

时间序列分析——模型的建立和定阶

李逢君 2016060601010

初步平稳化数据

上上次作业使用最小二乘法拟合多项式去除了时间序列的趋势项, 使用滑动平均去除时间序列的周期项。得到以下数据:

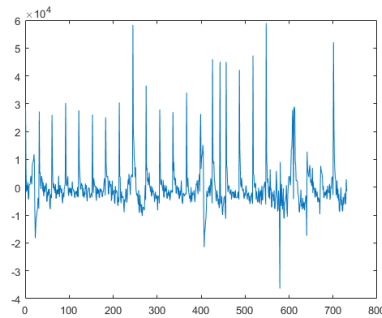


图 1 去除二次曲线趋势去周期项数据及自相关系数

数字特征估计

上次作业使用书上的算法对该数据的相关数字特征进行估计, 并与 matlab 自带的函数结果进行了对比, 结果如下:

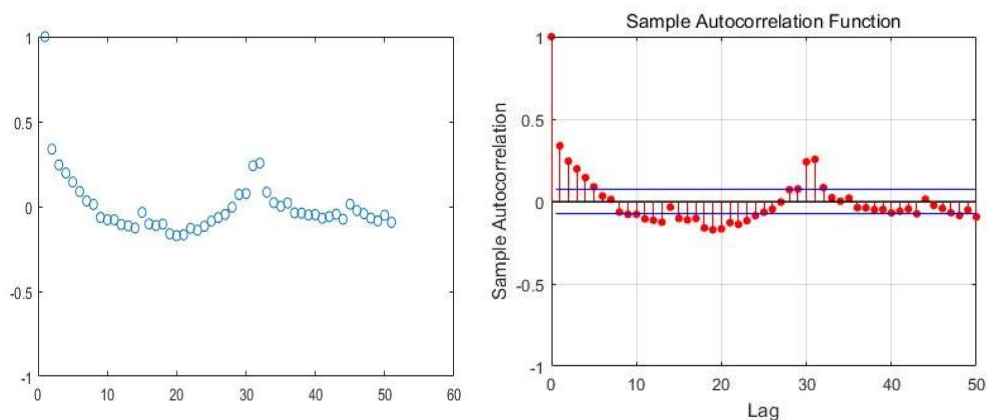


图 2 自相关函数图

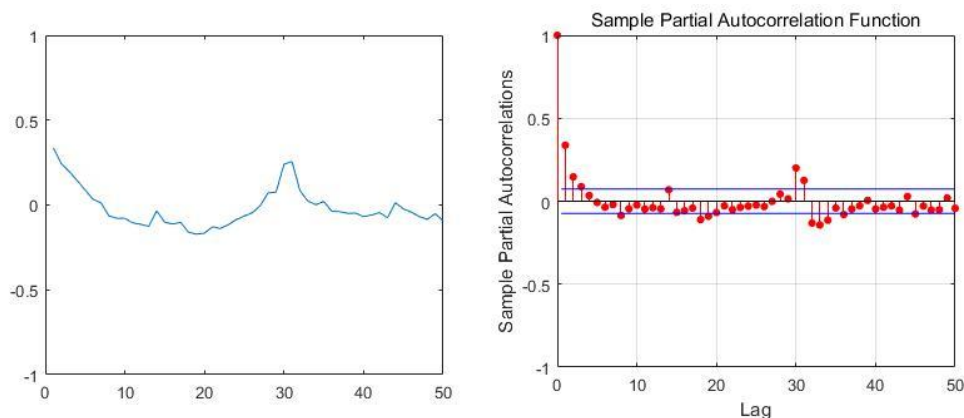


图 3 偏相关函数图

平稳性检验

在进行模型的建立和定阶之前，我们需要对该预处理后的动态数据的平稳性进行检验。

首先，通过观察自相关函数和偏相关函数可以发现两者都具备“拖尾性”，可以初步建立为 ARMA 模型。

另外使用非常数检验方法——逆序检验法：提取样本的前 720 项并将其分为 $K=20$ 组计算统计量

$$Z = \frac{A - \frac{1}{4}K(K-1)}{\sqrt{(2K^3 + 3K^2 - 5K/72)}}$$

其中

$$A = \sum_{i=1}^{K-1} \sum_{j=i+1}^K a_{ij} = \sum_{i=1}^{K-1} A_i$$

$$a_{ij} = \begin{cases} 1, & \text{当 } i < j \text{ 时 } \bar{X}_i < \bar{X}_j \\ 0, & \text{其他} \end{cases}$$

$$X_i = \frac{1}{M} \sum_{j=1}^M X_{i,j}, \quad i = 1, 2, \dots, K$$

得到 $Z = -0.2920$ ，在给定显著性水平 $\alpha = 0.05$ ， $|Z| = 0.2920 < 1.96$ ，认为序列无明显的趋势，为平稳序列。

代码实现如下：

主函数：

```
clear, clc, close all
data = load('data.mat');
data = data.y4;
data_length = length(data);
[A,B,u,r] = nixujianyan(data(1:720),20)
```

逆序检验函数：

```

function [A,B,u,r] = nixujianyan (ser,K)
%NIXUJIANYAN Summary of this function goes here
% Detailed explanation goes here
y=cell(K,1);
length=size(ser')/K;
for k=1:K
    y{k}=ser(1+(k-1)*length:k*length);
end
for k=1:K
    yeve(k)=mean(y{k});
end
for k=1:K
    yvar(k)=var(y{k});
end
a=zeros(K,K);
for j=1:K
    for i=1:j
        if yeve(i)<yeve(j);
            a(i,j)=1;
        end
    end
end
b=zeros(K,K);
for j=1:K
    for i=1:j
        if yvar(i)<yvar(j);
            b(i,j)=1;
        end
    end
end
[A,an]=size(find(a==1));
[B,bn]=size(find(b==1));
u=(A+0.5-K*(K-1)/4)/sqrt(K*(2*K*K+3*K-5)/72);
r=(B+0.5-K*(K-1)/4)/sqrt(K*(2*K*K+3*K-5)/72);
end

```

零均值检验

为应用零均值 ARMA 模型，需检验对给定(或处理过)的动态数据能否判定为零均值过程？

检验方法：将样本均值与其标准差进行比较，若样本均值落在区间内 $[-\sqrt{\text{Var}[\bar{X}]}, \sqrt{\text{Var}[\bar{X}]}$],

则认为序列是零均值过程。

计算原均值为27312，原标准差为10716，故需要对原序列进行零均值化

$$X' = X - \text{mean}(X)$$

新的序列均值为 10263，标准差为 10716 满足零均值检验，因此可以建立零均值的 ARMA 模型。

模型的定阶与参数估计

利用相关函数定阶法可以初步估计序列的阶数为 5~6，但是这种方法往往有偏差。

可以调用 matlab 的 arima 函数计算 Akaike 和 bayesian 信息准则

表 1 AIC 准则计算，行为 p，纵为 q

	1	2	3	4	5	6	7	8	9	10
1	15637.88	15417.67	15393.51	15385.11	15387.82	15735.04	15531.11	15698.01	15741.02	15382.59
2	15561.35	15401.9	15377.98	15347.9	15351.25	15724.04	15724.28	15334.73	15337.78	15729.28
3	15514.71	15355.88	15361.89	15345.83	15337.12	15722.17	15713.03	15319.19	15723.93	15727.64
4	15370.25	15719.29	15337.07	15340.05	15338.36	15719.65	15721.43	15719.91	15724.33	15725.15
5	15565.07	15719.83	15335.76	15336.98	15345.31	15721.87	15720.81	15723.37	15723.63	15724.06
6	15360.89	15330.05	15313.06	15316.61	15314.93	15721.58	15720.35	15721.98	15719.37	15719.17
7	15361.46	15722.15	15313.25	15337.06	15339.64	15723.51	15723.78	15722.18	15723.33	15728.45
8	15731.66	15329.61	15315.49	15339.48	15339.16	15724.15	15717.32	15721.3	15727.87	15724.31
9	15351.82	15328.96	15335.49	15339.22	15339.44	15725.58	15720.01	15735.6	15728.9	15725.75
10	15353.24	15726.7	15335.51	15341	15341.39	15710.17	15727.56	15728.68	15731.89	15729.25

表 2 BIC 准则计算，行为 p，纵为 q

	1	2	3	4	5	6	7	8	9	10
1	15656.26	15440.64	15421.08	15417.27	15424.58	15776.39	15577.05	15748.55	15796.15	15442.31
2	15584.32	15429.47	15410.14	15384.66	15392.6	15769.98	15774.82	15389.86	15397.5	15793.6
3	15542.28	15388.04	15398.64	15387.18	15383.07	15772.71	15768.16	15378.92	15788.26	15796.56
4	15402.41	15756.05	15378.42	15385.99	15388.9	15774.79	15781.16	15784.23	15793.25	15798.66
5	15601.82	15761.18	15381.7	15387.52	15400.44	15781.59	15785.13	15792.28	15797.14	15802.17
6	15402.24	15376	15363.6	15371.74	15374.65	15785.9	15789.26	15795.49	15797.48	15801.87
7	15407.4	15772.68	15368.38	15396.78	15403.96	15792.42	15797.29	15800.28	15806.03	15815.74
8	15782.2	15384.75	15375.22	15403.8	15408.08	15797.66	15795.43	15804	15815.16	15816.2
9	15406.96	15388.69	15399.81	15408.14	15412.95	15803.68	15802.71	15822.9	15820.79	15822.23
10	15412.97	15791.02	15404.43	15414.51	15419.49	15792.87	15814.86	15820.57	15828.37	15830.33

两者皆在 p=7，q=4 时取得最小，因此可以建立模型 ARMA(7,4)，各项参数估计如下：

Parameter	Value	Standard Error	t Statistic
Constant	-10	88.2295	-0.113341
AR{1}	-1.90485	0.295883	-6.43784
AR{2}	-1.5	0.486253	-3.08481
AR{3}	-0.824613	0.457087	-1.80406
AR{4}	-0.524969	0.296212	-1.77228
AR{5}	-0.314328	0.221108	-1.4216
AR{6}	-0.202157	0.151306	-1.33609
AR{7}	-0.0892595	0.0636626	-1.40207
MA{1}	0.369539	0.293198	1.26037
MA{2}	-0.905159	0.288638	-3.13597
MA{3}	-0.443595	0.210755	-2.10478
MA{4}	0.130132	0.21126	0.615983
Variance	1.01347e+08	0.00133512	7.59084e+10

R=7, M=4, AIC=15337.055881, BIC=15396.783256

图 4 各项参数估计

实现代码如下：

```
for i = 0:10
    for j = 0:10
        % 2012b 之前
        % spec = garchset('R',i,'M',j,'Display','off'); % 指定模型的结构
        % [coeffX,errorsX,LLFX]=garchfit(spec,x); % 拟合参数
        % num=garchcount(coeffX); % 计算拟合参数的个数
        Mdl = arima(i,0,j);
        [EstMdl,EstParamCov,LLFX] = estimate(Mdl, data');
        num = sum(any(EstParamCov));
        % 计算 Akaike 和 bayesian 信息准则
        [aic,bic]=aicbic(LLFX, num, data_length);
        fprintf('R=%d, M=%d, AIC=%f, BIC=%f\n',i,j,aic,bic); %显示计算结果
        aics(i+1,j+1)=aic;
        bics(i+1,j+1)=bic;
    end
end
```

极大似然估计

正态性检验

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2$$

$$G_3^* = \frac{1}{N} \sum_{t=1}^N \left(\frac{X_t - \bar{X}}{\hat{\sigma}} \right)^3$$

$$G_4^* = \frac{1}{N} \sum_{t=1}^N \left(\frac{X_t - \bar{X}}{\hat{\sigma}} \right)^4 - 3$$

检验方法：如果上述表达式中 G_3^* 、 G_4^* 近似为 0，则可认为序列具有正态性
峰度计算

```
sk = skewness(data);
```

偏度计算

```
ku = kurtosis(data);
```

得到峰度为 3.0240，偏度为 17.6314，因此认为该序列不具有正态性。

正态白噪声的检验

统计量

$$G_3 = \sqrt{\frac{N}{6}} G_3^*$$

$$G_4 = \sqrt{\frac{N}{24}} G_4^*$$

渐进服从标准正态分布

判断准则，若 G_3 和 G_4 的值有一个超过 2，就否定序列是正态白噪声的假设。
经过计算 $G_3=33.378$ ， $G_4=80.749$ ，因此否定序列是正态白噪声的假设。

结论：由于不通过正态检验，因此无法进行极大似然估计。