**时间序列预测问题**

时间序列预测，可以让我们在理解事物变化，随时间的关系，找到事物间的某种规律从而运用到生活实际当中去。本文介绍了由**ARIMA**模型，和**SARIMA**模型，对相关优拟合度的数据经行时间序列预测

**关键词：时间序列预测 p值 q值 平稳性 ARIMA SARIMA**

1 问题重述

1.1 问题一：列车运行数量预测时间序列预测问题

某车站连续5年60个月列车运行数量的时间序列数据，试建立合适的时间序列模型描述该时间序列数据的变化规律，并预测后一年12个月的列车运行数量.如果可以精确预测相关列车运行数量，可以在列车轨道布局，车站修建等方面有更好的保障道路交通网络的正常运行。

1.2 问题二：国际航空公司旅客数量时间序列预测问题

建立合适的时间序列模型描述该时间序列数据的变化规律，并预测后一年12个月的旅客数量，可以更好的机场规划，预算有利于未来发展，航空公司的计划。

2 问题分析

2.1 问题一分析

本文通过先对数据进行预处理，如平稳性检验。随后使用ARIMA等模型对每天的列车运行数量进行建模和预测。最后对题目要求进行预测展示结果。

2.2 问题二分析

使用SARIMA模型对时间序列进行预测，预测出一年12个月的旅客数量，最后对题目要求进行预测展示结果。

3 模型假设

3.1 问题一模型假设：

1.没有其他相关因素如 政策，疫情，科技等额外影响，只通过时间预测

3.2 问题二模型假设：

1.旅客暂时没有上限饱和

2.航空公司没有最大运算限制

4 符号说明

5 问题一模型建立与求解

5.1 模型准备——时间序列预测法

5.1.1 移动平均时间序列预测

步骤如下：

设观测序列为时间序列，取移动平均的项数。一次移动平均的平均值计算公

二次移动平均值计算公式

当预测目标的基本趋势是在某一水平上下波动时，我们可以用一次移动平均方法建立模型进行预测分析，即

，

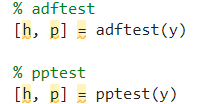
其预测标准误差为

5.2 matlab时间序列预测

5.2.1 平稳性分析

首先对目标数据进行平稳性分析，使用adftest函数和pptest函数进行平稳性检验，

原数据平稳性分析（图1）



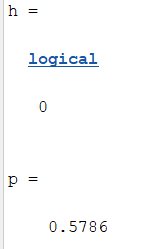
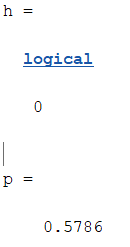


图1

可知平稳性不好，因此进行一阶差分得出（图2）



图 2

此时的平稳性较好，可以进行下一步操作。

5.2.2 定阶

在此处使用AIC进行定阶通过循环查找拟合度最高的p，q值AIC表格如下所示（表1）

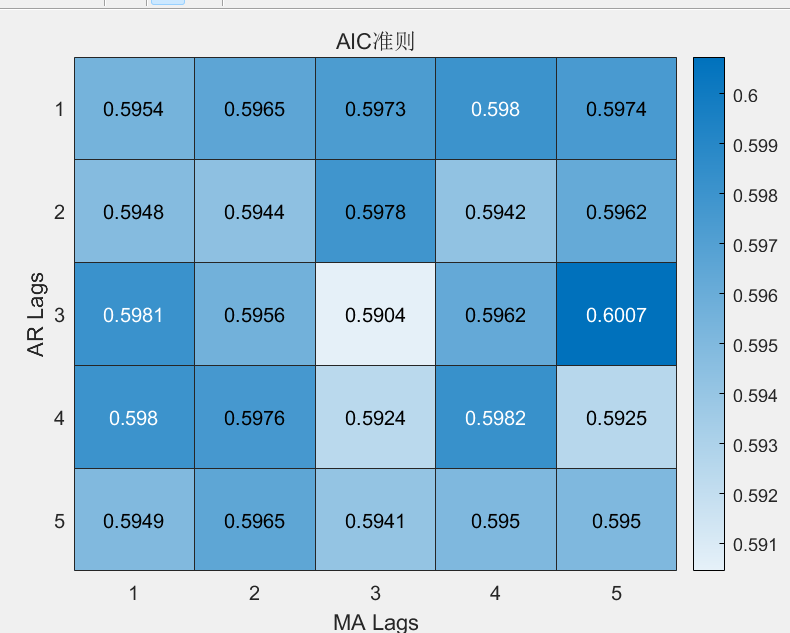


表 1

由图可得，最小的值在（3，3）处取到，即p = 3， q = 3。

5.2.2 创建模型

创建arima模型，通过命令行输出结果可得，数据基本上平稳。（图3）

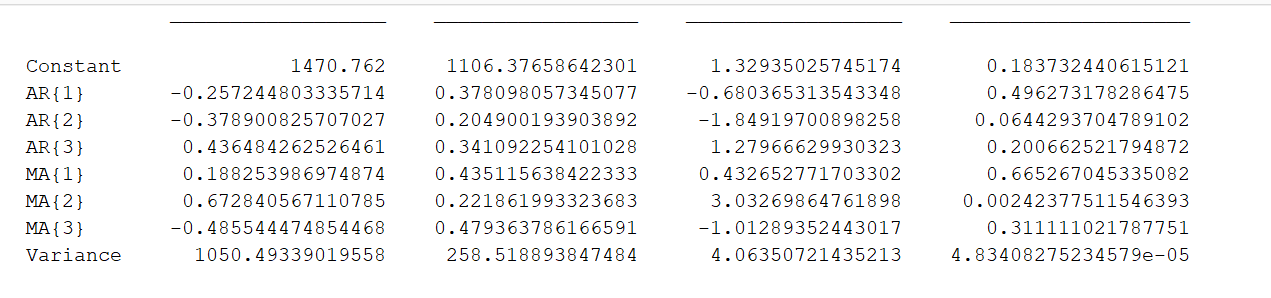


图 3

5.3 预测结果 （图4）

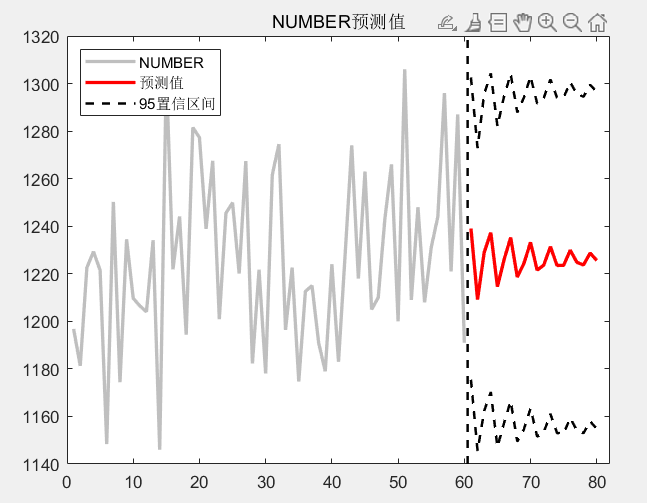


图 4

6 问题二模型建立与求解

6.1 模型准备——时间序列预测法

6.2 matlab时间序列预测

6.2.1 读入数据 （图5）

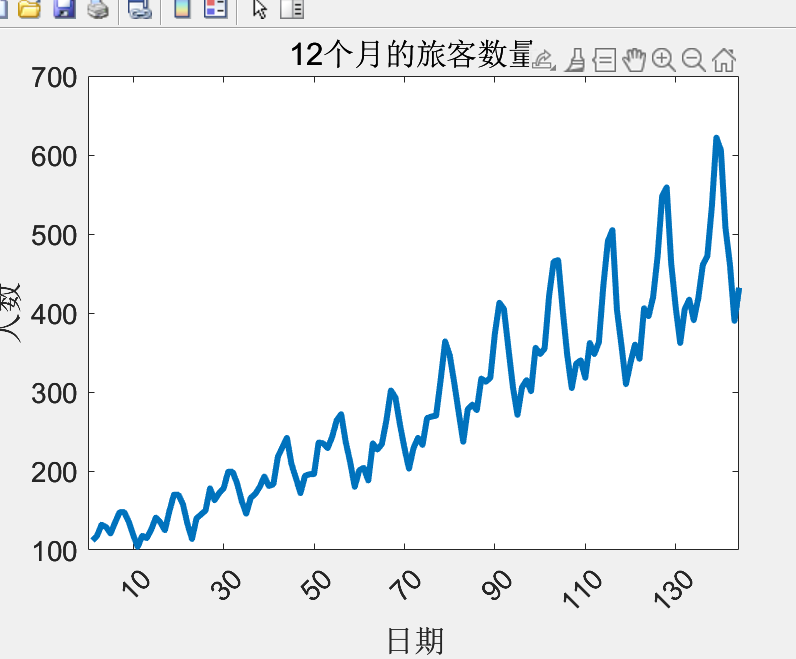


图 5

可以看出数据具有季节性的特点，故选择采用SARIMA进行建模。

6.2.2 差分与平稳性分析

差分并判断平稳性。一阶差分（图6）和二阶差分（图7）的图分别如下所示

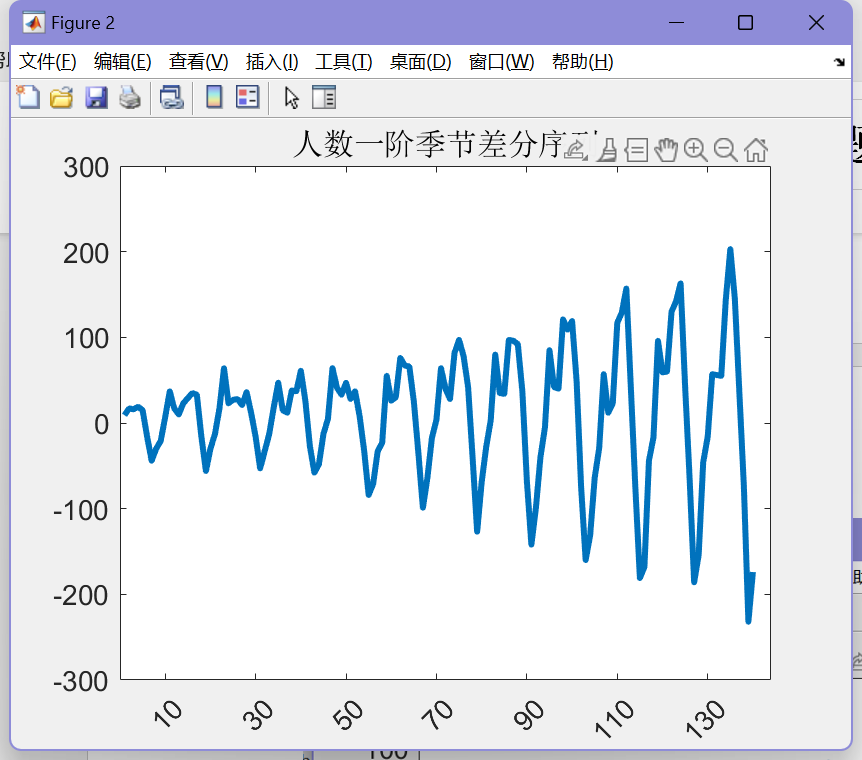


图 6

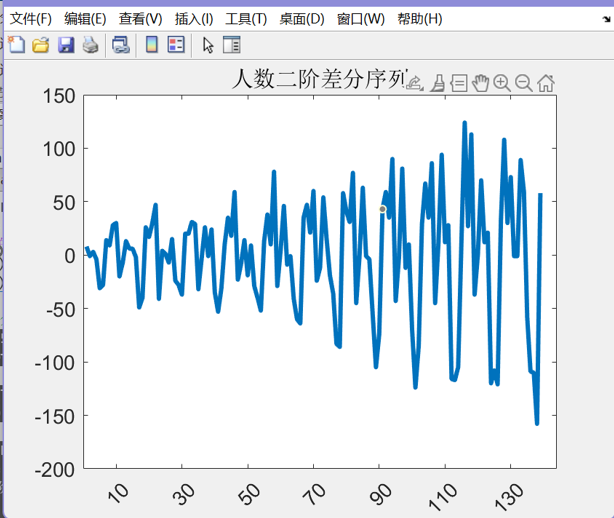


图 7



故此时的数据足够平稳，可以选择使用二阶差分进行后面的操作。

6.2.3 定阶

对二阶差分后的结果进行平稳性检验，得到的数据如下图

使用AIC进行定阶操作 得（图8）

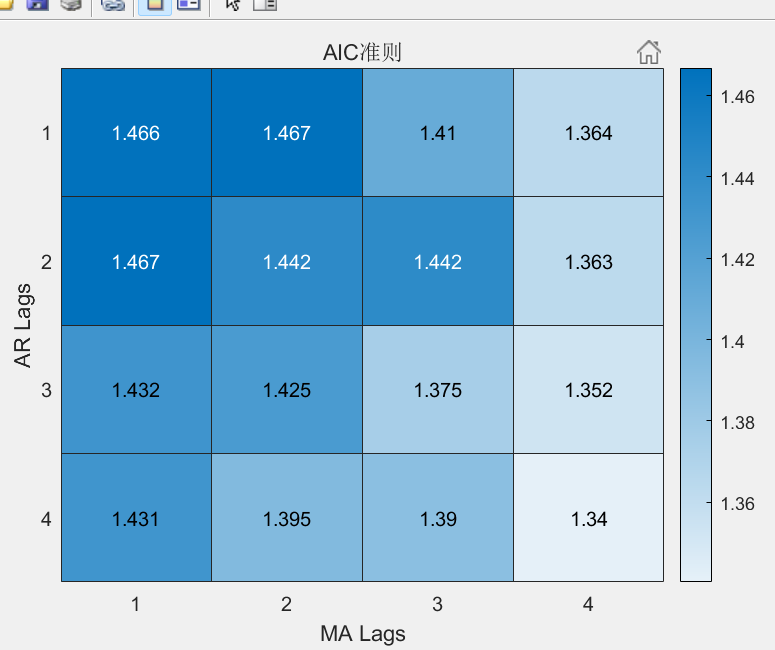
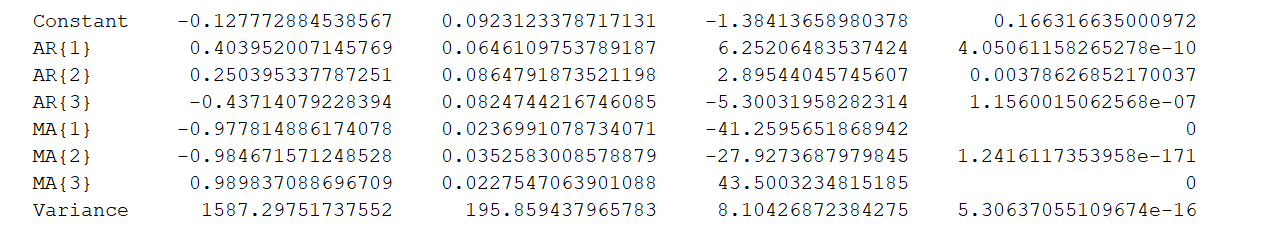


图 8

按照图示，应当选择（4，4）作为p，q的值，可是由于p，q的值过大，会导致arima使用错误，故选择较好的（3，3）作为p，q的数据，得到以下输出：



6.2.4 可视化分析

通过infer函数得到数据的残差res，得到图像如下所示 （图9）

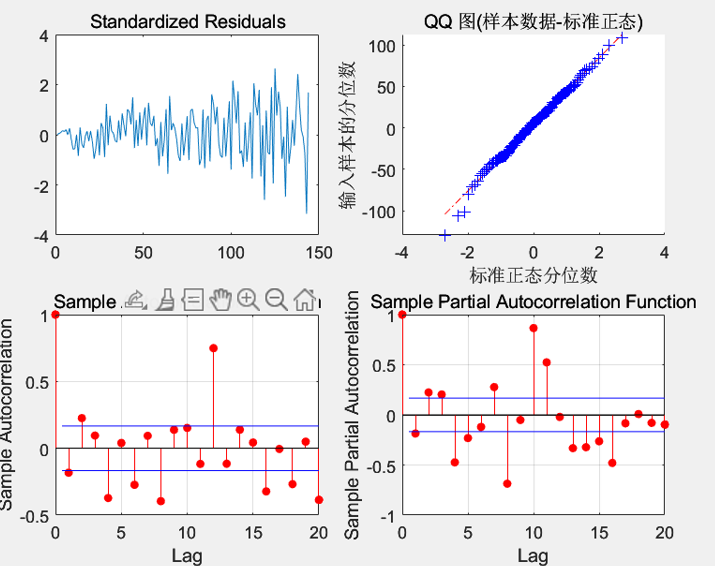
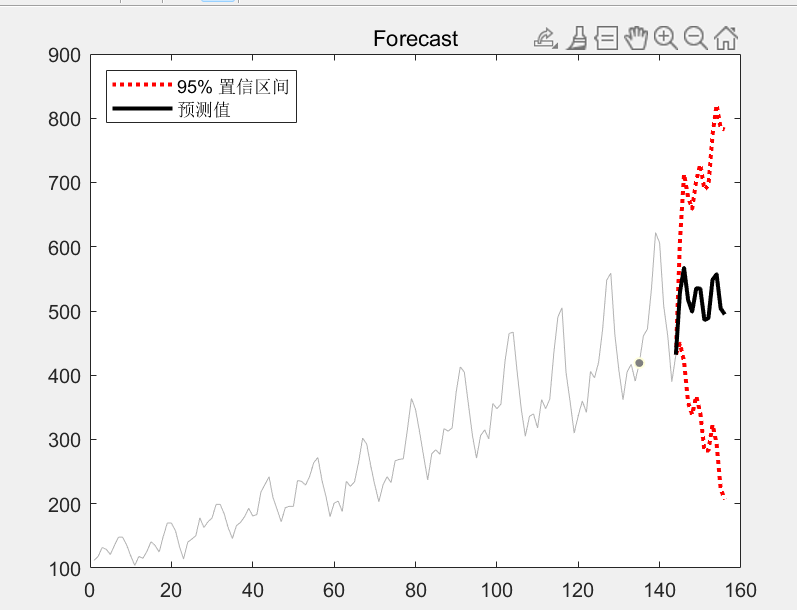


图 9

6.3 预测



附录：

clear

%% ARIMA(p,d,q)消费指数时间序列数据

% Australian Consumer Price Index (CPI) measured from 1972 and 1991,

load Data\_JAustralian

[a,b,c]=xlsread("exercise1data.xlsx")

y = a(:,2);

T = length(y);

% 可视化

figure

plot(y,'linewidth',3)

title('liecheyunxingshulang')

axis tight

%% 平稳性的单位根检验

% adftest

[h, p] = adftest(y)

% pptest

[h, p] = pptest(y)

%% 差分

diff\_y = diff(y); % 一阶差分运算

% figure(4), plot( diff\_y, 'o-', 'linewidth', 1, 'markersize', 6)

figure(2),

plot( diff\_y, 'linewidth', 2)

h1 = gca;

h1.FontSize = 14;

xlabel('t')

title('CPI的一阶差分序列')

%% 平稳性的单位根检验

[h, p] = adftest(diff\_y) % 平稳性adf检验

%% autocorr

figure(3), autocorr(diff\_y); %自相关系数图

%% parcorr

figure(4), parcorr(diff\_y); %偏自相关系数图

maxLags = 4;

%% 建立选定的较优模型ARIMA(2,1,0)

mdl = arima(3,0,3);

EstMdl = estimate(mdl, y);

%% 残差检验

% 可视化分析

res = infer(EstMdl,y);

figure(5)

subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))

title('Standardized Residuals')

subplot(2,2,2)

qqplot(res)

subplot(2,2,3)

autocorr(res)

subplot(2,2,4)

parcorr(res)

% Ljung-Box Q (lbq) 检验

[h, p] = lbqtest(res)

%% 预测

[yF,yMSE] = forecast(EstMdl,20,'Y0',y);

UB = yF + 1.96\*sqrt(yMSE); % 95%置信区间上限

LB = yF - 1.96\*sqrt(yMSE); % 95%置信区间下限

figure(6)

h4 = plot(y,'Color',[.75,.75,.75],'LineWidth',2);

hold on

h5 = plot(T+1:T+20,yF,'r','LineWidth',2);

h6 = plot(T+1:T+20,UB,'k--','LineWidth',1.5);

plot(T+1:T+20,LB,'k--','LineWidth',1.5);

h7 = gca;

plot((T+0.5)\*[1;1], h7.YLim, 'k--','LineWidth',1.5), % 样本点与预测点的时间分割线

h7.XLim = [0,T+22];

legend([h4,h5,h6],'NUMBER','预测值',...

'95置信区间','Location','Northwest')

title('CPI预测值')

hold off

clear

%% SARIMA(p,d,q)中国季度GDP数据建模1995---2017

[a,b,c]=xlsread("exercise2data.xlsx")

y = a(:,2);

T = length(y);

% 可视化

figure

plot(y,'linewidth',3)

title('liecheyunxingshulang')

axis tight

T = length(y);

figure(1)

plot(y,'linewidth',3)

h1 = gca;

h1.XLim = [0,T];

h1.XTick = 13:20:T;

h1.XTickLabel = 1995:5:2017;

h1.XTickLabelRotation = 45;

h1.FontSize = 14;

title('1992-2017年中国GDP')

xlabel('日期')

ylabel('GDP')

%% 差分运算

% 季节性趋势差分

diffS1\_y = y(5:end)-y(1:end-4);

figure(2),plot(diffS1\_y,'linewidth',3)

h1 = gca;

h1.XLim = [0,T];

h1.XTick = 13:20:T;

h1.XTickLabel = 1995:5:2017;

h1.XTickLabelRotation = 45;

h1.FontSize = 14;

title('GDP一阶季节差分序列')

% 增长趋势差分

diffS1D1\_y = diff(diffS1\_y);

figure(3),plot(diffS1D1\_y,'linewidth',3)

h1 = gca;

h1.XLim = [0,T];

h1.XTick = 13:20:T;

h1.XTickLabel = 1995:5:2017;

h1.XTickLabelRotation = 45;

h1.FontSize = 14;

title('GDP二阶差分序列')

%% 平稳性的单位根检验

% adftest

[h, p] = adftest(diffS1D1\_y)

% pptest

[h, p] = pptest(diffS1D1\_y)

figure(4), autocorr(diffS1D1\_y); %自相关系数图

figure(5), parcorr(diffS1D1\_y); %偏自相关系数图

%%

%% AIC BIC定阶

maxLags = 3;

AICSet = zeros(maxLags, maxLags);

BICSet = zeros(maxLags, maxLags);

for i = 1:maxLags

for j = 1:maxLags

mdl = arima('ARLags', [1:i], 'MALags', [1:j]);

[EstMdl, EstParamCov, LogL, info] = estimate(mdl, diffS1D1\_y);

[AICSet(i, j), BICSet(i, j)]= aicbic(LogL, length(info.X), length(diffS1D1\_y));

end

end

figure(6), heatmap(AICSet/1000)

xlabel('MA Lags')

ylabel('AR Lags')

title('AIC准则')

figure(7), heatmap(BICSet/1000)

xlabel('MA Lags')

ylabel('AR Lags')

title('BIC准则')

%% 建立最优SARIMA(3,1,3)模型

%% (5) 拟合SARIMA(3,1,3)模型

mdl = arima('Seasonality', 4, ...

'D', 1, ...

'ARLags', 1:3, ...

'MALags', 1:3)

[EstMdl, EstParamCov, LogL, info] = estimate(mdl, y);

[AICSets]= aicbic(LogL, length(info.X))

% 存在许多不显著项，首先考虑剔除p值最大的项AR{3}，然后再重新观察p值情况。

mdl = arima('Seasonality', 4, ...

'D', 1, ...

'ARLags', 1:5, ...

'MALags', 1:5)

[EstMdl, EstParamCov, LogL, info] = estimate(mdl, y);

[AICSets]= aicbic(LogL, length(info.X))

%%%%%%%% 其他系数都变得显著

mdl = arima('Constant', 0, ...

'Seasonality', 4, ...

'D', 1, ...

'ARLags', 1:2, ...

'MALags', 1:3)

[EstMdl, EstParamCov, LogL, info] = estimate(mdl, y);

[AICSets]= aicbic(LogL, length(info.X))

%% 残差检验

% 可视化分析

res = infer(EstMdl,y);

% res(60:90)=[]; [h, p] = lbqtest(res)

figure

subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))

title('Standardized Residuals')

subplot(2,2,2)

qqplot(res)

subplot(2,2,3)

autocorr(res)

subplot(2,2,4)

parcorr(res)

% Ljung-Box Q (lbq) 检验

[h, p] = lbqtest(res)

%% 预测

[yF,yMSE] = forecast(EstMdl,20,'Y0',y);

UB = yF + 1.96\*sqrt(yMSE); % 95%置信区间上限

LB = yF - 1.96\*sqrt(yMSE); % 95%置信区间下限

figure

h4 = plot(y,'Color',[.75,.75,.75],'LineWidth',2);

hold on

h5 = plot(T+1:T+20,yF,'r','LineWidth',1.5);

h6 = plot(T+1:T+20,UB,'k:','LineWidth',1.5);

plot(T+1:T+20,LB,'k:','LineWidth',1.5);

h7 = gca;

plot((T+0.5)\*[1;1], h7.YLim, 'k--','LineWidth',1.5), % 样本点与预测点的时间分割线

h7.XLim = [0,T+22];

h7.XTick = 13:20:T+20;

h7.XTickLabel = 1995:5:2022;

h7.XTickLabelRotation = 45;

h7.FontSize = 14;

title('中国季度GDP预测')

xlabel('日期')

ylabel('GDP')

legend([h4,h5,h6],'CPI','预测值',...

'95置信区间','Location','Northwest')

hold off