第八章. 模型诊断

原俊青，浙江工业大学 2022-11-22

http://homepage.zjut.edu.cn/yjq/

对于一组给定的时间序列$X_1, X_2, \cdots, X_n$, 经过适当的变换或者差分之后，转化为平稳的时间序列，记为$Y_1, Y_2, \cdots, Y_n$， 利用样本acf、pacf、eacf、ARMA最优子集等进行模型识别定阶。例如AR(2)模型

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + e_t, \qquad \text{其中 } e_t \sim WN(0, \sigma_e^2).$$

并利用极大似然（或者最小二乘，矩估计）有效地估计模型的参数，$\hat{\mu}, \hat{\phi}_1, \hat{\phi}_2$，相应的残差定义为

$$\hat{e}_t = Y_t - (1 - \hat{\phi}_1 - \hat{\phi}_2)\hat{\mu} - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2}$$

类比于回归模型$Y_t = f(\cdot) + e_t$, 残差=实际观测值-模型预测值

-e_t 不可观测，残差可以;

-残差为e_t的估计值;

-如果模型被正确识别，那么残差应该近似具有白噪声的性质。

残差分析

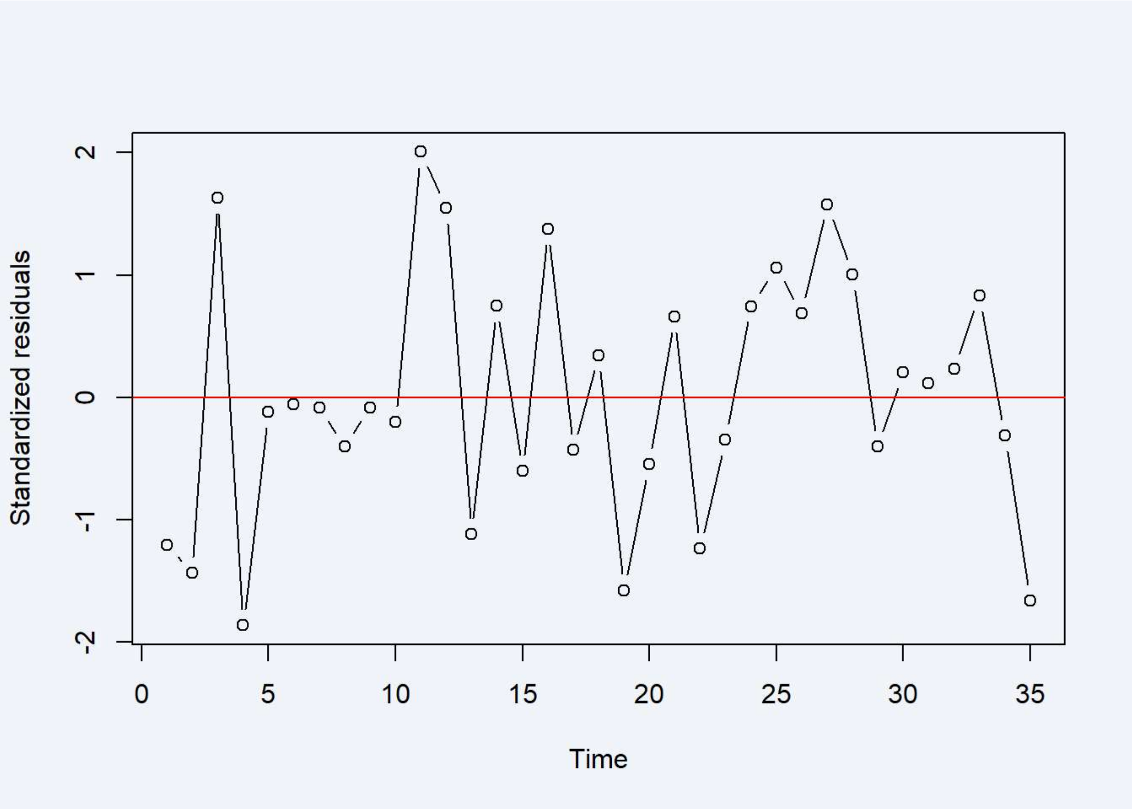
- 绘制残差图： plot(rstandard(arima(data,order=c(p,d,q))),type= ‘b’)

- 残差的正态性： hist( )； qqnorm( )； qqline( )； shapiro.test( ) ##H_0:正态

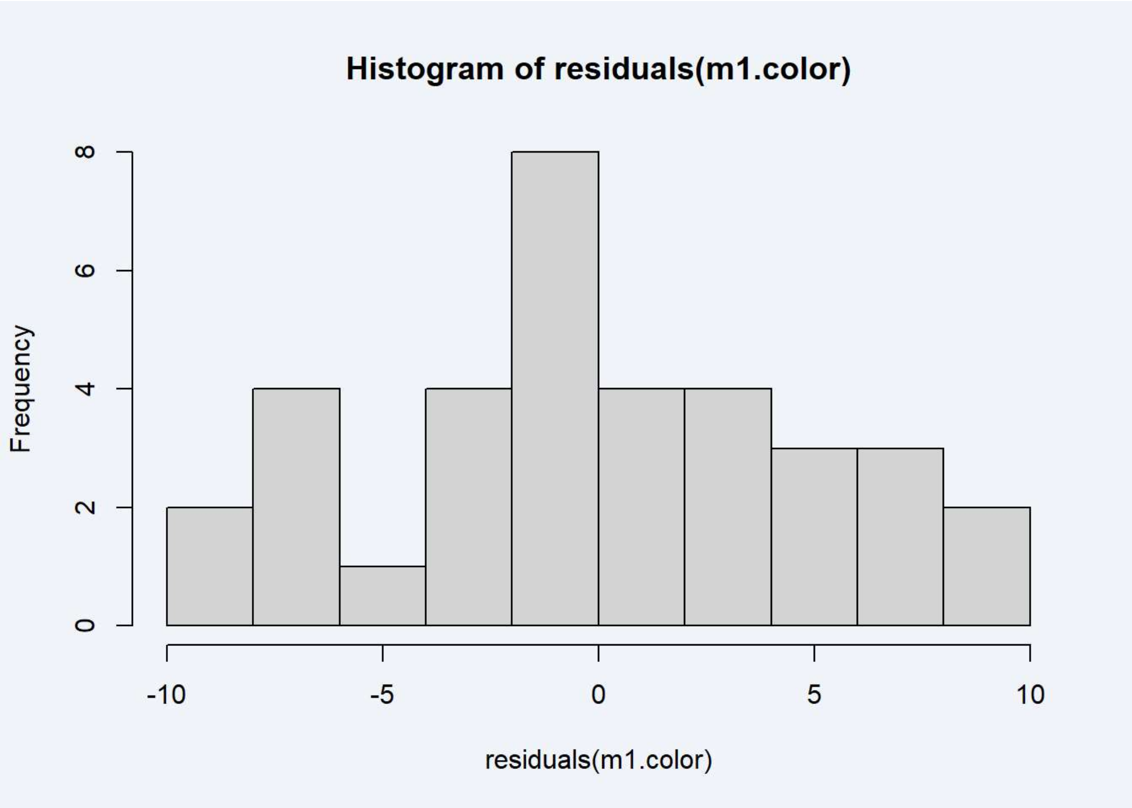
案例分析

化工颜色属性序列

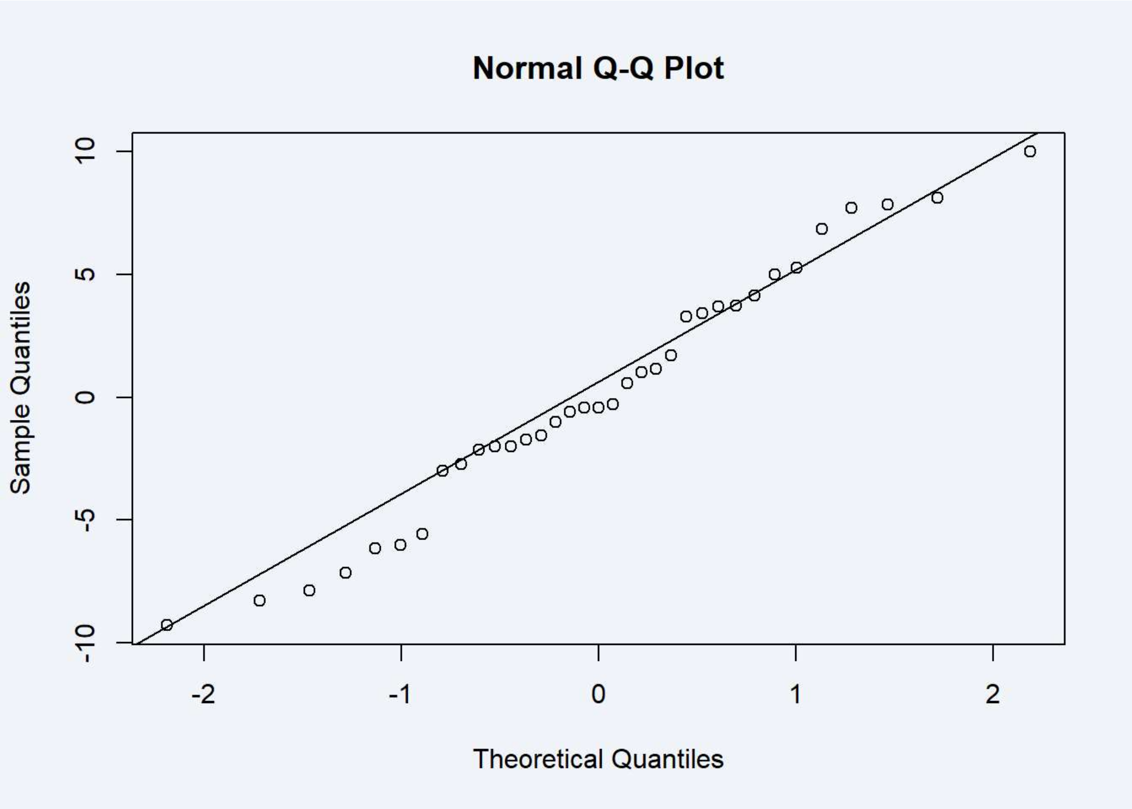
```
> library(TSA)
> data(color)
> m1.color=arima(color,order=c(1,0,0)) ##AR(1)
> m1.color

Call:
arima(x = color, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.5705    74.3293
s.e.  0.1435     1.9151

sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15
>
> plot(rstandard(m1.color),ylab='Standardized residuals',type='b')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=0,col='red')
```



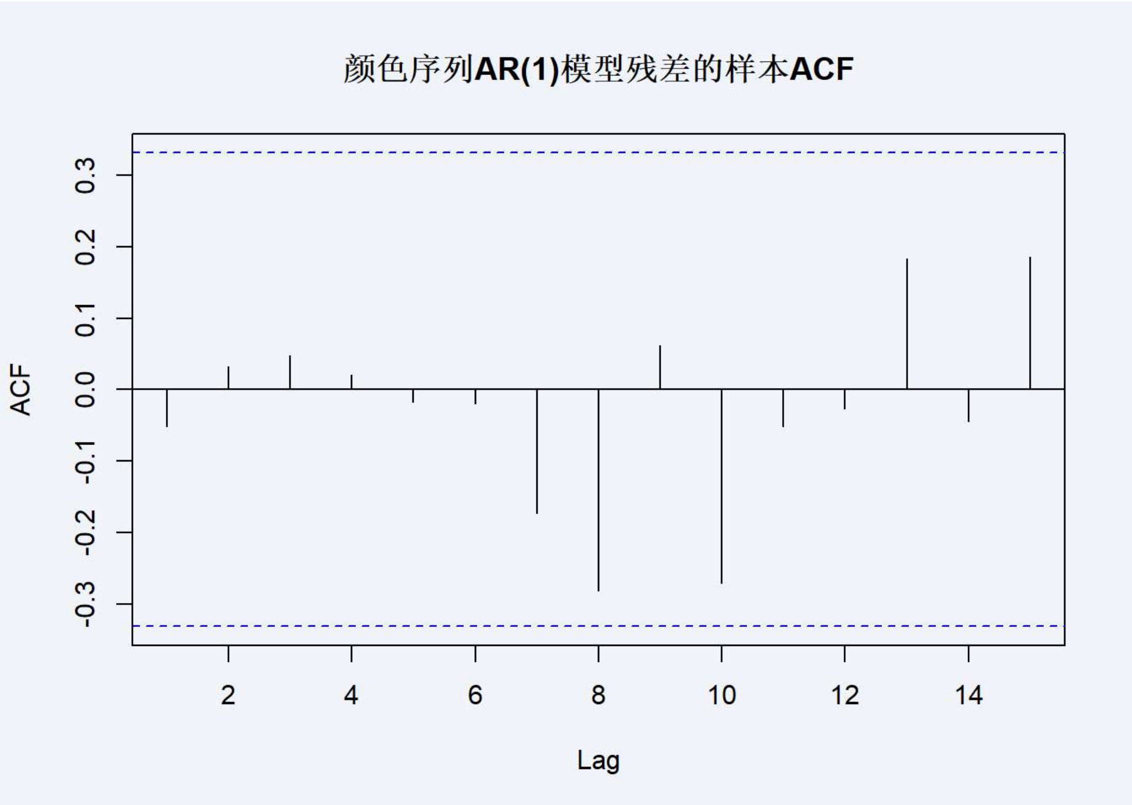

```
>
> ###残差正态性检验
> hist(residuals(m1.color))
Error : The fig.showtext code chunk option must be TRUE
```




```
> qqnorm(residuals(m1.color))
Error : The fig.showtext code chunk option must be TRUE
> qqline(residuals(m1.color))
```




```
> shapiro.test(residuals(m1.color)) #Shapiro-Wilk检验: 正态性

	Shapiro-Wilk normality test

data:  residuals(m1.color)
W = 0.97536, p-value = 0.6057
>
> ###残差相关性检验
> acf(residuals(m1.color),main='颜色序列AR(1)模型残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
```



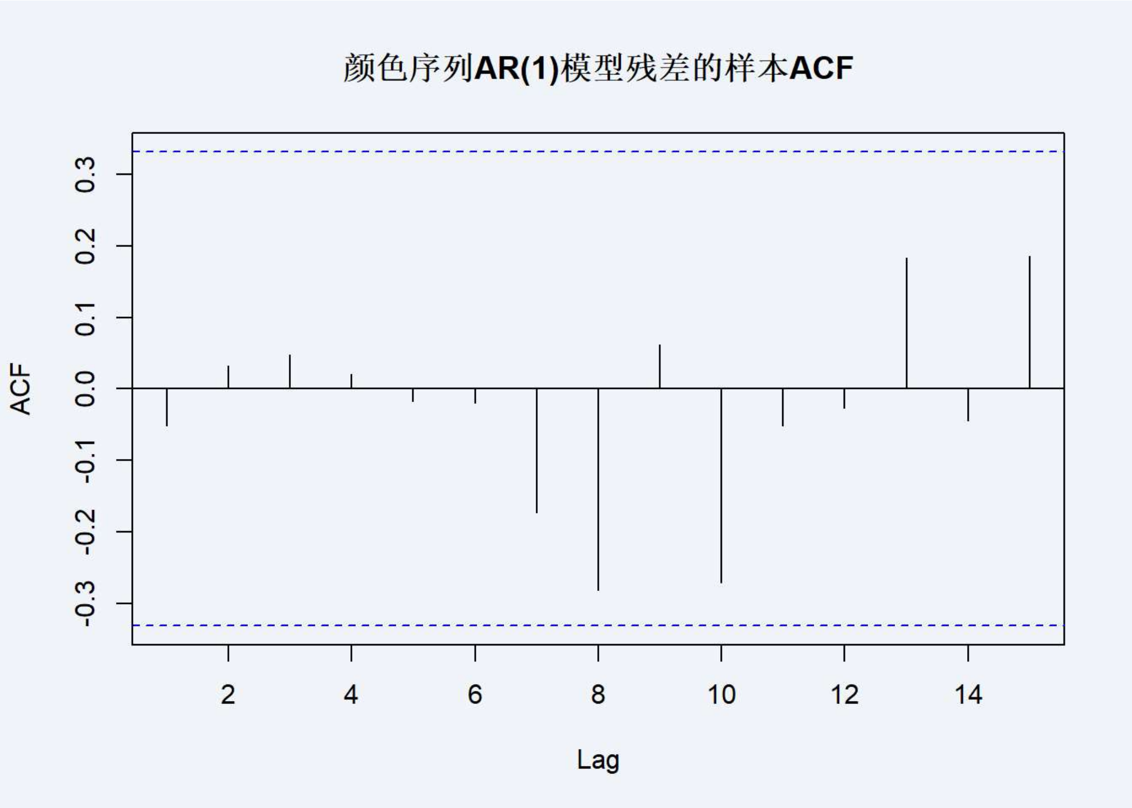

```
> HatPhi<-m1.color$coef[1]; HatPhi
      ar1
0.5705478
> n<-m1.color$nobs; n
[1] 35
```

化工颜色属性时间序列拟合的AR(1)模型中，$\hat{\phi}=0.5705478$，$n=35$, 残差项ACF值的近似标准差为

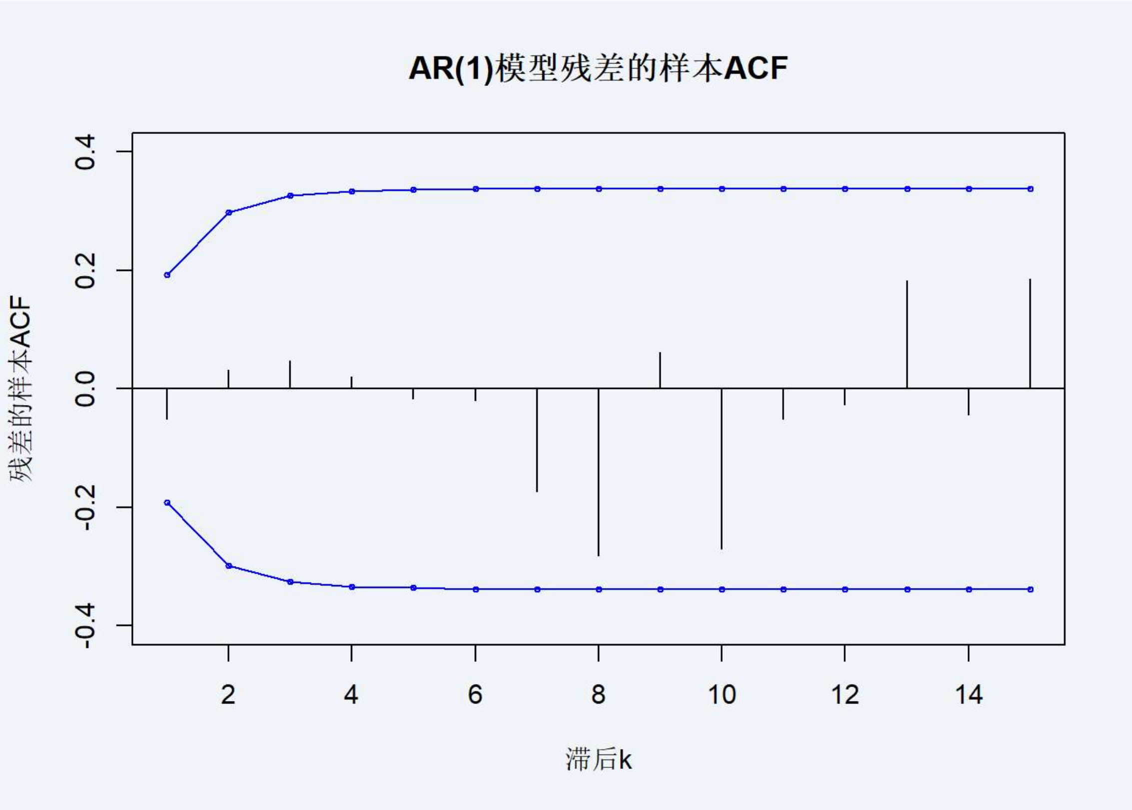
表8-3 颜色序列AR(1)模型残差的自相关函数的估计及标准差

滞后k	1	2	3	4	5	6
残差ACF	-0.051	0.032	0.047	0.021	-0.017	-0.019
$\sqrt{\widehat{Var}(\hat{r}_k)}$	0.096	0.149	0.163	0.167	0.168	0.169

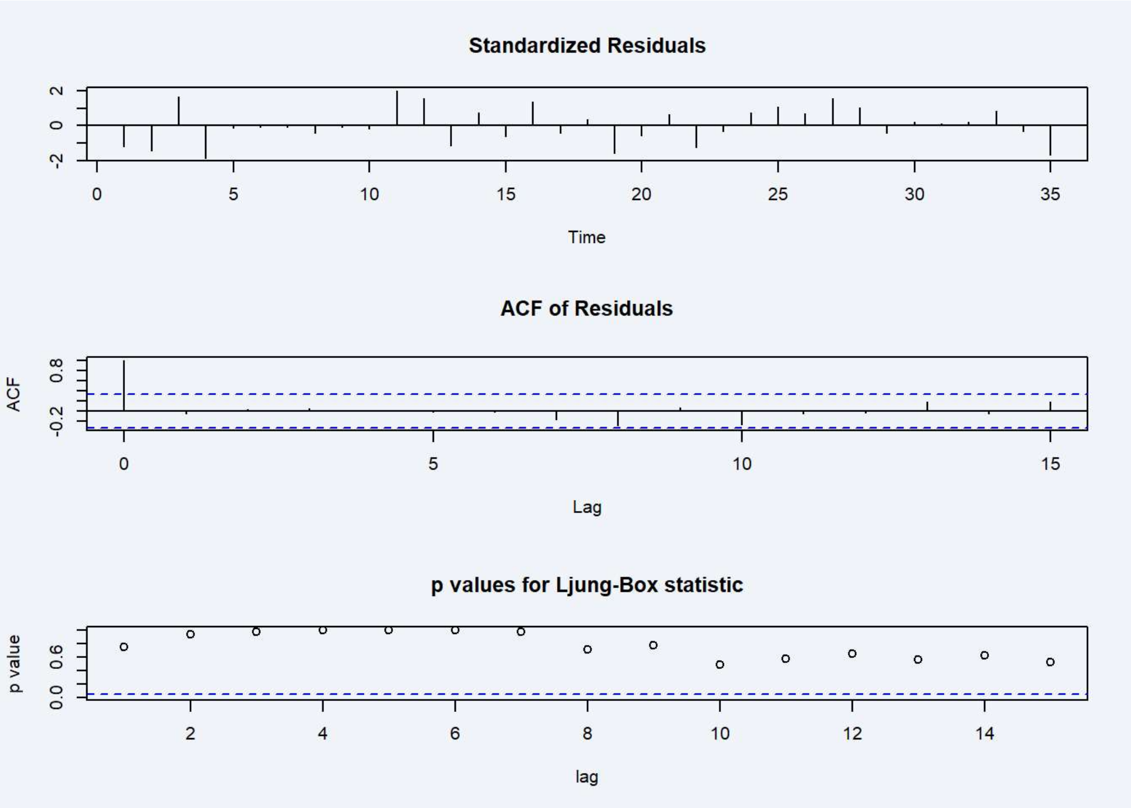
```
> HatAcf<-signif(acf(residuals(m1.color),main='颜色序列AR(1)模型残差的样本ACF')$acf,3)
Error : The fig.showtext code chunk option must be TRUE
```




```
> plot(HatAcf,type='h',  main='AR(1)模型残差的样本ACF', xlim=c(1,15),ylim=c(-0.4,0.4),xlab='滞后k', ylab='残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=0)
> k=1:15
> bd<-c(std,rep(0.169,9))
> par(new=TRUE)
> plot(x=k, y=-2*bd, type='o', cex=0.5, col='blue', xlim=c(1,15), ylim=c(-0.4,0.4), axes=FALSE, xlab='', ylab='')
Error : The fig.showtext code chunk option must be TRUE
> par(new=TRUE)
> plot(x=k, y=2*bd, type='o', cex=0.5, col='blue', xlim=c(1,15), ylim=c(-0.4,0.4), axes=FALSE, xlab='', ylab='')
Error : The fig.showtext code chunk option must be TRUE
```



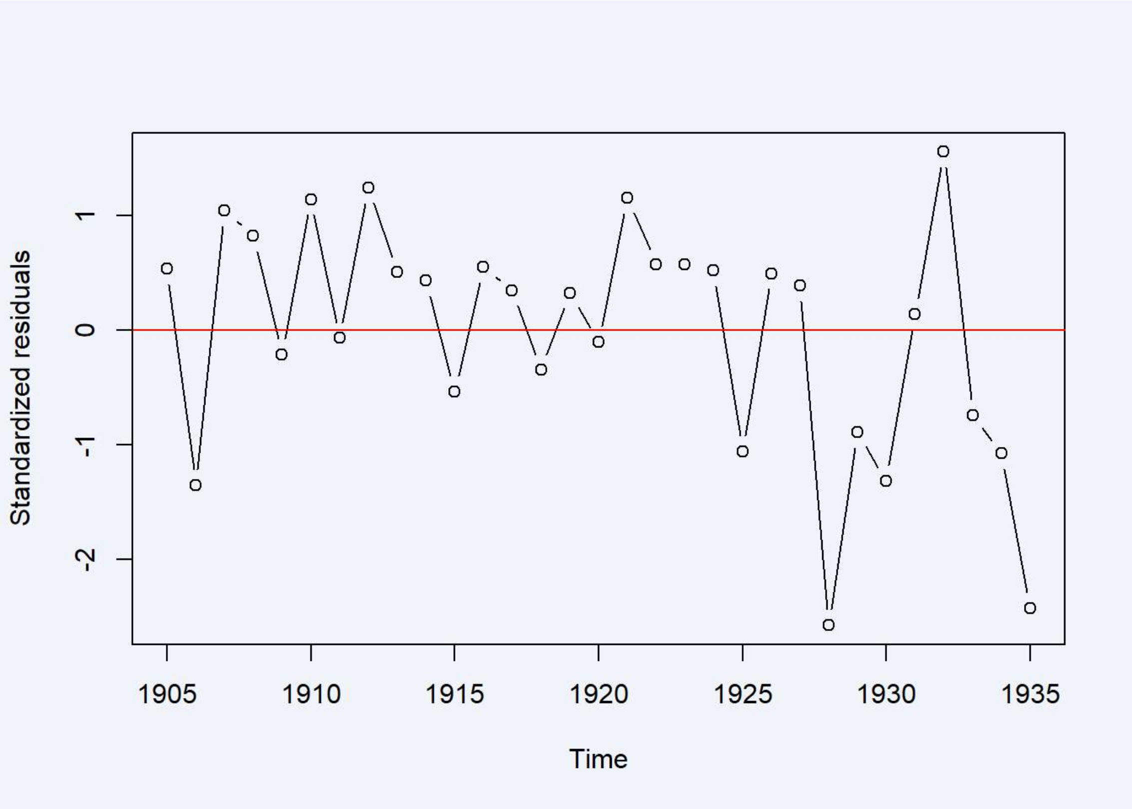

```
>
> ###残差相关性检验
> HatAcf<-signif(acf(residuals(m1.color),plot=F)$acf[1:6],2);HatAcf # Exhibit 8.11
[1] -0.051  0.032  0.047  0.021 -0.017 -0.019
> # to display the first 6 acf to 2 significant digits.
>
> #Ljung-Box检验 H_0: r_1=r_2=...=r_lag=0, 不相关
> LB.test(m1.color, lag=6) #LB-test the residuals of m1.color

    Box-Ljung test

data:  residuals from  m1.color
X-squared = 0.28032, df = 5, p-value = 0.998
> #LB.test(model, lag = 12, type = c("Ljung-Box", "Box-Pierce")
>
> tsdiag(m1.color,gof=15,omit.initial=F) # Exhibit 8.12
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
```



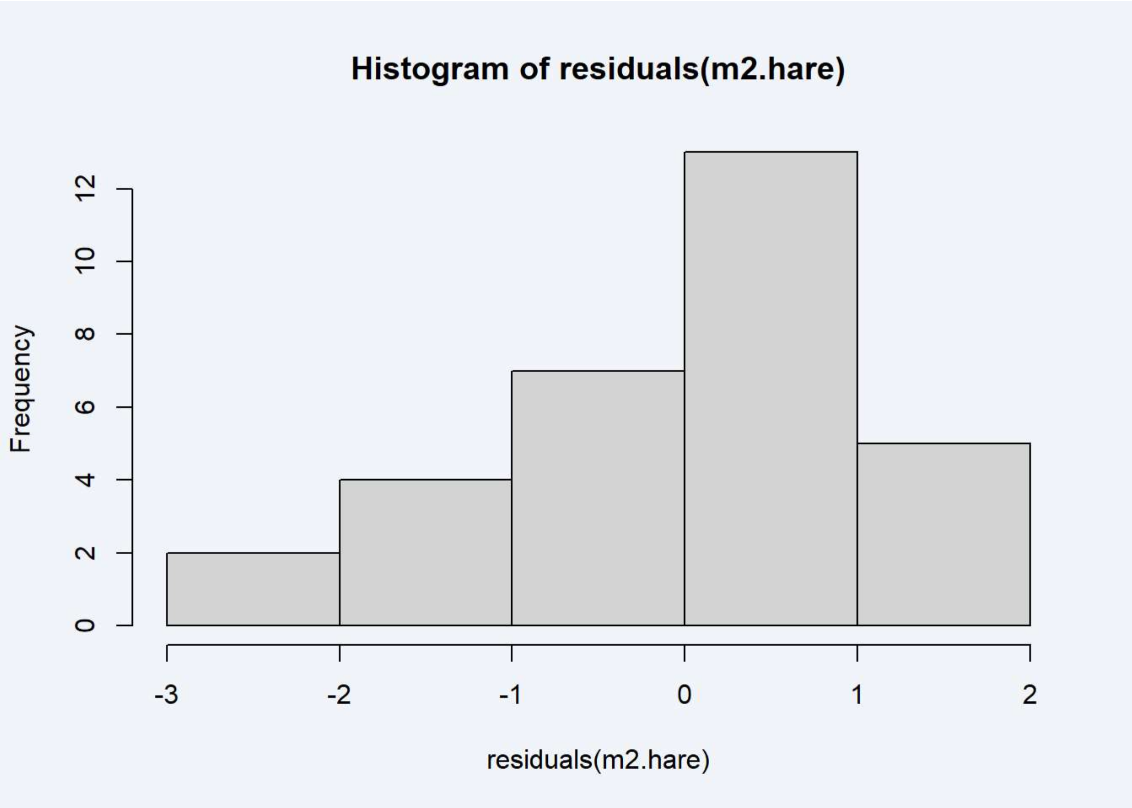

```
> # the tsdiag function is modified from that in the stats package of R.
>
> #runs(residuals(m1.color))  ##游程检验 H_0: 独立性
> #runs(rstandard(m1.color))  ##非参数检验，标准化没有影响
```

加拿大野兔丰度序列

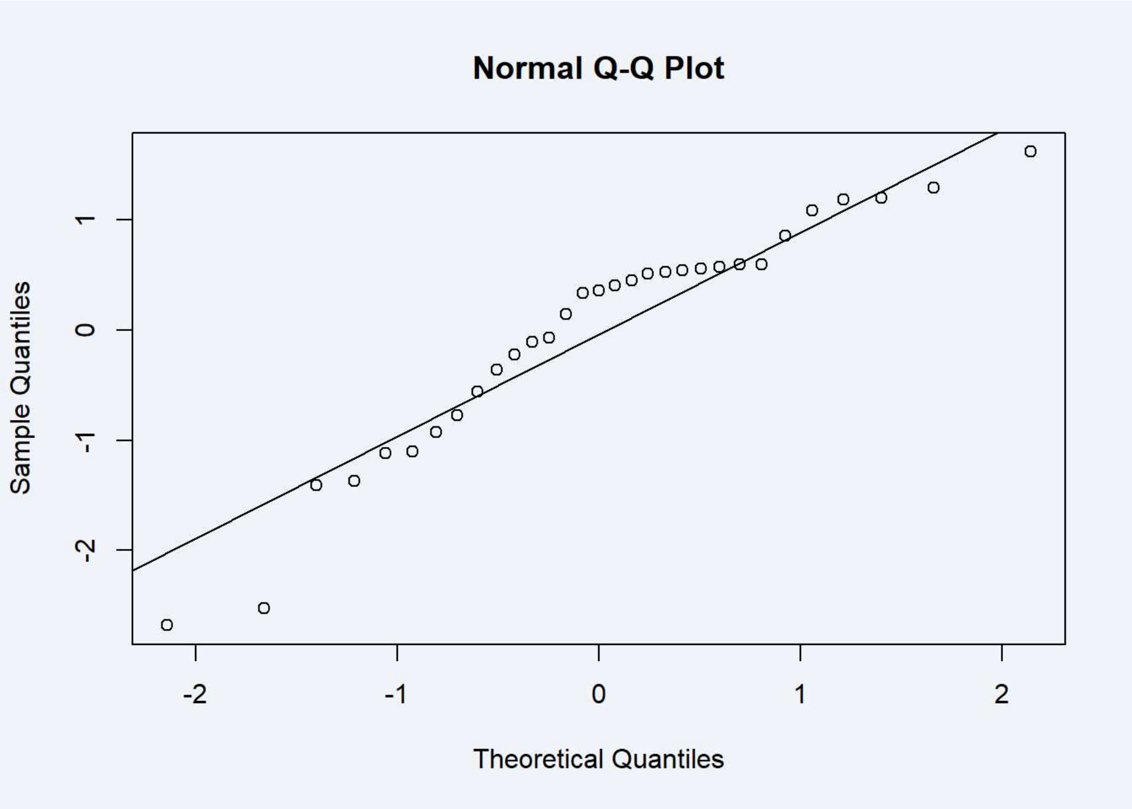
```
> library(TSA)
> data(hare)
> m1.hare=arima(sqrt(hare),order=c(3,0,0))
> m1.hare

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

Coefficients:
         ar1      ar2      ar3  intercept
      1.0519  -0.2292  -0.3931     5.6923
s.e.  0.1877   0.2942   0.1915     0.3371

sigma^2 estimated as 1.066:  log likelihood = -46.54,  aic = 101.08
> m2.hare=arima(sqrt(hare),order=c(3,0,0),fixed=c(NA,0,NA,NA))
> m2.hare

Call:
arima(x = sqrt(hare), order = c(3, 0, 0), fixed = c(NA, 0, NA, NA))

Coefficients:
         ar1  ar2      ar3  intercept
      0.9190    0  -0.5313     5.6889
s.e.  0.0791    0   0.0697     0.3179

sigma^2 estimated as 1.088:  log likelihood = -46.85,  aic = 99.69
>
> plot(rstandard(m2.hare),ylab='Standardized residuals',type='b')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=0,col='red')
```



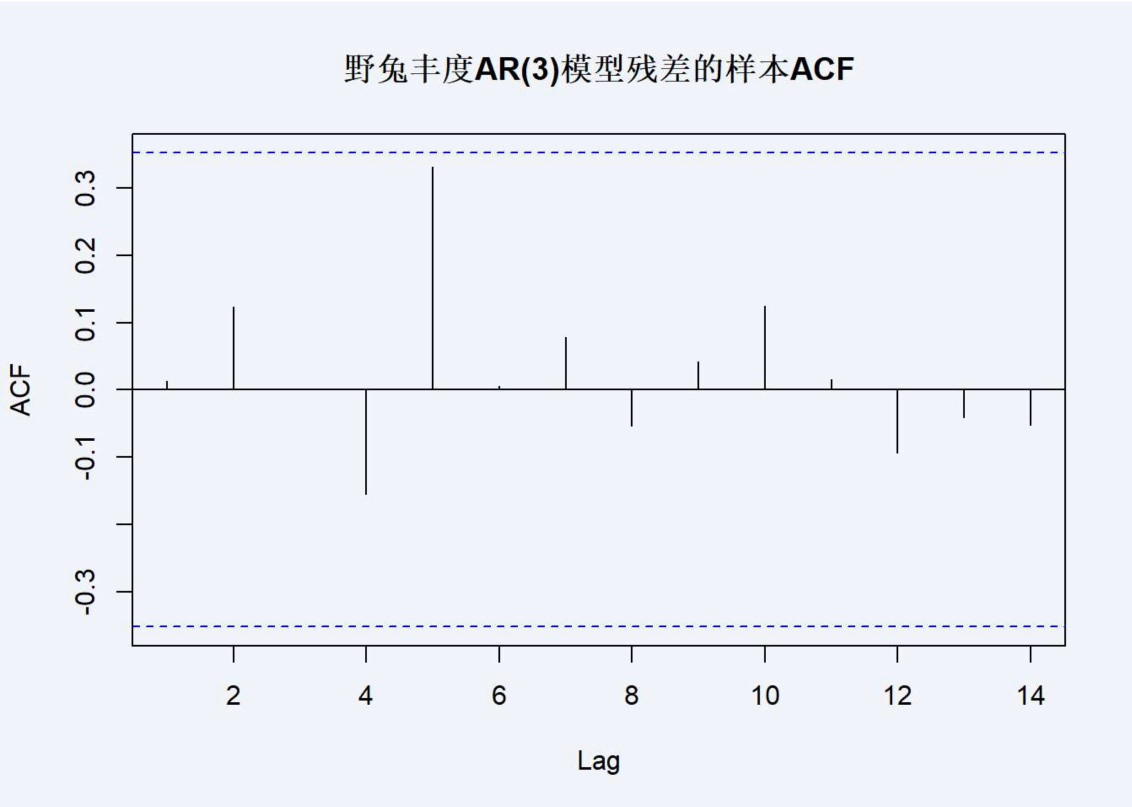

```
>
> ###残差的正态性检验
> hist(residuals(m2.hare))
Error : The fig.showtext code chunk option must be TRUE
```



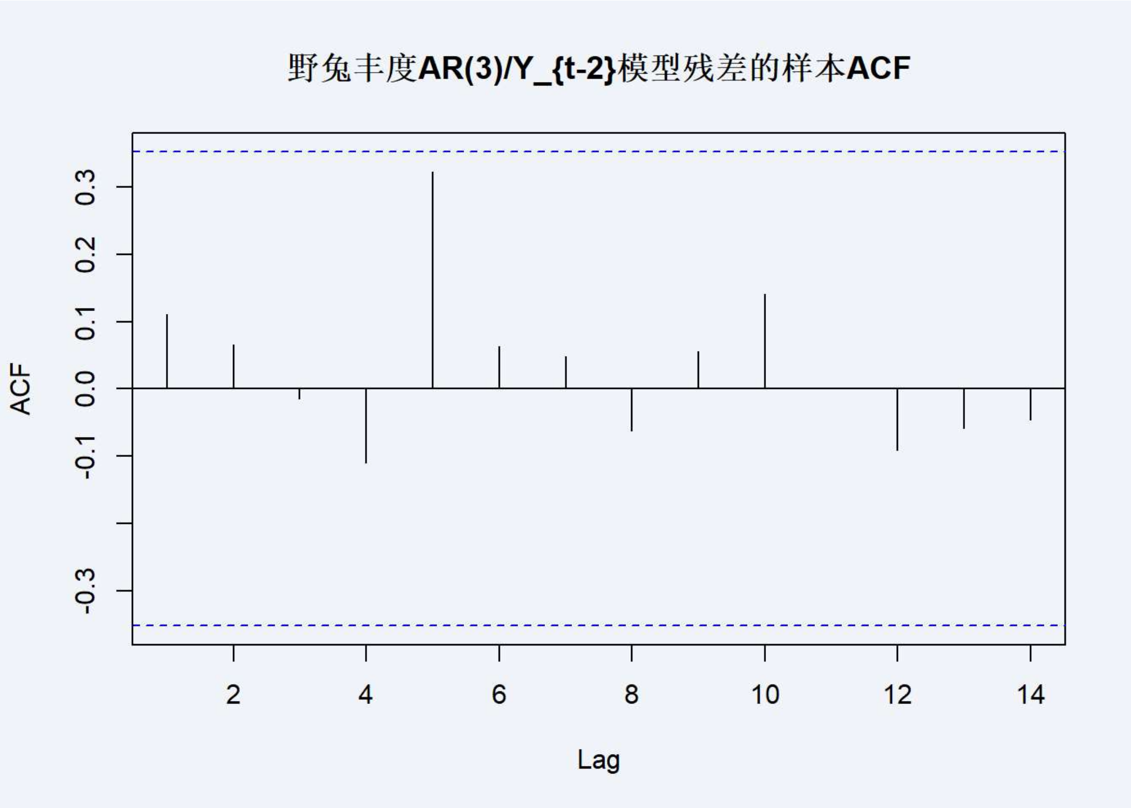

```
> qqnorm(residuals(m2.hare))
Error : The fig.showtext code chunk option must be TRUE
> qqline(residuals(m2.hare))
```



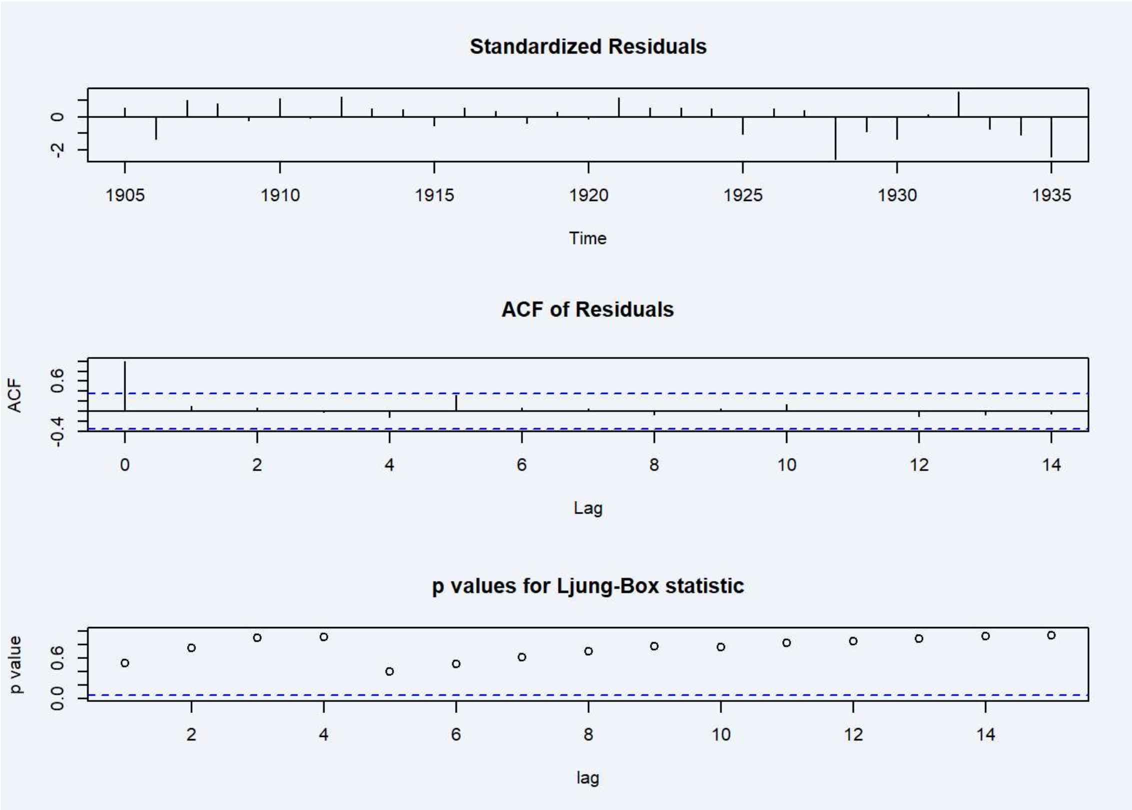

```
> shapiro.test(residuals(m2.hare)) ##H_0:正态性

    Shapiro-Wilk normality test

data:  residuals(m2.hare)
W = 0.92523, p-value = 0.03257
>
> ###残差的相关性检验
> acf(residuals(m1.hare),main='野兔丰度AR(3)模型残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
```




```
> acf(residuals(m2.hare),main='野兔丰度AR(3)/Y_{t-2}模型残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
```




```
>
> HatAcf<-signif(acf(residuals(m2.hare),plot=F)$acf[1:6],2);HatAcf
[1]  0.110  0.065 -0.014 -0.110  0.320  0.063
> # to display the first 6 acf to 2 significant digits.
>
> #Ljung-Box检验 H_0: r_1=r_2=...=r_lag=0, 不相关
> LB.test(m2.hare, lag=6) #LB-test the residuals of m1.color

    Box-Ljung test

data:  residuals from  m2.hare
X-squared = 5.2802, df = 4, p-value = 0.2597
> #LB.test(model, lag = 12, type = c("Ljung-Box", "Box-Pierce")
>
> tsdiag(m2.hare,gof=15,omit.initial=F) # Exhibit 8.12
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
```




```
> # the tsdiag function is modified from that in the stats package of R.
>
> runs(residuals(m2.hare))  ##游程检验 H_0: 独立性
$pvalue
[1] 0.602

$observed.runs
[1] 18

$expected.runs
[1] 16.09677

$n1
[1] 13

$n2
[1] 18

$k
[1] 0
> #runs(rstandard(m2.hare))  ##非参数检验，标准化没有影响
```

对原始数据进行开方，然后以AR(3)模型进行拟合，即

$$\sqrt{Y_t} - \mu = \phi_1(\sqrt{Y_{t-1}} - \mu) + \phi_2(\sqrt{Y_{t-2}} - \mu) + \phi_3(\sqrt{Y_{t-3}} - \mu) + e_t$$

参数的极大似然估计为$\tilde{\mu} = 5.6923, \tilde{\phi}_1 = 1.0519, \tilde{\phi}_2 = -0.2292, \tilde{\phi}_3 = -0.3930$, 代入AR(3)模型，整理得

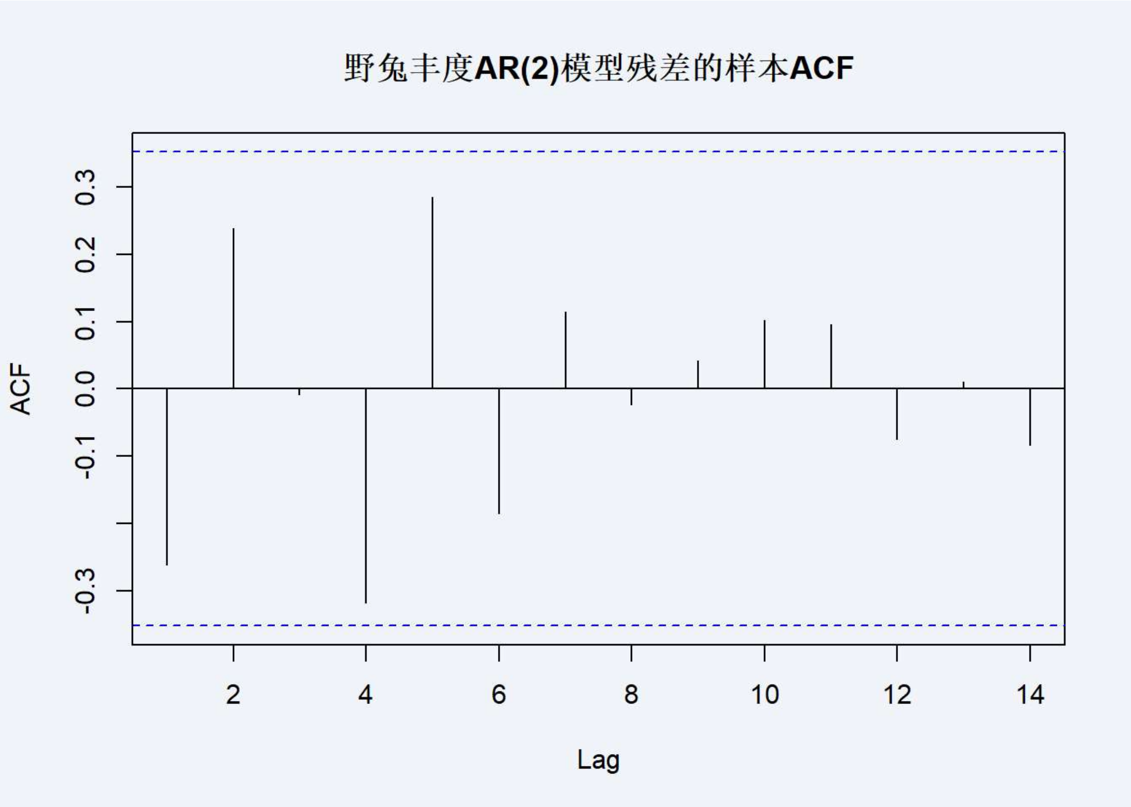
$$\sqrt{Y_t} = 3.2463 + 1.0519\sqrt{Y_{t-1}} - 0.2292\sqrt{Y_{t-2}} - 0.3930\sqrt{Y_{t-3}} + e_t$$

ϕ_2的95%置信区间为$[-0.2292 - 2 * 0.2942]$, 不显著不为零，故可以去掉该项。令$\phi_2 = 0$, 参数的极大似然估计为$\tilde{\mu} = 5.6889, \tilde{\phi}_1 = 0.9190, \tilde{\phi}_3 = -0.5313$, 代入AR(3)模型，整理得

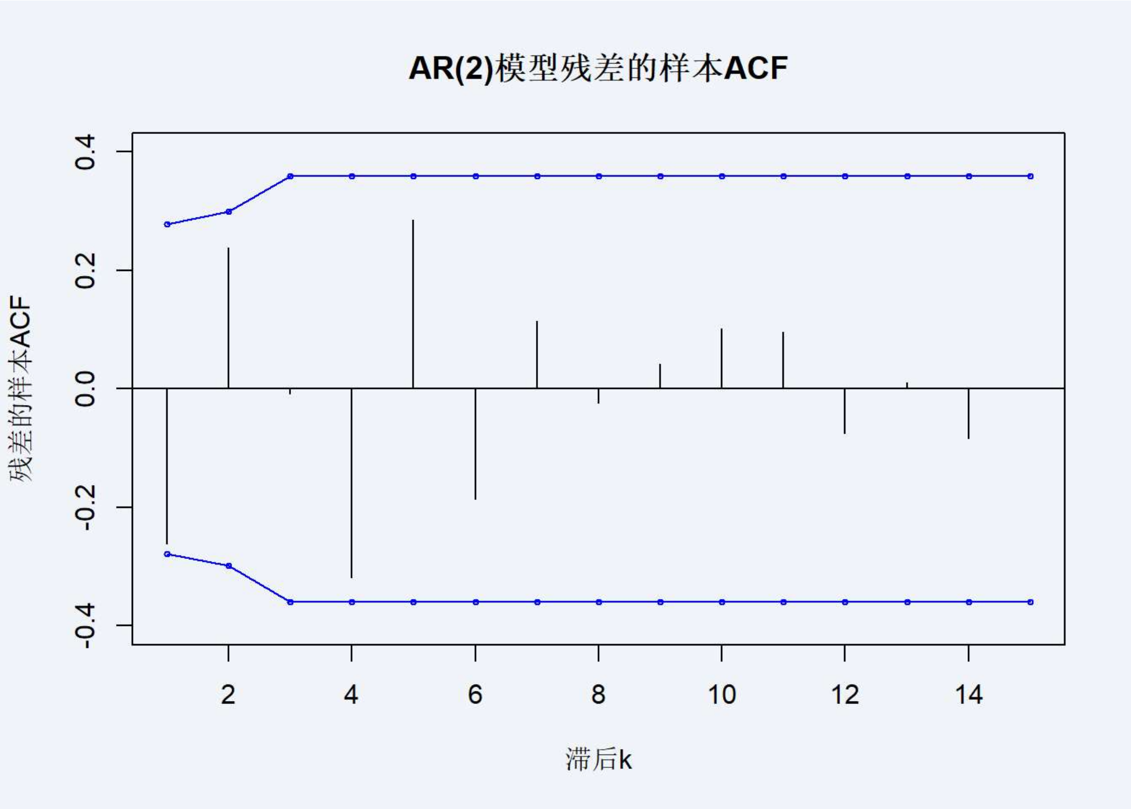
$$\sqrt{Y_t} = 3.4833 + 0.9190\sqrt{Y_{t-1}} - 0.5313\sqrt{Y_{t-3}} + e_t$$

```
> m3.hare<-arima(sqrt(hare),order=c(2,0,0))
> m3.hare

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

Coefficients:
         ar1      ar2  intercept
      1.3514  -0.7763     5.7134
```

```
s.e.   0.1286   0.1242     0.4753

sigma^2 estimated as 1.223:  log likelihood = -48.46,  aic = 102.91
> HatAcf<-acf(residuals(m3.hare),main='野兔丰度AR(2)模型残差的样本ACF')$acf
Error : The fig.showtext code chunk option must be TRUE
```

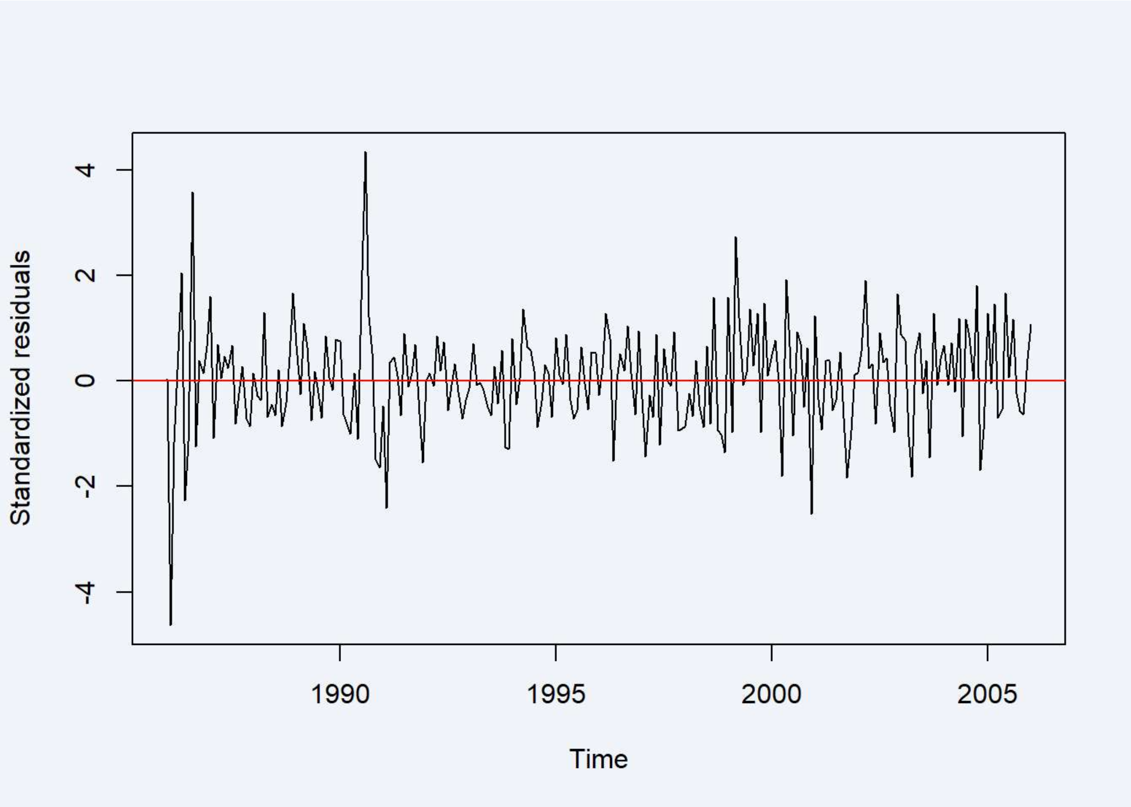



```
> HatAcf<-round(HatAcf,4)
>
> plot(HatAcf,type='h',  main='AR(2)模型残差的样本ACF', xlim=c(1,15),ylim=c(-0.4,0.4),xlab='滞后k', ylab='残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=0)
>
> HatPhi_1<-m3.hare$coef[1]; HatPhi_1
     ar1
1.351401
> HatPhi_2<-m3.hare$coef[2]; HatPhi_2
       ar2
-0.7762724
> n<-m3.hare$nobs; n
[1] 31
> std1<-round(abs(HatPhi_2)/sqrt(n),4); std1 ##\hat(sd (\hat r_1))
   ar2
0.1394
> std2<-round(sqrt(HatPhi_2^2+HatPhi_1^2*(1+HatPhi_2)^2)/sqrt(n),4); std2 ##\hat(sd (\hat r_2))
   ar2
0.1496
> k=1:15
> bd<-c(std1,std2,rep(round(1/sqrt(n),4),13));bd
   ar2    ar2
0.1394 0.1496 0.1796 0.1796 0.1796 0.1796 0.1796 0.1796 0.1796 0.1796 0.1796

0.1796 0.1796 0.1796 0.1796
> par(new=TRUE)
> plot(x=k, y=-2*bd, type='o', cex=0.5, col='blue', xlim=c(1,15), ylim=c(-0.4,0.4), axes=FALSE, xlab='', ylab='')
Error : The fig.showtext code chunk option must be TRUE
> par(new=TRUE)
> plot(x=k, y=2*bd, type='o', cex=0.5, col='blue', xlim=c(1,15), ylim=c(-0.4,0.4), axes=FALSE, xlab='', ylab='')
Error : The fig.showtext code chunk option must be TRUE
```

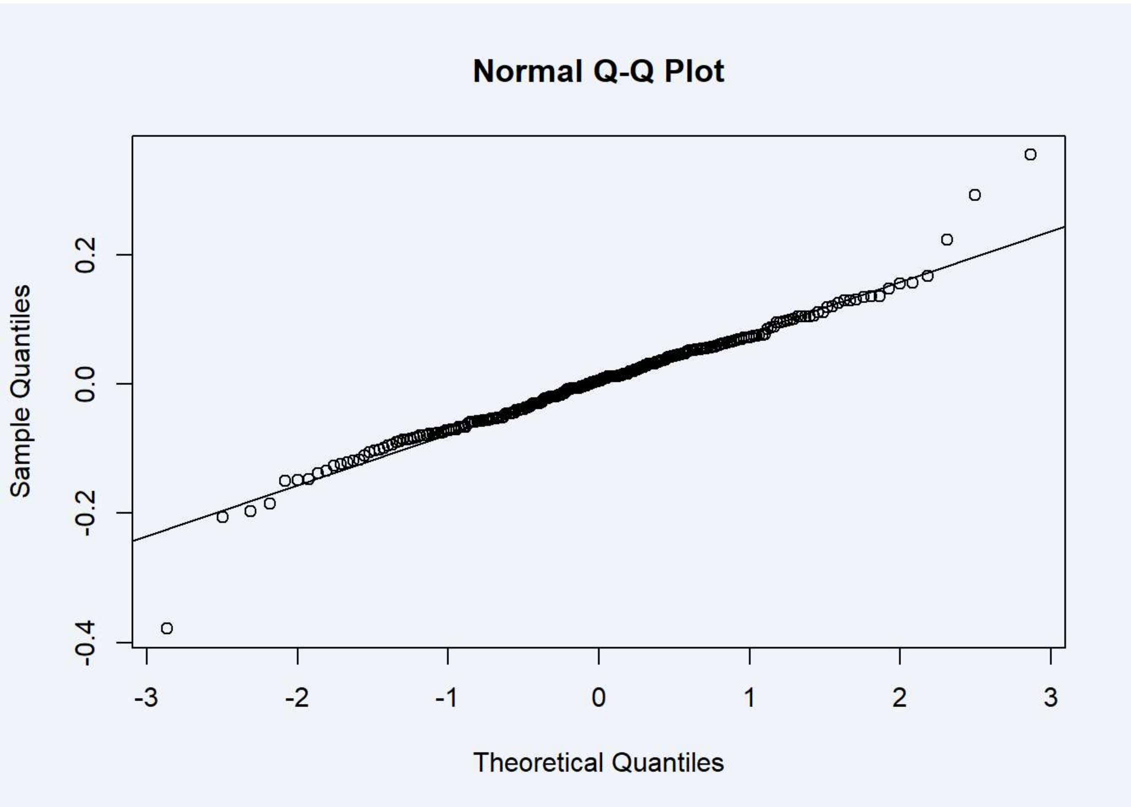


石油价格序列

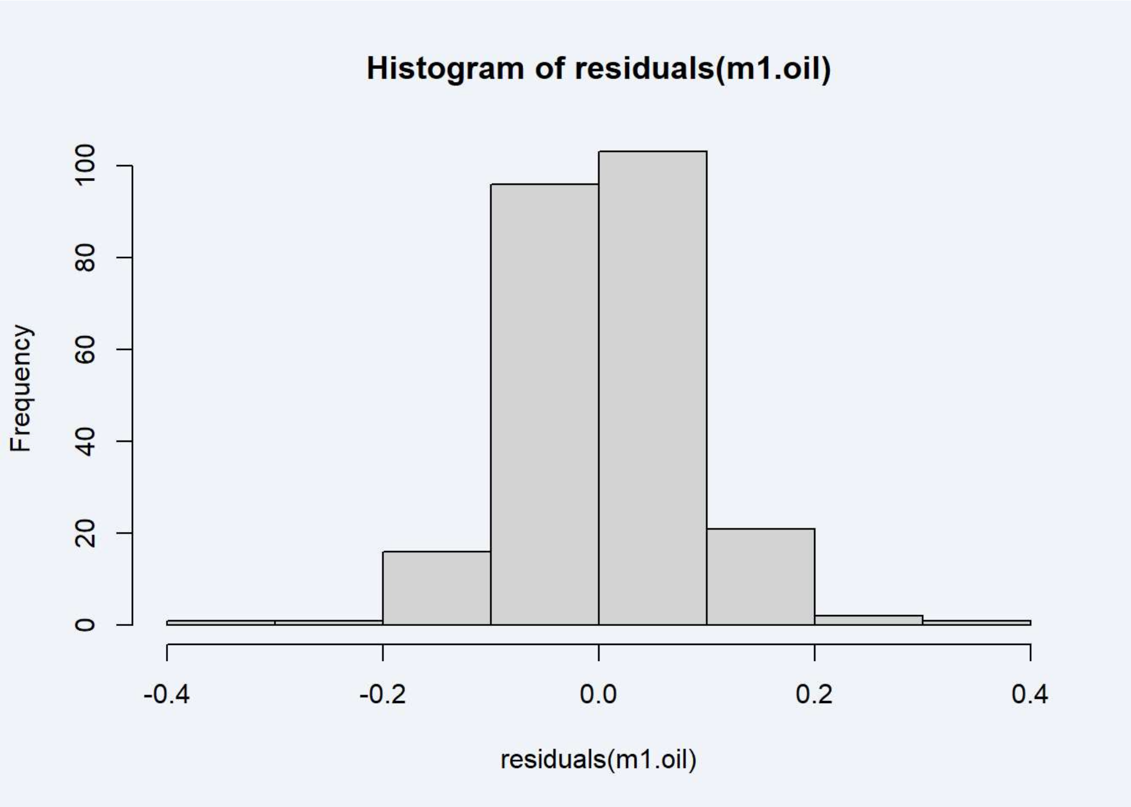
```
> library(TSA)
> data(oil.price)
> m1.oil=arima(log(oil.price),order=c(0,1,1)) ###IMA (1,1)
> m1.oil

Call:
arima(x = log(oil.price), order = c(0, 1, 1))

Coefficients:
         ma1
      0.2956
s.e.  0.0693

sigma^2 estimated as 0.006689:  log likelihood = 260.29,  aic = -518.58
>
> plot(rstandard(m1.oil),ylab='Standardized residuals',type='l')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=0,col='red')
```



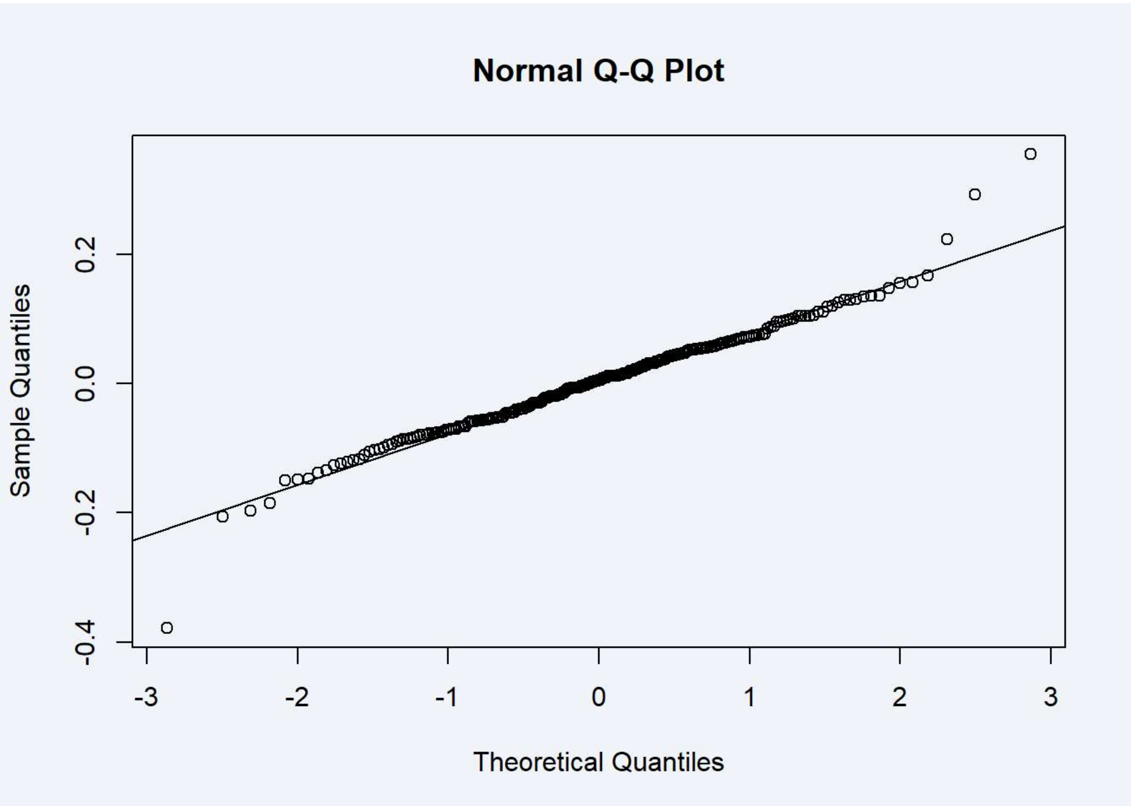

```
>
> ###正态性检验
> qqnorm(residuals(m1.oil))
Error : The fig.showtext code chunk option must be TRUE
> qqline(residuals(m1.oil))
```



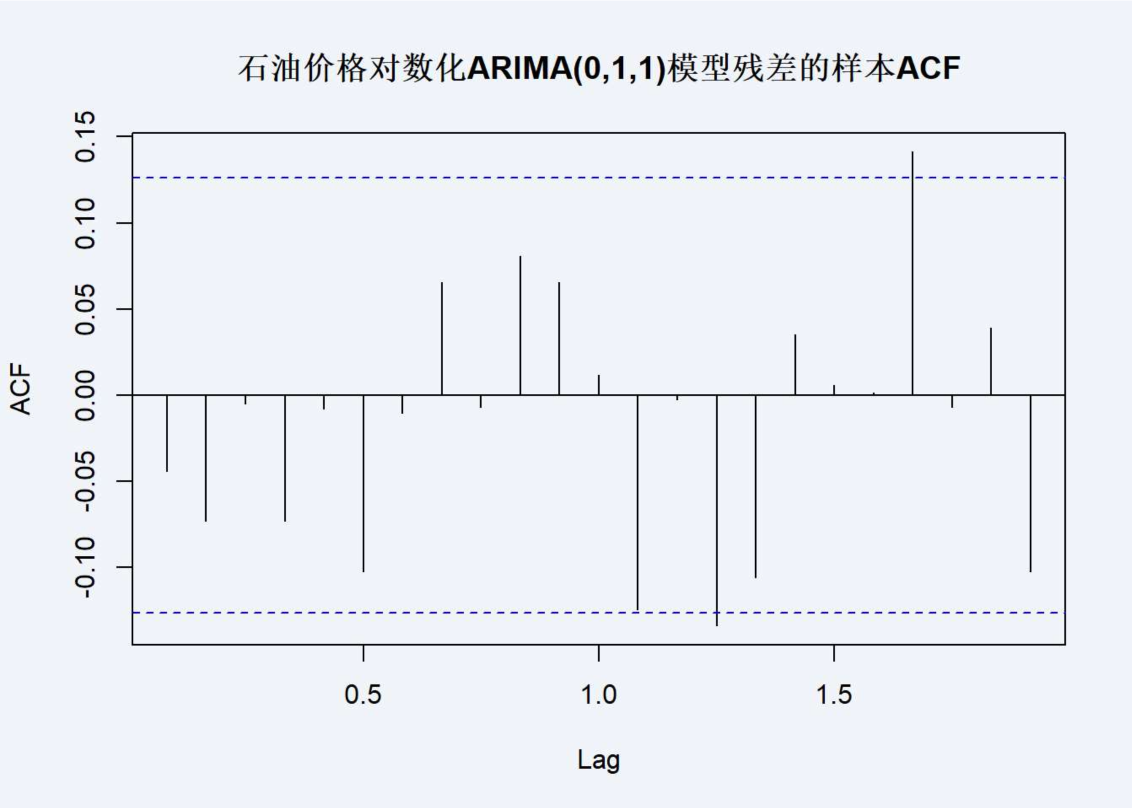

```
> shapiro.test(residuals(m1.oil)) ##H_0:正态性

    Shapiro-Wilk normality test

data:  residuals(m1.oil)
W = 0.96883, p-value = 3.937e-05
>
> ###残差的正态性检验
> hist(residuals(m1.oil))
Error : The fig.showtext code chunk option must be TRUE
```



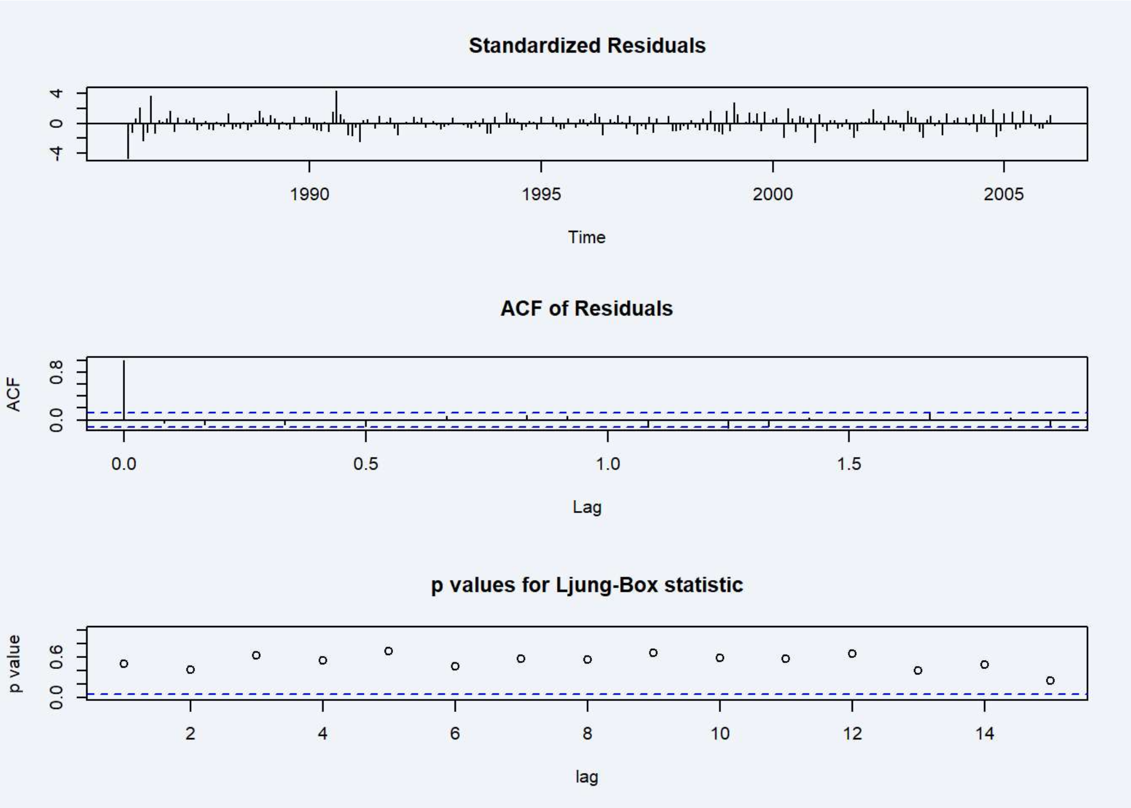

```
> qqnorm(residuals(m1.oil))
Error : The fig.showtext code chunk option must be TRUE
> qqline(residuals(m1.oil))
```




```
> shapiro.test(residuals(m1.oil)) ##H_0:正态性

    Shapiro-Wilk normality test

data:  residuals(m1.oil)
W = 0.96883, p-value = 3.937e-05
>
> ###残差的相关性检验
> acf(residuals(m1.oil),main='石油价格对数化ARIMA(0,1,1)模型残差的样本ACF')
Error : The fig.showtext code chunk option must be TRUE
```




```
>
> HatAcf<-signif(acf(residuals(m1.oil),plot=F)$acf[1:6],2);HatAcf
[1] -0.0440 -0.0730 -0.0050 -0.0730 -0.0078 -0.1000
> # to display the first 6 acf to 2 significant digits.
>
> #Ljung-Box检验 H_0: r_1=r_2=...=r_lag=0, 不相关
> LB.test(m1.oil, lag=6) #LB-test the residuals of m1.color

    Box-Ljung test

data:  residuals from  m1.oil
X-squared = 5.6847, df = 5, p-value = 0.3381
> #LB.test(model, lag = 12, type = c("Ljung-Box", "Box-Pierce")
>
> tsdiag(m1.oil,gof=15,omit.initial=F) # Exhibit 8.12
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
Error : The fig.showtext code chunk option must be TRUE
```




```
> # the tsdiag function is modified from that in the stats package of R.
>
> runs(residuals(m1.oil))  ##游程检验 H_0: 独立性
$pvalue
[1] 0.341

$observed.runs
[1] 129

$expected.runs
[1] 121.1494

$n1
[1] 114

$n2
[1] 127

$k
[1] 0
> #runs(rstandard(m2.hare))  ##非参数检验，标准化没有影响
```

对原始数据进行对数变换，然后以IMA(1，1)模型进行拟合，即

$$\nabla \log Y_t - \mu = e_t - \theta e_{t-1}$$

极大似然估计的结果表明，AIC最小的是零均值的IMA(1，1)模型。

$$\log Y_t = \log Y_{t-1} + e_t + 0.2956 * e_{t-1}.$$

案例分析

- 残差的自相关性：残差的自相关函数记为$\hat{r}_k$, 对于较大的n,

  * 白噪声的样本自相关函数$r_k \sim AN(0, \frac{1}{n}), Corr(r_k, r_j) \approx 0.$

  * AR(1)模型残差 $\hat{e}_t = Y_t - \hat{\phi} Y_{t-1}$ 的自相关函数

$$\begin{cases} Var(\hat{r}_1) \approx \frac{\phi^2}{n} \\ Var(\hat{r}_k) \approx \frac{1-(1-\phi^2)\phi^{2k-2}}{n}, \quad k>1 \\ Corr(\hat{r}_1,\hat{r}_k) \approx -sign(\phi)\frac{(1-\phi^2)\phi^{k-2}}{1-(1-\phi^2)\phi^{2k-2}} \end{cases}$$

图表

表8-1：标准差

表8-1：AR(1)模型残差自相关函数$\sqrt{nVar(\hat{r}_k)}$的逼近

k	$\phi=0.3$	$\phi=0.5$	$\phi=0.7$	$\phi=0.9$
1	0.3	0.5	0.7	0.9
2	0.96	0.9	0.87	0.92
3	1	0.98	0.94	0.94
4	1	0.99	0.97	0.95
5	1	1	0.99	0.96
6	1	1	0.99	0.97
7	1	1	1	0.97
8	1	1	1	0.98
9	1	1	1	0.98

表8-2：相关系数

表8-2：AR(1)模型残差自相关函数$Corr(\hat{r}_1,\hat{r}_k)$的逼近

k	$\phi=0.3$	$\phi=0.5$	$\phi=0.7$	$\phi=0.9$
1	1	1	1	1
2	-0.99	-0.92	-0.68	-0.22
3	-0.28	-0.39	-0.41	-0.2
4	-0.08	-0.19	-0.27	-0.17
5	-0.02	-0.09	-0.18	-0.15
6	-0.01	-0.05	-0.12	-0.13
7	0	-0.02	-0.09	-0.12
8	0	-0.01	-0.06	-0.11
9	0	-0.01	-0.04	-0.09

图表

* AR(2)模型残差 $\hat{e}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2}$ 的自相关函数

$$\begin{cases} Var(\hat{r}_1) \approx \frac{\phi_2^2}{n} \\ Var(\hat{r}_2) \approx \frac{\phi_2^2+\phi_1^2(1+\phi_2)^2}{n} \\ Var(\hat{r}_k) \approx \frac{1}{n}, \quad k\geq 3 \end{cases}$$

- 自相关检验

  残差的自相关函数记为$r_1, r_2, \cdots$, 自相关检验问题

  * 原假设 $H_0: r_1 = r_2 = \cdots = r_K = 0$

  * 备择假设 $H_1:$ 至少存在某个$r_k \neq 0, 1 \leq k \leq K$

* Box-Pierce检验:

$$Q=n(\hat{r}_1^2+\hat{r}_2^2+\cdots+\hat{r}_K^2)\overset{ARMA(p,q)}{\backsim_{appr.}}\chi^2(_{K-p-q})$$

* Ljung-Box检验:

$$Q_*=n(n+2)(\frac{\hat{r}_1^2}{n-1}+\frac{\hat{r}_2^2}{n-2}+\cdots+\frac{\hat{r}_K^2}{n-K})\overset{ARMA(p,q)}{\backsim_{appr.}}\chi^2(_{K-p-q})$$

过度拟合和参数冗余

- 过度拟合

识别并拟合出一个初步合适的模型之后，在此基础上，寻找一些更一般的包含初始模型的扩展模型，然后进行比较，**检查额外的参数是否显著地不为零，共同的参数估计是否有显著的改变。**

* 显著性t检验 $H_0:\beta_j=0 \qquad \text{vs.} \qquad H_1:\beta_j\neq 0$

- 统计量 $t_j=\frac{\hat{\beta}_j}{\widehat{sd(\hat{\beta}_j)}}\overset{H_0}{\sim}t(n-m), \quad n\text{样本容量},m\text{参数个数}$

- 拒绝域 $\left\{t_j:|t_j|\geq t_{\alpha/2}(n-m)\right\}$

- p-值 $\Pr(|t_{n-m}|\geq t_j|\beta_j=0)$

- 参数冗余

考虑下面的ARMA(2,3)模型

$$Y_t-(\phi+c)Y_{t-1}+\phi cY_{t-2}=e_t-(\theta_1+c)e_{t-1}-(\theta_2-c\theta_1)e_{t-2}+c\theta_2e_{t-3}$$

- AR参数多项式 $1-(\phi+c)x+\phi cx^2=(1-\phi x)(1-cx)$

- MA参数多项式 $1-(\theta_1+c)x-(\theta_2-c\theta_1)x^2+c\theta_2x^3=(1-\theta_1x-\theta_2x^2)(1-cx)$

$$(1-\phi B)(1-cB)Y_t=(1-\theta_1B-\theta_2B^2)(1-cB)e_t$$

ARMA(2,3)模型中的参数c是可以取任意值的，是不可识别的。等价于ARMA(1,2)模型

$$Y_t-\phi Y_{t-1}=e_t-\theta_1e_{t-1}-\theta_2e_{t-2}$$

- 案例分析（化工颜色属性序列）

Model	AIC	LogLik	Sig2	AR_1	AR_2	MA_1	MU
AR(1)	216.1471	-106.0735	24.83407	0.5705478	NA	NA	74.32928
AR(2)	217.8428	-105.9214	24.59941	0.5173004	0.1004908	NA	74.15508
ARMA(1,1)	217.8847	-105.9423	24.63363	0.6720801	NA	-0.1467323	74.17298
ARMA(2,1)	219.8202	-105.9101	24.57902	0.2188729	0.2735330	0.3036481	74.16528

```
Call:
arima(x = color, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.5705    74.3293
s.e.  0.1435     1.9151

sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15

Call:
```

```
arima(x = color, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.5173  0.1005    74.1551
s.e.  0.1717  0.1815     2.1463

sigma^2 estimated as 24.6:  log likelihood = -105.92,  aic = 217.84

Call:
arima(x = color, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.6721  -0.1467    74.1730
s.e.  0.2147   0.2742     2.1357

sigma^2 estimated as 24.63:  log likelihood = -105.94,  aic = 217.88

Call:
arima(x = color, order = c(2, 0, 1))

Coefficients:
         ar1     ar2     ma1  intercept
      0.2189  0.2735  0.3036    74.1653
s.e.  2.0056  1.1376  2.0650     2.1121

sigma^2 estimated as 24.58:  log likelihood = -105.91,  aic = 219.82
```

消除掉冗余参数之后，化工颜色属性序列的合适模型为AR(1)：

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$$

- 最小均方误差预测

基于序列可获得的直到时间t的历史数据，即$Y_1, Y_2, \cdots, Y_t$， 预测未来m期的值Y_{t+m}， 称时间t为**预测起点**，m为预测**前置时间**，而用$\widehat{Y}_t(m)$代表预测值。最小均方误差预测为

$$\widehat{Y}_t(m) = E(Y_{t+m}|Y_1, Y_2, \cdots, Y_t)$$

* 考虑AR(1)模型未来1期预测的问题:

$$Y_{t+1} - \mu = \phi(Y_t - \mu) + e_{t+1}$$

给定$Y_1, Y_2, \cdots, Y_t$，取条件期望得

$$\begin{aligned}\widehat{Y}_t(1) &= E(\mu|Y_1, \cdots, Y_t) + \phi E((Y_t - \mu)|Y_1, \cdots, Y_t) + E(e_{t+1}|Y_1, \cdots, Y_t) \\ &= \mu + \phi(Y_t - \mu)\end{aligned}$$

* 考虑AR(1)模型未来2期预测的问题:

$$Y_{t+2} - \mu = \phi(Y_{t+1} - \mu) + e_{t+2}$$

给定$Y_1, Y_2, \cdots, Y_t$，取条件期望得

$$\begin{aligned}\widehat{Y}_t(2) &= E(\mu|Y_1, \cdots, Y_t) + \phi E((Y_{t+1} - \mu)|Y_1, \cdots, Y_t) + E(e_{t+2}|Y_1, \cdots, Y_t) \\ &= \mu + \phi(\widehat{Y}_t(1) - \mu)\end{aligned}$$

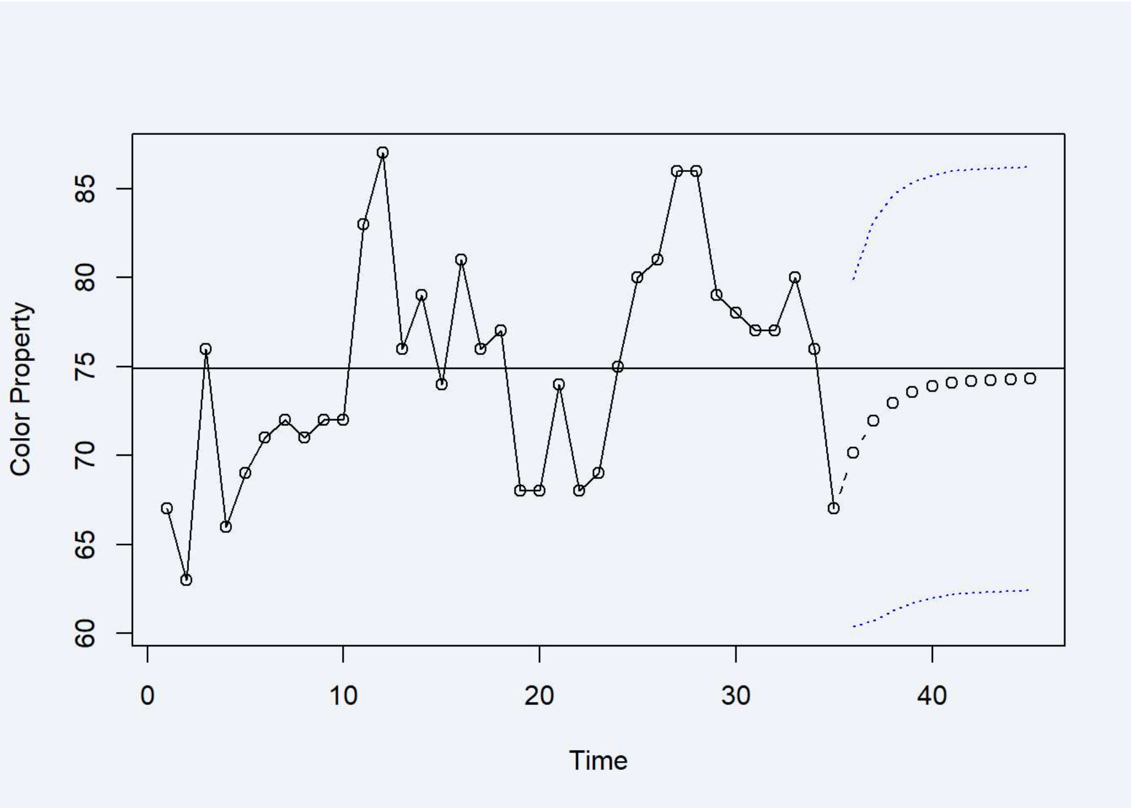
* 考虑AR(1)模型未来m期预测的问题:

$$Y_{t+m} - \mu = \phi(Y_{t+(m-1)} - \mu) + e_{t+m}$$

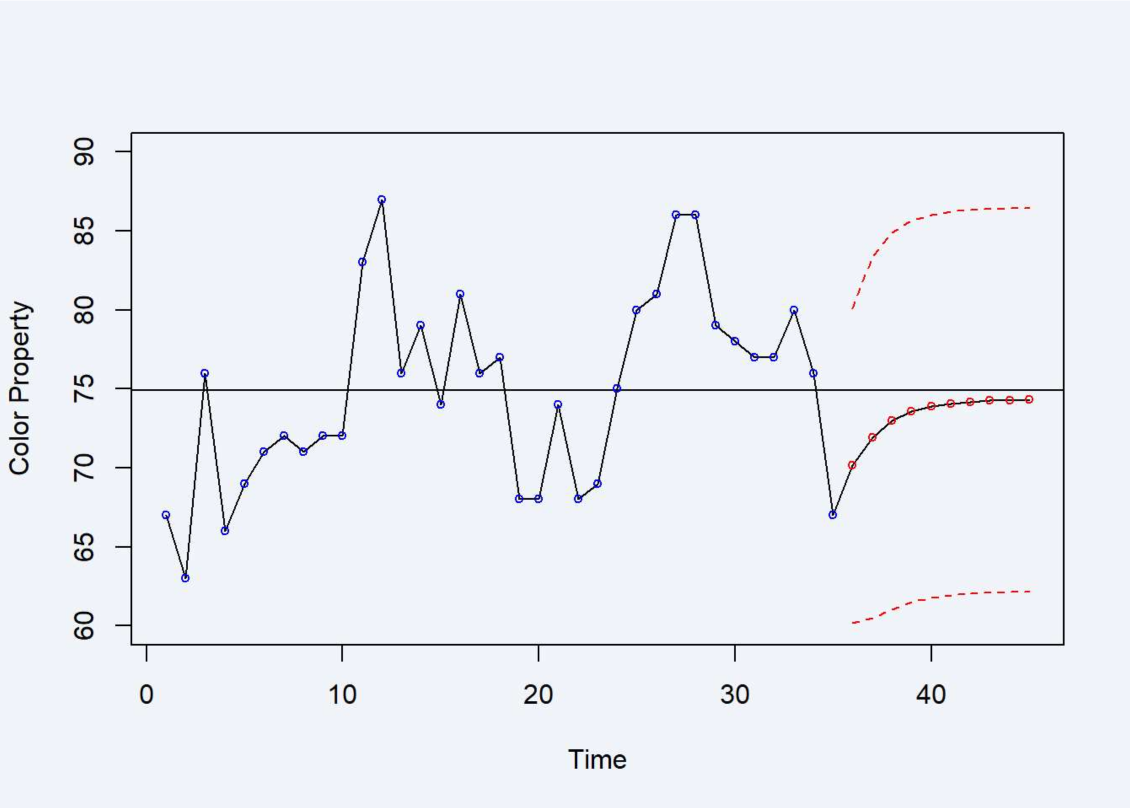
给定$Y_1, Y_2, \cdots, Y_t$，取条件期望得

$$\begin{aligned}\widehat{Y}_t(m) &= E(\mu|Y_1,\cdots,Y_t) + \phi E((Y_{t+(m-1)} - \mu)|Y_1,\cdots,Y_t) + E(e_{t+m}|Y_1,\cdots,Y_t)\\ &= \mu + \phi\left(\widehat{Y}_t(m-1) - \mu\right) = \mu + \phi\left[\left(\mu + \phi(\widehat{Y}_t(m-2) - \mu)\right) - \mu\right]\\ &= \mu + \phi^2\left(\widehat{Y}_t(m-2) - \mu\right) = \cdots\\ &= \mu + \phi^m\left(Y_t - \mu\right)\end{aligned}$$

```
> m1.color

Call:
arima(x = color, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.5705    74.3293
s.e.  0.1435     1.9151

sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15
> plot(m1.color, n.ahead=10,type='b',xlab='Time',ylab='Color Property', col='blue')
Error : The fig.showtext code chunk option must be TRUE
> abline(h=mean(color))
```




```
>
> m=10
> pre<-predict(m1.color,n.ahead=m)$pred
> se<-predict(m1.color,n.ahead=m)$se
> lowBd<-pre-2*se
> upBd<-pre+2*se
>
> plot(ts(c(color,pre)), type='l', cex=0.5, xlab='Time',ylab='Color Property', ylim=c(60,90))
Error : The fig.showtext code chunk option must be TRUE
> abline(h=mean(color))
> points(color, col='blue',cex=0.7)
> points(pre, col='red',cex=0.7)
> lines(x=time(pre),y=upBd, col='red', type='l', lty=2)
> lines(x=time(pre),y=lowBd, col='red', type='l', lty=2)
```



- 总结：

1. 小心识别一个初始模型；
2. 扩展模型时，不要同时增加AR和MA部分的阶数；
3. 按照残差分析的建议进行模型扩展；
4. 利用显著性检验识别冗余参数，消除过度拟合问题。
