}_k^2, \quad (83)$$

在模型适合的条件下(83)不仅近似为 $\chi^2(L(N)-n-m)$ 分布，而且有

$$E(\tilde{Q}) = L(N) - n - m,$$

因此，利用 Ljung-Box-Pierce 统计量进行模型适应性检验效果更好一些。

对例 14.1 拟合 AR(2) 模型，模型适应性检验的 Ljung-Box-Pierce 统计量计算结果见表 14.16。

表 14.16 Ljung-Box-Pierce 统计量计算结果

$L(250)$	\tilde{Q}	上 0.05 分位数	相伴概率 (p 值)
6	3.9017	12.5916	0.6900
12	10.0163	21.0261	0.6145
18	12.0012	28.8693	0.8472

p 值表明 AR(2) 模型是合适的。残差自相关如图 14.38 所示，残差自相关检验也表明 AR(2) 模型是适合的。

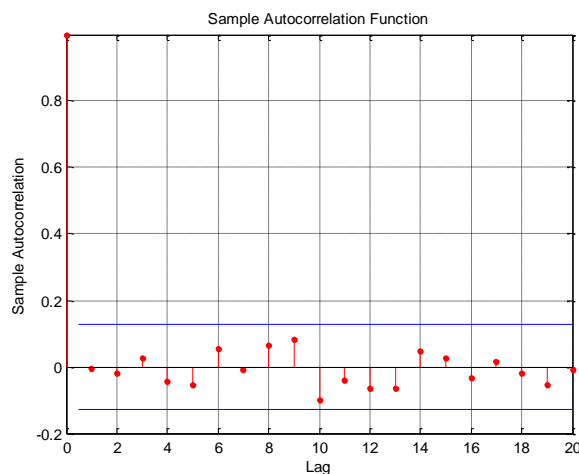


图 14.38 例 14.1 的残差自相关图

计算的 MATLAB 程序如下

```

clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
N=length(yt); %样本点的个数
n=2; %AR 模型的阶数
m=ar(xt,n,'ls') %拟合模型
res=resid(m,xt); %计算残差向量
r=autocorr(res); %计算自相关函数
r(1)=[]; %删除 r(0) 的值
L=[6, 12, 18];
for i=1:3
    Q(i)=0;
    for j=1:L(i)
        Q(i)=Q(i)+N*(N+2)*r(j)^2/(N-j);
    end
end
chi2alpha=chi2inv(0.95,L), %计算上 alpha 分位数的值
%下面是使用 MATLAB 工具箱的命令 lbqtest 做检验

```

```

[h1,pValue1,st1] = lbqtest(res,'lags',6);
[h2,pValue2,st2] = lbqtest(res,'lags',12);
[h3,pValue3,st3] = lbqtest(res,'lags',18);
h=[h1,h2,h3] %显示是否拒绝原假设的标志值, 0 是接受, 1 是拒绝
pValue=[pValue1,pValue2,pValue3] %MATLAB 工具箱计算的概率
P=1-chi2cdf(Q,L) %计算统计量对应的概率
st=[st1,st2,st3] %MATLAB 工具箱计算的 Q 统计量值
Q %显示三个统计量
autocorr(res) %画自相关函数图

```

12.5.5 Pandit-Wu 方法

Box-Jenkins 用样本自相关函数、偏自相关函数作为理论自相关函数、偏自相关函数的近似, 不可避免地会产生一定误差。Pandit-Wu 在 Box-Jenkins 方法的基础上, 经过实践和进一步法展, 与 1977 年提出了系统建模的新方法。这种方法的依据是: 任一平稳序列总可以用 $ARMA(n, n-1)$ 模型来表示。而 $AR(n)$ 模型、 $MA(m)$ 模型以及 $ARMA(n, m)$ ($m \neq n-1$) 模型都是 $ARMA(n, n-1)$ 的特例。其建模思想可以概括为: 逐渐增加模型的阶数, 拟合较高阶的 $ARMA(n, n-1)$ 模型, 直到再增加模型的阶数而剩余平方和不显著减小为止。

主要建模步骤如下:

(1) 零均值化

注意: 如果将序列的均值作为参数估计, 则可省去该过程。

(2) 从 $n=1$ 开始, 逐渐增加模型阶数, 拟合 $ARMA(2n, 2n-1)$ 模型, 即两阶两阶地增加模型阶数。

两阶两阶地增加模型阶数的第一个原因是为了经济起见。以 2 增加模型的阶数不会遗漏模型阶数确实为奇数的模型。事实上, 如果一个模型的阶数为奇数时, 则拟合较高阶数为偶数的模型的第 $2n$ 个自回归系数和第 $2n-1$ 个移动平均系数的绝对值必定相当接近于零, 所以可以删掉较高阶数的偶数阶模型的小参数得到较低奇数阶模型。

(3) 模型适应性检验

模型适应性检验的具体方法已在前面详细阐述, 这里不再重复。

(4) 求最优模型

系统意义上的最优模型不仅是一个适应模型, 而且是一个经济模型。因此还需要检验模型是否包含小参数, 若有, 可用 F 检验判断是否可以删去, 拟合较低阶模型, 进而得到系统意义上的最优模型。

对例 14.1 用 Pandit-Wu 方法建模, 计算输出结果如表 14.17。

表 14.17 对磨轮剖面资料拟合 $ARMA(2n, 2n-1)$ 模型的计算机输出结果

ARMA 阶次	FPE	AIC	剩余平方和	残差方差
(2, 1)	6.4340	1.8619	1473.2	6.0132
(4, 3)	6.5259	1.8774	1427.3	5.9719
(6, 5)	6.6744	1.9024	1393.9	5.9826
(8, 7)	7.0088	1.9551	1401.4	6.1736
(10, 9)	6.9405	1.9505	1311.3	5.9336

F 检验表明 $ARMA(4, 3)$ 和 $ARMA(2, 1)$ 没有显著差异, 从表 14.17 可以看出合适的模型为 $ARMA(2, 1)$, 又 $ARMA(2, 1)$ 中系数 θ_1 非常小, F 检验表示 θ_1 是多余的参数, 故合适的模型为 $AR(2)$ 。

计算的 MATLAB 程序如下

```

clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
N=length(yt); %样本点的个数

```

```

for i=1:5
    order=[2*i,2*i-1];
    m=arimax(xt,order);
    pa=predict(m,xt);
    fp(i)=fpe(m);
    myaic(i)=aic(m);
    cancha(i)=sum((pa-xt).^2);
    fangcha(i)=cancha(i)/(250-6*i+1);
end
fp, myaic, cancha, fangcha
F=(cancha(2)-cancha(1))/4/(cancha(2)/(250-11)) %计算 n=2 时的 F 统计量
Falp=finv(0.95,4,250-11) %计算 F 分布的 alpha 分位数
zm=ar(xt,2)
zres=resid(zm,xt); %计算残差向量
zcancha=sum(zres.^2)
F2=(zcancha-cancha(1))/(cancha(1)/(250-5))
Falp2=finv(0.95,1,250-5)

```

14.5.6 实例

下面所涉及的时间序列都是由非平稳时间序列经平稳化而得到的。

例 14.2 某地区 1984—2005 年各季度的实际国内生产总值数据见表 14.18，对该数据进行平稳化处理（先取对数，再进行一阶差分），平稳化后序列记为 X_t ，共 87 个数据。如图 14.39 所示。

表 14.18 某地区 1984—2005 年各季度的实际国内生产总值（逐行排列）

5699.8	5797.9	5854.3	5902.4	5956.9	6007.8	6101.7	6148.6
6207.4	6232	6291.7	6323.4	6365	6435	6493.4	6606.8
6639.1	6723.5	6759.4	6848.6	6918.1	6963.5	7013.1	7030.9
7112.1	7130.3	7130.8	7076.9	7040.8	7086.5	7120.7	7154.1
7228.2	7297.9	7369.5	7450.7	7459.7	7497.5	7536	7637.4
7715.1	7815.7	7859.5	7951.6	7973.7	7988	8053.1	8112
8169.2	8303.1	8372.7	8470.6	8536.1	8665.8	9773.7	8838.4
8936.2	8995.3	9098.9	9237.1	9315.5	9392.6	9502.2	9671.1
9695.6	9847.9	9836.6	9887.7	9875.6	9905.9	9871.1	9910
9977.3	10031.6	10090.7	10095.8	10138.6	10230.4	10410.9	10502.6
10612.5	10704.1	10808.9	10897.1	10999.3	11089.2	11202.3	11233.5

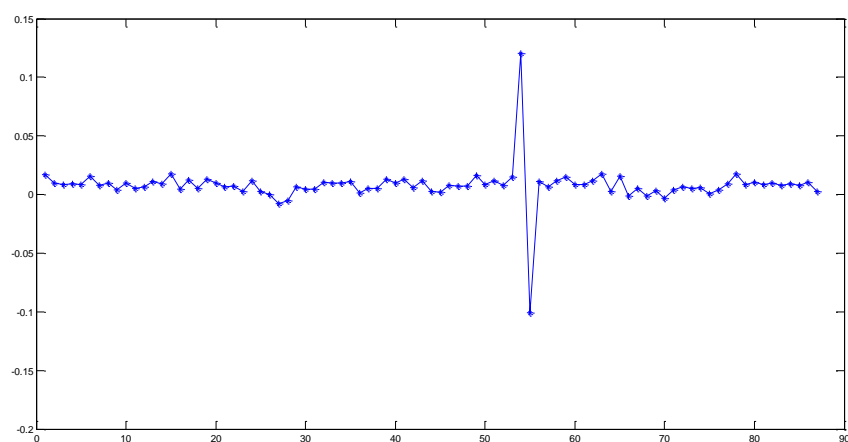


图 14.39 平稳化后的某地区国内生产总值

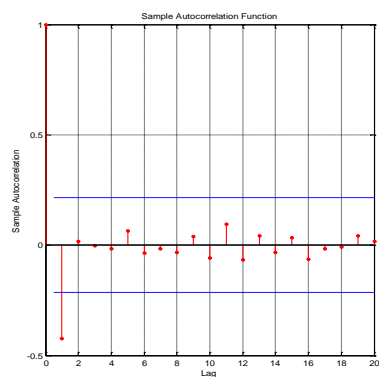


图 14.40 自相关函数图

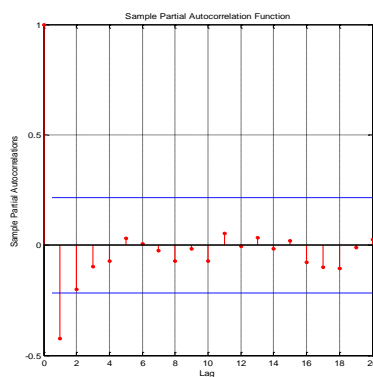


图 14.41 偏自相关函数图

从数据图 14.39 可以看出，该序列是比较平稳的。用 Box-Jenkins 方法建模， X_t 的自相关函数和偏自相关函数如图 14.40 和图 14.41 所示，可以看出 $\hat{\rho}_k$ 序列在 2 步截尾， $\hat{\phi}_{kk}$ 序列缓慢衰减，呈现拖尾现象，故可初步断定序列 X_t 适合 MA (2) 模型，对 X_t 拟合 MA (2) 模型，结果为

$$X_t = 0.0078 + a_t - 0.5363a_{t-1} + 0.02586a_{t-2},$$

其中， $\hat{\sigma}_a^2$ 为 2.4366×10^{-4} 。模型适应性检验的 Ljung-Box-Pierce 统计量见表 14.19。 χ^2 检验表明 MA (2) 模型是适应的。残差自相关函数如图 14.42 所示，残差自相关函数满足 $|\hat{\rho}_k| \leq 1.96/\sqrt{87} = 0.21$ ，残差自相关检验也表明 MA (2) 模型是适合的。

表 14.19 例 14.2 的 Ljung-Box-Pierce 统计量计算结果

$L(87)$	\tilde{Q}	相伴概率 (p 值)
6	0.3581	0.9992
12	1.6296	0.9998

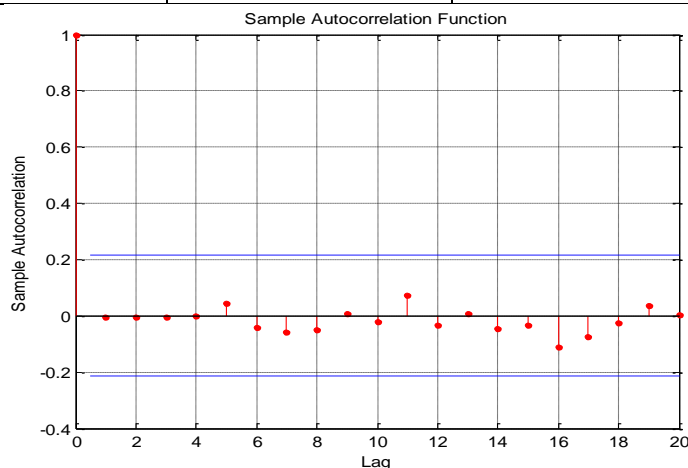


图 14.42 残差自相关函数图

计算的 MATLAB 程序如下

```
clc, clear
a0=load('data1418.txt'); a=a0'; a=a(:); %按照时间序列展开成列向量
a=log(a); xt=diff(a); %平稳化变换，先取对数，再做差分运算
plot(xt,'*-')
mu=mean(xt), yt=xt-mu; N=length(xt);
figure, subplot(121), autocorr(yt)
subplot(122), parcorr(yt)
m=armax(yt,[0,2]) %拟合模型
```

```

figure, myres=resid(m,yt); % 计算残差向量
sigma2=sum(myres.^2)/(N-3) % 计算残差方差
[h1,pValue1,st1] = lbqtest(myres,'lags',6);
[h2,pValue2,st2] = lbqtest(myres,'lags',12);
h=[h1,h2] % 显示是否拒绝原假设的标志值, 0 是接受, 1 是拒绝
pValue=[pValue1,pValue2]
st=[st1,st2]
cr=autocorr(myres) % 计算残差自相关函数
figure, autocorr(myres) % 画残差自相关函数

```

14.6 平稳时间序列预测

下面讨论平稳时间序列的预测问题。设当前时刻为 t ，我们已经知道平稳时间序列 X_t 在时刻 t 及以前时刻的观察值 $X_t, X_{t-1}, X_{t-2}, \dots$ ，现在用序列 X_t 对时刻 t 以后的观察值 X_{t+l} ($l > 0$) 进行预测，这种预测称为以 t 为原点，向前期(或步长)为 l 的预测，预测值记为 $\hat{X}_t(l)$ 。预测误差为

$$e_t(l) = X_{t+l} - \hat{X}_t(l), \quad (84)$$

预测误差的均方值为

$$E[e_t^2(l)] = E\{[X_{t+l} - \hat{X}_t(l)]^2\}. \quad (85)$$

下面讨论使上式达到最小的预测 $\hat{X}_t(l)$ 。

14.6.1 条件期望预测

我们要解决的问题是在时刻 t 用 X_t, X_{t-1}, \dots 对 X_{t+l} 的取值进行预测，而 X_{t+l} 是一个未知的随机变量，由于 X_t 之间具有相关性，因而 X_{t+l} 的概率分布是有条件的（即在 X_t, X_{t-1}, \dots 已给定的条件下），其期望也是有条件的，一个直观的想法是用其条件期望值作为预测值，即

$$\hat{X}_t(l) = E(X_{t+l} | X_t, X_{t-1}, \dots). \quad (86)$$

有关 X_t 和 a_t 的条件期望具有以下性质。

(1) 常量的条件期望是其本身

对 ARMA 序列而言，现在时刻与过去时刻的观察值及扰动的条件期望是其本身，即

$$E(X_k | X_t, X_{t-1}, \dots) = X_k, \quad (k \leq t),$$

$$E(a_k | X_t, X_{t-1}, \dots) = a_k, \quad (k \leq t).$$

(2) 未来扰动的条件期望为零，即

$$E(a_{t+l} | X_t, X_{t-1}, \dots) = 0, \quad (l \geq 1).$$

(3) 未来取值的期望为未来取值的预测值，即

$$E(X_{t+l} | X_t, X_{t-1}, \dots) = \hat{X}_{t+l}, \quad (l \geq 1).$$

14.6.2 预测的三种形式

前面介绍了 ARMA 模型的三种表示形式。

(1) 差分方程形式

$$X_t - \phi_1 X_{t-1} - \dots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \dots - \theta_m a_{t-m}. \quad (87)$$

(2) 传递形式

$$X_t = \sum_{j=0}^{\infty} G_j a_{t-j}. \quad (88)$$

(3) 逆转形式

$$X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + a_t. \quad (89)$$

与之相对应，也有预测的三种形式。

1. 由 ARMA 模型的传递形式进行预测

利用 ARMA 模型的传递形式(88)可将 X_{t+l} 表示为

$$X_{t+l} = \sum_{j=0}^{\infty} G_j a_{t+l-j}. \quad (90)$$

利用条件期望的性质对 (90) 求条件期望, 有

$$\hat{X}_t(l) = E(X_{t+l} | X_t, X_{t-1}, \dots) = \sum_{j=0}^{\infty} G_j E(a_{t+l-j} | X_t, X_{t-1}, \dots) = \sum_{j=0}^{\infty} G_{l+j} a_{t-j} \quad (91)$$

预测误差为

$$e_t(l) = X_{t+l} - \hat{X}_t(l) = G_0 a_{t+l} + G_1 a_{t+l-1} + \dots + G_{l-1} a_{t+1} \quad (92)$$

预测误差的均方值

$$E[e_t^2(l)] = E\{[X_{t+l} - \hat{X}_t(l)]^2\} = \sigma_a^2(G_0^2 + G_1^2 + \dots + G_{l-1}^2).$$

X_{t+l} 的条件方差为

$$\text{Var}(X_{t+l}) = E[X_{t+l} - E(X_{t+l})]^2 = E\{[X_{t+l} - \hat{X}_t(l)]^2\} = \sigma_a^2(G_0^2 + G_1^2 + \dots + G_{l-1}^2) \quad (93)$$

于是 \hat{X}_{t+l} 预测 95% 的置信区间为

$$\hat{X}_{t+l} \pm 1.96 \sigma_a (G_0^2 + G_1^2 + \dots + G_{l-1}^2)^{\frac{1}{2}} \quad (94)$$

下面说明以上条件期望预测是一种最小均方误差预测。考虑到平稳序列的可逆性, 也可以将 X_{t+l} 的预测表示为扰动 a_t, a_{t-1}, \dots 的线性组合

$$\hat{X}_t(l) = \sum_{j=0}^{\infty} G_j^* a_{t-j}, \quad (95)$$

其中 G_j^* 待定, 预测误差的方差为

$$\begin{aligned} E[e_t^2(l)] &= E\{[X_{t+l} - \hat{X}_t(l)]^2\} = E[(G_0 a_{t+l} + G_1 a_{t+l-1} + \dots + G_{l-1} a_{t+1}) + \sum_{j=0}^{\infty} (G_{l+j} - G_j^*) a_{t-j}]^2 \\ &= \sigma_a^2 [G_0^2 + G_1^2 + \dots + G_{l-1}^2 + \sum_{j=0}^{\infty} (G_{l+j} - G_j^*)^2]. \end{aligned}$$

要使上式达到最小, 必须且只须

$$G_{l+j} = G_j^*, \quad j = 0, 1, 2, \dots,$$

式(95)变为

$$\hat{X}_t(l) = \sum_{j=0}^{\infty} G_{l+j} a_{t-j} \quad (96)$$

即为(91)式, 这说明条件期望预测与最小均方误差预测是一致的。

2. 用 ARMA 模型的逆转形式预测

由 ARMA 模型的逆转形式(89)可将 X_{t+l} 表示为

$$X_{t+l} = \sum_{j=1}^{\infty} I_j X_{t+l-j} + a_{t+l}, \quad (97)$$

利用条件期望的性质对(97)求条件期望, 有

$$\hat{X}_t(l) = E(X_{t+l} | X_t, X_{t-1}, \dots) = \sum_{j=1}^{\infty} I_j E(X_{t+l-j} | X_t, X_{t-1}, \dots) = \sum_{j=1}^{l-1} I_j \hat{X}_t(l-j) + \sum_{j=l}^{\infty} I_j X_{t+l-j}. \quad (98)$$

从上式可以看出, 预测要用到所有过去 X_t 的信息。实际上, ARMA 模型的可逆性保证了 I_j 构成收敛级数, 按预定的精度要求, 可取某个 k 值, 当 $j > k$ 时, 令 $I_j = 0$, 即忽略 X_{t+l} 对 X_{t+l-j} 的依赖性, 进而得出预测值。

3. 用 ARMA 模型 (即差分方程形式) 进行预测

利用模型的差分方程形式预测, 计算相对简单, 实际中最常用。

(1) AR (1) 模型

设序列 X_t 适合如下 AR (1) 模型, 则 X_{t+l} 可表示为

$$X_{t+l} = \phi_1 X_{t+l-1} + a_{t+l},$$

$l=1, 2$ 时求条件期望, 有

$$\hat{X}_t(1) = \phi_1 X_t,$$

$$\hat{X}_t(2) = E[(\phi_1 X_{t+1} + a_{t+2}) | X_t, X_{t-1}, \dots] = \phi_1 \hat{X}_t(1) = \phi_1^2 X_t.$$

一般的, 当 $l > 1$ 有

$$\hat{X}_t(l) = E[(\phi_1 X_{t+l-1} + a_{t+l}) | X_t, X_{t-1}, \dots] = \phi_1 \hat{X}_t(l-1),$$

即预测值满足模型差分方程形式的自回归部分

$$\hat{X}_t(l) - \phi_1 \hat{X}_t(l-1) = 0,$$

因而

$$\hat{X}_t(l) = \phi_1^l X_t.$$

(2) ARMA(1, 1) 模型

设序列 X_t 适合 ARMA(1, 1) 模型

$$X_t - \phi_1 X_{t-1} = a_t - \theta_1 a_{t-1},$$

X_{t+l} 可表示为

$$X_{t+l} = \phi_1 X_{t+l-1} + a_{t+l} - \theta_1 a_{t+l-1} \quad (99)$$

$l=1$ 时求条件期望, 有

$$\hat{X}_t(1) = E[(\phi_1 X_t + a_{t+1} - \theta_1 a_t) | X_t, X_{t-1}, \dots] = \phi_1 X_t - \theta_1 a_t, \quad (100)$$

其中 $a_t = X_t - \hat{X}_{t-1}(1) = X_t - \phi_1 X_{t-1} + \theta_1 a_{t-1}$ 。就是说, a_t 需要递推计算。但实际中数据有限, 过于靠前的 a_{t-j} 是未知的, 因而我们往往给定初始值, 取以前某时刻 $a_{t-j} = 0$, 即假定

$X_{t-j} = \hat{X}_{t-j-1}(1)$, 这样就可以递推计算出 a_t , 进而得到 $\hat{X}_t(1)$ 。

两步预测

$$\hat{X}_t(2) = E[(\phi_1 X_{t+1} + a_{t+2} - \theta_1 a_{t+1}) | X_t, X_{t-1}, \dots] = \phi_1 \hat{X}_t(1).$$

l 步预测

$$\hat{X}_t(l) = E[(\phi_1 X_{t+l-1} + a_{t+l} - \theta_1 a_{t+l-1}) | X_t, X_{t-1}, \dots] = \phi_1 \hat{X}_t(l-1), \quad (101)$$

即 $l > 1$ 时, 预测值满足模型差分方程形式的自回归部分

$$\hat{X}_t(l) - \phi_1 \hat{X}_t(l-1) = 0, \quad (102)$$

故预测值为

$$\hat{X}_t(l) = b'_0 \phi_1^l \quad (l > 0) \quad (103)$$

又由 (100) 有

$$\hat{X}_t(1) = b'_0 \phi_1 = \phi_1 X_t - \theta_1 a_t,$$

故 (103) 中的待定系数

$$b'_0 = X_t - \frac{\theta_1}{\phi_1} a_t \quad (104)$$

因此, 当 $l > 0$ 时, 预测值 (差分方程 (102) 的特解) 为

$$\hat{X}_t(l) = (X_t - \frac{\theta_1}{\phi_1} a_t) \phi_1^l \quad (l > 0).$$

说明: 如果把预测值 $\hat{X}_t(l)$ 看作 l 的函数, 则预测函数的形式是由模型的自回归部分决定的, 滑动平均部分用来确定预测函数中的待定系数, 使得预测函数“适应”于观测数据。

(3) MA(1) 模型

设序列 X_t 适合 MA(1) 模型

$$X_t = a_t - \theta_1 a_{t-1},$$

X_{t+l} 可表示为

$$X_{t+l} = a_{t+l} - \theta_1 a_{t+l-1}.$$

$l=1$ 时求条件期望, 有

$$\hat{X}_t(1) = E[(a_{t+1} - \theta_1 a_t) | X_t, X_{t-1}, \dots] = -\theta_1 a_t,$$

其中 $a_t = X_t - \hat{X}_{t-1}(1) = X_t + \theta_1 a_{t-1}$, 需要递推计算。 $l=2$ 时求条件期望, 有

$$\hat{X}_t(2) = E[(a_{t+2} - \theta_1 a_{t+1}) | X_t, X_{t-1}, \dots] = 0,$$

一般地, 有

$$\hat{X}_t(l) = 0 \quad (l \geq 2).$$

说明: 对于 MA(m) 模型而言, 超过 m 步的预测值均为零。这与 MA 序列的短期记忆性是

吻合的。

(4) ARMA(n, m)模型预测的一般结果

由 ARMA 模型 (87) 可将 X_{t+l} 表示为

$$X_{t+l} = \phi_1 X_{t+l-1} + \cdots + \phi_n X_{t+l-n} + a_{t+l} - \theta_1 a_{t+l-1} - \cdots - \theta_m a_{t+l-m}, \quad (105)$$

利用条件期望的性质对(105)求条件期望, 当 $l \leq \max\{m, n\}$ 时, 有

$$\begin{aligned} \hat{X}_t(1) &= E[(\phi_1 X_t + \cdots + \phi_n X_{t+1-n} + a_{t+1} - \theta_1 a_t - \cdots - \theta_m a_{t+1-m}) | X_t, X_{t-1}, \cdots] \\ &= \phi_1 X_t + \cdots + \phi_n X_{t+1-n} - \theta_1 a_t - \cdots - \theta_m a_{t+1-m}, \\ \hat{X}_t(2) &= E[(\phi_1 X_{t+1} + \cdots + \phi_n X_{t+2-n} + a_{t+2} - \theta_1 a_{t+1} - \cdots - \theta_m a_{t+2-m}) | X_t, X_{t-1}, \cdots] \\ &= \phi_1 \hat{X}_t(1) + \phi_2 X_t + \cdots + \phi_n X_{t+2-n} - \theta_1 a_t - \cdots - \theta_m a_{t+2-m}, \\ &\vdots \\ \hat{X}_t(l) &= \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \cdots + \phi_{l-1} \hat{X}_t(1) + \phi_l X_t + \cdots + \phi_n X_{t+l-n} - \theta_1 a_t - \cdots - \theta_m a_{t+l-m}. \end{aligned}$$

当 $l > m$ 时, 滑动平均部分全部消失, 有

$$\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \cdots + \phi_n \hat{X}_t(l-n), \quad (106)$$

其中对 $j \geq 0$ 应理解为 $\hat{X}_t(-j) = X_{t-j}$ 。因此, 各步的预测结果满足差分方程的自回归部分。(106)的通解(即预测函数形式)为

$$\hat{X}_t(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{n-1}^{(t)} f_{n-1}(l), \quad (107)$$

这里 $l > \max\{m, n\}$, 其中 $f_0(l), f_1(l), \cdots, f_{n-1}(l)$ 的形式由模型特征方程

$$\lambda^n - \phi_1 \lambda^{n-1} - \cdots - \phi_n = 0 \quad (108)$$

的特征根来决定。当预测原点 t 给定时, 系数 $b_0^{(t)}, b_1^{(t)}, \cdots, b_{n-1}^{(t)}$ 都是常数, 并由模型的滑动平均部分决定。随着预测原点的变化, 这些系数也将改变, 以使预测值适应于序列已观测部分的特性。

综上所述, 对于 ARMA(n, m)模型, 自回归部分决定了预测函数的形式, 而滑动平均部分则用于确定预测函数中的系数。

例 14.3 设 X_t 适合以下的 ARMA(2, 1)模型

$$X_t - 0.8X_{t-1} + 0.5X_{t-2} = a_t - 0.3a_{t-1},$$

已知 $X_{t-3}, X_{t-2}, X_{t-1}, X_t$ 分别为 -1, 2, 2.5, 0.6, $a_{t-2} = 0$, 求 $\hat{X}_t(1), \hat{X}_t(2)$ 和预测函数 $\hat{X}_t(l)$ 。

解 先求 a_{t-1}, a_t , 由于

$$X_{t-1} - 0.8X_{t-2} + 0.5X_{t-3} = a_{t-1} - 0.3a_{t-2},$$

故

$$a_{t-1} = X_{t-1} - 0.8X_{t-2} + 0.5X_{t-3} + 0.3a_{t-2} = 2.5 - 0.8 \times 2 - 0.5 \times (-1) + 0.3 \times 0 = 0.4,$$

$$a_t = X_t - 0.8X_{t-1} + 0.5X_{t-2} + 0.3a_{t-1} = 0.6 - 0.8 \times 2.5 + 0.5 \times 2 + 0.3 \times 0.4 = -0.28.$$

所以

$$\hat{X}_t(1) = E[(0.8X_t - 0.5X_{t-1} + a_{t+1} - 0.3a_t) | X_t, X_{t-1}, \cdots] = 0.8 \times 0.6 - 0.5 \times 2.5 - 0.3 \times (-0.28) = 0.07$$

$$\hat{X}_t(2) = E[(0.8X_{t+1} - 0.5X_t + a_{t+2} - 0.3a_{t+1}) | X_t, X_{t-1}, \cdots] = 0.8 \times 0.07 - 0.5 \times 0.6 = -0.244.$$

当 $l > 1$ 时, 预测值满足模型自回归部分决定的差分方程

$$\hat{X}_t(l) - 0.8\hat{X}_t(l-1) + 0.5\hat{X}_t(l-2) = 0,$$

特征方程 $\lambda^2 - 0.8\lambda + 0.5 = 0$ 的根为 $0.4 \pm 0.5831i$, 故预测函数形式为

$$\hat{X}_t(l) = (\sqrt{0.4^2 + 0.5831^2})(c_1 \sin \theta l + c_2 \cos \theta l) \quad (l > 0),$$

其中 $\theta = \arctan \frac{0.5831}{0.4} = 55.55^\circ$, c_1, c_2 的确定要用到 $\hat{X}_t(1) = 0.07$, $\hat{X}_t(2) = -0.244$, 经过计算得

到 $c_1 = -0.2915$, $c_2 = 0.6$ 。可以看出, 随着 l 的增大, 预测值将振荡衰减趋于零(序列均值)。

计算的 MATLAB 程序如下

```
clc, clear
lamda=roots([1,-0.8,0.5]), lamda=lamda(1)
rho=abs(lamda)
theta=atand(imag(lamda)/real(lamda))
```

```
syms c1 c2
eq1=rho*(c1*sind(theta)+c2*cosd(theta))-0.07;
eq2=rho^2*(c1*sind(theta*2)+c2*cosd(theta*2))+0.244;
[c1,c2]=solve(eq1,eq2);
c1=double(c1), c2=double(c2)
```

例 14.4 利用例 14.1 所建的模型 $X_t = 0.7852X_{t-1} - 0.2241X_{t-2} + a_t$ 在 $t = 250$ 做一步、两步和三步预测，这里 $\sigma_a = 2.4378$ 。

首先由 ϕ_1, ϕ_2 求出 Green 函数 G_0, G_1, G_2 。

$$G_0 = 1, \quad G_1 = \phi_1 = 0.7852, \quad G_2 = \phi_1 G_1 + \phi_2 = 0.3925.$$

然后利用模型进行预测

$$\hat{X}_{250}(1) = 0.7852X_{250} - 0.2241X_{249} = 2.7500,$$

$$\hat{X}_{250}(2) = 0.7852\hat{X}_{250}(1) - 0.2241X_{250} = 1.1329,$$

$$\hat{X}_{250}(3) = 0.7852\hat{X}_{250}(2) - 0.2241\hat{X}_{250}(1) = 0.2734,$$

这些预测的 95% 的置信区间为

$$\hat{X}_{250}(1) \pm 1.96\sigma_a = 2.7500 \pm 4.7782,$$

$$\hat{X}_{250}(2) \pm 1.96\sigma_a \sqrt{1 + G_1^2} = 1.1329 \pm 6.0752,$$

$$\hat{X}_{250}(3) \pm 1.96\sigma_a \sqrt{1 + G_1^2 + G_2^2} = 0.2734 \pm 6.3581.$$

再利用 $Y_t = X_t + \bar{Y}$ ($\bar{Y} = 9.42$) 得到原序列 Y_t 的预测分别为

$$\hat{Y}_{250}(1) = 12.1688 \pm 4.7782,$$

$$\hat{Y}_{250}(2) = 10.5517 \pm 6.0752,$$

$$\hat{Y}_{250}(3) = 9.6922 \pm 6.3581.$$

已知数据的预测值和未来 3 个时刻的预测值如图 14.43 所示。

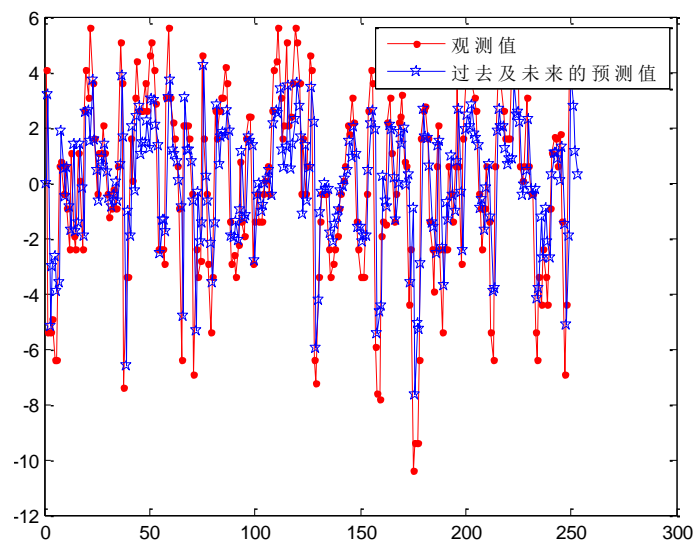


图 14.43 预测图示

计算的 MATLAB 程序如下

```
clc, clear
yt=load('data1415.txt');
yt=yt'; yt=yt(:); %原始数据按照顺序展开成列向量
mu=mean(yt), xt=yt-mu; %数据零均值化
m=ar(xt,2,'ls') %拟合模型
a=m.a %提出 AR 部分的系数
fai1=-a(2), fai2=-a(3)
g0=1, g1=fai1, g2=fai1*g1+fai2
```

```

xt0=[xt;0]; %为了进行一步预测，需要在已知值后加一个任意数
pa=predict(m,xt); %预测过去各时刻的值
yuce=forecast(m,xt,3) %预测未来 3 个时刻的取值
sigma=sqrt(m.NoiseVariance)
green=[g0,g1,g2];
banjing=1.96*sigma*sqrt(cumsum(green.^2)) %求置信区间的半径
yhat=yuce+mu %求得 yt 未来的三个预测值
plot(xt,'r-') %画观测值
hold on, plot([pa;yuce], 'p-') %画过去及未来的预测值
legend('观测值','过去及未来的预测值')

```

14.7 趋势性时间序列模型

14.7.1 随机时间序列的趋势性检验

通常是从平稳性的直观意义出发或利用间接反推的方法来判断序列的平稳性。对判断出来的非平稳性还需要结合具体系统的机理来分析。

1. 数据图法

在 $t - X_t$ 平面直角坐标系中将所研究的时间序列绘成连线图，观察其是否存在趋势性或周期性。

优点：简单、直观、运用方便等。

缺点：主观性强，依赖分析者的实际经验，不同的分析者可能会得出不同的结论。

例 14.5 表 14.20 是 1978—2004 年陕西省国内生产总值序列数据，共 27 个，图 14.44 是对应的数据图。

表 14.20 1978—2004 年陕西省国内生产总值序列数据（横读）

81.07	94.52	94.91	102.09	111.95	123.39
149.35	180.87	208.31	244.96	314.48	358.37
404.3	466.84	538.43	661.42	816.58	1000.03
1175.92	1300.03	1381.53	1487.61	1660.92	1844.27
2035.96	2398.58	2883.51			

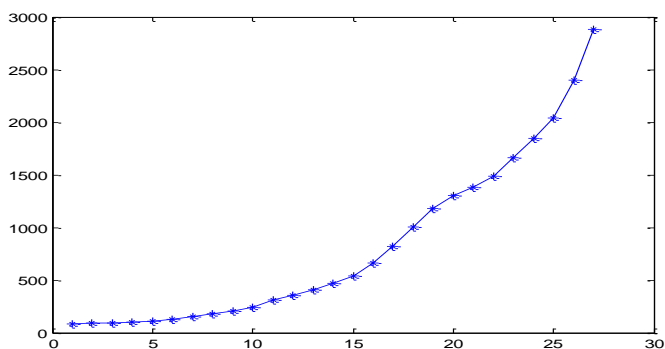


图 14.44 1978—2004 年陕西省国内生产总值序列数据图

画图的 MATLAB 程序如下

```

clc, clear
a=textread('data1420.txt');
a=a'; a=nonzeros(a);
plot(a,'*-')

```

从图中可以看出，数据具有一定的上升趋势，周期性不很明显，可以初步断定序列是非平稳的。

2. 自相关、偏自相关函数法

一个零均值平稳序列的自相关和偏自相关函数要么是截尾的，要么是拖尾的。如果一个序列零均值化以后的自相关函数或偏自相关函数既不截尾，也不拖尾，则可断定该序列是非

平稳的。

图 14.45 是例 14.5 中序列的自相关与偏相关函数图，可以看出偏自相关函数是截尾的，但自相关函数缓慢衰减，这说明序列存在一定的非平稳性。

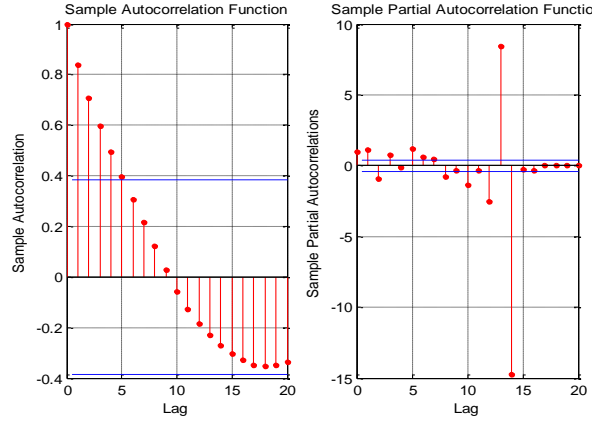


图 14.45 相关函数图

画图的 MATLAB 程序如下

```
clc, clear
a=textread('data1420.txt'); a=a'; a=nonzeros(a);
b=a-mean(a);
subplot(121),autocorr(b)
subplot(122),parcorr(b)
```

3. 特征根检验法

先拟合序列的适应模型，然后求由该适应模型的参数组成的特征方程的特征根。如果所有的特征根 λ_i 都满足平稳性条件，即 $|\lambda_i| < 1$ ，则可认为该序列是平稳的，否则该序列就是非平稳的。

4. 逆序检验法.

逆序检验法可以检验出均值或方差存在的某种趋势。一般步骤为：

第一步，由时间序列求出一个大致不相关的均值或方差值的序列。具体地可将整个序列分成 M 段，求出每段数据按时间平均的均值（或方差）。设所得到的序列为 y_1, y_2, \dots, y_M 。

第二步，计算均值序列（或方差序列）的逆序总数。

对于 y_i ($i=1, 2, \dots, M-1$) 来说，若后边有一个值大于它，即 $y_j > y_i$ ($j > i$)，则称为一个逆序，记 y_i 的逆序数为 A_i (即)，则逆序总数为 $A = \sum_{i=1}^M A_i$ 。

第三步，计算统计量进行检验。

一个随机序列的逆序数具有如下的期望和方差：

$$E(A) = \frac{1}{4} M(M-1), \quad (109)$$

$$D(A) = \frac{M(2M^2 + 3M - 5)}{72}, \quad (110)$$

其中， M 为数据个数，并且统计统计量

$$D(A) = \frac{M(2M^2 + 3M - 5)}{72} \quad (111)$$

渐近服从 $N(0,1)$ 分布。

在显著性水平 $\alpha = 0.05$ 的情况下，若 $|z| < 1.96$ ，则认为序列无明显的趋势，否则认为序列是非平稳的。若 A 过大，则序列有上升的趋势；若 A 过小，表明序列有下降的趋势。

说明：逆序检验法在检验序列是否存在单调趋势方面是有效的，但在升降均衡时是无效的。

例 14.6 判断例 14.5 中序列的平稳性。

将序列中的 27 个数据分为 6 段，前 3 段每段 5 个数据，后 3 段每段 4 个数据，可得均值序列如下：

96.908 181.376 416.484 913.4875 1457.5225 2290.58

逆序总数为

$$A = \sum_{i=1}^5 A_i = 5 + 4 + 3 + 2 + 1 = 15,$$

由(109)和(110)可得

$$E(A) = \frac{1}{4} \times 6 \times (6-1) = 7.5,$$

$$D(A) = \frac{1}{72} [6 \times (2 \times 6^2 + 3 \times 6 - 5)] = 7.0833,$$

于是

$$Z = \frac{15 + \frac{1}{2} - 7.5}{\sqrt{7.0833}} = 3.0059 > 1.96.$$

序列有上升的趋势，是非平稳的。

计算的 MATLAB 程序如下

```
clc, clear
a=textread('data71.txt'); a=a'; a=nonzeros(a);
for k=1:3
    b(k)=mean(a((k-1)*5+1:5*k)); %提取均值序列
end
for k=4:6
    b(k)=mean(a(15+4*(k-4)+1:15+4*(k-3))); %提取均值序列
end
b,M=length(b);
nx=[];
for k=1:M
    nx=[nx,length(find(b(k:end)>b(k)))]; %计算逆序数
end
nx
nxsum=sum(nx) %求逆序的总数
EA=M*(M-1)/4 %计算数学期望
DA=M*(2*M^2+3*M-5)/72 %计算方差
Z=(nxsum+1/2-EA)/sqrt(DA) %计算统计量
```

5. 游程检验法

(1) 游程 设序列 X_t 的均值是 \bar{X} 。对序列中比 \bar{X} 小的观测值记“-”号，其余的记“+”号，这样相对于原序列就可得到一个符号序列。在记号序列中每一段连续相同的记号序列就叫做一个游程。

(2) 游程检验法的基本思想

作为一个随机平稳序列，各观测值应在均值附近随机波动，因而高于平均数和低于平均数的变化应该适当。如果游程太少，则观测值连续高于或低于平均数。这说明序列中可能存在某种趋势性或周期性波动。同样，如果游程过多，也可能存在某种趋势。如 n 个数据有 n 个游程，说明序列中存在着此小彼大、此大彼小的必然趋势。

(3) 游程检验

设序列的长度为 $N = N_1 + N_2$ ，其中 N_1 和 N_2 分别是记号序列中“+”号与“-”号出现的次数。设游程的总数为 r ，对于随机序列， r 具有以下的期望和方差

$$E(r) = \frac{2N_1N_2}{N} + 1, \quad D(r) = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)},$$

统计量

$$Z = \frac{r - E(r)}{\sqrt{D(r)}}$$

渐近服从于 $N(0,1)$ 分布。

因此，对于所检验序列可计算出统计量 Z ，在给定显著水平 α 下，若 $|Z| < 1.96$ ，则可认为序列是平稳的，否则，序列是非平稳的。

例 14.7 判断例 14.5 中序列的平稳性。

该序列有 $N = 27$ ， $N_1 = 10$ ， $N_2 = 17$ ， $r = 2$ ，

$$E(r) = \frac{2N_1N_2}{N} + 1 = 13.5926，$$

$$D(r) = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)} = 5.6146，$$

$$Z = \frac{r - E(r)}{\sqrt{D(r)}} = -4.8924，$$

在显著性水平 $\alpha = 0.05$ 下， Z 超出了临界值 ± 1.96 的范围，故该序列是非平稳的。

计算的 MATLAB 程序如下

```
clc, clear
a=textread('data1420.txt'); a=a'; a=nonzeros(a);
n=length(a), mu=mean(a);
b=zeros(1,n); % 记号初始化，‘-’ 记为 0;
b(a>mu)=1; % ‘+’ 记为 1
n1=sum(a>=mu) % 求 ‘+’ 的个数
n2=sum(a<mu) % 求 ‘-’ 的个数
r=1+length(findstr(b,[0,1]))+length(findstr(b,[1,0])) % 求游程的个数
Er=2*n1*n2/n+1 % 求数学期望
Dr=2*n1*n2*(2*n1*n2-n)/n^2/(n-1) % 求方差
Z=(r-Er)/sqrt(Dr) % 求 Z 统计量的值
[h,p2,stats]=runstest(a) % 直接使用 MATLAB 工具箱的命令进行检验
```

14.7.2 平稳化方法

由于序列非平稳性表现的多样性和复杂性，平稳化的方法也是多样的。

1. 差分运算

(1) 如果序列 X_t 具有线性的趋势，则可以通过对其进行一次差分 ($\nabla X_t = X_t - X_{t-1}$ ，称为一阶差分) 将线性趋势剔除掉，然后对差分后的序列拟合 ARMA 模型进行分析和预测，最后通过差分的逆运算得到 X_t 的有关结果。

(2) 如果序列 X_t 具有二次趋势，则两次差分 ($\nabla^2 X_t$ ，两次以上的差分称为高阶差分) 可变为平稳序列；如果序列 X_t 具有 d 次多项式趋势，则通过 d 次差分可变为平稳序列。下面讨论 X_t 与 $\nabla^d X_t = W_t$ 的关系。

当 $d = 2$ 时，有

$$W_t = \nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}，$$

因而

$$X_t = W_t + 2X_{t-1} - X_{t-2}。$$

一般情形

$$W_t = \nabla^d X_t = (1-B)^d X_t = \left[1 + \sum_{k=1}^d (-1)^k C_d^k B^k \right] X_t，$$

因而

$$X_t = W_t - \sum_{k=1}^d (-1)^k C_d^k X_{t-k}。$$

(3) k 步差分运算

相距 k 期的两个序列值之间的减法运算称为 k 步差分运算。记 $\nabla_k x_t$ 为 x_t 的 k 步差分

$$\nabla_k x_t = x_t - x_{t-k} = (1 - B^k)x_t \tag{112}$$

2. 差分方式的选择

- (1) 序列蕴含着显著的线性趋势，一阶差分就可以实现趋势平稳。
- (2) 序列蕴含着曲线趋势，通常低阶（二阶或三阶）差分就可以剔除曲线趋势的影响。

例 14.8 对例 14.5 的序列进行平稳化。

从图 14.44 看出该序列含有曲线递增的长期趋势。首先对该序列进行一阶差分，其序列趋势图如图 14.46 所示。

从图 14.46 可以看出，一阶差分消除了序列中部分长期趋势，但长期趋势信息提取不充分，一阶差分后序列含着长期递增的趋势，于是对一阶差分后序列再做一次差分，其趋势图见图 14.47。

从图 14.47 可以看出，二阶差分比较充分地提出了原序列中蕴含的长期趋势，使得差分后序列不再呈现确定性趋势了。

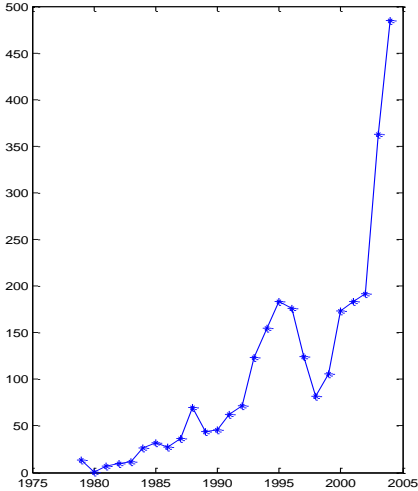


图 14.46 陕西省国内生产总值 1 阶差分序列时序图

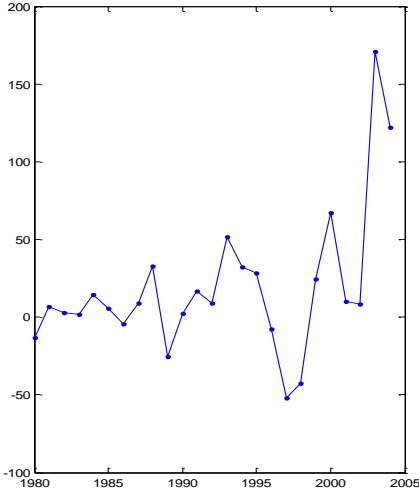


图 14.47 陕西省国内生产总值 2 阶差分序列时序图

(3) 对于具有周期性和长期趋势的序列，我们利用季节差分和一阶差分就可以提取周期和长期趋势。

例 14.9 对中国月度商业零售额序列（数据见表 14.21）进行平稳化。

表 14.21 中国商业零售额月度序列（1993.1—2004.11）（横读）

1038.3	983.4	1067.8	1071.2	1094.4	1140.8	1087.6	1070.8	1132.3	1160.2	1195.6	1550.2
1192.2	1158.5	1166.6	1170	1214	1281	1252	1286	1396	1444	1554	1932.2
1602.2	1491.5	1533.3	1548.7	1585.4	1639.7	1623.6	1637.1	1756	1818	1935.2	2389.5
1909.1	1911.2	1860.1	1854.8	1898.3	1966	1888.7	1916.4	2083.5	2148.3	2290.1	2848.6
2288.5	2213.5	2130.9	2100.5	2100.8	2164.7	2102.5	2104.4	2239.6	2348	2454.9	2881.7
2549.5	2306.4	2279.7	2252.7	2265.2	2326	2286.1	2314.6	2443.1	2536	2652.2	3131.4
2662.1	2538.4	2403.1	2356.8	2364	2428.8	2380.3	2410.9	2604.3	2743.9	2859.1	3383
2962.9	2804.9	2626.6	2571.5	2636.9	2645.2	2596.9	2636.3	2854.3	3029.3	3107.8	3680
3332.8	3047.1	2876.1	2820.9	2929.6	2908.7	2851.4	2889.4	3136.9	3347.3	3421.7	4033.3
3552.2	3416	3197.4	3163.3	3320.5	3302.8	3244.2	3284.4	3627.2	3815.2	3831.1	4404.4
4087.1	3856.6	3645.2	3535.2	3536.6	3732	3718.4	3769.2	4139	4365.4	4202.7	4735.7
4569.4	4211.4	4049.8	4001.8	4166.1	4251	4209.2	4262.7	4717.7	4983.2	4965.6	

首先画出样本数据图，其结果见图 14.48。从图 14.48 看出该序列含有曲线递增的长期趋势和周期性变动趋势。记原序列为 X_t ($t=1,2,\dots,144$)，首先对序列进行季节差分， $Y_t = X_{t+12} - X_t$ ($t=1,2,\dots,132$)，其序列趋势图见图 14.49，从图上可以看出，经季节差分的序列基本上消除了季节影响，但还有上升趋势，再对序列进行一阶差分，序列就变成了平稳

序列（见图 14.50）。

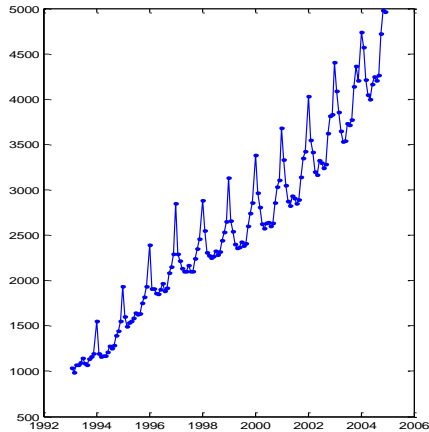


图 14.48 商业零售额序列图

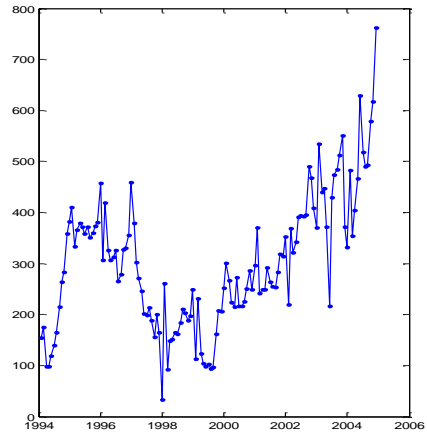


图 14.49 商业零售额序列季节差分图

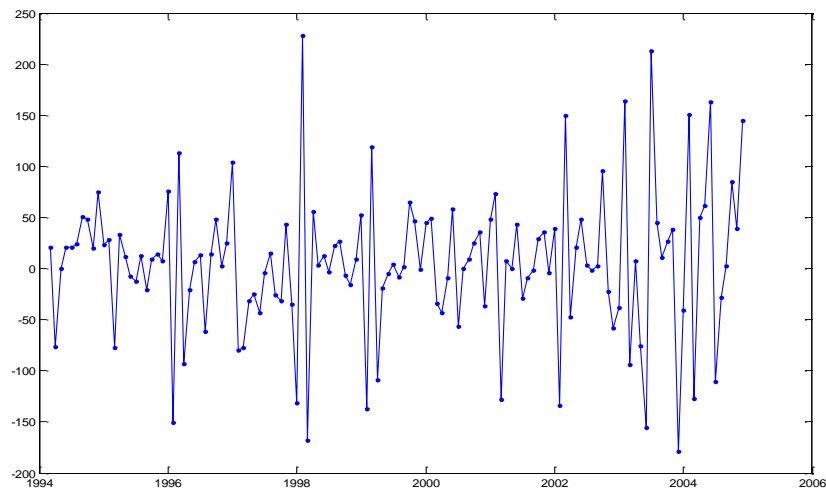


图 14.50 商业零售额序列季节差分与一阶差分图

计算和画图的 MATLAB 程序如下

```
clc, clear
a=textread('data1421.txt'); a=a'; a=nonzeros(a); n=length(a);
t=[1993+1/12:1/12:2004+11/12]';
subplot(121), plot(t,a,'-')
for i=13:n
    ja(i-12)=a(i)-a(i-12);
end
subplot(122), plot(t(13:end),ja,'-')
da=diff(ja); figure, plot(t(14:end),da,'-')
```

（4）序列含有指数趋势，则可以进行对数化的差分方法，就可以提取其确定性趋势。

如果序列含有指数趋势，则可以通过取对数将指数趋势转化成线性趋势，然后再进行差分以消除线性趋势。

表 14.22 我国社会消费品零售总额（横读）

262.7	328.8	356.1	364.0	424.0	441.6	481.2
556.5	595.4	537.7	543.7	544.8	572.7	590.1
632.8	679.1	649.2	698.2	728.8	776.9	853.5
917.7	967.4	1046.4	1099.0	1174.3	1264.9	1476.0
1794.0	2002.5	2181.5	2426.1	2899.2	3801.4	4374.0
5115.0	6534.6	7074.2	7250.3	8245.7	9704.8	12462.1

16264.7						
---------	--	--	--	--	--	--

例 14.10 表 14.22 是我国 1952—1994 年社会消费品零售总额，时序图见图 14.51。可以看出，有一定指数趋势。取对数后如图 14.52 所示，具有明显的非线性增长趋势，前期和后期增长速度较快，而中期稍慢。一阶差分后如图 14.53 所示，前期和后期数据偏大，而中期稍小，具有一定的非平稳性。二阶差分后如图 14.54 所示。

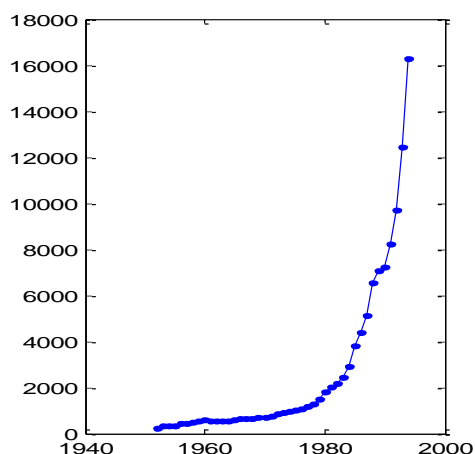


图 14.51 社会消费品零售额数据图

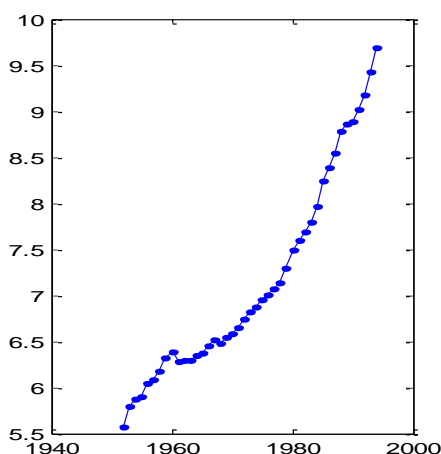


图 14.52 社会消费品零售额数据对数化图

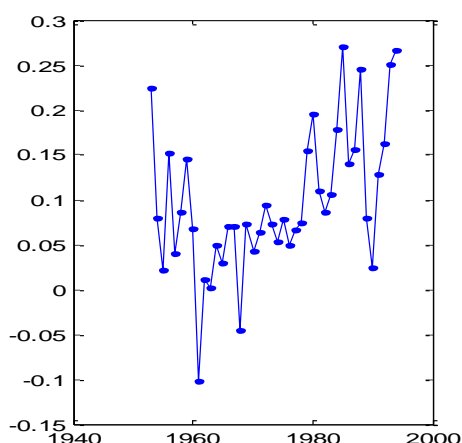


图 14.53 对数化数据一阶差分图

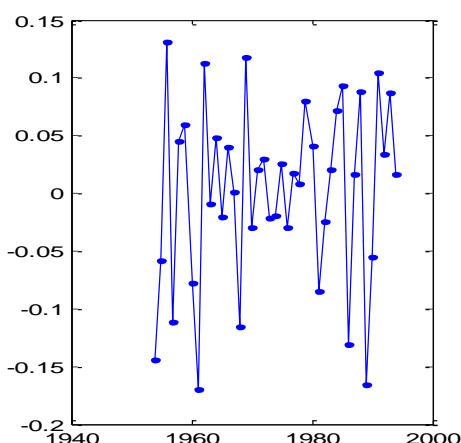


图 14.54 对数化数据二阶差分图

差分虽然是平稳化的有效方法，但不能过度差分，这样会导致有效信息的损失，从而导致预测精度的降低。

画图的 MATLAB 程序如下

```
clc, clear
a=textread('data1422.txt'); a=a'; a=nonzeros(a);
t=[1952:1994]';
subplot(121), plot(t,a,'-')
dsa=log(a); subplot(122), plot(t,dsa,'-')
d1=diff(dsa); figure, subplot(121), plot(t(2:end),d1,'-')
d2=diff(d1), subplot(122), plot(t(3:end),d2,'-')
```

14.7.3 趋势模型

对于趋势性时间序列，通常可以通过两种处理方法来建立相应的模型。一类是通过某些数学方法剔除 X_t 中所包含的趋势性 μ_t ，余下的 Y_t 可按平稳过程进行分析和建模，最后再经

反运算由 Y_t 的结果得出 X_t 有关结果, 此种方法是通过建立求和自回归移动平均模型来实现。另一类方法是具体求出 μ_t 的拟合形式 $\hat{\mu}_t$, 然后对残差序列 $\{X_t - \hat{\mu}_t\}$ 进行分析, 该残差序列可以认为是平稳的。利用前述方法可求 \hat{Y}_t , 最后综合可得 $\hat{X}_t = \hat{\mu}_t + \hat{Y}_t$, 这种方法是建立组合模型。下面分别介绍这两种模型。

1. 求和自回归移动平均模型 (ARIMA 模型)

(1) ARIMA 模型的结构

具有结构

$$\begin{cases} \Phi(B)\nabla^d X_t = \Theta(B)a_t, \\ E(a_t) = 0, \text{Var}(a_t) = \sigma_a^2, E(a_t a_s) = 0, s \neq t, \\ E(X_s a_t) = 0, \forall s < t, \end{cases} \quad (113)$$

的模型为求和自回归移动平均模型, 简记为 ARIMA (p, d, q) 模型, 式中, $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ 为平稳可逆 ARIMA (p, d, q) 模型的自回归系数多项式, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ 为平稳可逆 ARIMA (p, d, q) 模型的移动平均系数多项式。(113) 可以简记为

$$\nabla^d X_t = \frac{\Theta(B)}{\Phi(B)} a_t \quad (114)$$

式中, $\{a_t\}$ 为零均值白噪声序列。

ARIMA 模型的实质是差分运算和 ARMA 模型的组合。这说明任何非平稳序列只要通过适当阶数的差分, 实现差分后平稳, 就可以对差分后序列进行 ARMA 模型拟合了。

求和自回归移动平均模型这个名字的由来是因为 d 阶差分后序列可表示为

$$\nabla^d X_t = \sum_{i=0}^d (-1)^i C_d^i X_{t-i} \quad (115)$$

式中, $C_d^i = \frac{d!}{i!(d-i)!}$, 即差分后序列等于原序列的若干序列值的加权和, 而对它又可以拟合

自回归移动平均 ARMA 模型, 所以称它为求和自回归移动平均模型。

当 $d=0$, ARIMA (p, d, q) 模型实质上是 ARMA (p, q) 模型;

当 $p=0$, ARIMA (p, d, q) 模型实质上是 IMA (d, q) 模型;

当 $q=0$, ARIMA (p, d, q) 模型实质上是 ARI (p, d) 模型;

当 $d=1, p=0, q=0$, ARIMA (p, d, q) 模型实质上是 ARIMA $(0, 1, 0)$ 模型

$$\begin{cases} X_t = X_{t-1} + a_t, \\ E(a_t) = 0, \text{Var}(a_t) = \sigma_a^2, E(a_t a_s) = 0, s \neq t, \\ E(X_s a_t) = 0, \forall s < t, \end{cases} \quad (116)$$

该模型被称为随机游动模型 (random walk), 或醉汉模型。

(2) ARIMA 模型的性质

i) 平稳性

假如 $\{X_t\}$ 服从 ARIMA (p, d, q) 模型 (113), 记 $A(B) = \Phi(B)\nabla^d$, $A(B)$ 被称为广义自回归系数多项式, 显然 ARIMA 模型的平稳性完全由 $A(B)=0$ 的根来决定。

因为 $\{X_t\}$ 服从 ARIMA (p, d, q) 模型, 所以不妨设

$$\Phi(B) = \prod_{i=1}^p (1 - \lambda_i B), \quad |\lambda_i| < 1, \quad i=1, 2, \dots, p, \quad (117)$$

$$A(B) = \Phi(B)\nabla^d = \left[\prod_{i=1}^p (1 - \lambda_i B) \right] (1-B)^d, \quad (118)$$

由 (117)、(118) 式容易判断, 显然 ARIMA 模型的广义自回归系数多项式共有 $p+d$ 个根, 其中 p 个在单位圆外, d 个根在单位圆上。ARIMA 模型的广义自回归系数多项式的根为特征根的倒数。

因为 d 个特征根在单位圆上而非在单位圆内，所以 $d \neq 0$ 时，ARIMA (p, d, q) 模型不平稳。

ii) 方差齐性

对于 ARIMA (p, d, q) 模型，当 $d \neq 0$ 时，不仅均值非常数，序列方差也是非常数。以最简单的随机游走模型 ARIMA (0, 1, 0) 为例

$$X_t = X_{t-1} + a_t = X_{t-2} + a_t + a_{t-1} = \cdots = X_0 + a_t + a_{t-1} + \cdots + a_1, \quad (119)$$

则

$$\text{Var}(X_t) = \text{Var}(X_0 + a_t + a_{t-1} + \cdots + a_1) = t\sigma_a^2, \quad (120)$$

这是一个时间 t 的递增函数，随着时间趋向无穷，序列 $\{X_t\}$ 的方差也趋向无穷。

(3) 常见的 ARIMA 模型

i) ARIMA (0, 1, 1)

$$(1-B)X_t = (1-\theta_1 B)a_t,$$

即

$$X_t - X_{t-1} = a_t - \theta_1 a_{t-1}.$$

这就是说， X_t 是一阶齐次非平稳序列，一次差分后适合 MA(1) 模型。值得注意的是，不能认为 X_t 是平稳的 ARMA (1, 1) 模型，因为其特征根 $\lambda = 1$ 不在单位圆内。

ii) ARIMA (0, 2, 2) 模型

$$(1-B)^2 X_t = (1-\theta_1 B - \theta_2 B^2)a_t$$

即序列 X_t 两次差分后适合 MA(2) 模型。

3) ARIMA (1, 1, 1) 模型

$$(1-\phi_1 B)(1-B)X_t = (1-\theta_1 B)a_t$$

即序列 X_t 经一次差分后适合 ARMA (1, 1) 模型。

(4) ARIMA 模型建模步骤

ARIMA 模型建模步骤如下

- i) 对所分析的时间序列进行平稳化检验；
- ii) 如果是平稳时间序列，就建立 ARMA 模型；否则就进行下一步；
- iii) 对该时间序列进行平稳化；
- iv) 对平稳后的序列进行白噪声检验，如果是白噪声序列，建模就结束，否则就进行下一步；
- v) 对平稳白噪声序列拟合 ARMA 模型；
- vi) 模型的应用。

(5) 案例分析

例 14.11 对 1949—2006 年陕西省人口总数 (见表 14.23) 序列建模，并通过 2004—2006 年的数据验证模型的拟合效果。

表 14.23 1949—2006 年陕西省人口总数 (横读)

1317.3	1399.5	1485.5	1527.9	1615.2	1651.4	1715.8	1755.9	1802.9	1831.7
1880.8	1944.2	1968.5	2007.8	2056.0	2099.9	2144.3	2193.4	2232.8	2287.1
2360.5	2427.9	2492.1	2553.6	2610.2	2655.4	2692.1	2721.8	2751.1	2779.5
2807.1	2831.4	2864.5	2904.1	2930.9	2965.7	3001.7	3042.6	3088.2	3140.0
3198.0	3316.3	3363.1	3404.6	3442.5	3480.5	3513.1	3543.0	3570.1	3595.7
3618.1	3644.0	3658.6	3673.7	3689.5	3705.2	3720.0	3735.0		

(1) 判断序列的平稳性

该序列时序图如图 14.55 所示。时序图显示该序列有着显著的趋势，为典型的非平稳序列。

(2) 对原序列进行平稳化

因为原序列呈现出近似线性趋势，所以选择一阶差分，一阶差分后序列时序图如图 14.56 所示。

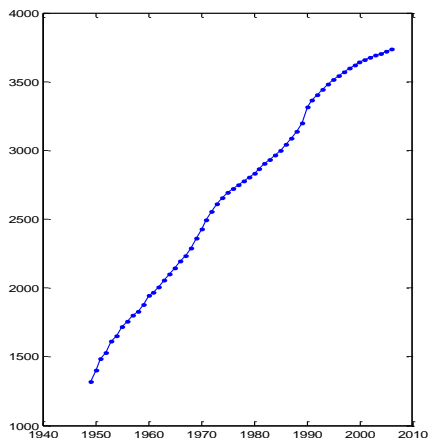


图 14.55 陕西省人口总数序列图

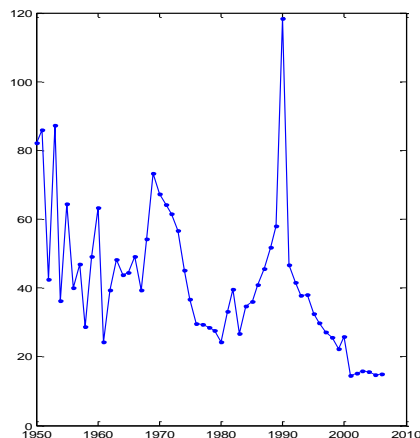


图 14.56 1 阶差分序列图

时序图显示出差分后序列仍然还具有趋势性，该序列不平稳，进一步平稳化，得到序列二阶差分结果 $W_t = \nabla^2 x_t$ 见图 14.57。时序图显示出差分后序列在均值附近比较稳定地波动。为了进一步确定平稳性，考察差分后序列的自相关图，如图 14.58 所示。自相关图显示序列很强的短期相关性，所以可以初步认为二阶差分后序列平稳。

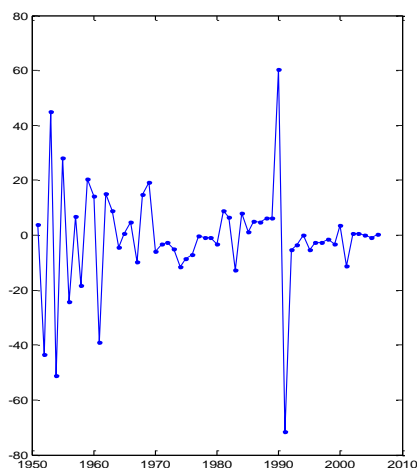


图 14.57 2 阶差分序列图

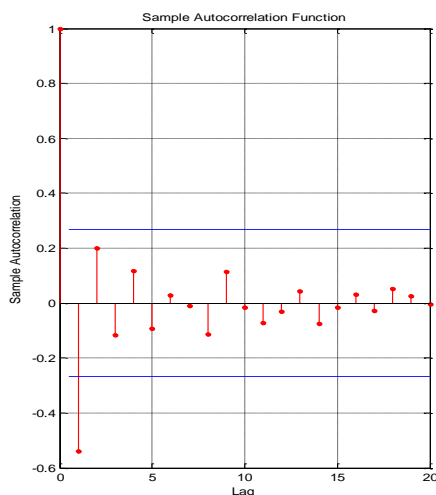


图 14.58 2 阶差分序列自相关图

(3) 对平稳的二阶差分序列进行白噪声检验

白噪声检验结果见表 14.24 所示，在检验的显著性水平取为 0.05 的条件下，由于延迟 6 阶的 χ^2 检验统计量的 P 值为 0.013，小于 0.05，所以该差分后序列不是白噪声序列，即差分后序列蕴含着不容忽视的相关信息。

表 14.24 二阶差分序列白噪声检验结果

延迟阶数	χ^2 统计量	上 0.05 分位数	相伴概率 (P 值)
6	21.9181	12.5916	0.0013
12	24.1604	21.0261	0.0193
18	25.1918	28.8693	0.1543

(4) 对平稳非白噪声差分序列拟合 ARMA 模型

二阶差分后序列的自相关图 (图 14.58) 已经显示出该序列有自相关系数拖尾的性质，考察其偏自相关系数的性质，如图 14.59 所示，偏自相关图显示在一阶具有截尾性，所以考虑用 ARMA (1, 0) 模拟拟合二阶差分后序列。就是用 ARIMA (1, 2, 0) 模型拟合原序列，其拟

合的模型为

$$W_t = -0.5333W_{t-1} + a_t, \quad \text{Var}(a_t) = 276.6875.$$

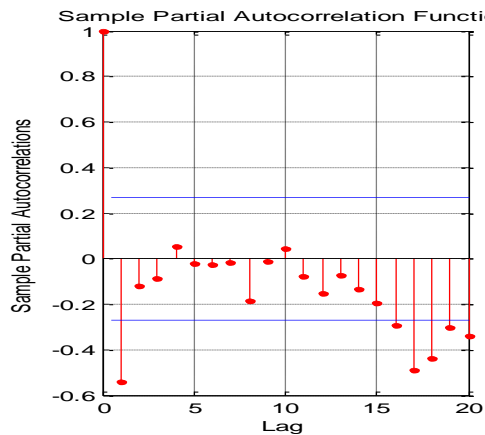


图 14.59 二阶差分序列偏自相关图

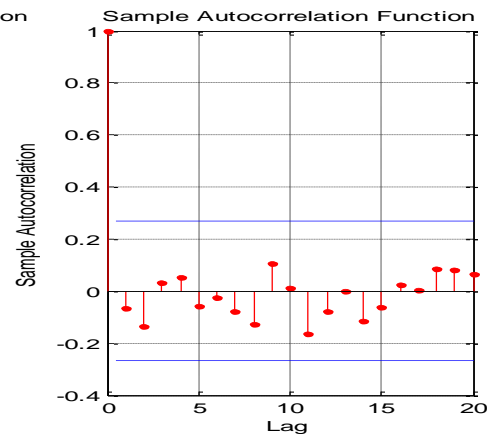


图 14.60 残差的自相关图

(5) 模型是适应性检验

残差序列的自相关图见图 14.60，可以认为残差序列为白噪声序列。

对残差序列进行 Q 检验，检验结果见表 14.25。显然，拟合检验统计量的 P 值都大于显著性检验水平 0.05，也说明可以认为残差序列为白噪声序列。

表 14.25 残差序列白噪声检验结果

延迟阶数	χ^2 统计量	相伴概率 (P 值)
6	1.8633	0.9318
12	6.5829	0.8839
18	8.5978	0.9683

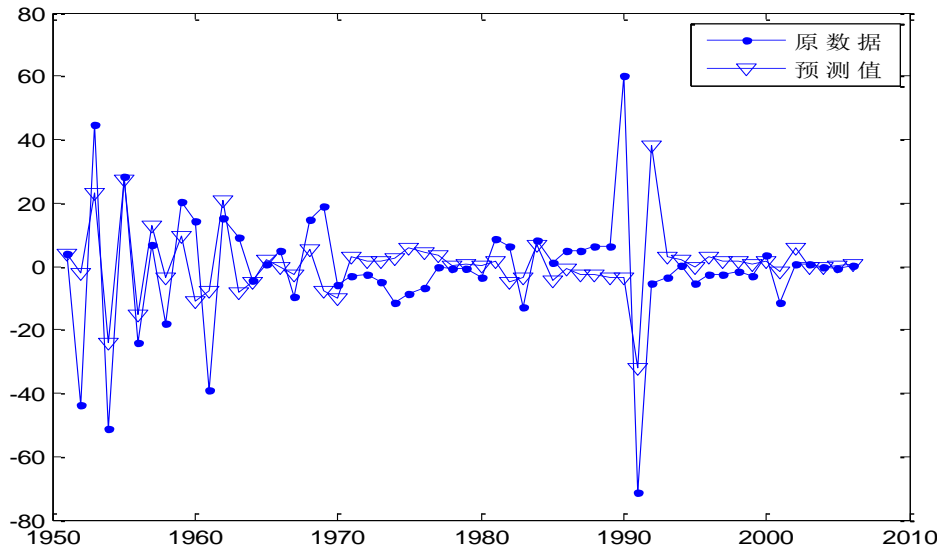


图 14.61 预测序列与观测序列比较图

预测序列与观测序列比较图见图 14.61，从图 14.61 可以看出，预测值与观测值比较接近，总之，说明 ARIMA (1, 2, 0) 模型适应该序列。

(6) 利用所建立的模型对陕西省 2004—2006 年的人口总数进行预测，其结果见表 14.26。从表 14.26 可以看出模型拟合比较好。

表 14.26 模型的预测值

年份	2004	2005	2006
----	------	------	------

实际人口数（万人）	3705.20	3720.00	3735.00
拟合预测数（万人）	3704.9267	3720.9533	3735.2799

```

clc, clear
a=textread('data1423.txt'); a=a'; a=nonzeros(a);
t=[1949:2006]'; n=length(t);
subplot(121), plot(t,a,'.-')
da=diff(a); subplot(122), plot(t(2:end),da,'.-')
dda=diff(da);
figure, subplot(121), plot(t(3:end),dda,'.-')
subplot(122), autocorr(dda)
[h1,pValue1,st1] = lbqtest(dda,'lags',6) %进行 chi2 检验
[h2,pValue2,st2] = lbqtest(dda,'lags',12)
[h3,pValue3,st3] = lbqtest(dda,'lags',19)
fw1=chi2inv(0.95,6), fw2=chi2inv(0.95,12), fw3=chi2inv(0.95,18) %计算分位数
figure, subplot(121), parcorr(dda)
m=ar(dda,1)
m.NoiseVariance %显示残差的方差
pdda=predict(m,dda) %计算预测值
myres=dda-pdda; %计算残差向量
[hc1, pvaluec1,stc1]=lbqtest(myres,'lags',6)
[hc2, pvaluec2,stc2]=lbqtest(myres,'lags',12)
[hc3, pvaluec3,stc3]=lbqtest(myres,'lags',18)
subplot(122), autocorr(myres)
figure, plot(t(3:end),dda,'.-')
hold on, plot(t(3:end),pdda,'V-')
legend('原数据','预测值')
xhat(1)=a(1); xhat(2)=a(2);
for i=3:n
    xhat(i)=pdda(i-2)+2*a(i-1)-a(i-2); %恢复到原始数据的预测值
end
xhhat=xhat(end-2:end) %提出最后 3 个时刻的预测值

```

例 14.12（续例 14.10） 对我国社会消费品零售总额数据建模。

（1）对序列进行平稳化

从例 14.10 可知，该序列 X_t 具有指数趋势，并经过二次指数差分，得到序列 $\nabla^2 \ln X_t$ 是基本平稳的。

表 14.27 二阶差分序列白噪声检验结果

延迟阶数	χ^2 统计量	P 值
3	7.3483	0.0616
6	9.0936	0.1684
9	10.3055	0.3263

（2）对平稳的 $\nabla^2 \ln X_t$ 序列进行白噪声检验。

白噪声检验结果如表 14.27 所示。由于延迟 3 阶的 χ^2 检验统计量的 P 值为 0.0616，在检验的显著性水平取为 0.10 的条件下，可以认为序列不是白噪声，即差分后序列蕴含着不容忽视的相关信息。

（3）对平稳非白噪声差分序列拟合 ARMA 模型。

二次指数差分后序列的自相关图和偏自相关图（见图 14.62），可以看出此序列的自相关和偏自相关函数具有拖尾性，用低阶模型 ARMA 模型对其进行拟合。通过残差方差和 AIC 值比较分析，该序列的模型为

$$\nabla^2 \ln X_t = Y_t,$$

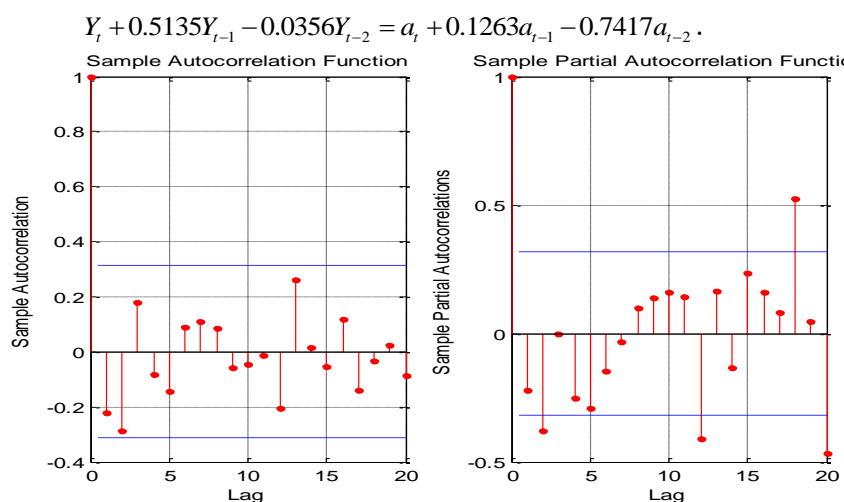


图 14.62 二次指数差分自相关图和偏自相关图

(4) 模型的适应性检验

主要对残差序列进行检验，残差序列的自相关函数值比较小，可以认为该残差序列为白噪声序列，说明所建模型适应该序列。

计算的 MATLAB 程序如下

```
clc, clear
a=textread('data1422.txt'); a=a'; a=nonzeros(a);
t=[1952:1994]'; n=length(t);
xt=log(a); dxt=diff(xt); ddx=diff(dxt);
[h1,pValue1,st1] = lbqtest(ddx,'lags',3) %进行 chi2 检验
[h2,pValue2,st2] = lbqtest(ddx,'lags',6)
[h3,pValue3,st3] = lbqtest(ddx,'lags',9)
subplot(121), autocorr(ddx)
subplot(122), parcorr(ddx)
order=[2 2; 2 1; 1 2; 1 1];
for i=1:4
    m{i}=armax(ddx,order(i,:));
    myaic(i)=aic(m{i}); %计算 AIC 准则函数的取值
    canchafc(i)=m{i}.NoiseVariance; %计算残差方差
end
myaic, canchafc
mm=m{1};
figure, myres=resid(mm,ddx); %计算残差向量
figure, autocorr(myres) %画残差的自相关函数值
```

2. 组合模型

组合模型基本思想和步骤

一个非平稳序列可表示为 $X_t = \mu_t + Y_t$ ，其中 μ_t 是 X_t 随时间变化的均值，刻画非平稳序列中的确定性组成部分； Y_t 是一个平稳序列，可以拟合一个 ARMA 模型，描述 X_t 中的随机组成部分。具体的处理方法分成两大类：第一类是设法消除确定性因素（长期趋势、周期趋势）的作用，然后对剩余序列拟合一个 ARMA 模型。第二类是从系统分析的观点出发，用确定性模型描述序列中确定性因素的变动规律，用 ARMA 模型刻画序列中随机因素的一般规律性，从而建立起确定性加随机的组合模型。

建立组合模型的基本步骤如下

(1) 根据时间序列的特性，用一定的函数形式拟合序列的确定性组成部分，直到剩余序列平稳为止；

(2) 对剩余序列拟合适应的 ARMA 模型；

(3) 将分别拟合的确定性模型和 ARMA 模型结合起来并以其参数作为初始值, 用非线性最小二乘法估计组合模型的参数, 得到最终的组合模型。

确定性趋势的判定

判定是否存在确定性趋势的最简单的方法, 是画出序列的趋势图, 然后进行观察分析。再一个方法是根据特征根来判别。当所有特征根的模都小于 1 时, 序列是平稳的, 否则就是非平稳的。当序列不平稳时, 特征根还可以告诉我们存在某种确定性趋势。因为格林函数决定了序列的具体模式。

如果存在绝对值接近于 1 或大于 1 的实特征根, 则说明序列中可能存在长期增减趋势。具体有以下几种情形:

(1) 常数趋势

如果一个序列中存在着常数趋势, 由于系统的随机性, 序列呈现在某一常数水平上下波动。这种系统在时刻 t 的响应, 如果没有随机扰动, 就等于前一时刻的值, 即 $X_t = X_{t-1}$; 但由于随机扰动的作用, 有 $X_t = X_{t-1} + a_t$ 。这个模型实际上是 $\phi_1 = 1$ 时的 AR(1) 模型, 即 AR(1) 模型的极限形式——随机游动。

对于 AR(1) 系统来说, 若 $\lambda_1 = \phi_1 \rightarrow 1$, 则序列中存在着常数趋势, 此时格林函数 $G_j = \phi_1^j = \lambda_1^j \rightarrow 1$; 当 $\lambda_1 = \phi_1 = -1$ 时, 有 $X_t = -X_{t-1} + a_t$, 此时

$$G_j = \phi_1^j = (-1)^j,$$

事实上, 这时系统响应为两个常数序列

$$\begin{cases} X_t = X_0 + a_t, & t \text{ 为偶数,} \\ X_t = -X_0 + a_t, & t \text{ 为奇数.} \end{cases}$$

综上所述, 当 $|\lambda_1| \rightarrow 1$ 时, 系统可能存在常数趋势。对于高价模型来说, 若其中有一个实根的绝对值接近于 1, 则系统存在常数趋势。

(2) 线性趋势

以 ARMA(2, 1) 模型为例进行说明。当特征根 $\lambda_1 = \lambda_2 = 1$ 时, 格林函数为

$$G_j = (1 - \theta_1)j + 1,$$

X_t 可以表示为

$$X_t = \sum_{j=0}^{\infty} G_j a_{t-j} = \sum_{j=0}^{\infty} [(1 - \theta_1)j + 1] a_{t-j},$$

这说明随着 j 的增大, X_t 以线性形式 $1 + (1 - \theta_1)j$ 而变化。

对于一个高价模型, 若其中有两个实特征根的绝对值接近于 1, 序列中可能存在线性趋势。推而广之, 若有三个实特征根的绝对值接近于 1, 序列中可能存在二次趋势; 若有 $n+1$ 个实特征根的绝对值接近于 1, 序列中可能存在 n 次多项式趋势, 因此可以用

$$X_t = \beta_0 + \beta_1 t + \cdots + \beta_n t^n + Y_t$$

来描述。

(3) 指数趋势模型

若有 l 个绝对值大于 1 的实特征值, 则序列中可能存在 l 个指数增减趋势, 可以用

$$X_t = \sum_{j=1}^l R_j e^{\lambda_j t} + Y_t$$

来拟合。

组合模型的建立

当我们判明序列中存在的确定性趋势后, 就要用合适的函数形式来拟合确定性部分, 直到剩余序列平稳为止, 然后对剩余序列拟合适应的 ARMA 模型, 最后建立起组合模型。

$$\begin{cases} X_t = f(t) + Y_t, \\ Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_n Y_{t-n} + a_t - \theta_1 a_{t-1} - \cdots - \theta_m a_{t-m}. \end{cases} \quad (121)$$

下面我们来看一个实例。

例 14.13 对 1946—1970 年美国各季生产者耐用品支出建模, 数据见表 14.28。

表 14.28 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

(1) 首先对此序列进行观察分析

图 14.63 为数据曲线图，可以看出具有指数上升趋势，因此，我们对确定性部分先拟合一个指数增长模型，即

$$X_t = \mu_t + Y_t, \quad \mu_t = R_1 e^{r_1 t},$$

这里各季度依次编号为 $t=1, 2, \dots, 100$ 。

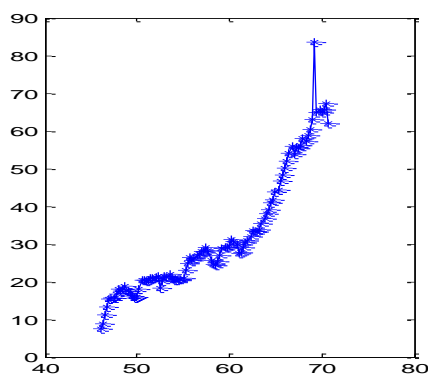


图 14.63 耐用消费品支出曲线图

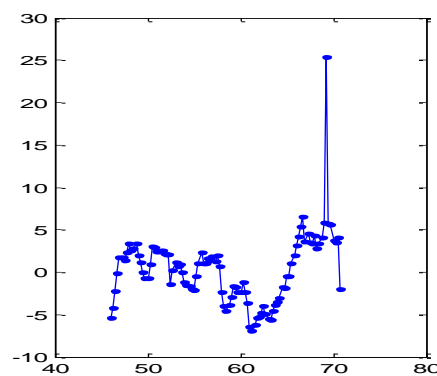


图 14.64 剩余序列数据图

(2) 确定性趋势的拟合

为了能用线性最小二乘法估计参数 R_1 和 r_1 ，把 $\mu_t = R_1 e^{r_1 t}$ 两边取对数得

$$\ln \mu_t = \ln R_1 + r_1 t,$$

利用观测数据求得 $\hat{R}_1 = 12.6385$ ， $\hat{r}_1 = 0.0162$ 。剩余平方和为 1683.5371。剩余序列 Y_t 如图 14.64 所示，可以认为是平稳的。

(3) 对剩余序列拟合 ARMA 模型

Y_t 的自相关与偏自相关如图 14.65 所示，可初步断定 Y_t 的适应模型为 AR 模型，逐步增加

AR 模型阶数进行拟合，其残差方差图如图 14.66 所示，因此，合适的模型为 AR (2)，即

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t,$$

参数估计为 $\hat{\phi}_1 = 0.5451$ ， $\hat{\phi}_2 = 0.2478$ 。

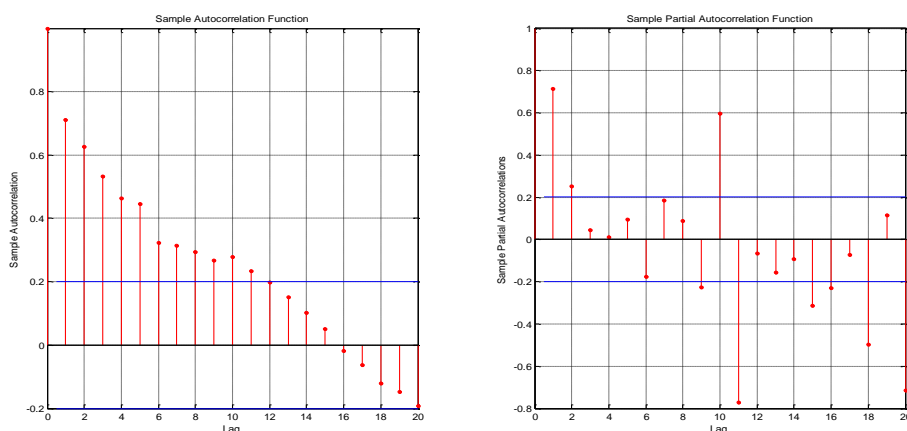


图 14.65 剩余序列的相关函数图

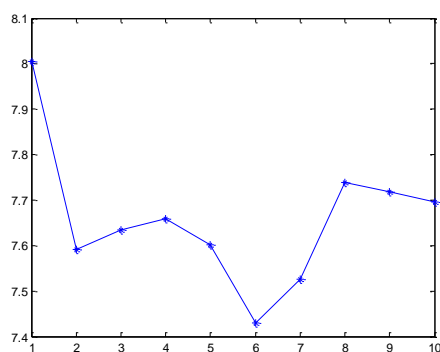


图 14.66 残差方差图

上面进行计算和画图的 MATLAB 程序如下

```
clc, clear
a=load('data1428.txt');
a=a'; a=a(:); %把数据变成列向量
n=length(a);
t0=[46:1/4:71-1/4]';
t=[1:100]';
xishu=[ones(n,1),t];
cs=xishu\log(a);
cs(1)=exp(cs(1))
ahat=cs(1)*exp(cs(2)*t);
cha=a-ahat;
res=sum(cha.^2)
subplot(121), plot(t0,a,'*-')
subplot(122), plot(t0,cha,'.-')
figure, subplot(121), autocorr(cha)
subplot(122), parcorr(cha)
for i=1:10
    cs2{i}=ar(cha,i); %拟合模型
    lin=cs2{i}; %临时变量，新版 MATLAB 中 resid 不认细胞数组
    cha2=resid(lin,cha); %计算残差向量
    myvar(i)=sum(cha2.^2)/(100-i); %计算残差方差
end
```

figure, plot(myvar,'*-')

(4) 建立组合模型

最后我们要以已估计出来的 R_1, r_1, ϕ_1, ϕ_2 的值为初始值用非线性最小二乘法对模型参数进行整体估计，模型整体可写为

$$X_t = \mu_t + Y_t = R_1 e^{r_1 t} + \phi_1 (X_{t-1} - R_1 e^{r_1(t-1)}) + \phi_2 (X_{t-2} - R_1 e^{r_1(t-2)}) + a_t$$

最终的参数整体估计为

$$\hat{R}_1 = 12.1089, \quad \hat{r}_1 = 0.017, \quad \hat{\phi}_1 = 0.517, \quad \hat{\phi}_2 = 0.2397.$$

残差平方和为 738.4402，残差自相关图 14.67 表明整体模型是适应的。我们用所建的模型以 1970 年第 4 季度即 $t=100$ 为原点进行预测，结果如表 14.29 所示。

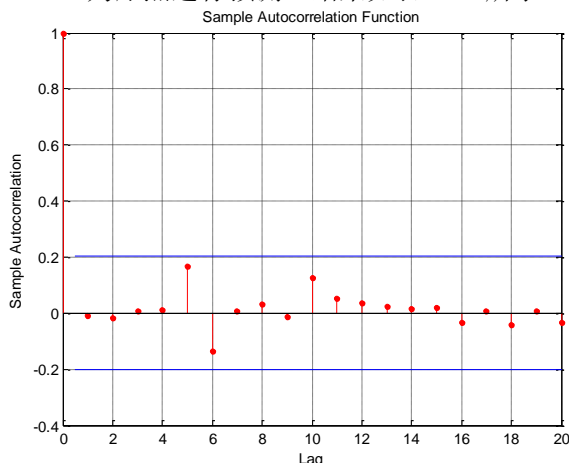


图 14.67 拟合整体模型后的残差自相关图

表 14.29 耐用品支出预测表

l	0	1	2	3	4	5	6	7	8
$t=100$	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
$\hat{X}_t(l)$	62.1	65.8298	66.8384	68.562	70.0083	71.4879	72.9238	74.3507	75.768

计算的 MATLAB 程序如下

```

clc, clear
xt=@(cs,x) cs(1)*(exp(cs(2)*x(:,3))-cs(3)*exp(cs(2)*(x(:,3)-1))-...
    cs(4)*exp(cs(2)*(x(:,3)-2)))+cs(3)*x(:,1)+cs(4)*x(:,2);
cs0=[12.6385,0.0162,0.5451,0.2478]';
a=load('data1428.txt');
a=a'; a=a(:); %把数据变成列向量
x=[a(2:end-1),a(1:end-2),[3:100]'];
cs=lsqcurvefit(xt,cs0,x,a(3:end))
res=a(3:end)-xt(cs,x);
Q=sum(res.^2) %计算残差平方和
autocorr(res) %画残差自相关图
xhat=a;
for j=101:108
    xhat(j)=cs(1)*(exp(cs(2)*j)-cs(3)*exp(cs(2)*(j-1))-...
        cs(4)*exp(cs(2)*(j-2)))+cs(3)*xhat(j-1)+cs(4)*xhat(j-2);
end
xhat101_108=xhat(101:108)

```

参考文献

- [1] 王振龙主编，应用时间序列分析（第二版），中国统计出版社，2010 年。