时间序列模型的应用及探究

时间序列：将同一统计指标的数值按其发生的时间先后顺序排列而成的数列。

时间序列模型也成为回归模型，一方面承认事物发展的延续性，运用过去时间序列的数据进行统计就可以推测事物的发展趋势；另一方面，充分考虑到偶然因素影响的随机性，使用历史数据，进行统计分析对数据进行适当处理来消除随机波动的影响。简单易行，便于掌握并充分运用时间序列的各项数据，计算速度较快，能够比较精确的确定模型的动态参数；但是不能够反映事物的内在联系，结合因素的相互联系，只适用于短期预测。

1.1 确定性时间序列分析方法

时间序列预测技术是通过对预测目标自身时间序列的处理，来研究其变化趋势，一个时间序列的格式往往是由以下几种变化形式的叠加或耦合：

1. 长期趋势变动。指的是时间序列朝着一定的方向持续上升或者下降，或者停留在某一水平的倾向，反映了客观事物的主要变化趋势。
2. 季节变动。
3. 循环变动指的是周期一年以上，由非季节因素引起的涨落起伏波形相似的波动。
4. 不对则变动，分为突然变动和随机变动。
    1) 加法模型：

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

    2) 乘法模型：

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

    3) 混合模型：

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

其中，y_t表示观测目标的观测记录，均值$E\left(R_t\right)=0$方差$Var\left(R_t\right)=\sigma^2$。

若在预测时间范围之内，无变动且随便变动方差$Var\left(R_t\right)=\sigma^2$较小，则可以使用一些方法来检验过去和现在的演变趋势。

1.1.1 移动平均法

假设观测序列为$y_1,\cdots,y_T$，取移动平均的项数$N<T$。一次移动的平均值计算公式为

$$\begin{aligned} M_t^{(1)} &= \frac{1}{N}\left(y_t+y_{t-1}+\cdots+y_{t-N-1}\right) \\ &= \frac{1}{N}\left(y_{t-1}+\cdots+y_{t-N}\right)+\frac{1}{N}\left(y_t-y_{t-N}\right)=M_{t-1}^{(1)}+\frac{1}{N}\left(y_t-y_{t-N}\right), \end{aligned} \tag{0.1}$$

则二次移动平均值计算公式为：

$$\begin{aligned} M_t^{(2)} &= \frac{1}{N}\left(M_t^{(1)}+\cdots+M_{t-N-1}^{(1)}\right) \\ &= M_{t-1}^{(2)}+\frac{1}{N}\left(M_t^{(1)}-M_{t-N}^{(1)}\right), \end{aligned} \tag{0.2}$$

若预测目标的趋势实在某一水平上下波动时，可以使用一次移动平均方法建立预测模型，即

$$\hat{y}_{t+1}=M_t^{(1)}=\frac{1}{N}\left(y_t+y_{t-1}+\cdots+y_{t-N-1}\right)=N,N+1,\cdots,T, \tag{0.3}$$

其预测标准误差为

$$S=\sqrt{\frac{\sum_{t=N+1}^{T}\left(\hat{y}_t-y_t\right)^2}{T-N}}。 \tag{0.4}$$

最近 N 期序列值的平均值作为未来各期的预测结果。N 的取值随着历史序列的基本趋势变动而变动。在有确定的机械变动周期的资料中，移动平均的项数应该取周期长度，选择最佳 N 值的方法：比较若干模型的预测误差，预测标准误差最小者为好。

当预测目标的基本趋势与某一线性模型相吻合时，常使用二次移动平均法，但是序列同时存在线性趋势与周期波动时，可以使用趋势移动平均法建立预测模型：

$$\hat{y}_{T+m} = a_T + b_T m, m = 1, 2, \cdots, m,$$

其中， $a_T = 2M_T^{(1)} - M_T^{(2)}$； $b_T = \dfrac{2}{N-2}\left(M_T^{(1)} - M_T^{(2)}\right)$。

1.1.2 指数平滑法

一次移动平均实际上认为最近 N 期数据对未来值影响相同，都加权 1/N 而 N 期以前的数据对未来值没有影响，加权为 0。二次及更高次移动平均的加权却不是 1/N，次数越高，权数的结构就越复杂，但是永远保持对称的权数，即两端项权数小，中间项权数大，将不符合系统的动态性。历史数据对未来值的影响都是随着时间间隔的增长而递减，所以更切合实际的方法应是对各期观测值依时间顺序进行加权平均作为预测值。指数平滑法可以满足这一要求，但是需要具有简单的递推形式。

指数平滑法根据次数的不同展示为以下形式。

1．一次指数平滑法

1) 预测模型

假设时间序列为 $y_1, \cdots, y_T$， α 为加权系数， $0 < \alpha < 1$ 一次指数平滑公式为一次移动的平均值计算公式为

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha\left(y_t - S_{t-1}^{(1)}\right) \tag{0.5}$$

(0.5)式是根据移动平均公式改进而来，则移动平均数的递推公式为

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

以 $M_{t-1}^{(1)}$ 作为 y_{t-N} 最佳估计，则有

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

使 $\alpha = \dfrac{1}{N}$， 以 S_t 代替 $M_t^{(1)}$ 可得(0.5)式，即

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

展开式(0.5)，则有

$$S_t^{(1)}=\alpha y_t+(1-\alpha)\left[\alpha y_{t-1}+(1-\alpha)S_{t-2}^{(1)}\right]=\cdots=\alpha\sum_{j=0}^{\infty}(1-\alpha)^j y_{t-j}。\quad(0.6)$$

(0.6)式表明$S_t^{(1)}$是全部历史数据的加权平均，加权系数分别是α、$\alpha(1-\alpha)$、$\alpha(1-\alpha)^2$、$\cdots$、则

$$\alpha\sum_{j=0}^{\infty}(1-\alpha)^j=\frac{\alpha}{1-(1-\alpha)}=1,\quad(0.7)$$

因为加权系数符合指数规律且具有平滑数据的功能，所以称之为指数光滑。以这种平滑值进行预测，就是一次指数光滑法，预测模型为

$$\hat{y}_{t+1}=S_t^{(1)}\quad(0.8)$$

即

$$\hat{y}_{t+1}=\alpha y_t+(1-\alpha)\hat{y}_t\quad(0.9)$$

(0.9)式表示以第 t 期指数平滑值作为 t+1 期预测值。

2) 加权系数的选择

加权系数α越大，则新数据所占的比重就越来越大，原预测值所占的比重就越小，反之亦然。新预测值是根据预测误差对原预测值进行修正而得到的。加权系数α体现了修正的幅度，加权系数越大，修正幅度就越大；加权系数越小，修正幅度就越小。

加权系数α应该根据时间序列的具体性质进行选择，具体选择可以根据以下原则：一、若时间序列波动不大，则应当取小一点，以减小修正幅度，使得预测模型能够包含较长时间序列的信息；二、若时间序列波动具有迅速且明显的变动倾向，则应该取大一点，使得预测模型灵敏度更高，以迅速跟上数据的变化。

3) 初始值的确定

初始值是由预测者估计或者指定的，使用一次指数平滑法进行预测，除了选择合适的加权系数α外，还需要确定初始值$s_0^{(1)}$当时间序列的数据较多，比如数量在 50 个以上，初始值对以后的影响很小，就可以选择第一期数据作为初始值，若初始值数量在 5 个以下，初始值对以后的预测值影响很大，就必须研究如何正确确定初始值。一般来说是选取最初几期实际值的平均值作为初始值。

2. 二次指数平滑法

当使用一次指数平滑法进行预测时，若时间序列的变动出现直线趋势时，就会存在明显的滞后偏差，所以必须加以修正。再次做出二次指数平滑，利用之后偏差的规律建立直线趋势模型，这就是而辞职书平滑法。计算公式为

$$\begin{cases} S_t^{(1)} = \alpha y_t + \left(1-\alpha\right) S_{t-1}^{(1)}, \\ S_t^{(2)} = \alpha S_t^{(1)} + \left(1-\alpha\right) S_{t-1}^{(2)}, \end{cases} \tag{0.10}$$

其中，$S_t^{(1)}$ 表示一次指数的平滑值；$S_t^{(2)}$ 表示二次指数的平滑值。

当时间序列 $\left\{y_t\right\}$ 从某时期开始具有直线趋势时，可以使用直线趋势模型

$$\hat{y}_{t+m} = a_t + b_t m, m = 1, 2, \mathrm{L} \ , \tag{0.11}$$

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)}, \\ b_t = \dfrac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right), \end{cases} \tag{0.12}$$

进行预测。

3. 三次指数平滑法

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

$$\begin{cases} S_t^{(1)} = \alpha y_t + \left(1-\alpha\right) S_{t-1}^{(1)}, \\ S_t^{(2)} = \alpha S_t^{(1)} + \left(1-\alpha\right) S_{t-1}^{(2)}, \\ S_t^{(3)} = \alpha S_t^{(2)} + \left(1-\alpha\right) S_{t-1}^{(3)}, \end{cases} \tag{0.13}$$

其中，$S_t^{(3)}$ 表示三次指数的平滑值。

三次指数平滑法的预测模型为：

$$\hat{y}_{t+m} = a_t + b_t m + c m^2, m = 1, 2, \mathrm{L} \ , \tag{0.14}$$

其中，

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)}, \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}\right], \\ c_t = \dfrac{\alpha^2}{2(1-\alpha)^2}\left[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right] \end{cases} \tag{0.15}$$

1.1.3 差分指数平滑法

若时间序列的变动具有直线趋势，使用一次指数光滑法会出现滞后偏差，原因在于数据不能够满足模型要求，那么可以从数据变换的角度来考虑改进措施，即在运用指数平滑法对先前的技术做出一些处理，使其能够适合于一次指数平化模型，接着在对处理完的结果作技术上的返回处理，使之恢复为原变量的形态。差分方法是改变数据变动趋势的简易方法。以下是使用差分方法改进指数平滑法的步骤。

当时间序列呈直线增加时，运用一阶差分指数模型进行预测。模型为

$$\nabla y_t = y_t - y_{t-1} \tag{0.16}$$

$$\nabla \hat{y}_{t+1} = \alpha \nabla y_t + (1-\alpha)\nabla \hat{y}_t \tag{0.17}$$

$$\hat{y}_{t+1} = \nabla y_{t+1} + y_t \tag{0.18}$$

其中，∇ 为差分记号。

(0.16)式表现对呈现直线增加的序列作一阶差分，构建成一个平稳的序列；(0.18)式表示把经过一阶差分后的新序列的指数平滑预测值与变量的实际值叠加，作为变量下一期的预测值，公式为

$$y_{t+1} = y_{t+1} - y_t + y_t = \nabla y_{t+1} + y_t \tag{0.19}$$

指数平滑值指的是一种加权平均数，把序列中逐期增加的加权平均数加上当前值得实际数进行预测，比一次指数平滑法只用变量以往取值的加权平均数作为下一期的预测更加合理，从而使得预测值始终围绕实际值上下波动，从根本上解决了在有直线增长趋势的情况下，用一次指数平滑法所得出的结果始终落后于是机制的问题。

差分方法和指数平滑法的联合运用可以克服一次指数平滑法的滞后偏差以及改进初始值的问题。在经过数据的差分处理之后，对于所产生的新序列基本上是平稳的。基于这种情况，首先，初始值取新序列的第一期数据对于未来预测值不会产生多大的影响；其次，它拓展了指数平滑发的适用范围。但是，对于指数平滑法的加权系数的选择、只能够逐期预测问题，差分指数预测模型未能将其进行改进。

1.1.4 具有季节性特点的时间序列的预测

标题中国提到的季节不仅指的是自然季节，还可能指的是商品销售的机械。在现实经济活动中，对于季节性时间序列的预测，需要从数学上完全拟合其变化曲线非常困难，但是从拟合曲线目的出发，预测能够找到时间序列的变化趋势，尽可能精确，可以使用季节系数法。

步骤如下：

1) 首先，收集m年的每年各季度或者各年份的时间序列样本数据a_{ij}其中，i 表示年份的序号；j表示季度或者月份的序号。$i,j=1,2,\cdots,m$

2) 计算出每年所有季度或者所有月份的算数平均值$\bar{a}=\frac{1}{k}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij},k=mn$。。

3) 计算同季度或者同月份数据的算术平均值$\bar{a}_{.j}=\frac{1}{m}\sum_{i=1}^{m}a_{ij},j=1,2,\cdots,m$。

4) 计算系度系数或者月份系数$b_j=\bar{a}_{.j}/\bar{a}$。

5) 预测计算。当时间序列式按季度列出是，先求出预测年份（下一年）的年加权平均

$$y_{m+1}=\frac{\sum_{i=1}^{m}w_iy_i}{\sum_{i=1}^{m}w_i}, \tag{0.20}$$

其中，$y_i=\sum_{j=1}^{n}a_{ij}$表示第i年的年合计数；w_i表示第i年的权数。

再计算预测年份的季度平均值 $\bar{y}_{m+1}:\bar{y}_{m+1}=y_{m+1}/n$。最后，预测年份第 j 季度的预测值为

$$y_{m+1,j}=b_j\bar{y}_{m+1}\text{。}$$

1.2 ARIMA 序列与季节性序列

在应用时间序列解决实际问题时，不难发现时间序列往往有三个特征：趋势性、季节性、非平稳性。

1.2.1 ARIMA 序列及其预报

差分运算可以使一类非平稳序列平文化，若一阶差分不能够使时间序列平稳化，则进行二阶差分、三阶差分，直至第 d 阶差分。最后平稳序列。

一阶差分：

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

二阶差分：

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

一般来说，d 阶差分为

$$\nabla^d X_t=(1-B)^d X_t, \tag{0.21}$$

其中，∇^d 成为 d 阶差分算子，则

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

假设 $\{X_t,t=0,\pm1,\pm2,\cdots\}$ 是非平稳序列，若存在正整数 d 使得

$$\nabla^d X_t=W_t,$$

而 $\{X_t,t=0,\pm1,\pm2,\cdots\}$ 是 ARMA(p,q) 序列，则称 X_t 是 ARMA(p,d,q) 序列。则 X_t 满足

$$\varphi(B)\nabla^d X_t=\theta(B)\varepsilon_t, \tag{0.22}$$

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

$$\varphi(B)\left(\nabla^d X_t - \mu\right) = \theta(B)\varepsilon_t, t > d \tag{0.23}$$

那么称X_t是 ARMA(p,d,q)序列，若μ未知，则用$\nabla^d X_t$的平均值$\bar{x}$进行估计。

若初值$X_1, X_2, \cdots, X_d$ 已知，则

$$W_t = \nabla^d X_t, t = d+1, \cdots, n,$$

就可以复原X_t。

以下简单介绍 ARIMA 序列的预报

假设$\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$是 ARMA$(p,d,q)$序列，则基于$d = 1, 2$的情况，

1) 当$d = 1$时，

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

则

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

所以

$$\hat{X}_k(m) = \hat{X}_k(m-1) + \hat{W}_k(m) = X_k + \sum_{j=1}^{m} \hat{W}_k(J) \tag{0.24}$$

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

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

复原$\hat{X}_k(m)$，可得

$$\hat{X}_k(m) = \hat{X}_k(m-1) + \hat{W}_k(m) = X_k + \sum_{j=1}^{m} \hat{W}_k(J) \tag{0.25}$$

1.2.2 季节性序列及其预报

在一些实际问题中，时间序列有很明显的周期规律性，如气温、用电量等。由季节性因素或者其他因素引起的周期性变化的时间序列，称为季节性时间序列，

行营的模型为季节性模型。

一般地，对周期s的序列，可先进行差分运算，即

$$\begin{aligned}\nabla_s X_t &= \left(1-B^S\right)X_t,\\ \nabla_s^d &= \left(1-B^S\right)^d X_t,\end{aligned} \tag{0.26}$$

然后再进行 ARIMA 建模。

1.3 纺织生产布料问题研究

1.3.1 问题重述

我国 1974 年——1981 年布的产量如表 1.1 所示。

表 1.1 我国 1974 年——1981 年布的产量

年份	1974	1975	1976	1977	1978	1979	1980	1981
产量	80.8	94	88.4	101.5	110.3	121.5	134.7	142.7

1) 试用趋势移动平均法来建立布的年产量预测模型。
2) 分别取α=0.3,α=0.6，$S_0^{(1)}=S_0^{(2)}=\dfrac{y_1+y_2+y_3}{3}=87.7$建立布的直线指数平滑预测模型。
3) 计算模型拟合误差，比较三个模型的优劣
4) 用最优的模型预测 1982 年和 1985 年布的产量。

1.3.2 符号规定与基本假设

1. 符号规定

1) $t=1,2,\cdots 12$ 分别代表 1974 年——1985 年；
2) $y_t\left(t=1,2,\cdots,8\right)$ 表示已知的 8 个预测值；
3) $\hat{y}_t\left(t=9,10,11\right)$ 表示y_t递推预测的预测值； bar
4) S_1 表示预测的标准误差。

2. 基本假设

1) 假设布的产量不受机器生产技术影响；

2) 假设只考虑布的产量不考虑除生产以外的其余因素影响。

1.3.3 模型的分析与建立

1) 建立布的年产量预测模型

$$\hat{y}_{t+1}=\frac{y_t+y_{t-1}+y_{t-2}}{3},t=3,4,\cdots 11 \tag{0.27}$$

预测的标准误差为

$$S_1=\sqrt{\frac{\sum_{t=3}^{8}\left(\hat{y}_t-y_t\right)^2}{8-3}}=19.3542$$

2) 二次指数平滑法的计算公式为：

$$\begin{cases} S_t^{(1)}=\alpha y_t+\left(1-\alpha\right)S_{t-1}^{(1)}, \\ S_t^{(2)}=\alpha S_t^{(1)}+\left(1-\alpha\right)S_{t-1}^{(2)}, \end{cases} \tag{0.28}$$

其中，$S_t^{(1)}$表示一次指数的平滑值；$S_t^{(2)}$表示二次指数的平滑值

当时间序列$\{y_t\}$从某时期开始具有直线趋势时，可以使用直线趋势模型

$$\hat{y}_{t+m}=a_t+b_t m, m=1,2,\cdots, \tag{0.29}$$

$$\begin{cases} a_t=2S_t^{(1)}-S_t^{(2)}, \\ b_t=\dfrac{\alpha}{1-\alpha}\left(S_t^{(1)}-S_t^{(2)}\right), \end{cases} \tag{0.30}$$

进行预测。

若$\alpha=0.3$则预测的标准误差为

$$S_2=\sqrt{\frac{\sum_{t=2}^{8}\left(\hat{y}_t-y_t\right)^2}{8-1}}=11.7966$$

若$\alpha=0.6$则预测的标准误差为

$$S_3=\sqrt{\frac{\sum_{t=3}^{8}\left(\hat{y}_t-y_t\right)^2}{8-3}}=7.0136$$

3) 从标准差的角度考虑，选择$\alpha=0.6$的二次指数平滑模型。

4) 本题选用$\alpha=0.6$的二次指数平滑模型作为最优化的模型得到 1982 年和 1985 年的产量预测值为 152.9452 亿米和 182.825 亿米。

1.3.4 模型的求解

1) 由第一问编写如下所示 MATLAB 代码：

```
yt=[80.8 94.0 88.4 101.5 110.3 121.5 134.7 142.7];
m=length(yt);
n=3;
for i=n+1:m+1
       ytt(i)=sum(yt(i-n:i-1))/n;
end
ytt
for i=m+1:m+3
       yt(i)=ytt(i);
       ytt(i+1)=sum(yt(i-n+1:i))/n;
end
yhat=ytt(end-3:end)
s1=sqrt(mean((yt(n+1:m)-ytt(n+1:m)).^2))
```

2) 运行结果如下：

```
ytt =

           0           0           0     87.7333     94.6333    100.0667    111.1000
122.1667    132.9667


yhat =

    132.9667    136.7889    137.4852    135.7469
```

3) 由第二问编写如下所示 MATLAB 代码

```
  function xiti81
[sigma1,yhat21]=yuce(0.3)
[sigma2,yhat22]=yuce(0.6)
function [sigma,yhat2]=yuce(alpha);
yt=[80.8 94.0 88.4 101.5 110.3 121.5 134.7 142.7];
n=length(yt);st1(1)=mean(yt(1:3));st2(1)=st1(1);
```

```
for i=2:n
    st1(i)=alpha*yt(i)+(1-alpha)*st1(i-1);
    st2(i)=alpha*st1(i)+(1-alpha)*st2(i-1);
end
at=2*st1-st2;
bt=alpha/(1-alpha)*(st1-st2);
yhat=at+bt;
sigma=sqrt(mean((yt(2:end)-yhat(1:end-1)).^2));
m=1:4;
yhat2=at(end)+bt(end)*m;
```

4) 第二问运行结果：

```
 sigma1 =
   11.7966
yhat21 =
   143.6959    150.0191    156.3423    162.6655
sigma2 =
     7.0136
yhat22 =
   152.9452    162.9176    172.8900    182.8625
```

1.3.5 结果分析

本题第一问建立出布的年产量预测模型为(0.27)。本题第二问建立布的直线指数预测模型为(0.29)。本题第三问从标准差的角度考虑，选择$\alpha=0.6$的二次指数平滑模型。本题第四问选用$\alpha=0.6$的二次指数平滑模型作为最优化的模型得到 1982 年和 1985 年的产量预测值为 152.9452 亿米和 182.825 亿米

1.4 商品零售额问题研究

1.4.1 问题重述

1960 年—1985 年全国社会商品零售额如图 1.2 所示。

表 1.2 全国社会商品零售额数据

年份	1960	1961	1962	1963	1964	1965	1966	1967
零售总额	696.6	607.7	604	604.5	638.2	670.3	732.8	770.5

年份	1968	1969	1970	1971	1972	1973	1974	1975
零售总额	737.3	801.5	858	929.2	10233	1106.7	1163.6	12711

年份	1976	1977	1978	1979	1980	1981	1982
零售总额	1339.4	1432.8	1558.6	1800	2140	2350	2570

试用三次指数平滑法预测 1983 年和 1958 年全国社会商品零售额。

1.4.2 符号规定与基本假设

1. 符号规定

1) $y_t\left(t=1,2,\cdots,22\right)$ 表示已知的 22 个预测值；

2) $\hat{y}_t\left(t=9,10,11\right)$ 表示 y_t 递推预测的预测值；

3) $S_t^{(i)}, i=1,2,3$ 表示第 i 次指数平滑值。

4) $\hat{y}_{23+m}\left(m=1,2,3\right)$ 表示 1983，1984，1985 年销售额的预测值

2. 基本假设

1) 假设本问题考虑全社会商品零售额数据；
2) 假设本问题只考虑销售，不考虑其余因素
3) 假设本问题只考虑销售额总额，不考虑其余分支

1.4.3 模型的分析与建立

令加权系数 $\alpha=0.3$，则计算公式为

$$\begin{cases} S_t^{(1)}=\alpha y_t+\left(1-\alpha\right)S_{t-1}^{(1)}, \\ S_t^{(2)}=\alpha S_t^{(1)}+\left(1-\alpha\right)S_{t-1}^{(2)}, \\ S_t^{(3)}=\alpha S_t^{(2)}+\left(1-\alpha\right)S_{t-1}^{(3)}, \end{cases} \tag{0.31}$$

其中，$S_t^{(1)}$ 表示一次指数的平滑值；$S_t^{(2)}$ 表示二次指数的平滑值；$S_t^{(3)}$ 表示三次指数的平滑值。

初始值为

$$S_0^{(1)}=S_0^{(2)}=S_0^{(3)}=\frac{y_1+y_2+y_3}{3}=636.2$$

三次指数平滑法的预测模型为：

$$\hat{y}_{t+m}=a_t+b_t m+cm^2, m=1,2,\mathrm{L}\ , \tag{0.32}$$

其中，

$$\begin{cases} a_t=3S_t^{(1)}-3S_t^{(2)}+S_t^{(3)}, \\ b_t=\dfrac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)S_t^{(1)}-2(5-4\alpha)S_t^{(2)}+(4-3\alpha)S_t^{(3)}\right], \\ c_t=\dfrac{\alpha^2}{2(1-\alpha)^2}\left[S_t^{(1)}-2S_t^{(2)}+S_t^{(3)}\right] \end{cases} \tag{0.33}$$

根据(0.33)式可知，当$t=23$时，存在

$$a_{23}=2572.2613, b_{23}=259.3374, c_{23}=8.9819$$

则预测模型为

$$\hat{y}_{23+m}=8.9819m^2+259.3374m+2572.2613$$

最后求得 1983，1985 年销售额的预测值分别是 240.5806 亿元，3431.1106 亿元。

1.4.4 模型的求解

使用MATLAB软件进行编程，将表1.2中的数据保存到纯文本文件data02.txt中。代码为

```
clc,clear
dd=textread('data02.txt');
yt=dd([2:2:end],:);yt=yt';yt=nonzeros(yt);
n=length(yt);alpha=0.4;st0=mean(yt(1:3))
st1(1)=alpha*yt(1)+(1-alpha)*st0;
st2(1)=alpha*st1(1)+(1-alpha)*st0;
st3(1)=alpha*st2(1)+(1-alpha)*st0;
for i=2:n
st1(i)=alpha*yt(i)+(1-alpha)*st1(i-1);
st2(i)=alpha*st1(i)+(1-alpha)*st2(i-1);
st3(i)=alpha*st2(i)+(1-alpha)*st3(i-1);
```

```
end
xlswrite('lingshou.xls',[st1',st2',st3'])
at=3*st1-3*st2+st3;
bt=0.5*alpha/(1-alpha)^2*((6-5*alpha)*st1-2*(5-4*alpha)*st2+(4-3*alpha)*st3);
ct=0.5*alpha^2/(1-alpha)^2*(st1-2*st2+st3);
yhat=at+bt+ct;
xlswrite('lingshou.xls',yhat','sheet1','D2')
plot(1:n,yt,'D',2:n,yhat(1:end-1),'*')
legend('实际值','预测值',2)
xishu=[ct(end),bt(end),at(end)];
yuce=polyval(xishu,[1:3])
```

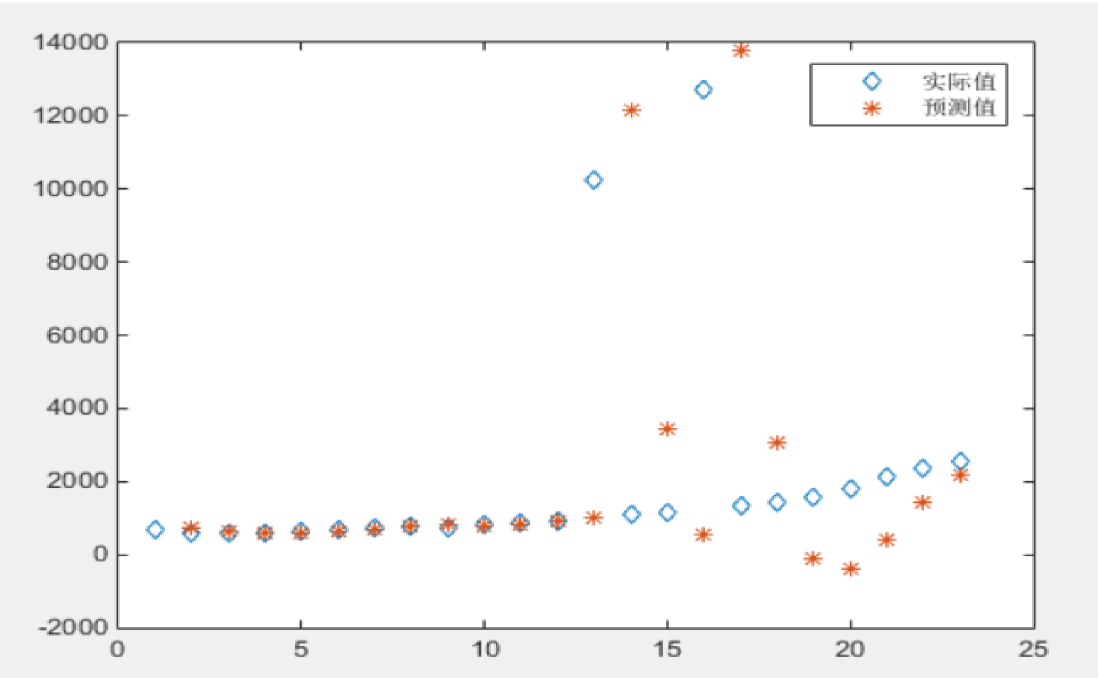
1.4.5 结果分析

上述代码运算出结果为：

```
st0 =
   636.1000
yuce =
     1.0e+03 *
       2.7251       3.0329       3.4101
```

即 1983，1985 年销售额的预测值分别是 240.5806 亿元，3431.1106 亿元。

图 1.1 拟合图像



1.5 耐用品支出资料问题研究

1.5.1 问题重述

1946—1970 年美国各系耐用品支出资料如表 1.3 所示。

1) 对所给时间序列建模；

2) 对时间序列进行两年的预报

表 1.3 1946—1970 年美国各系耐用品支出资料

年度	一季	二季	三季	四季	年度	一季	二季	三季	四季
1946	7.5	8.9	11.1	13.4	1959	27	28.7	29.1	29
1947	15.5	15.7	15.6	16.7	1960	29.6	31.2	30.6	29.8
1948	18	174	17.9	18.8	1961	27.6	27.7	29	30.3
1949	17.6	17	16.1	15.7	1962	31	32.1	33.5	33.2
1950	159	17.9	20.3	20.4	1963	33.2	33.8	35.5	36.8
1951	20.2	20.5	20.9	20.9	1964	37.9	39	40	416
1952	21.1	21.4	18.2	20.1	1965	43.7	44.4	46.6	48.3
1953	21.4	21.3	21.9	21.3	1966	50.2	52.1	54	56
1954	20.4	20.4	20.7	20.7	1967	53.9	55.6	55.4	56.2
1955	20.9	23	24.9	26.5	1968	57.9	57.3	58.8	60.4
1956	25.6	26.1	27	27.2	1969	63.1	83.5	64.8	65.7
1957	28.1	28	29.1	28.3	1970	64.8	65.6	67.2	62.1
1958	25.7	24.5	24.4	25.5					

1.5.2 符号规定与基本假设

1. 符号规定

1. $\hat{R}_1$ 表示参数估计

2. $t\left(t=1,2,\cdots m\right)$ 表示周期

2. 基本假设

1. 假设本题只考虑耐用品的支出情况；

2. 假设耐用品支出的影响只受到销售的影响

1.5.3 模型的分析与建立

1) 对所给时间序列建模

1. 首先对序列进行观察分析，对确定部分拟合一个指数增长模型，即

$$X_t = \mu_t + Y_t, \mu_t = R_1 e^{r_1 t}, t = 1, 2, \cdots, 100。\tag{0.34}$$

2. 确定性趋势的拟合

3. 对剩余序列拟合 ARMA 模型。Y_t 的自相关与偏自相关如图 1.2 所示，Y_t 的自相关适应模型为 AR 模型，逐步增加 AR 模型阶数进行拟合，残差方差如图 1.5 所示，所以合适的模型为 AR（2），即

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t,\tag{0.35}$$

则参数估计为 $\hat{\varphi}_1 = 0.5451, \hat{\varphi}_2 = 0.2478$

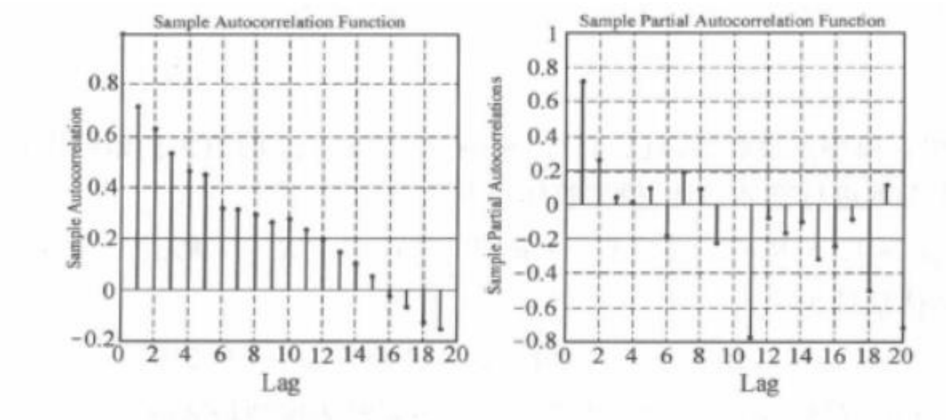


图 1.2 剩余序列的相关函数图

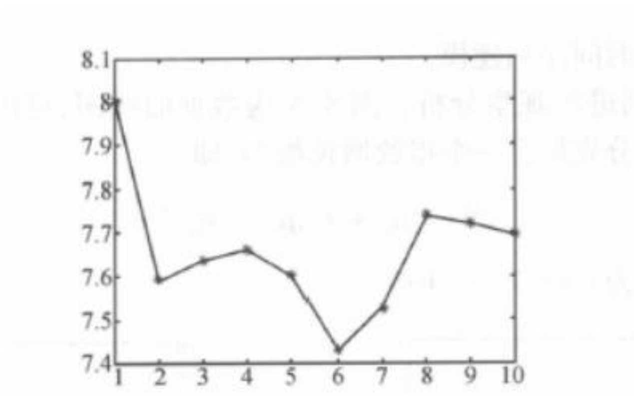


图 1.3 残差方差图

4. 建立组合模型。将已估计出的 $R_1, r_1, \varphi_1, \varphi_2$ 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，模型为

$$X_t = \mu_t + Y_t = R_1 e^{r_1 t} + \varphi_1\left(X_{t-1} - R_1 e^{r_1(t-1)}\right) + \varphi_2\left(X_{t-2} - R_1 e^{r_1(t-2)}\right) + a_t \quad (0.36)$$

5. 最终可得参数整体估计为

$$\hat{R}_1 = 12.1089, \hat{r}_1 = 0.017, \hat{\varphi}_1 = 0.517, \hat{\varphi}_2 = 0.2397$$

6. 残差平方和为 738.4402，残差自相关图 1.4 表明整体模型是适应的。

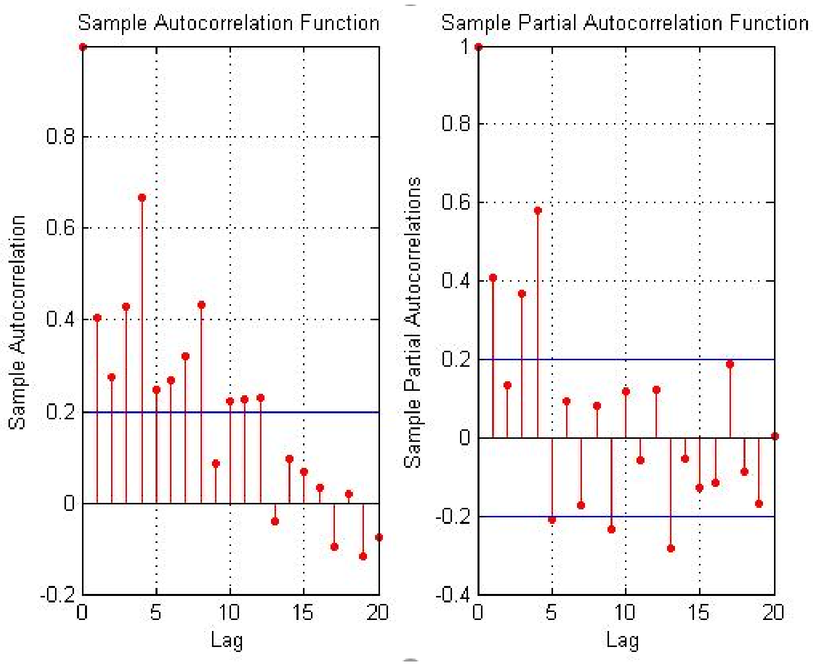


图 1.4 拟合整体模型后的残差自相关图

2) 对所给时间序列进行两年的

使用所给的模型从 1970 年第 4 季度开始预测，结果如下表 1.4 所示。

t	0	1	2	3	4	5	6	7	8
$t+l$	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	668384	68.562	70.0083	71.4879	72.9238	74.3507	75.768

1.5.4 模型的求解

使用 MATLAB 软件进行编程，编写代码如下所示：

```
clc,clear
a=load('data02 .txt');
a=a';a=a(:);
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)
figure,subplot(121),autocorr(cha)
subplot(121),parcorr(cha)
for i=1:10
   cs2{i}=ar(cha,i);%拟合模型
   cha2=resid(cs2{i},cha)
   myvar(i) = sum(cha2.^2)/(100 - i);%计算残差方差
end
fugure,plot(myvar,'* -')


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('data03.txt');
 a=a',a=a(:);
 x=[a(2:end-1),a(1:end-2),[3:100]'];
 cs=lsqcurefit(xt,cs0,x,a(3;end))
res=a(3:end)-xt(cs,x);
Q=sum(res.^2)
autocorr(res)
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)
```

1.5.5 结果分析

表 1.4 预测出之后八个季度的结果

t	0	1	2	3	4	5	6	7	8
$t+l$	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	668384	68.562	70.0083	71.4879	72.9238	74.3507	75.768

1.6 GDP 问题研究

1.6.1 问题重述

1952 年—1997 年我国人均国内生产总值（单位：元）数据如下所示。

表 1.5 1952 年—1997 年我国人均国内生产总值

年代	人均生产总值	年代	人均生产总值	年代	人均生产总值
1952	119	1968	222	1984	682
1953	142	1969	243	1985	853
1954	144	1970	275	1986	956
1955	150	1971	288	1987	1104
1956	165	1972	292	1988	1355
1957	168	1973	309	1989	1512
1958	200	1974	310	1990	1634
1959	216	1975	327	1991	1879
1960	218	1976	316	1992	2287

1961	185	1977	339	1993	2939
1962	173	1978	379	1994	3923
1963	181	1979	417	1995	4854
1964	208	1980	460	1996	5576
1965	240	1981	489	1997	6079
1966	254	1982	525		
1967	235	1983	580		

（1） 使用 ARIMA（2，1，1）（p,d,q）模拟拟合，求模型参数的估计值。

（2） 求数据的 10 步预报值。

1.6.2 符号规定与基本假设

1. 符号规定

1. $\{x_t\}$ 表示原始数据序列

2. $\{y_t\}$ 表示经过一阶差分变换后的序列

2. 基本假设

1. 假设本题数据经过完全统计
2. 假设国民生产总值数据非常合理

1.6.3 模型的分析与建立

经过 MATLAB 软件处理数据，可得

$$y_t = 1.253y_{t-1} - 0.352y_{y-2} + \varepsilon_t + 0.5022\varepsilon_{t-1}$$

未来十年的预测值分别为 6419.4474，6668.7704，6861.1915，7014.4250，7198.6091，7240.2050，7323.7357，7392.7357，7449.4293，7496.3755。

1.6.4 模型的求解

通过 MATLAB 编写的代码如下所示：

将表格中的数据保存到纯文本文件中 data6.txt 中。

```
clc,clear
a=textread('data6.txt');
xt=a(:,[2:2:end]);xt=nonzeros(xt);
yt=diff(xt);
m = armax(yt,[2,1])
yd = yt;
for i=1:10
    tt1=predict(m,[.;0]);
    tt2=tt1{:}(end);
    ythat(i)=tt2;
    yd=[yd;tt2];
end
ythat
xthat=xt(end)+cumsum(ythat)
```

1.6.5 结果分析

经过 MATLAB 软件处理数据，可得

$$y_t = 1.253 y_{t-1} - 0.352 y_{y-2} + \varepsilon_t + 0.5022 \varepsilon_{t-1}$$

未来十年的预测值分别为 6419.4474，6668.7704，6861.1915，7014.4250，7198.6091，7240.2050，7323.7357，7392.7357，7449.4293，7496.3755。

1.7 山猫数量问题研究

1.7.1 问题重述

某地区山猫的数量在前连续 114 年的统计数据如表 1.6 所示。分析该数据，得出山猫的生长规律，并预测出以后两个年度山猫的数量。

表 1.6 山猫数据

269	321	585	871	1475	2821	3928	5943	4950	2577	523	98

184	279	409	2285	2685	3409	1824	409	151	45	68	213
546	1033	2129	2536	957	361	377	225	360	731	1638	2725
2871	2119	684	299	236	245	552	1623	3311	6721	4254	687
255	473	358	784	1594	1676	2251	1426	756	299	201	229
469	736	2042	2811	4431	2511	389	73	39	49	59	188
377	1292	4031	3495	537	105	153	387	758	1307	3465	6991
6313	3794	1836	345	382	808	1388	2713	3800	309	2985	3790
674	71	80	108	229	399	1132	2432	3575	2935	1537	529
485	662	1000	1520	2657	3396						

1.7.2 符号规定与基本假设

1. 符号规定

1. $\{x_t\}$ 表示原始序列；
2. T　表示周期；
3. $\{y_t\}$ 表示消除季节趋势，得到序列。

2. 基本假设

1. 假设该地区山猫生长只受自然规律影响；
2. 假设山猫的数量统计为完全统计；

1.7.3 模型的分析与建立

1. 序列时序图。

   记原始序列为$\{x_t\}$，时序图表示该序列大致有 12 个周期变化，周期的长度大致为 9 年或者 10 年，那么该问题以周期$T=10$年来进行计算。

2. 差分平稳。

   对原序列作 10 步差分，消除季节趋势，得到序列$\{y_t\}$，其中$y_t = x_{t+10} - x_t$ 即可得到时序图。

3. 模型拟合。

   根据差分后的序列的自相关和偏自相关的性质，使用 ARMA 模型进行拟合，拟合的 ARMA（1，10）模型较为理想，且通过白噪声检验，则该模型不适合拟合这个序列。计量经济学

1.7.4 模型的求解

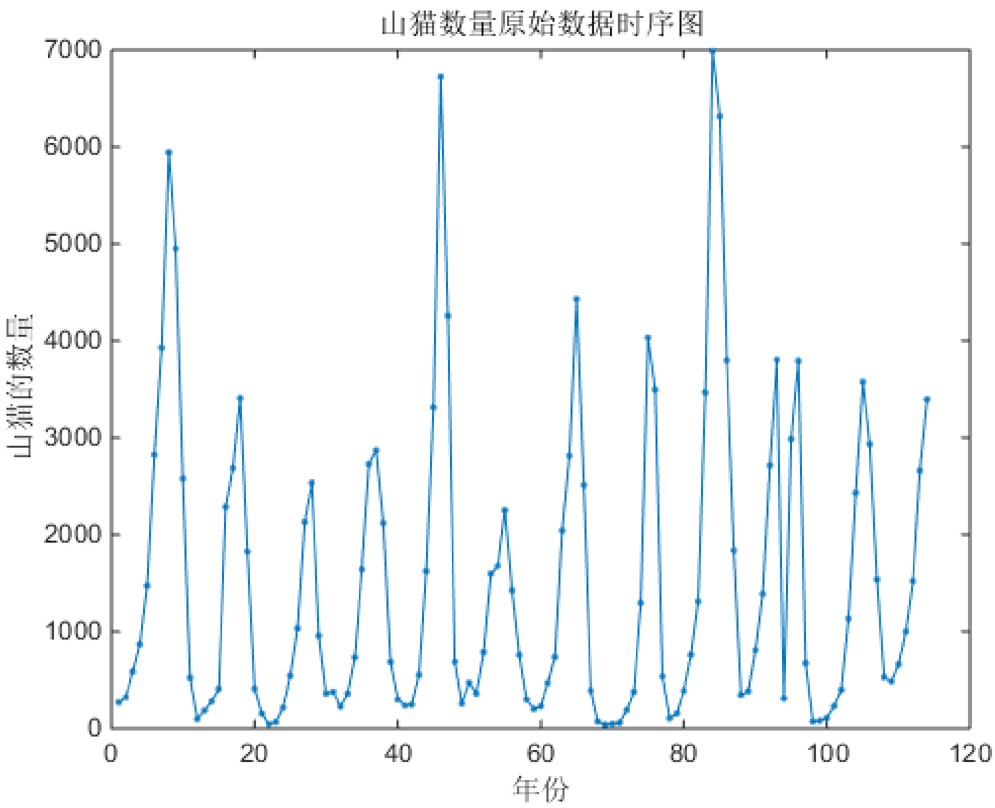
编写代码如下所示：

```
clc,clear
a=textread('data4.txt');
a=a';c=a(:);
c=nonzeros(c);n=length(c);
plot(c,'.-')
title('山猫数量原始数据时序图')
xlabel('年份')
ylabel('山猫的数量')
for i=11:n
    b(i-10)=c(i)-c(i-10);
end
b=b';
figure,plot(b,'.-')
title('消除季节趋势后的数据的时序图')
xlabel('年份')

figure,subplot(1,2,1),autocorr(b)
xlabel( '自相关函数图' )
```

```
subplot(122),parcorr(b)
xlabel('非自相关函数图')
cs=armax(b,[1,10])
figure,myres=resid(cs,b);
[h1,p1,st1]=lbqtest(myres,'lags',6)
[h2,p2,st2]=lbqtest(myres,'lags',12)
[h2,p2,st2]=lbqtest(myres,'lags',18)
bhat1=predict(cs,[b;0]);
bhat(1)=bhat1(end);
bhat2=predict(cs,[b;bhat(1);0]);
bhat(2)=bhat2(end);
ahat(1)=a(end-9)+bhat(1);
ahat(2)=a(end-8)+bhat(2)
```

1.7.5 结果分析



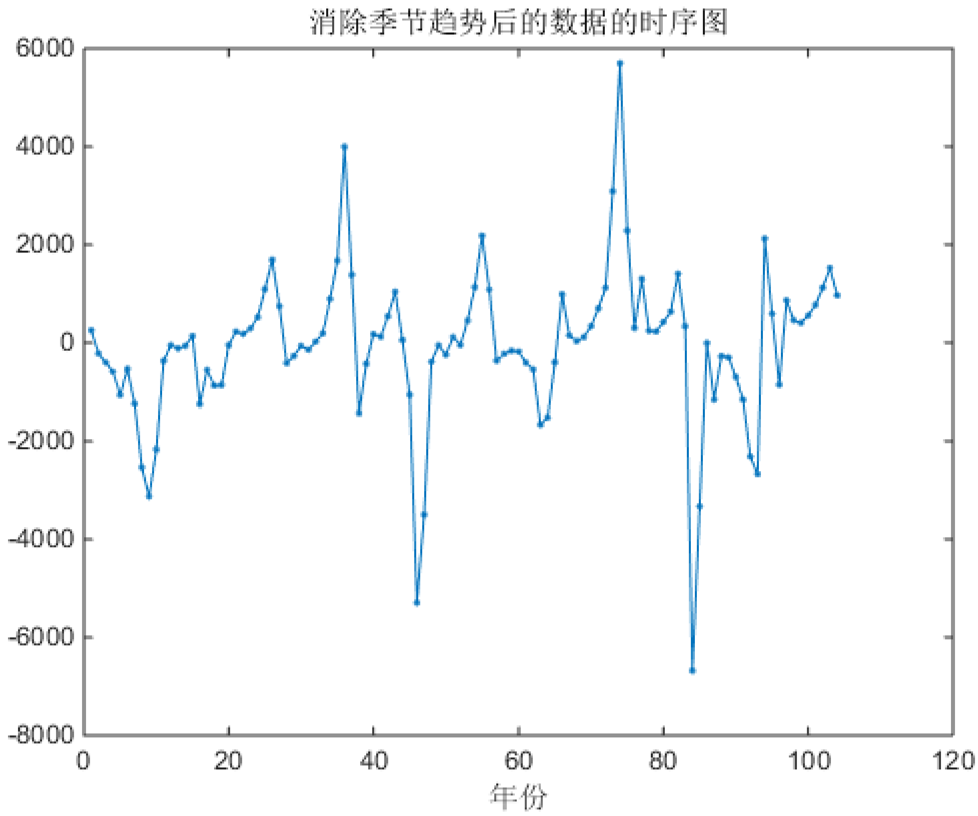


图 1.5 消除季节趋势后数据的时序图

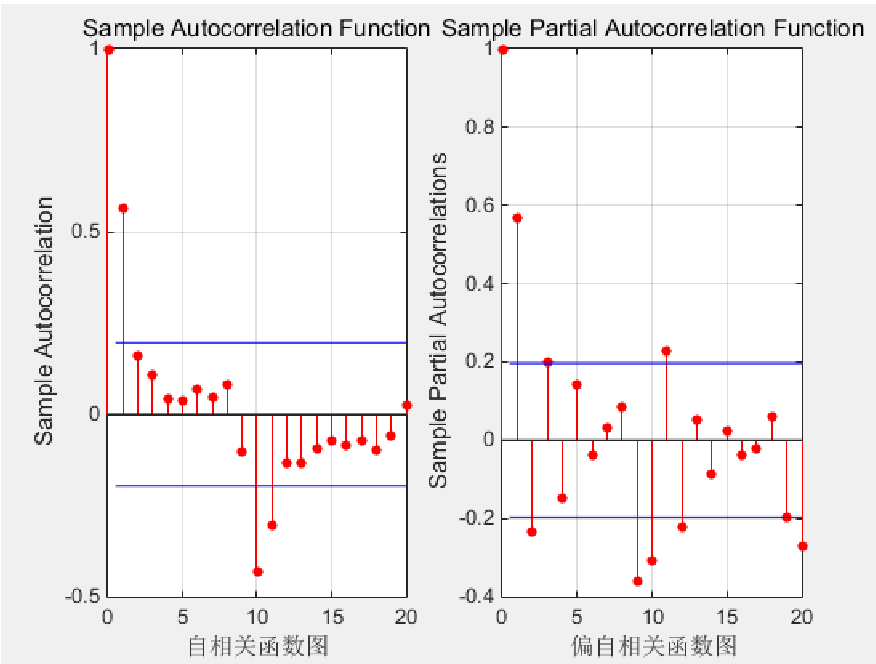


图 1.6 自相关函数图和偏自相关函数图

经过 MATLAB 软件编写程序算的结果为两个年度的预测值为 4296 和 3656。

1.8 小结

本次章节主要介绍时间序列模型及其相关问题的解决，时间序列模型也是一种回归模型，一方面承认事物发展的延续性，运用过去时间序列的数据进行统计分析来推测事物的发展趋势；另一方面充分考虑到偶然因素而统计分析，进行趋势预测，将该模型应用到日用品消耗、流浪猫数量、国民生产总值等方面，找到其连续几年的合理性数据（基于完全统计），运用该对应的时间序列模型进行预测，以达到预测信息的目的，有效的解决问题。