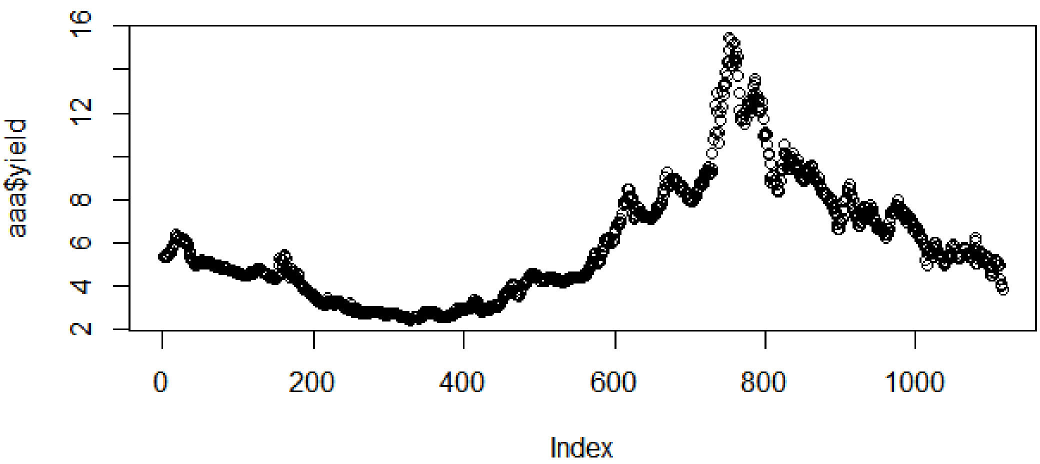
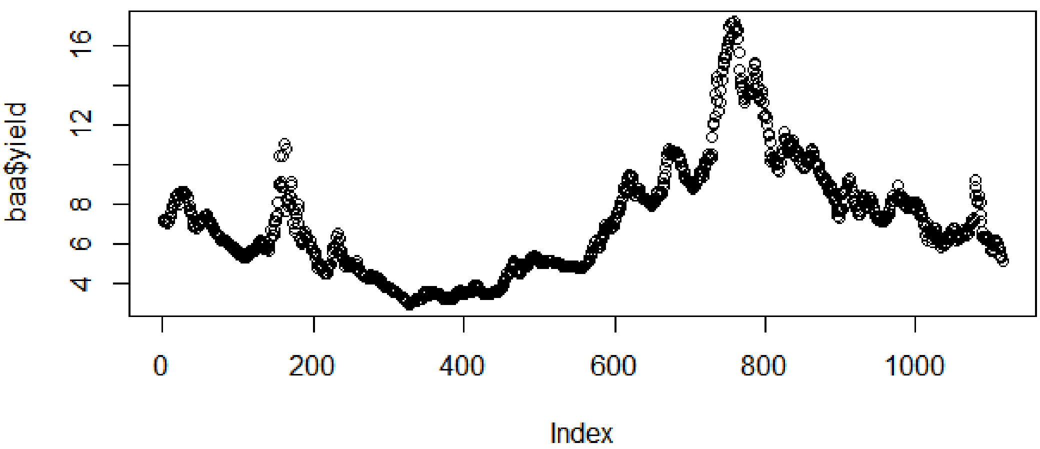
1. 考虑 Moody 公司 Aaa 和 Baa 级季度债券从从 1919 年 1 月到 2011 年 11 月的月收益率。

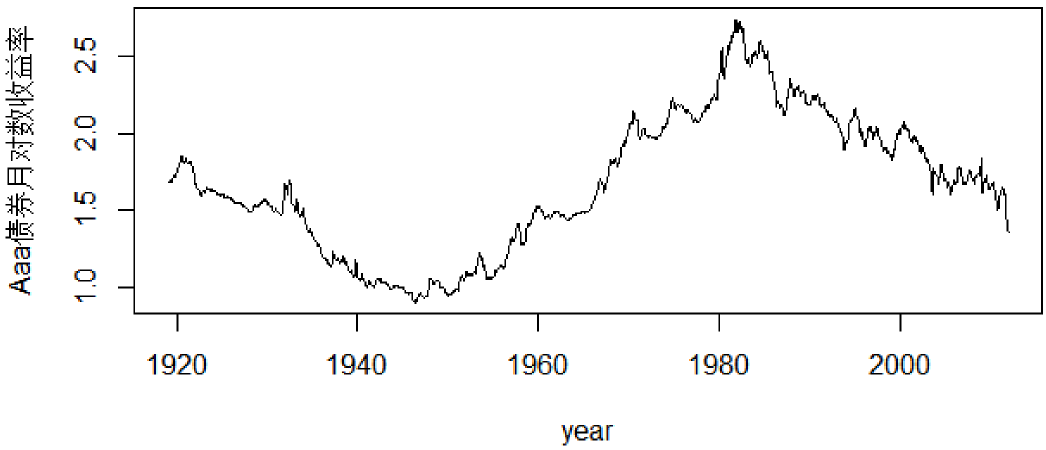
（1）判断 Aaa 和 Baa 的月对数收益率是否平稳时间序列

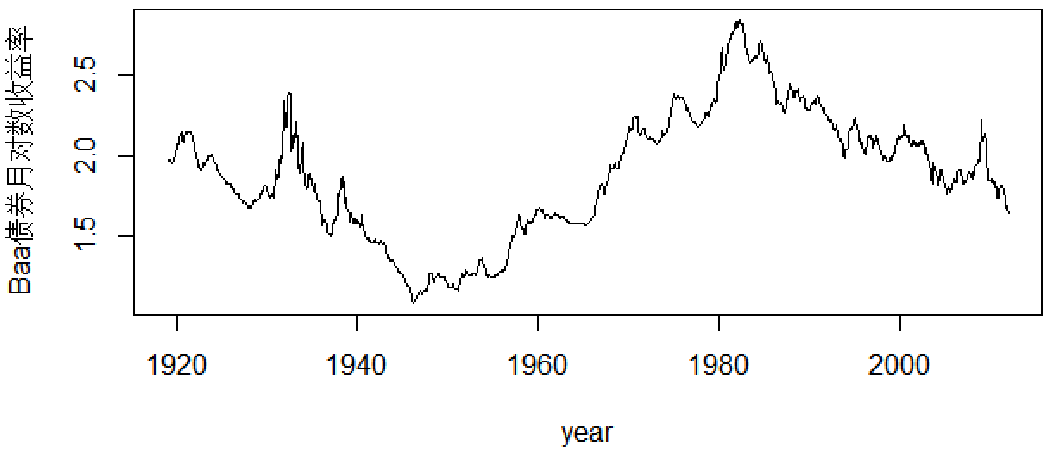
```
> aaa=read.table("E:/DATA/data mining/fts04/习题数据集/m-aaa-1911.txt",header=T)
> baa=read.table("E:/DATA/data mining/fts04/习题数据集/m-baa-1911.txt",header=T)
> head(aaa)
  year mon day yield
1 1919   1   1  5.35
2 1919   2   1  5.35
3 1919   3   1  5.39
4 1919   4   1  5.44
5 1919   5   1  5.39
6 1919   6   1  5.40
> head(baa)
  year mon day yield
1 1919   1   1  7.12
2 1919   2   1  7.20
3 1919   3   1  7.15
4 1919   4   1  7.23
5 1919   5   1  7.09
6 1919   6   1  7.04
> dim(aaa)
[1] 1115    4
> dim(baa)
[1] 1115    4
```

```
> plot(aaa$yield)
> plot(baa$yield)
> tdx=c(1:dim(aaa)[1])/12+1919
> aaa.log=log(aaa$yield)  #求对数
> baa.log=log(baa$yield)  #求对数
> par(mfcol=c(2,1))
> #画出Aaa的月对数收益率
> plot(tdx,aaa.log,xlab='year',ylab='Aaa债券月对数收益率',type='l')
> #画出Baa的月对数收益率
> plot(tdx,baa.log,xlab='year',ylab='Baa债券月对数收益率',type='l')
```










```
> library(fUnitRoots)
> m1 = ar(aaa.log,method='mle')
Warning message:
In arima0(x, order = c(i, 0L, 0L), include.mean = demean) :
  可能遇到了收斂问题: optim信息code=1
> m1$order
[1] 5
> adfTest(aaa.log,lags=5,type=c("c"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -0.9437
  P VALUE:
    0.708

Description:
 Mon Apr 27 20:33:51 2015 by user: Administrator
```

当 lag=5 时，p-value=0.708,ADF 检验统计量为-0.9437

零假设不能被拒绝，存在单位根

```
> m2 = ar(baa.log,method='mle')
> m2$order
[1] 10
> adfTest(baa.log,lags=10,type=c("c"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 10
  STATISTIC:
    Dickey-Fuller: -1.6819
  P VALUE:
    0.432

Description:
 Mon Apr 27 20:39:51 2015 by user: Administrator
```

当 lag=10 时，p-value=0.432，ADF 检验统计量为-1.6819

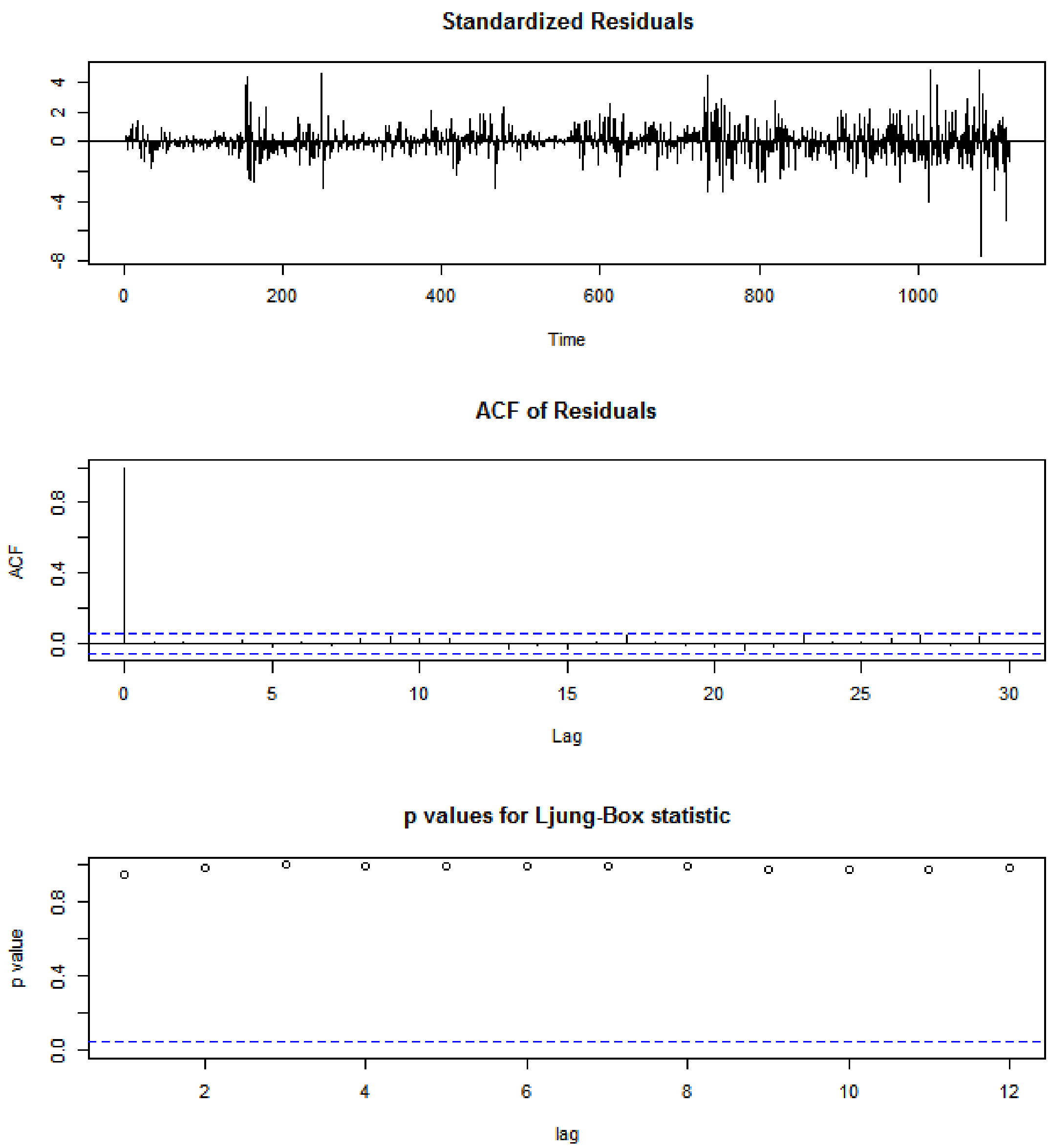
零假设不能被拒绝，存在单位根

（2）判断 Aaa 和 Baa 的月对数收益率是否白噪声

```
> library(forecast)
> auto.arima(aaa.log)
Series: aaa.log
ARIMA(2,1,3)

Coefficients:
          ar1      ar2     ma1     ma2     ma3
      -0.0644  -0.6485  0.4381  0.6547  0.1690
s.e.   0.1877   0.1943  0.1936  0.2161  0.0896

sigma^2 estimated as 0.0004623:  log likelihood=2696.54
AIC=-5381.08   AICc=-5381.01   BIC=-5350.99
```

选择的 ARIMA 模型是 ARIMA（2，1，3）

```
> m3=arima(aaa.log,order=c(2,1,3))
> m3

Call:
arima(x = aaa.log, order = c(2, 1, 3))

Coefficients:
          ar1      ar2     ma1     ma2     ma3
      -0.0644  -0.6485  0.4381  0.6547  0.1690
s.e.   0.1877   0.1943  0.1936  0.2161  0.0896

sigma^2 estimated as 0.0004623:  log likelihood = 2696.54,  aic = -5381.08
> tsdiag(m3,gof=12)   #检查残差是否白噪声
> Box.test(m3$residuals,lag=12,type='Ljung')

 Box-Ljung test

data:  m3$residuals
X-squared = 3.9581, df = 12, p-value = 0.9842
```

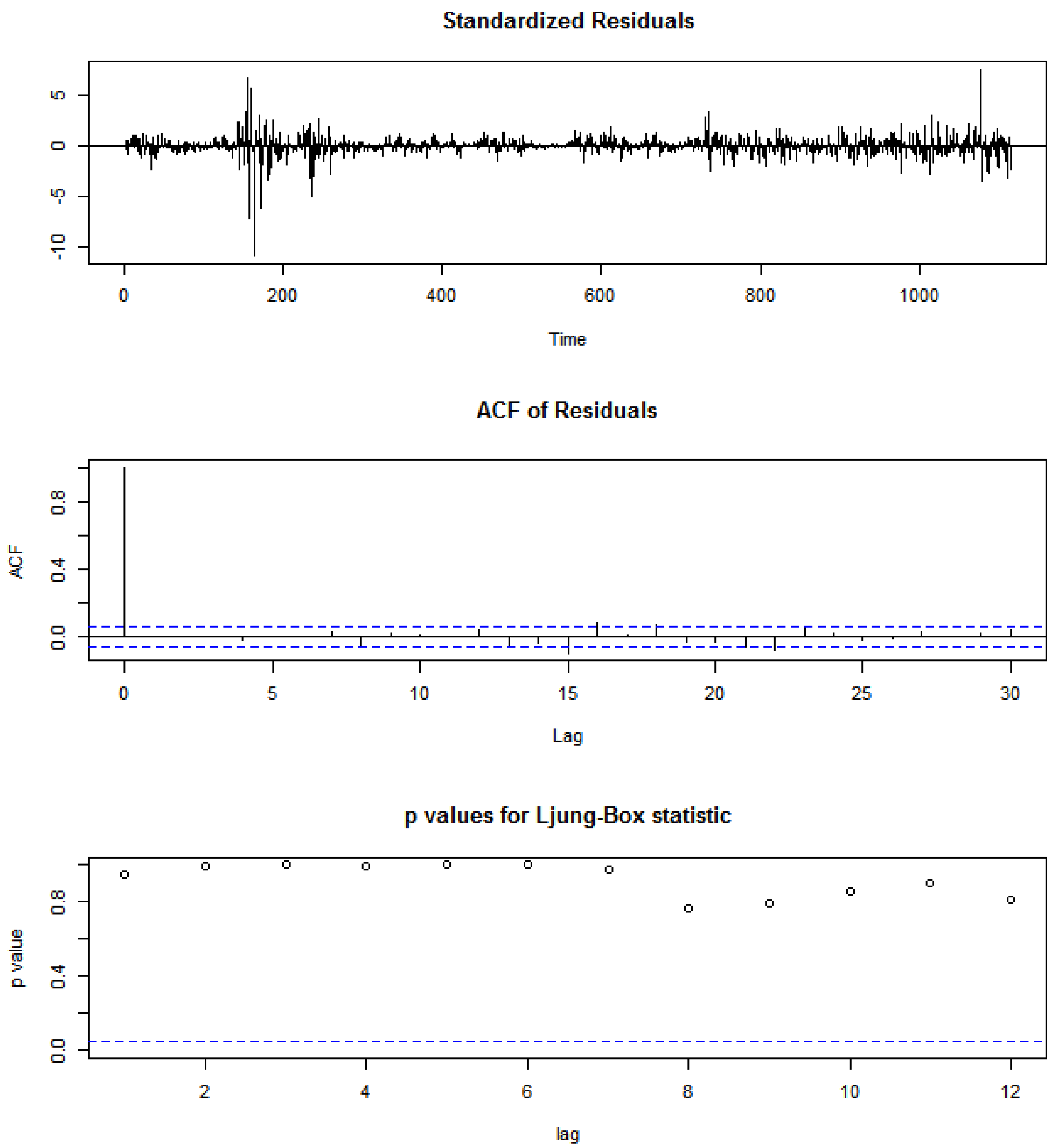


p-value=0.9842 显著，显然 Aaa 对数收益率是白噪声序列

```
> auto.arima(baa.log)
Series: baa.log
ARIMA(2,1,4)

Coefficients:
         ar1      ar2      ma1     ma2     ma3     ma4
      1.5043  -0.6627  -1.1895  0.2289  0.0479  0.1584
s.e.  0.0970   0.0839   0.0980  0.0773  0.0537  0.0305

sigma^2 estimated as 0.0006087:  log likelihood=2543.31
AIC=-5072.63   AICc=-5072.52   BIC=-5037.52
```

选择的 ARIMA 模型是 ARIMA(2,1,4)

```
> m4=arima(baa.log,order=c(2,1,4))
> m4

Call:
arima(x = baa.log, order = c(2, 1, 4))

Coefficients:
         ar1      ar2      ma1     ma2     ma3     ma4
      1.5043  -0.6627  -1.1895  0.2289  0.0479  0.1584
s.e.  0.0970   0.0839   0.0980  0.0773  0.0537  0.0305

sigma^2 estimated as 0.0006087:  log likelihood = 2543.31,  aic = -5072.63
> tsdiag(m4,gof=12)    #检查残差是否白噪声
> Box.test(m4$residuals,lag=12,type='Ljung')

 Box-Ljung test

data:  m4$residuals
X-squared = 7.667, df = 12, p-value = 0.8106
```



p-value=0.8106 显著，显然 Baa 对数收益率也是白噪声序列

（3）利用 4 中平滑法对 Aaa 和 Baa 的的月对数收益率做超前 1~3 步预测

```
> #######平滑法########
> #滑动平滑法
> sma.cal <- function( ts ) {
+   n <- length(ts)
+   for( t in 1 : 3 ) ts[n+t]= mean(ts[1:(n+t-1)])
+   return(tail(ts,3))
+ }
> sma.cal(aaa[,4])
[1] 5.890673 5.890673 5.890673
> sma.cal(baa[,4])
[1] 7.089336 7.089336 7.089336
```

```
> #加权平滑法
> wma.cal <- function( ts, weight ){
+   n <- length(ts)
+   k <- length(weight)
+   for( t in 1:3 ) ts[n+t]= sum( ts[(n+t-k):(n+t-1)] * weight )
+
+   return(tail(ts,3))
+ }
> wma.cal(aaa[,4],c(0.1,0.2,0.3,0.4))
[1] 3.99700 3.96480 3.95702
> wma.cal(baa[,4],c(0.1,0.2,0.3,0.4))
[1] 5.25700 5.24580 5.24042
```

```
> #k期移动平滑法
> kwma.cal <- function( ts, k=20 ) {
+   n <- length(ts)
+   for( t in 1:3) t[n+t]= mean(ts[(n+t-k):(n+t-1)])
+   return(tail(ts,3))
+ }
> kwma.cal(aaa[,4])
[1] 4.09 3.98 3.87
> kwma.cal(baa[,4])
[1] 5.27 5.37 5.14
```

```
> #指数平滑法
> ewma.cal <- function( ts, a=0.8 ){
+   n <- length(ts)
+   ewma <- c(ts[1])
+   for( t in 2:n ) ewma[t]= a*ts[t-1] + (1-a)*ewma[t-1]
+   for( t in 1:3 ) {
+       ewma[n+t]=a*ts[n+t-1]+(1-a)*ewma[n+t-1]
+       ts[n+t]=ewma[n+t]
+   }
+   return(tail(ts,3))
+ }
> ewma.cal(aaa[,4])
[1] 3.899556 3.899556 3.899556
> ewma.cal(baa[,4])
[1] 5.183362 5.183362 5.183362
```

2. 数据文件 wages 包含了 1981 年 7 月到 1987 年 6 月美国服装和纺织品行业工人的平均时薪（以美元计）的月度值。对该时间序列拟合线性趋势模型，并检验残差是否白噪声。

```
> library(TSA)
> wages = read.table("E:/DATA/data mining/fts04/习题数据集/wages.dat.txt", header=T)
> wages.lm=lm(wages$wages~time(wages$wages))
> summary(wages.lm)

Call:
lm(formula = wages$wages ~ time(wages$wages))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.23828 -0.04981  0.01942  0.05845  0.13136 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       7.9314358  0.0196657  403.31   <2e-16 ***
time(wages$wages) 0.0234234  0.0004682   50.03   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08257 on 70 degrees of freedom
Multiple R-squared:  0.9728,	Adjusted R-squared:  0.9724 
F-statistic:  2503 on 1 and 70 DF,  p-value: < 2.2e-16
```

上面给出了对该时间序列的线性拟合

下面将对残差进行白噪声检验

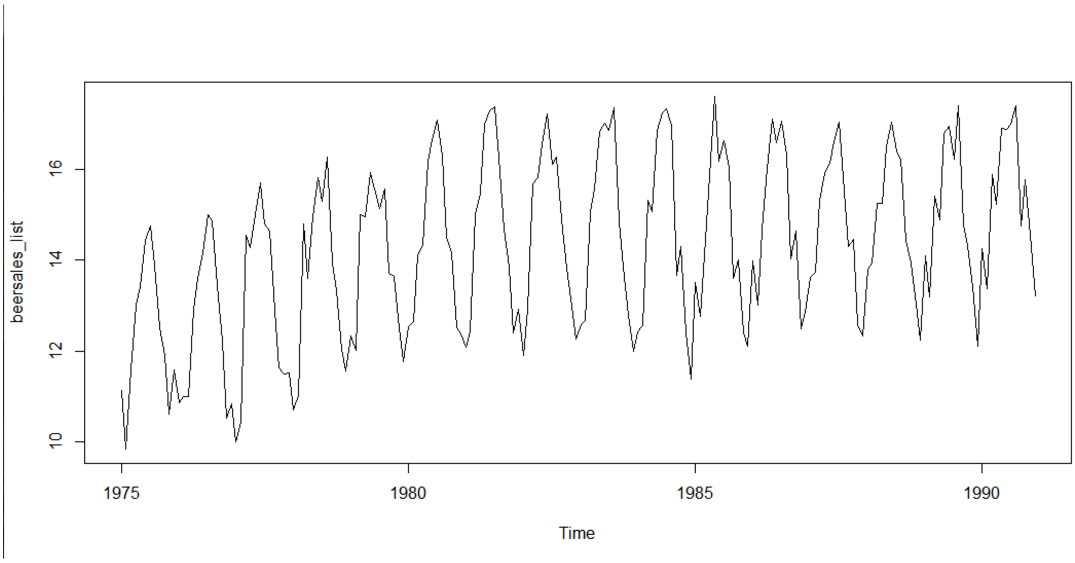
```
> Box.test(residuals(wages.lm),lag=12,type="Ljung")

	Box-Ljung test

data:  residuals(wages.lm)
X-squared = 111.3856, df = 12, p-value < 2.2e-16
```

p-value 非常小，拒绝零假设，不为白噪声

3. 数据文件 beersales 包含了从 1975 年 1 月到 1990 年 12 月美国月度啤酒销售量（单位：百万桶）

（1）画出该时间序列的时间序列图，并判断该时间序列是否有季节性波动

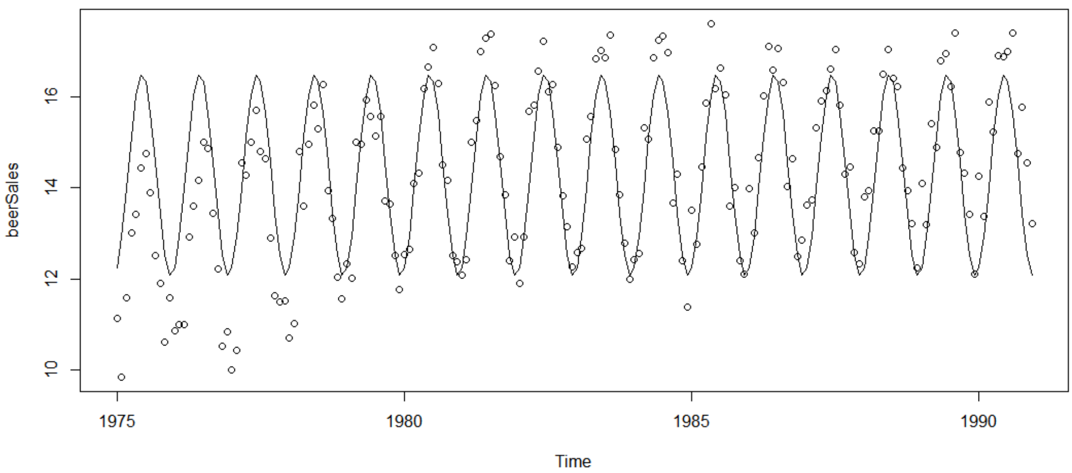
```
> beersales=read.table("E:/DATA/data mining/fts04/习题数据集/beersales.dat.txt", header=T)
> par(mfcol=c(1,1))
> beersales_list = ts(beersales$beersales,frequency=12,start=c(1975,1))
> plot.ts(beersales_list)
>
> month.=season(beersales_list)
> model2=lm(beersales_list~month.-1)
> summary(model2)

Call:
lm(formula = beersales_list ~ month. - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5745 -0.4772  0.1759  0.7312  2.1023 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
month.January    12.4857     0.2639   47.31   <2e-16 ***
month.February   12.3431     0.2639   46.77   <2e-16 ***
month.March      14.5679     0.2639   55.20   <2e-16 ***
month.April      14.8833     0.2639   56.39   <2e-16 ***
month.May        16.0846     0.2639   60.95   <2e-16 ***
month.June       16.3354     0.2639   61.90   <2e-16 ***
month.July       16.2543     0.2639   61.59   <2e-16 ***
month.August     16.0945     0.2639   60.98   <2e-16 ***
month.September  14.0585     0.2639   53.27   <2e-16 ***
month.October    13.7401     0.2639   52.06   <2e-16 ***
month.November   12.4377     0.2639   47.13   <2e-16 ***
month.December   12.0626     0.2639   45.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.056 on 180 degrees of freedom
Multiple R-squared:  0.995, Adjusted R-squared:  0.9946 
F-statistic:  2964 on 12 and 180 DF,  p-value: < 2.2e-16
```



（2）分别用季节均值法和余弦趋势法拟合数据，并比较两者的超期一步预测

季节均值法见上

```
> plot(ts(fitted(model4),freq=12,start=c(1975,1)),ylab='beerSales',type='l',
+      ylim=range(c(fitted(model4),beersales_list)))
> har.=harmonic(beersales_list,1)
> model4=lm(beersales_list~har.)
> summary(model4)

Call:
lm(formula = beersales_list ~ har.)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0812 -0.5975  0.1652  0.7376  2.3249 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     14.27898    0.07842 182.090  < 2e-16 ***
har.cos(2*pi*t) -2.04777    0.11090 -18.465  < 2e-16 ***
har.sin(2*pi*t)  0.83389    0.11090   7.519 2.16e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.087 on 189 degrees of freedom
Multiple R-squared:  0.6778, Adjusted R-squared:  0.6743 
F-statistic: 198.8 on 2 and 189 DF,  p-value: < 2.2e-16

> plot(ts(fitted(model4),freq=12,start=c(1975,1)),ylab='beerSales',type='l',
+      ylim=range(c(fitted(model4),beersales_list)))
> points(beersales_list)
```



beerSales
10
12
14
16
1975
1980
1985
1990
Time