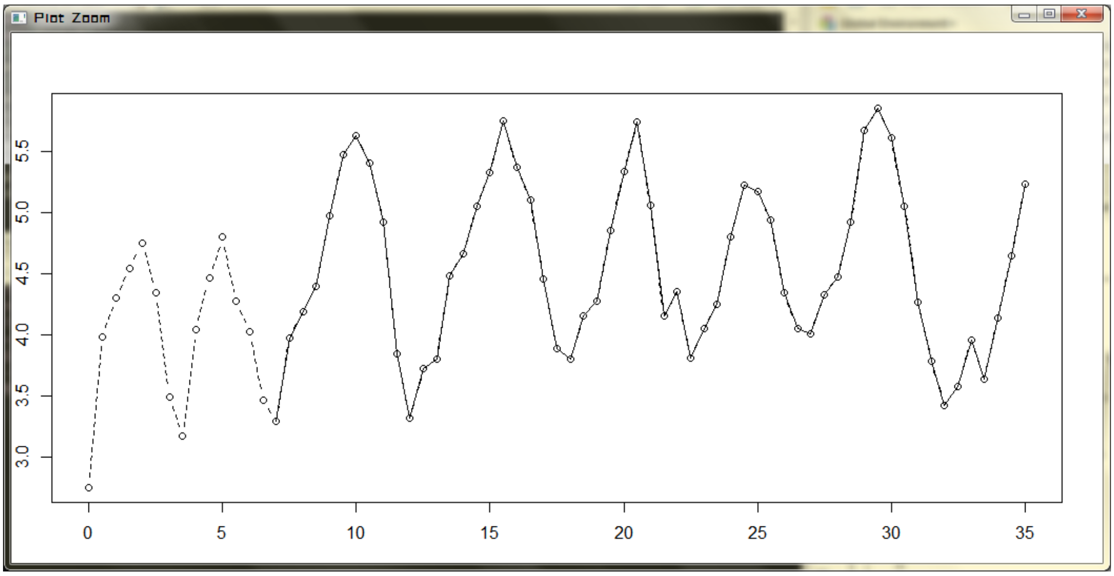
《时间序列分析及应用》

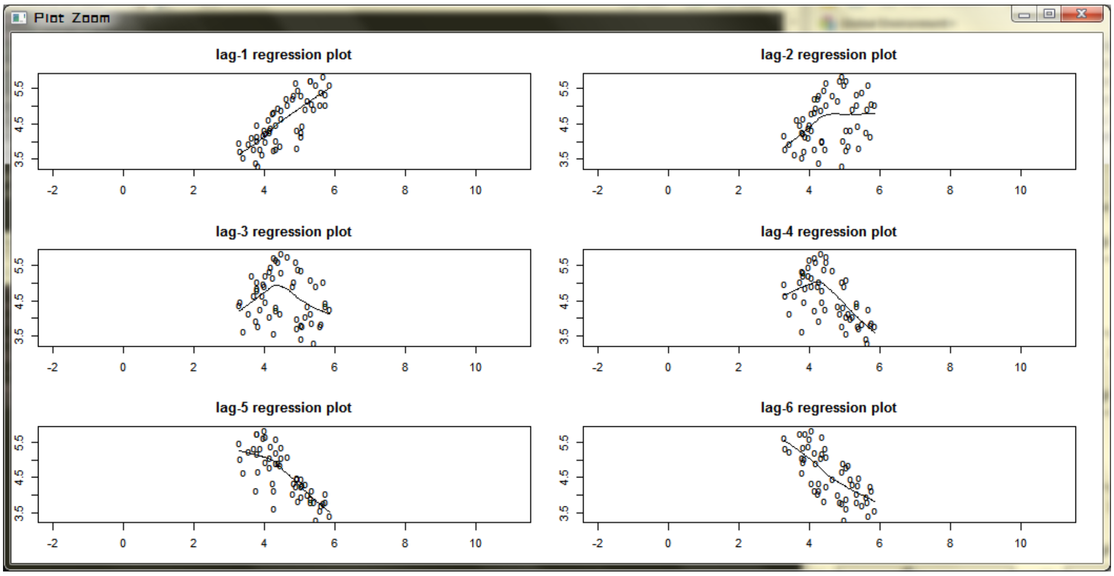
第 15 章课后作业：P298——15.1,15.2,15.3；15.6,15.7,15.8

15.1 对捕食者序列拟合一个延迟设定为 2 的 TAR 模型，并用 Stenseth 等（1998，1999）给出的框架解释结果（读者也许想先行检验他们的框架对 TAR 模型来说是否近似有效）另外，比较拟合模型与教材中延迟为 3 的 TAR（2；1，4）模型。（数据文件名为 veilleux）

```
> library(TSA)
> library(fGarch)
> data(veilleux)
> predator<-veilleux[,1]
> plot(log(predator),lty=2,type='b',xlab='Day',ylab='Log(predator)')
> predator.eq=window(predator,start=c(7,1))
> lines(log(predator.eq))
```

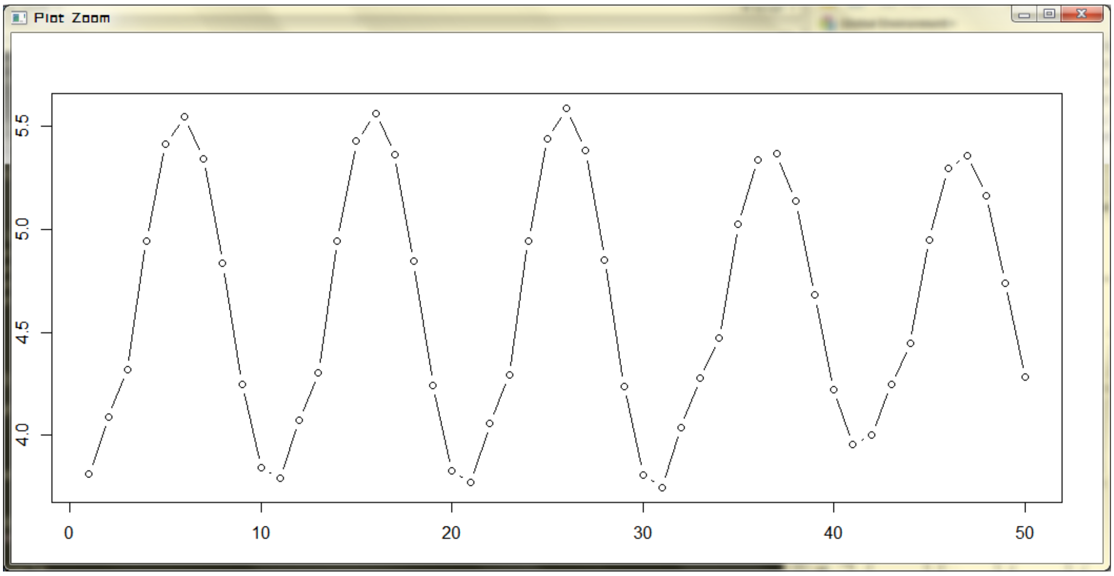



```
> lagplot(log(predator.eq))
```



```
> par(mfcol=c(1,1))
> #d=2
> predator.tar.3=tar(log(predator.eq),p1=4,p2=4,d=2,a=.1,b=.9,print=T)
time series included in this analysis is:  log(predator.eq)
SETAR(2, 1 , 4 ) model delay = 2
estimated threshold =  4.048  from a Minimum AIC  fit with thresholds
searched from the  17  percentile to the   81  percentile of all data.
The estimated threshold is the  26.4  percentile of
all data.
lower regime:
Residual Standard Error=0.2445
R-Square=0.997
F-statistic (df=2, 12)=2023.065
p-value=0

                           Estimate Std.Err t-value Pr(>|t|)
intercept-log(predator.eq)   0.9526  0.7856  1.2125   0.2487
lag1-log(predator.eq)        0.8230  0.2016  4.0817   0.0015



 (unbiased) RMS
0.05979
 with no of data falling in the regime being
log(predator.eq) 14


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
log(predator.eq) 0.05125


 upper regime:
Residual Standard Error=0.2526
R-Square=0.9976
F-statistic (df=5, 34)=2830.15
p-value=0
```

```
                           Estimate Std.Err t-value Pr(>|t|)
intercept-log(predator.eq)   4.0579  0.5716  7.0990   0.0000
lag1-log(predator.eq)        0.9122  0.1432  6.3692   0.0000
lag2-log(predator.eq)       -0.2616  0.2108 -1.2410   0.2231
lag3-log(predator.eq)       -0.1989  0.2024 -0.9825   0.3328
lag4-log(predator.eq)       -0.3178  0.1486 -2.1383   0.0398



 (unbiased) RMS
0.06382
 with no of data falling in the regime being
39


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
0.05564

 Nominal AIC is  12.15
> tar.skeleton(predator.tar.3,n=50)

 no limit cycle

 tail part of the skeleton:
 [1] 3.81 4.09 4.32 4.94 5.41 5.54 5.34 4.83 4.25 3.84 3.79 4.07 4.31 4.94 5.43 5.56 5.36 4.84 4.24
[20] 3.83 3.77 4.06 4.29 4.94 5.44 5.58 5.38 4.85 4.24 3.81 3.75 4.04 4.28 4.47 5.02 5.34 5.36 5.13
[39] 4.68 4.22 3.96 4.00 4.25 4.45 4.95 5.29 5.36 5.16 4.74 4.28
```



数据的初始部分由于瞬态效应貌似是非平稳的，因此从第 15 个数据开始使用延迟为 2 的 TAR（2；1，4）模型进行拟合。从结果可以看到，门限的估计值是 4.048，大约是第 26 百分位数，因此两区域中的样本分布比延迟为 3 的模型中更不均匀。两区域中噪声方差估计值分别为 0.05979 和 0.06382，比延迟为 3 的模型中估计值更大。因此，延迟为 3 的 TAR（2；1，4）模型比延迟为 2 的 TAR（2；1，4）模型拟合效果好。

15.2 对开方变换后的太阳黑子数据拟合一个 TAR 模型，并检验拟合优度。解释该拟合 TAR 模型。（数据文件名为 spots）

```
> data(spots)
> MAIC<-NULL
> for(d in 1:5){spots.tar<-tar(y=sqrt(spots),p1=5,p2=5,d=d,a=.1,b=.9)
+               MAIC<-rbind(MAIC,c(d,spots.tar$AIC,signif(spots.tar$thd,5),
+                                  spots.tar$p1,spots.tar$p2))}
> colnames(MAIC)<-c('d','nominal AIC','r','p1','p2')
> MAIC
     d nominal AIC      r p1 p2
[1,] 1       149.9 5.8822  5  5
[2,] 2       110.5 6.0581  3  5
[3,] 3       124.6 6.5955  2  5
[4,] 4       126.2 8.0436  3  5
[5,] 5       150.5 8.1548  4  5
```

选取最大自回归阶数为 5，MAIC 方法选定了延迟为 2 的 TAR（2；3，5）模型

```
> spots.tar.1<-tar(y=sqrt(spots),p1=5,p2=5,d=2,a=.1,b=.9,print=T)
time series included in this analysis is:  sqrt(spots)
SETAR(2, 3 , 5 ) model delay = 2
estimated threshold =  6.058  from a Minimum AIC  fit with thresholds
searched from the  20  percentile to the   79  percentile of all data.
The estimated threshold is the  35.7  percentile of
all data.
lower regime:
Residual Standard Error=0.9061
R-Square=0.9894
F-statistic (df=4, 16)=373.5257
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(spots)   9.5242  1.2796  7.4432   0.0000
lag1-sqrt(spots)        1.0479  0.1699  6.1679   0.0000
lag2-sqrt(spots)       -1.2022  0.2909 -4.1319   0.0008
lag3-sqrt(spots)       -0.5555  0.2441 -2.2752   0.0370



 (unbiased) RMS
0.8211
 with no of data falling in the regime being
sqrt(spots) 20


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
sqrt(spots) 0.6569
```

```

 upper regime:
Residual Standard Error=0.4627
R-Square=0.9977
F-statistic (df=6, 30)=2149.376
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(spots)   5.6852  0.8608  6.6049   0.0000
lag1-sqrt(spots)        0.3711  0.1070  3.4677   0.0016
lag2-sqrt(spots)        0.3847  0.1183  3.2527   0.0028
lag3-sqrt(spots)       -0.0656  0.1113 -0.5893   0.5601
lag4-sqrt(spots)       -0.3634  0.1088 -3.3407   0.0022
lag5-sqrt(spots)       -0.1118  0.0744 -1.5016   0.1437



 (unbiased) RMS
0.2141
 with no of data falling in the regime being
36


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
0.1784

 Nominal AIC is  110.5
```

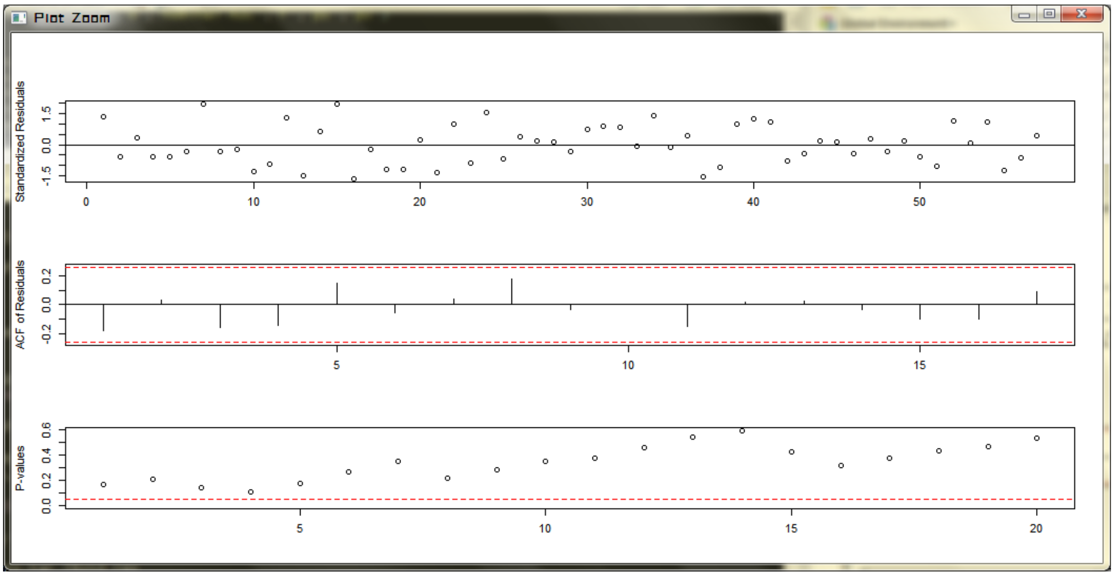
使用延迟为 2 的 TAR（2；3，5）模型拟合数据，其中上区域的 5 阶系数并没有通过参数检验

```
> spots.tar.2<-tar(y=sqrt(spots),p1=4,p2=4,d=2,a=.1,b=.9,print=T)
time series included in this analysis is:  sqrt(spots)
SETAR(2, 3 , 4 ) model delay = 2
estimated threshold =  6.058  from a Minimum AIC  fit with thresholds
searched from the  16  percentile to the   82  percentile of all data.
The estimated threshold is the  35.1  percentile of
all data.
lower regime:
Residual Standard Error=0.9061
R-Square=0.9894
F-statistic (df=4, 16)=373.5257
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(spots)   9.5242  1.2796  7.4432   0.0000
lag1-sqrt(spots)        1.0479  0.1699  6.1679   0.0000
lag2-sqrt(spots)       -1.2022  0.2909 -4.1319   0.0008
lag3-sqrt(spots)       -0.5555  0.2441 -2.2752   0.0370



 (unbiased) RMS
0.8211
 with no of data falling in the regime being
sqrt(spots) 20


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
sqrt(spots) 0.6569
```

```
 upper regime:
Residual Standard Error=0.4804
R-Square=0.9975
F-statistic (df=5, 32)=2542.773
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(spots)   4.8070  0.6422  7.4847   0.0000
lag1-sqrt(spots)        0.4025  0.1025  3.9278   0.0004
lag2-sqrt(spots)        0.4166  0.1203  3.4618   0.0015
lag3-sqrt(spots)       -0.0122  0.1104 -0.1109   0.9124
lag4-sqrt(spots)       -0.4886  0.0844 -5.7925   0.0000



 (unbiased) RMS
0.2308
 with no of data falling in the regime being
37


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
0.1996

 Nominal AIC is  113.7
```

使用延迟为 2 的 TAR（2；3，4）模型重新拟合数据

```
> shapiro.test(spots.tar.2$std.res)

 Shapiro-Wilk normality test

data:  spots.tar.2$std.res
W = 0.9744, p-value = 0.2667

> Box.test(spots.tar.2$std.res,lag=20,type="Ljung")

 Box-Ljung test

data:  spots.tar.2$std.res
X-squared = 15.4738, df = 20, p-value = 0.7487

> tsdiag(spots.tar.2,gof.lag=20)
```

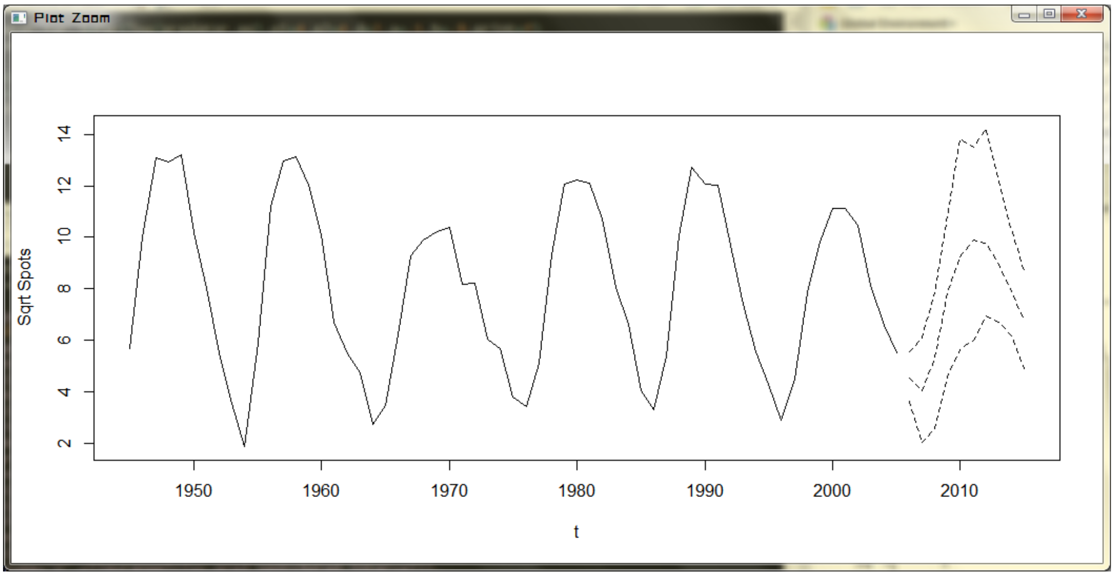


对模型进行检验，发现拟合残差为白噪声且为正态分布

15.3 使用习题 15.2 中的拟合模型预测未来 10 年里的年相对太阳黑子数。画出预测区间和预测的中位数（数据文件名为 spots）

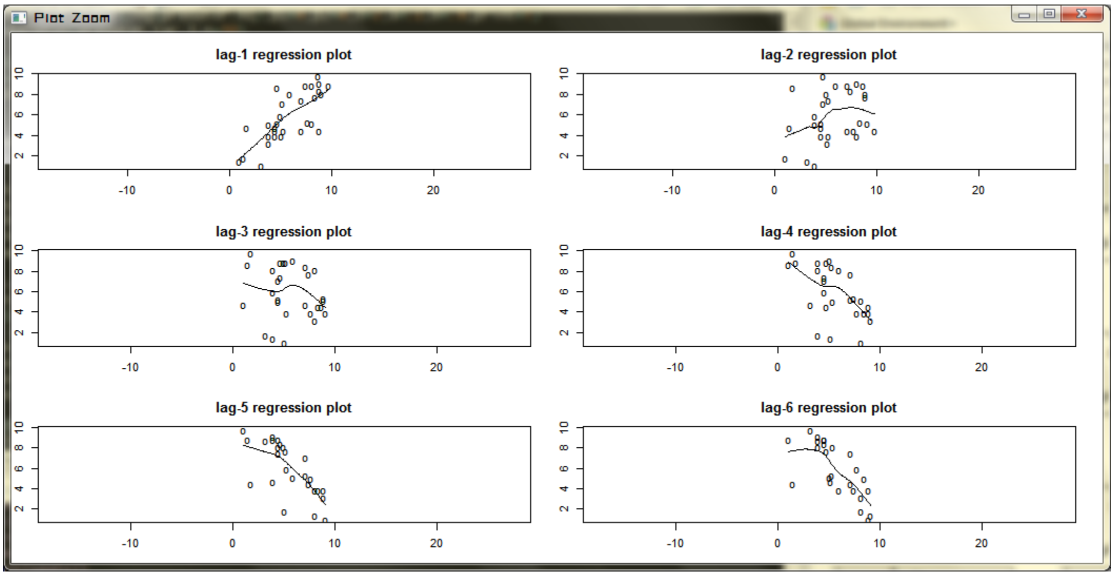
对未来 10 年的年相对太阳黑子数进行预测，并画图。

```
> par(mfcol=c(1,1))
> pred.spots<-predict(spots.tar.2,n.ahead=10,n.sim=1000)
> yy<-ts(c(sqrt(spots),pred.spots$fit),frequency=1,start=start(spots))
> plot(yy,type='n',ylim=range(c(yy,pred.spots$pred.interval)),
+      ylab='Sqrt Spots',xlab=expression(t))
> lines(sqrt(spots))
> lines(window(yy,start=end(spots)+c(0,1)),lty=2)
> lines(ts(pred.spots$pred.interval[2,],start=end(spots)+c(0,1),freq=1),lty=2)
> lines(ts(pred.spots$pred.interval[1,],start=end(spots)+c(0,1),freq=1),lty=2)
```



15.6 画出经开方变换的野兔数据的滞后回归图。有无迹象表明野兔数据是非线性的？（数据文件名为 hare）

```
> data(hare)
> lagplot(sqrt(hare))
```



从图中看出，2、3、6 阶滞后回归图显示出了很强的非线性

15.7 对野兔数据进行正式的非线性检验（Keenan 的检验、Tsay 的检验和门限似然比检验）野兔丰度过程是不是非线性的？解释你的结果（数据文件名为 hare）

```
> Keenan.test(sqrt(hare))
$test.stat
[1] 8.083568

$p.value
[1] 0.009207613

$order
[1] 3
```

Keenan 检验说明野兔丰度过程在 1%显著水平下是非线性的。

```
> Tsay.test(sqrt(hare))
$test.stat
[1] 2.135

$p.value
[1] 0.09923

$order
[1] 3
```

Tsay 检验说明野兔丰度过程在 10%的显著性水平下是非线性的。

```
> pvaluem<-NULL
> for(d in 1:5){
+     res<-tlrt(sqrt(hare),p=5,d=d,a=0.25,b=0.75)
+     pvaluem<-cbind(pvaluem,c(d,res$test.statistic,res$p.value))}
> rownames(pvaluem)<-c('d','test statistic','p-value')
> round(pvaluem,3)
                 [,1]   [,2]   [,3]   [,4]   [,5]
d               1.000  2.000  3.000  4.000  5.000
test statistic 71.558 54.964 16.807 39.359 41.657
p-value         0.000  0.000  0.083  0.000  0.000
>
```

门限似然比检验发现除延迟等于 3 的情况外，野兔丰度过程都是非线性的

15.8 假设野兔数据是非线性的，对其拟合一个 TAR 模型，并检验拟合优度（数据文件名为 hare）

```
> MAIC<-NULL
> for(d in 1:5){
+   hare.tar<-tar(y=sqrt(hare),p1=5,p2=5,d=d,a=.1,b=.9)
+   MAIC<-rbind(MAIC,c(d,hare.tar$AIC,signif(hare.tar$thd,3),hare.tar$p1,hare.tar$p2))}
> colnames(MAIC)<-c('d','nominal AIC','r','p1','p2')
> MAIC
     d nominal AIC    r p1 p2
[1,] 1       73.34 5.92  4  5
[2,] 2       76.22 5.29  4  4
[3,] 3       85.00 5.10  5  5
[4,] 4       80.85 5.10  5  5
[5,] 5       71.47 5.00  5  3
```

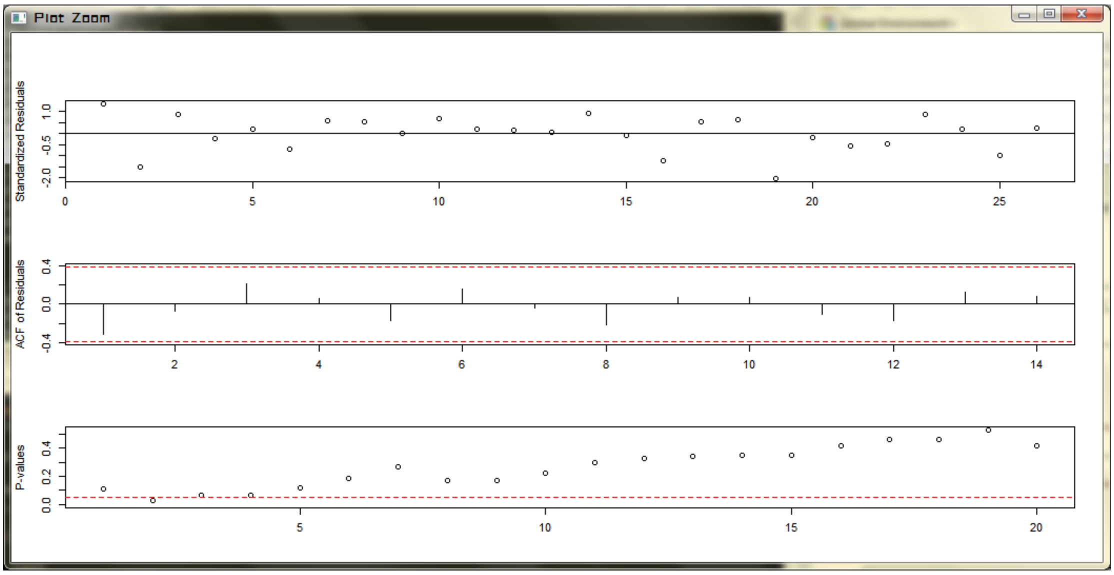
选取最大自回归阶数为 5，MAIC 方法选定了延迟为 5 的 TAR（2；5，3）模型

```
> hare.tar.1<-tar(y=sqrt(hare),p1=5,p2=5,d=5,a=.1,b=.9,print=T)
time series included in this analysis is:  sqrt(hare)
SETAR(2, 5 , 3 ) model delay = 5
estimated threshold =  5  from a Minimum AIC  fit with thresholds
searched from the  42  percentile to the   54  percentile of all data.
The estimated threshold is the  46.2  percentile of
all data.
lower regime:
Residual Standard Error=0.5612
R-Square=0.9976
F-statistic (df=6, 6)=413.3832
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(hare)    0.8673  0.9600  0.9035   0.4011
lag1-sqrt(hare)         1.0497  0.1763  5.9544   0.0010
lag2-sqrt(hare)         0.1153  0.2674  0.4314   0.6812
lag3-sqrt(hare)        -0.7305  0.2514 -2.9059   0.0271
lag4-sqrt(hare)        -0.5737  0.3358 -1.7088   0.1384
lag5-sqrt(hare)         1.2025  0.2632  4.5691   0.0038



 (unbiased) RMS
0.3149
 with no of data falling in the regime being
sqrt(hare) 12


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
sqrt(hare) 0.1575
```

```
 upper regime:
Residual Standard Error=1.0964
R-Square=0.9605
F-statistic (df=4, 10)=60.7245
p-value=0

                      Estimate Std.Err t-value Pr(>|t|)
intercept-sqrt(hare)    2.4250  1.0534  2.3021   0.0441
lag1-sqrt(hare)         1.0215  0.3385  3.0176   0.0129
lag2-sqrt(hare)         0.0162  0.4781  0.0338   0.9737
lag3-sqrt(hare)        -0.4920  0.3420 -1.4384   0.1809



 (unbiased) RMS
1.202
 with no of data falling in the regime being
14


 (max. likelihood) RMS for each series (denominator=sample size in the regime)
0.8586

 Nominal AIC is  71.47
```

使用延迟为 5 的 TAR（2；5，3）模型拟合数据，然后对模型进行检验

```
> shapiro.test(hare.tar.1$std.res)

 Shapiro-Wilk normality test

data:  hare.tar.1$std.res
W = 0.9491, p-value = 0.2215

> Box.test(hare.tar.1$std.res,lag=20,type="Ljung")

 Box-Ljung test

data:  hare.tar.1$std.res
X-squared = 19.4045, df = 20, p-value = 0.4957

> tsdiag(hare.tar.1,gof.lag=20)
```



拟合残差为白噪声且为正态分布