

书面作业

《金融数据分析导论》

1. P97——8

8. 考虑美国从 1947 年第 1 季度到 2011 年第 3 季度的季度实际 GNP，该数据存放于文件 q-GNPC96.txt 中，数据已做了季节调整，以 2005 年 GNP 为基础进行了通胀调整，以 10 亿美元为单位 (billions of chained 2005 dollars)。假设 x_t 代表 GNP 增长率的时间序列数据。
- (a) 通过 ar 命令，应用 AIC 准则，可以为 x_t 识别一个 AR(4) 模型。拟合这个模型，拟合的模型充分吗？为什么？
- (b) 数据 x_t 的样本 PACF 识别的是 AR(3) 时间序列模型。拟合这个模型，拟合的模型充分吗？为什么？
- (c) 如果用样本内法进行比较，哪个模型好？为什么？
- (d) 以 2004 年第 4 季度为预测原点，把数据分为估计样本和预测样本，应用回测检验，并以 MSFE 为标准，为 x_t 选择一个模型，并给出选择该模型的理由。

126

(a)

```
> gnp<-read.table("E:/DATA/data mining/fts03/data/q-GNPC96.txt",header=T)
> head(gnp)
  year mon day   gnp
1 1947   1   1 1780.4
2 1947   4   1 1778.1
3 1947   7   1 1776.6
4 1947  10   1 1804.0
5 1948   1   1 1833.4
6 1948   4   1 1867.6
> xt<-log(diff(gnp[,4])/gnp[-1,4]+1)#xt为GNP的对数增长率
```

```
> xt<-log(diff(gnp[,4])/gnp[-1,4]+1)#xt为GNP的对数增长率
> m1<-ar(xt,method='mle')
> m1$order
[1] 4
```

使用 ar() 命令为 x_t 识别了一个 AR (4) 模型

用 AR (4) 模型对 x_t 进行拟合

```
> t.test(xt)

One sample t-test

data:  xt
t = 12.4914, df = 257, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.00649595 0.00892741
sample estimates:
mean of x
0.00771168
```

```
> m1<-arima(xt,order=c(4,0,0))
> m1

Call:
arima(x = xt, order = c(4, 0, 0))

Coefficients:
      ar1      ar2      ar3      ar4  intercept
 0.3382  0.1484 -0.0983 -0.0892   0.0077
s.e.  0.0619  0.0652  0.0651  0.0619   0.0008

sigma^2 estimated as 8.158e-05:  log likelihood = 848.19,  aic = -1684.38
```

拟合得到 ar3 和 ar4 系数显著性不是很高

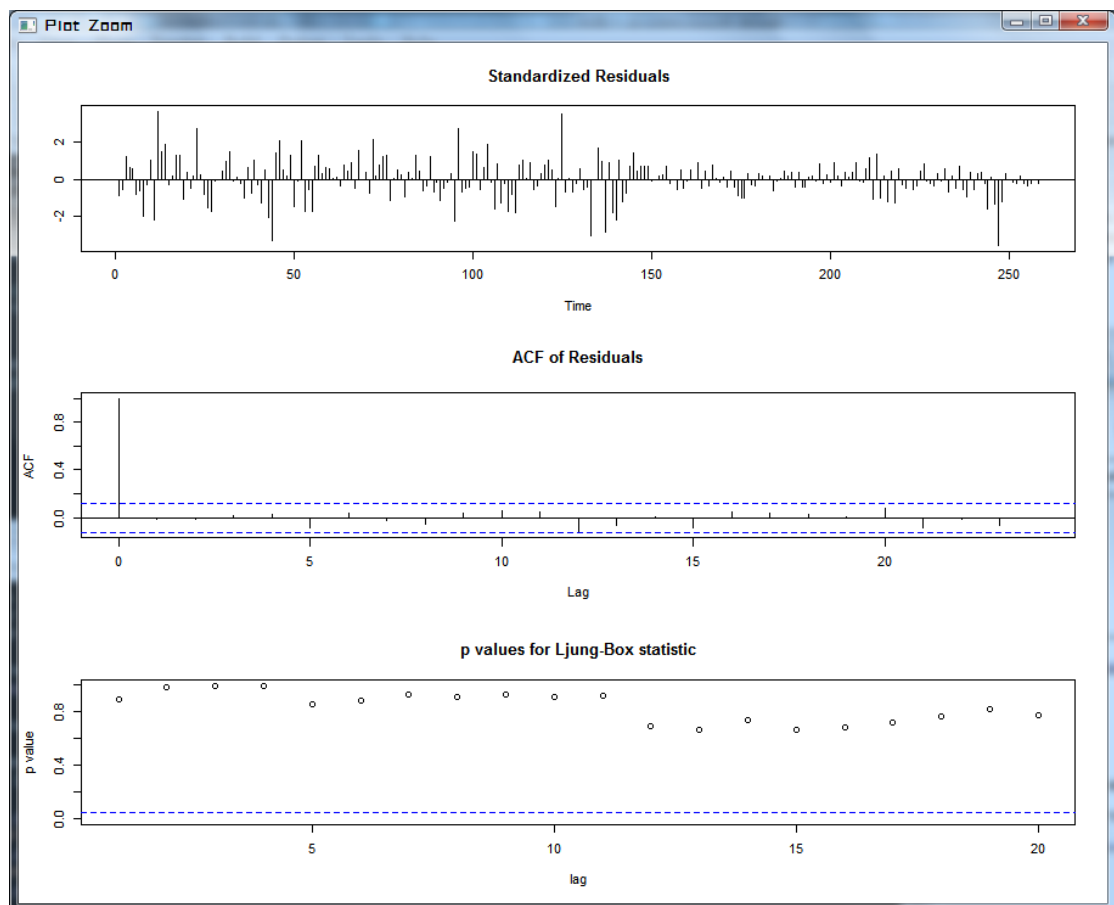
对模型进行残差检验

```
> Box.test(m1$residuals,lag=20,type='Ljung')

Box-Ljung test

data:  m1$residuals
X-squared = 15.0612, df = 20, p-value = 0.7729
```

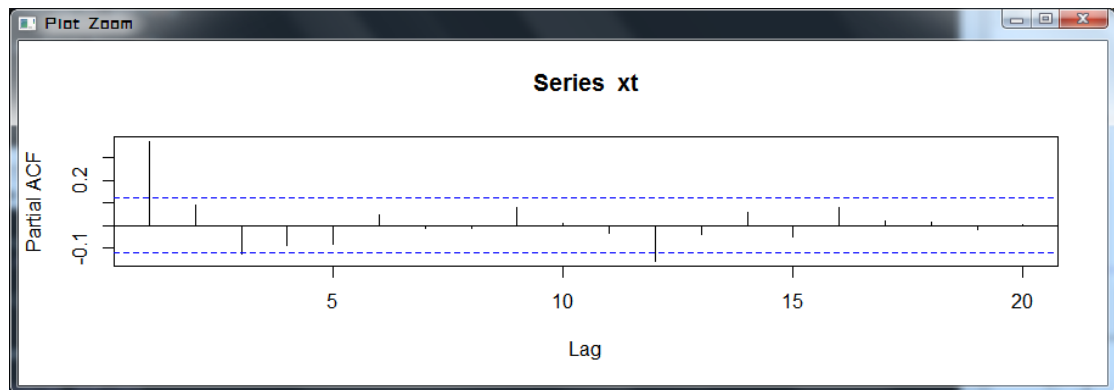
```
> tsdiag(m1,gof=20)
```



残差拟合为白噪声，拟合模型是充分的

(b)

```
> pacf(xt, lag=20)
```



数据 x_t 的样本 PACF 识别的是 AR (3) 时间序列模型

```
> m2<-arima(xt,order=c(3,0,0))
> m2

Call:
arima(x = xt, order = c(3, 0, 0))

Coefficients:
    ar1     ar2     ar3  intercept 
 0.3496  0.1358 -0.1294   0.0077 
s.e.    0.0616  0.0648  0.0617   0.0009 

sigma^2 estimated as 8.225e-05:  log likelihood = 847.15,  aic = -1684.31
```

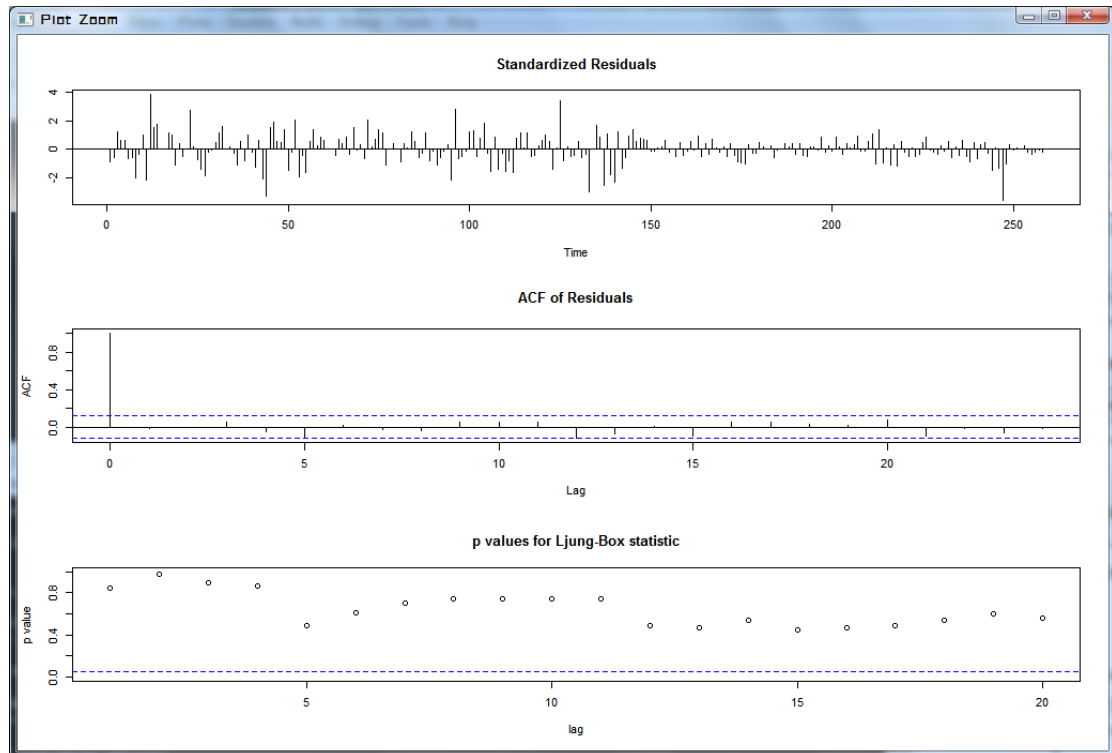
用 AR(3)模型对 x_t 进行拟合，拟合系数显著性很高
对模型进行检验

```
> Box.test(m2$residuals,lag=20,type='Ljung')

Box-Ljung test

data:  m2$residuals
X-squared = 18.4849, df = 20, p-value = 0.5555

> tsdiag(m2,gof=20)
```



残差拟合为白噪声，拟合模型是充分的

(c)

拟合得到的 AR (4) 模型的 AIC=-1684.38，拟合得到的 AR (3) 模型的 AIC=-1684.31，那么从 AIC 值来看，两个模型并没有很大的区别，AR (4) 模型稍微好一些。但 AR (4) 模型中有两个拟合系数的显著性并不是很好，从综合来看 AR (3) 模型要比 AR (4) 模型好。

(d)

```
> setwd("E:/DATA/data mining/fts07/fts07-2")
> source("backtest.txt")
> mm1<-backtest(m1,xt,232,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.008630469
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.005356595
> mm2<-backtest(m2,xt,232,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.008561698
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.005252466
```

拟合得到的 AR(4)模型的 RMSFE 值为 0.008630469，拟合得到的 AR(3)模型的 RMSFE 值为 0.008561698，说明 AR (3) 模型要优于 AR (4)

2. P135-136——1、2

1. 考虑从 1976 年 1 月到 2011 年 9 月加利福尼亚州和美国的月失业率，数据在文件 m-CAUS-7611.txt 中

给出(变量为: year、mon、CA 和 US)。

(a) 为加利福尼亚州月失业率建立一个纯时间序列模型，进行模型检验，并写出拟合的模型。

(b) 对于加利福尼亚州月失业率，建立一个时间序列模型，使用美国月失业率的 1 阶滞后作为解释变量，进行模型检验，并写出拟合的模型。

(c) 预测期从 2008 年 1 月到 2011 年 9 月，使用样本外预测，并比较这两个模型。

(a)

```
> unrate<-read.table("E:/DATA/data mining/fts03/data/m-CAUS-7611.txt",header=T)
> head(unrate)
  year mon  CA  US
1 1976   1  9.3  7.9
2 1976   2  9.1  7.7
3 1976   3  9.0  7.6
4 1976   4  8.9  7.7
5 1976   5  8.9  7.4
6 1976   6  8.9  7.6
> ca<-unrate$CA
```

```
> adfTest(ca,lags=24)

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 24
STATISTIC:
Dickey-Fuller: -0.157
P VALUE:
0.5663

Description:
Tue May 26 08:48:16 2015 by user: Administrator

> adfTest(ca,lags=0)

Title:
Augmented Dickey-Fuller Test

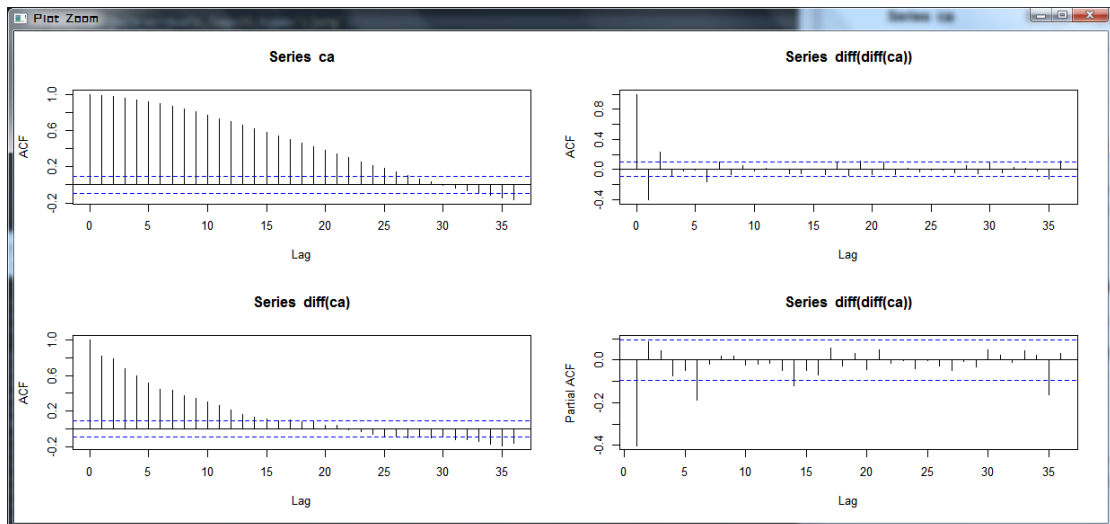
Test Results:
PARAMETER:
Lag Order: 0
STATISTIC:
Dickey-Fuller: 1.0937
P VALUE:
0.9259

Description:
Tue May 26 08:48:16 2015 by user: Administrator
```

不能拒绝零假设，非平稳的时间序列，需要对数据进行差分

从 acf 和 pacf 图上看，需要对数据进行二阶差分

```
> par(mfcol=c(2,2))
> acf(ca,lag=36)
> acf(diff(ca),lag=36)
> acf(diff(diff(ca)),lag=36)
> pacf(diff(diff(ca)),lag=36)
```



```
> ar(diff(diff(ca)))$order
[1] 6
```

```
> adfTest(diff(diff(ca)),lags=6)

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 6
STATISTIC:
Dickey-Fuller: -10.0802
P VALUE:
0.01

Description:
Tue May 26 08:54:16 2015 by user: Administrator

warning message:
In adfTest(diff(diff(ca)), lags = 6) : p-value smaller than printed p-value
```

```
> eacf(diff(diff(ca)))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o x x o o o o o o o
1 x x x o o x o o o o o o o
2 x x x o o x o o o o o o o
3 x x o o o x o o o o o o o
4 x x x o x x o o o o o o o
5 x x x x x x x o o o o o o
6 o x o o o x o o o o o o o
7 x x x o o o o o o o o o o
```

根据 acf 和 pacf 图定阶

最终选取 ARIMA(2, 2, 0)模型进行拟合

```
> m1=arima(ca,order=c(2,2,0))
> m1

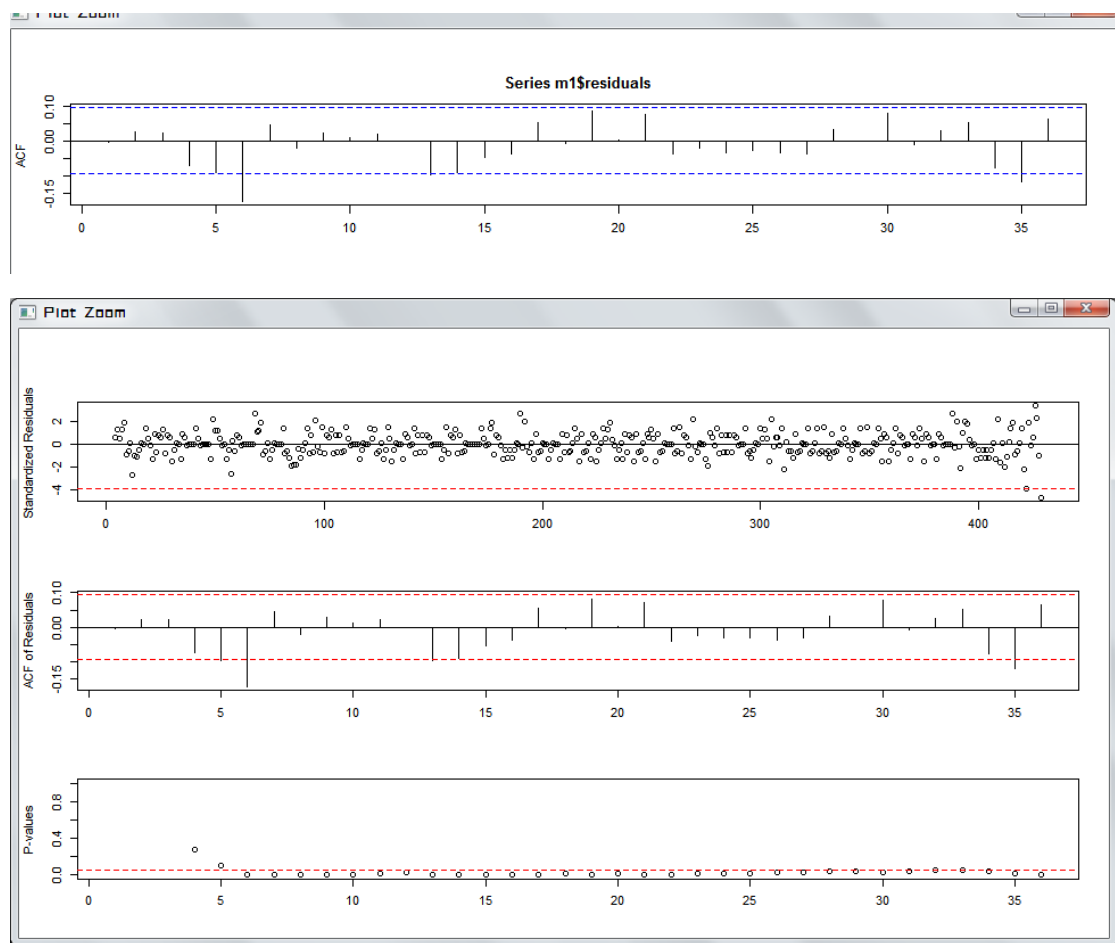
Call:
arima(x = ca, order = c(2, 2, 0))

Coefficients:
      ar1      ar2
    -0.3820    0.0773
s.e.    0.0495    0.0496

sigma^2 estimated as 0.005509: log likelihood = 504.52, aic = -1005.03
> acf(m1$residuals,lag=36)
> Box.test(m1$residuals,lag=36,type='Ljung')

Box-Ljung test

data: m1$residuals
X-squared = 56.7091, df = 36, p-value = 0.01534
> tsdiag(m1,gof=36)
```



表明模型拟合是充分的

(b)

对加利福尼亚州和美国的月收益率分别进行二次差分

```
> us<-unrate$US  
> ddca<-diff(diff(ca))  
> ddus<-diff(diff(us))
```

```

> m2<-lm(ddca[2:429]~ddus[1:428])
> summary(m2)

Call:
lm(formula = ddca[2:429] ~ ddus[1:428])

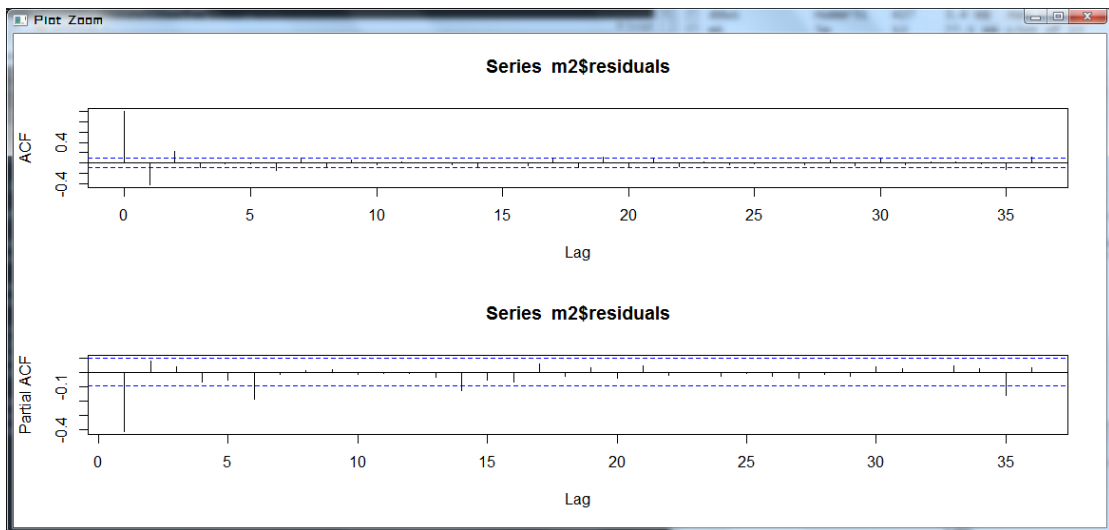
Residuals:
    Min       1Q   Median       3Q      Max
-0.302325 -0.086892  0.000247  0.063672  0.213108

Coefficients:
            Estimate Std. Error t value
(Intercept) -0.0002468  0.0039476  -0.063
ddus[1:428]  0.0257214  0.0170116   1.512
            Pr(>|t|)
(Intercept)    0.950
ddus[1:428]    0.131

Residual standard error: 0.08148 on 424 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.005363, Adjusted R-squared:  0.003017
F-statistic: 2.286 on 1 and 424 DF,  p-value: 0.1313

> par(mfcol=c(2,1))
> acf(m2$residuals,lag=36)
> pacf(m2$residuals,lag=36)

```



使用美国月失业率 1 阶滞后作为解释变量对加利福尼亚州月失业率进行回归，可以看到简单的回归模型并不能很好的拟合数据

```

> t.test(ddca)

One Sample t-test

data:  ddca
t = 0, df = 426, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.007766479  0.007766479
sample estimates:
mean of x
4.159896e-18

```

对残差构建 ARIMA(1, 0, 6)模型重新进行拟合，拟合得到的模型为

```
> m3<-arima(ddca[2:427],order=c(1,0,6),xreg=ddus[1:426],include.mean=F)
> m3

Call:
arima(x = ddca[2:427], order = c(1, 0, 6), xreg = ddus[1:426], include.mean = F)

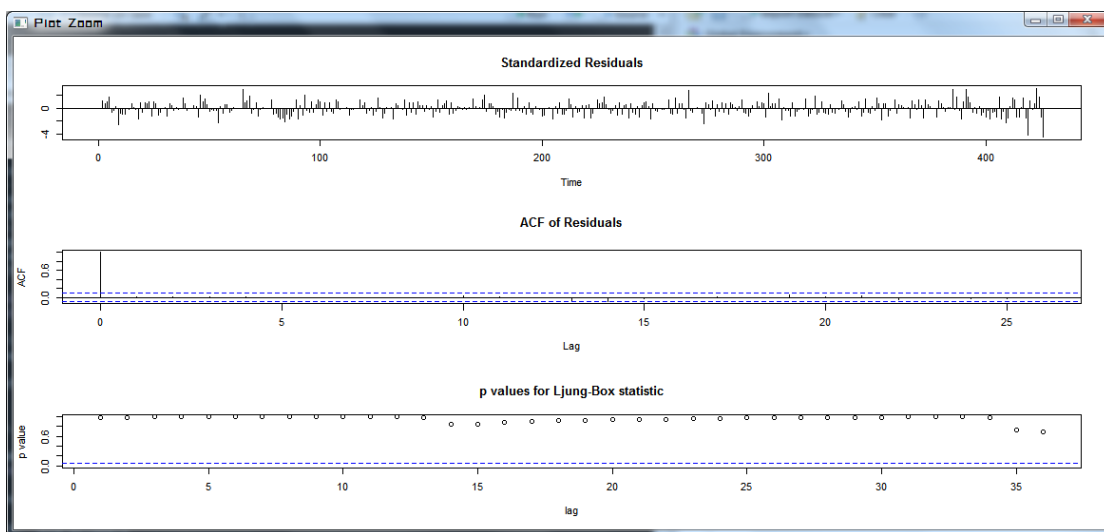
Coefficients:
      ar1      ma1      ma2      ma3      ma4      ma5      ma6      ddus[1:426]
-0.8133  0.3924 -0.1086  0.0840 -0.1187 -0.1326 -0.2233  0.0448
s.e.    0.0924  0.1001  0.0646  0.0564  0.0587  0.0572  0.0494  0.0179

sigma^2 estimated as 0.005108:  log likelihood = 519.2,  aic = -1020.4
> Box.test(m3$residuals,lag=36,type='Ljung')

Box-Ljung test

data:  m3$residuals
X-squared = 31.5502, df = 36, p-value = 0.6802

> tsdiag(m3,gof=36)
```



检验表明模型的拟合是充分的

(c)

```
> source("backtest.txt")
> mm1<-backtest(m1,ca,264,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.08423641
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.06306109
> mm3<-backtest(m3,ddca[2:427],263,1,inc.mean=F,xre=ddus[1:426])
[1] "RMSE of out-of-sample forecasts"
[1] 0.08713644
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.06527401
```

样本外预测发现这两个模型的效果差别不大

2. 考虑美国从 1971 年 4 月到 2011 年 11 月的 30 年期的月资产抵押率。数据来自 FRED，它们在文件 m-morgfed-7611.txt 中给出(变量为: year、mon、day、morg 和 fed)。

174

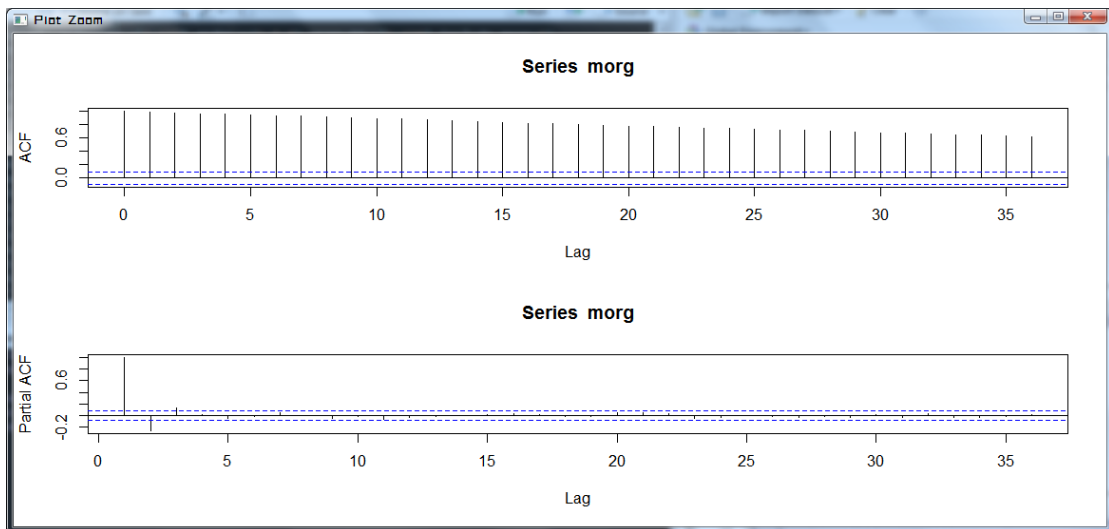
- 对于月抵押率数据，建立一个纯时间序列模型，进行模型检验，并写出拟合的模型。
- 众所周知，抵押率依赖于联邦基金利率。对抵押率建立一个时间序列模型，用联邦基金利率的 1 阶滞后后作为解释变量，进行模型检验，并写出拟合的模型。根据拟合的模型说明在 5% 的显著性水平下，抵押率是否依赖于联邦基金。
- 把 1971 年 4 月到 2011 年 11 月作为预测期，使用样本外预测并比较这两个模型。

(a)


```

> rate<-read.table("E:/DATA/data mining/fts03/data/m-morgfed-7111.txt",header=F)
> morg<-rate[,4]
> 
> par(mfcol=c(2,1))
> acf(morg,lag=36)
> pacf(morg,lag=36)

```

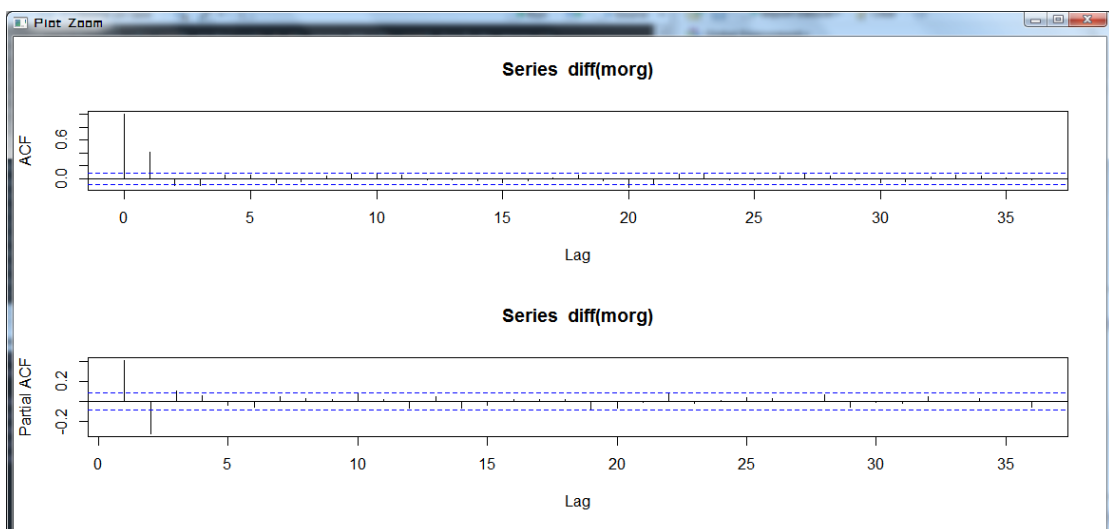


显然是个非平稳的时间序列，进行一阶差分

```

> acf(diff(morg),lag=36)
> pacf(diff(morg),lag=36)

```



从 acf 和 pacf 图上看，可构建一个 ARIMA (1, 1, 1) 模型
使用 ARIMA (1, 1, 1) 模型进行拟合

```

> m1<-arima(morg,order=c(1,1,1))
> m1

Call:
arima(x = morg, order = c(1, 1, 1))

Coefficients:
      ar1      ma1
    0.0426  0.5204
s.e.  0.0733  0.0592

sigma^2 estimated as 0.06632:  log likelihood = -30.53,  aic = 67.06

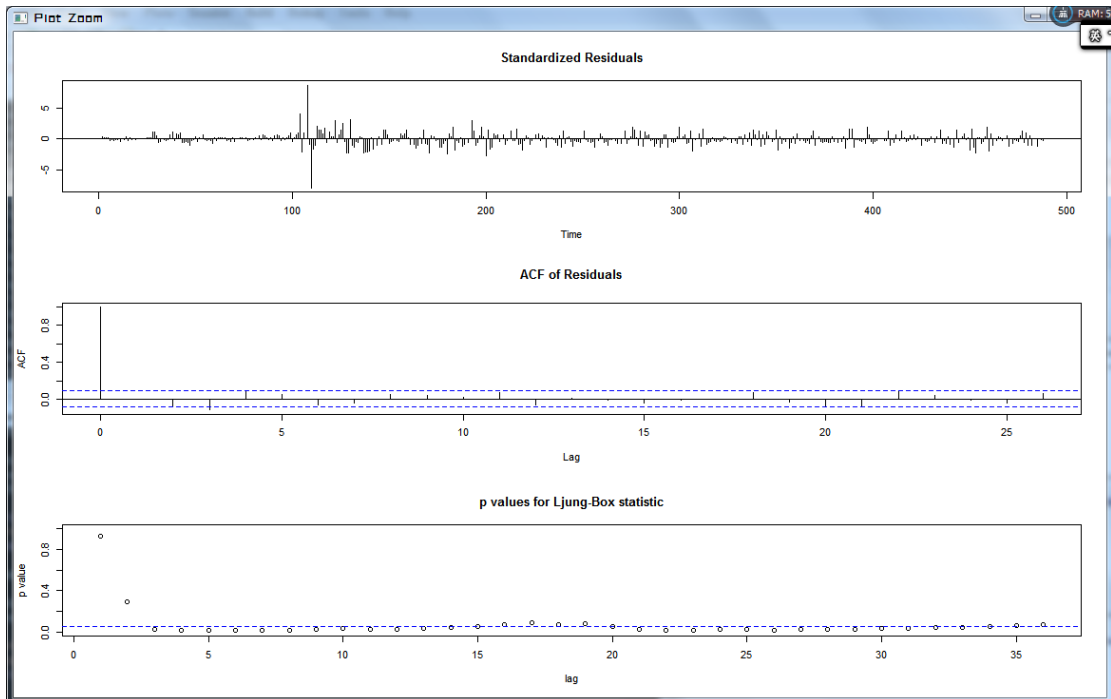
```

```
> Box.test(m1$residuals,lag=36,type='Ljung')

Box-Ljung test

data:  m1$residuals
X-squared = 49.133, df = 36, p-value = 0.07103

warning message:
display list redraw incomplete
> tsdiag(m1,gof=36)
```



检验表明模型的拟合是充分的

(b)

对抵押率和联邦基金利率分别进行一次差分

```
> fed<-rate[,5]
> dmorg<-diff(morg)
> dfed<-diff(fed)
```

使用联邦基金利率的 1 阶滞后作为解释变量对抵押率进行回归

```
> m2<-lm(dmorg[2:487]~-1+dfed[1:486])
> summary(m2)

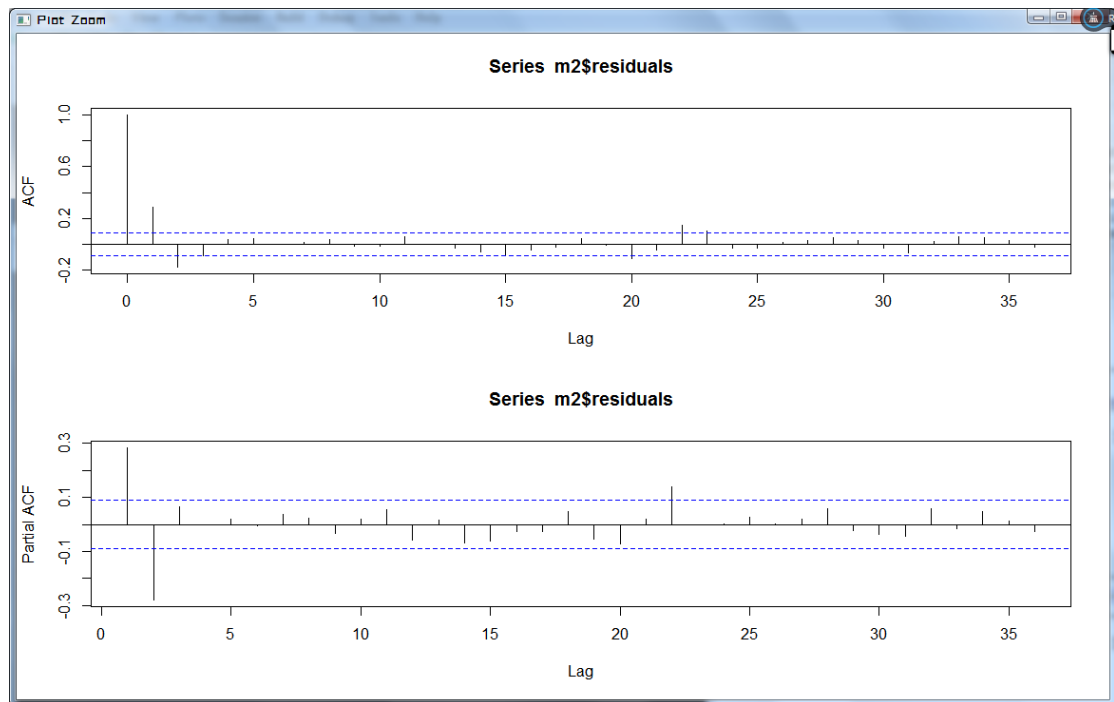
Call:
lm(formula = dmorg[2:487] ~ -1 + dfed[1:486])

Residuals:
    Min       1Q   Median       3Q      Max
-2.14814 -0.13645 -0.01930  0.09901  2.18233

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
dfed[1:486]   0.18604    0.02078   8.953  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2745 on 485 degrees of freedom
Multiple R-squared:  0.1418, Adjusted R-squared:  0.1401
F-statistic: 80.16 on 1 and 485 DF, p-value: < 2.2e-16

>
> par(mfcol=c(2,1))
> acf(m2$residuals,lag=36)
> pacf(m2$residuals,lag=36)
```



可见简单的回归模型并不能很好的拟合数据

```
> t.test(dmorg)

One Sample t-test

data:  dmorg
t = -0.5083, df = 486, p-value = 0.6115
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.03316867  0.01953418
sample estimates:
mean of x
-0.006817248
```

对残差构建 ARIMA (1, 0, 1) 模型重新进行拟合

```
> m3<-arima(dmorg[2:487],order=c(1,0,1),xreg=dfed[1:486],include.mean=F)
> m3

Call:
arima(x = dmorg[2:487], order = c(1, 0, 1), xreg = dfed[1:486], include.mean = F)

Coefficients:
      ar1      ma1  dfed[1:486]
    -0.0881  0.5515      0.1145
s.e.   0.0878  0.0672      0.0233

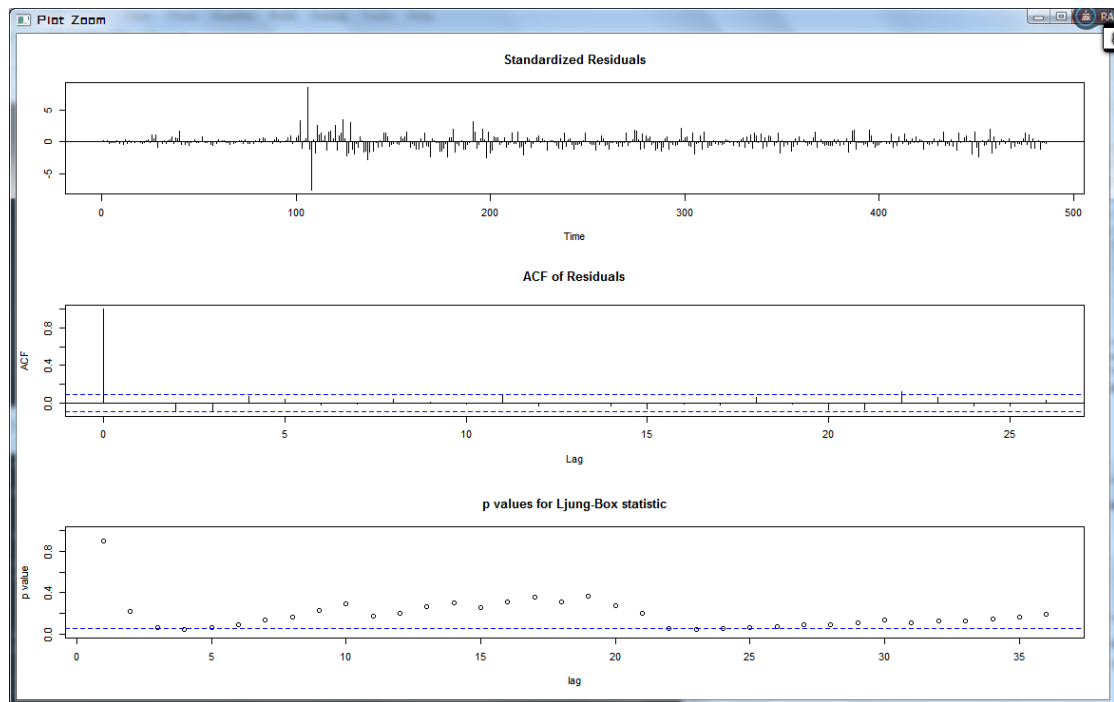
sigma^2 estimated as 0.06328:  log likelihood = -19,  aic = 46.01
> Box.test(m3$residuals, lag=36,type="Ljung")

Box-Ljung test

data:  m3$residuals
X-squared = 43.3178, df = 36, p-value = 0.1874

> tsdiag(m3,gof=36)
```

可以看到在 5% 的显著性水平下，抵押率确实是依赖于联邦基金的



检验表明模型的拟合是充分的

(c)

```
> source("backtest.txt")
> mm1<-backtest(m1,morg,393,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.1975722
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.1445895
> mm3<-backtest(m3,dmorg[2:487],392,1,inc.mean=F,xre=dfed[1:486])
[1] "RMSE of out-of-sample forecasts"
[1] 0.1914002
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.140103
```

样本外预测发现与联邦基金相关的模型更好