一. 导入数据

1.1. 从上海本地宝获取上海无症状感染者从 3.1 到 5.15 的数据记入 excel，读取 excel 数据并转换为时间序列。

```r
library(xlsxjars)
library(rJava)
library(xlsx)
#导入数据
mydata <- read.xlsx("D:\\MathModering\\RFunction\\上海新增3.1-5.13.xlsx",1)
no_sym <- mydata[,3]
no_sym=ts(no_sym,start=1)
```

图 1.1：读入数据

二. 拟合 ARIMA 模型前期准备

2.1. 数据平稳化

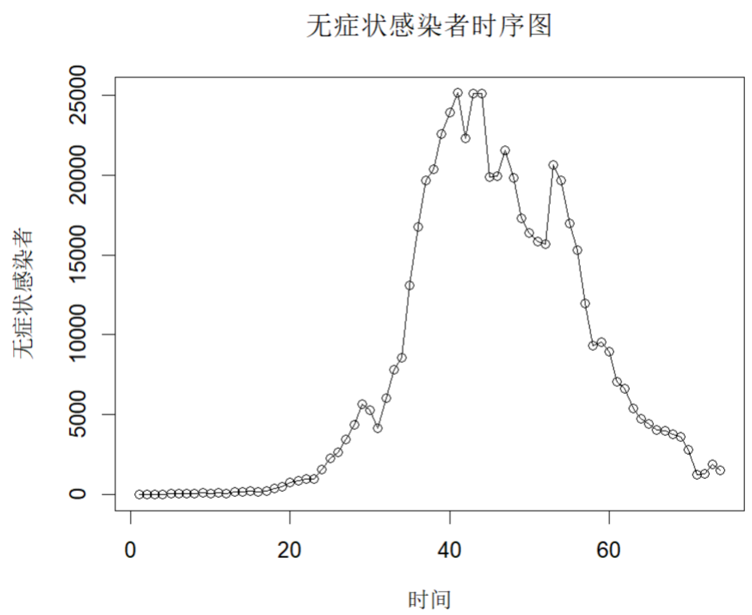

图 2.1.1：无症状感染者的时序图

根据无症状感染者的时序图，数据明显先增后减，具有不同的均值，因此无症状感染者时间序列为非平稳。考虑对数据进行差分运算。

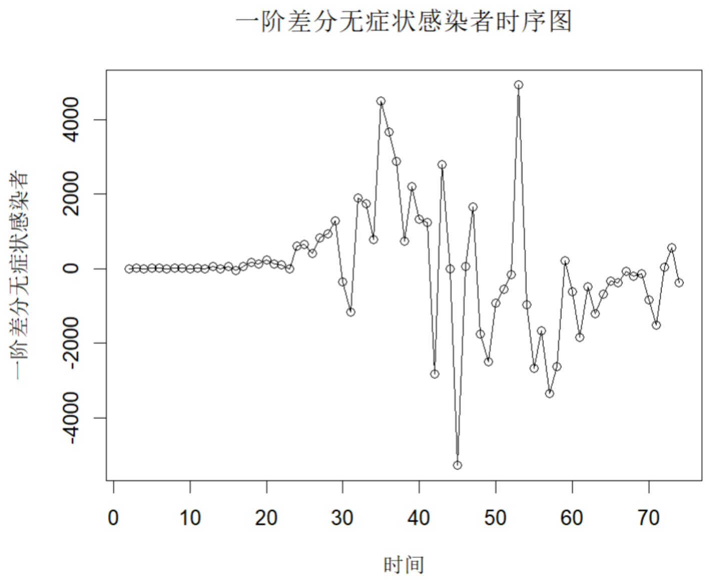

图 2.1.2：一阶差分无症状感染者的时序图

```
	Augmented Dickey-Fuller Test

data:  diff(no_sym)
Dickey-Fuller = -2.989, Lag order = 4, p-value = 0.1724
alternative hypothesis: stationary
```

图 2.1.3：一阶差分无症状感染者 ADF 检验

单纯通过时序图，难以区分是否平稳，进行单位根检验$p = 0.1724 < 0.05$，一阶差分的无症状感染者时间序列不平稳，进行二阶差分运算。

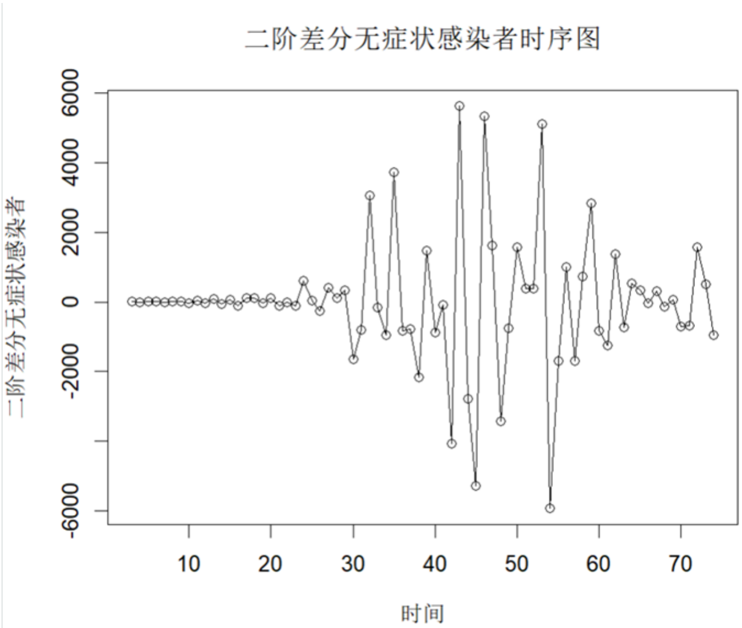

图 2.1.4：二阶差分无症状感染者的时序图

2.2. 白噪声检验

```
        Box-Pierce test

data:  df2
X-squared = 7.6598, df = 1, p-value = 0.005646


        Box-Pierce test

data:  df2
X-squared = 7.6598, df = 1, p-value = 0.005646


        Box-Pierce test

data:  df2
X-squared = 7.6598, df = 1, p-value = 0.005646
```

图 2.2.1：LB 检验

延迟 6,12,18 阶的统计量的 LB 统计量的 P 值$p = 0.005646 < 0.05$，二阶差分后的无症状感染者的时间序列不是白噪声序列，具有建模价值。

三. 拟合 ARIMA 模型

3.1. 通过自相关，偏相关函数对模型初步判断

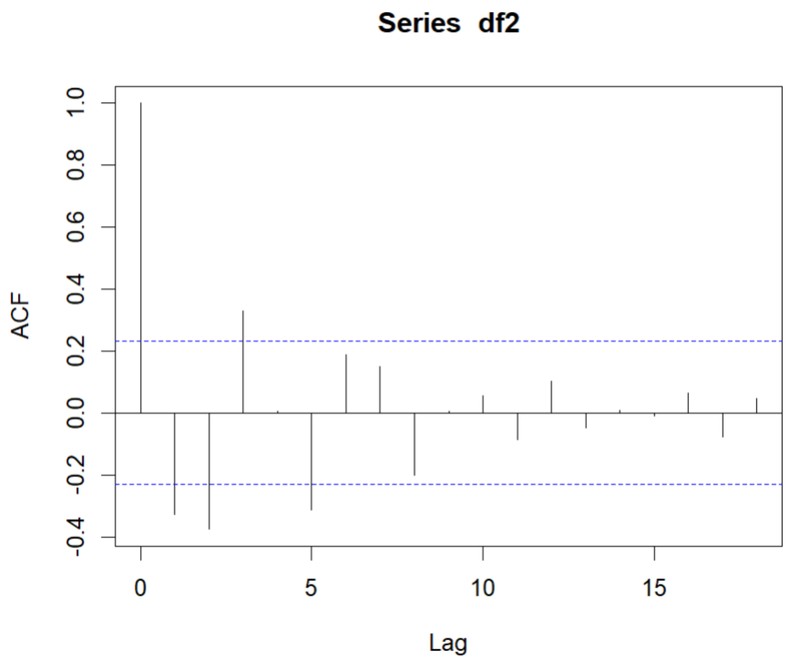

图 3.1.1：自相关函数

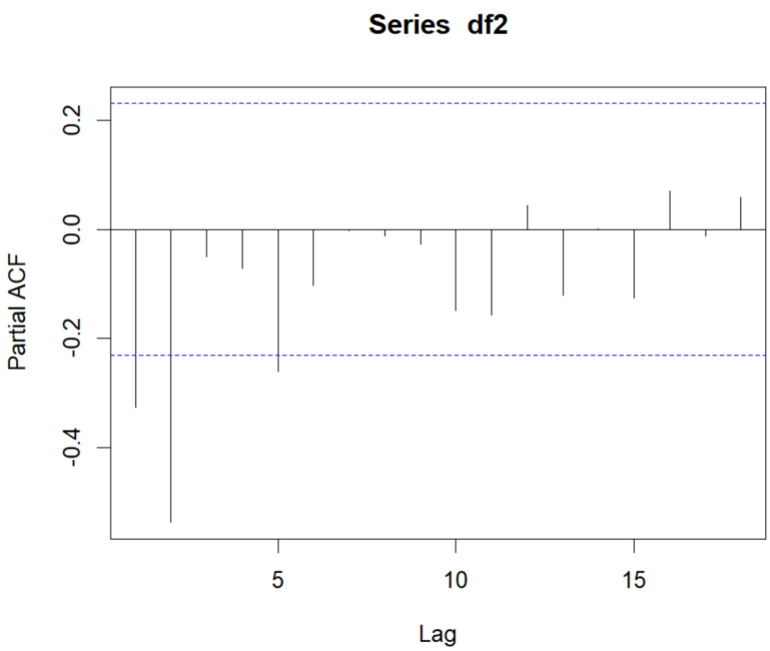

图 3.1.2：偏相关函数

根据自相关函数和偏自相关函数图，可以认为自相关函数 5 阶截尾，偏自相关函数 5 阶截尾，建立疏系数模型$ARIMA((1,1,0,0,1),2,(1,1,1,0,1))$，由于个别参数未通过显著性检验，对模型进行调整，最终建立了$ARIMA(2,2,(0,0,1))$模型。

```
Coefficients:
          ar1      ar2  ma1  ma2      ma3
      -0.6551  -0.7242    0    0  -0.3708
s.e.   0.1084   0.1099    0    0   0.1655

sigma^2 estimated as 2078527:  log likelihood = -626.44,  aic = 1258.88
```

模型： $\nabla^2 x_t = -0.6651 * x_{t-1} - 0.7242 * x_{t-2} + \varepsilon_t - 0.3708\varepsilon_{t-3}$

$$\varepsilon_t \sim N(0,2078527)$$

四. 对建立的$ARIMA$模型进行评估

4.1. 对残差进行白噪声检验

```
data:  model1$residual
X-squared = 2.0438, df = 6, p-value = 0.9156


        Box-Pierce test

data:  model1$residual
X-squared = 5.299, df = 12, p-value = 0.9472


        Box-Pierce test

data:  model1$residual
X-squared = 6.9157, df = 18, p-value = 0.9908
```

图 4.1.1：残差白噪声检验

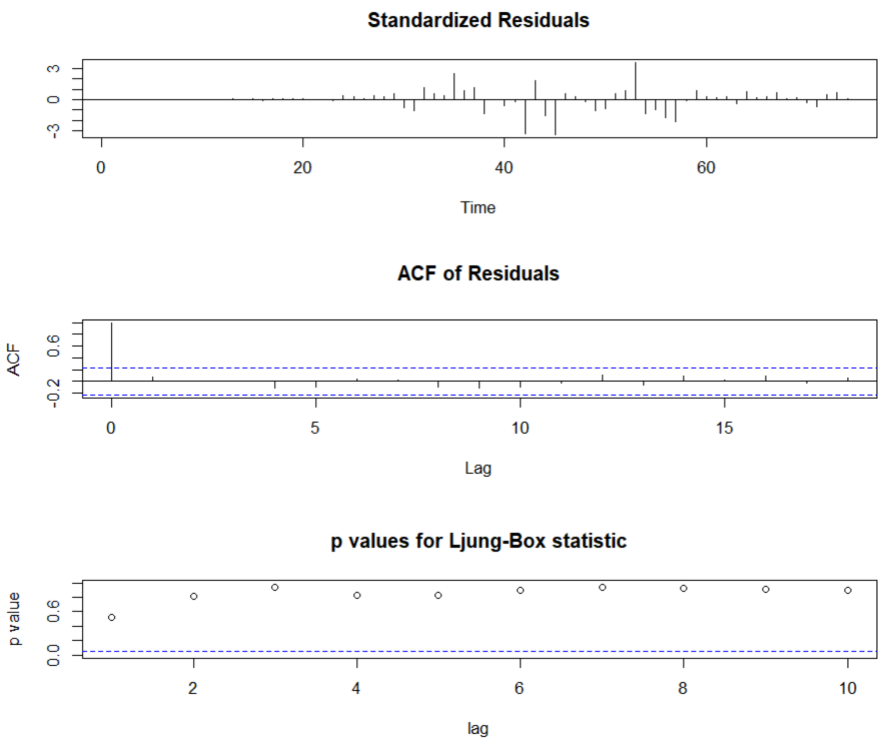

图 4.1.2：残差性质图

延迟 6，12，18 阶的残差的 LB 检验的 P 值都在 0.9 以上，显著>0.05，且残差的自相关函数没有显著的自相关性，因此残差是白噪声序列，观察标准化残差图，标准化残差存在中间大两侧小的特征，似乎存在波动聚集性，需进一步检验。

4.2 残差正态检验

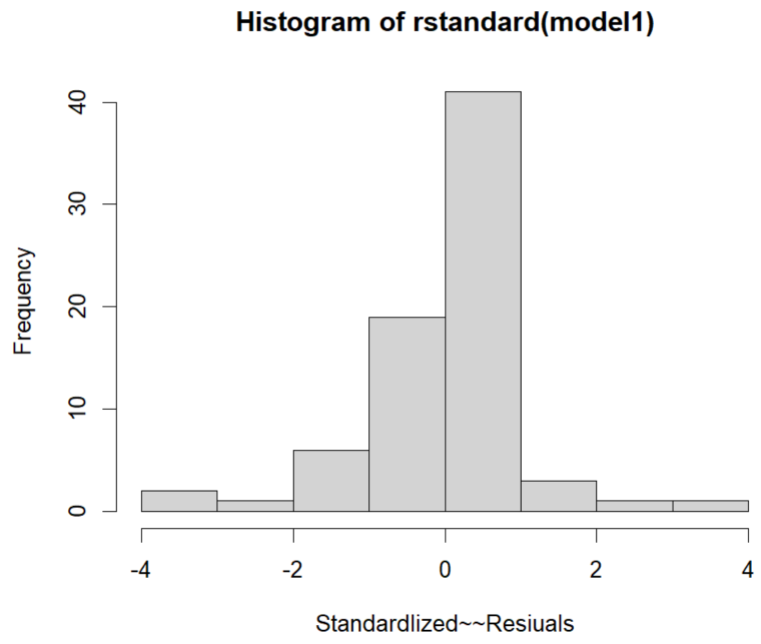

图 4.2.1：标准化残差直方图

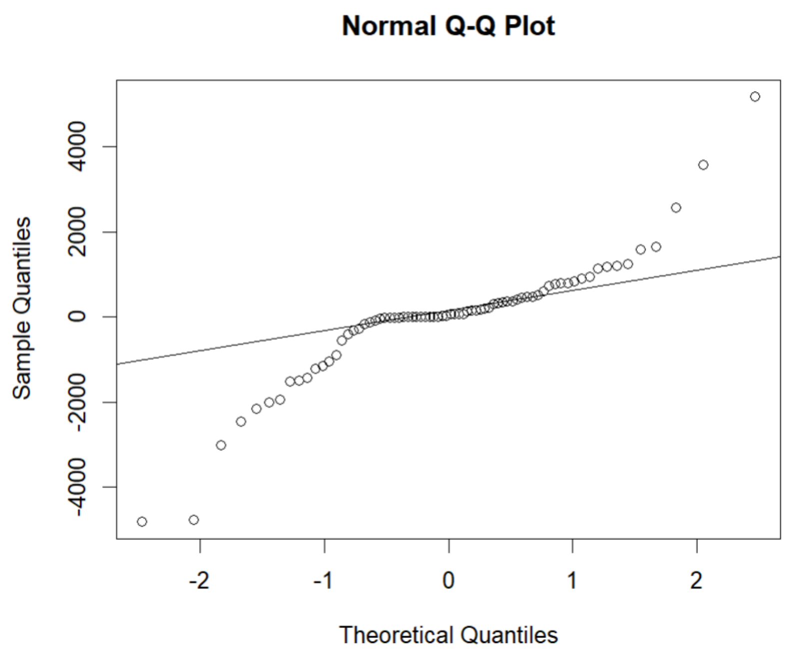

图 4.2.2：残差正态 QQ 图

标准化残差直方图呈现尖峰厚尾，为偏态分布，且 QQ 图显示两侧都明显偏离正态分布，初步判断该时间序列为非正态分布

```
> shapiro.test(residuals(model1))

	Shapiro-Wilk normality test

data:  residuals(model1)
W = 0.86009, p-value = 8.107e-07
```

图 4.2.3：残差正态检验

Shapiro-Wilk normality test 检验的 P 值显著小于 0.05，序列的残差不是正态分布，模型的残差似乎还蕴含更多信息，考虑进行 ARCH 效应检验与建模。

4.3. 参数显著性检验

```
> t1=-0.6551/0.1084
> pt(t1,df=69,lower.tail = T)
[1] 3.422543e-08
> t2=-0.7242/0.1099
> pt(t2,df=69,lower.tail = T)
[1] 3.655977e-09
> t3=-0.3708/0.1655
> pt(t3,df=69,lower.tail = T)
[1] 0.01414103
```

图 4.3.1：参数显著性检验

三个参数的 P 值都显著小于 0.05，因此三个参数都显著不为 0。

五. *ARIMA*模型预测

5.1. 模型预测

```
> x.fore=forecast(model1,h=5)
> x.fore
   Point Forecast        Lo 80    Hi 80     Lo 95     Hi 95
75      1077.7842   -769.8411 2925.410 -1747.915  3903.483
76      1037.4670 -2058.9628 4133.897 -3698.114  5773.048
77       721.9446 -3306.4900 4750.379 -5439.014  6882.904
78       319.5600 -4836.1792 5475.299 -7565.463  8204.583
79       173.3906 -6287.4641 6634.245 -9707.634 10054.415
```

图 5.1.1：未来 5 天无症状感染者预测

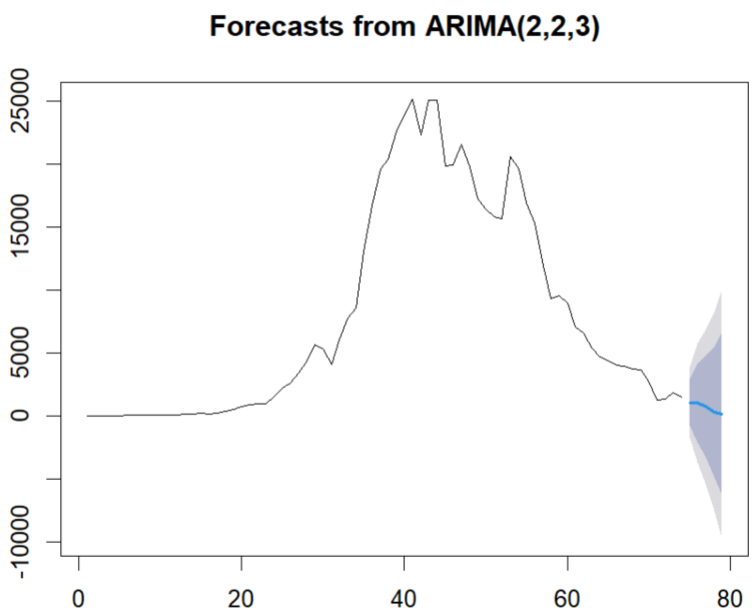

图 5.1.2：未来 5 天无症状感染者预测图

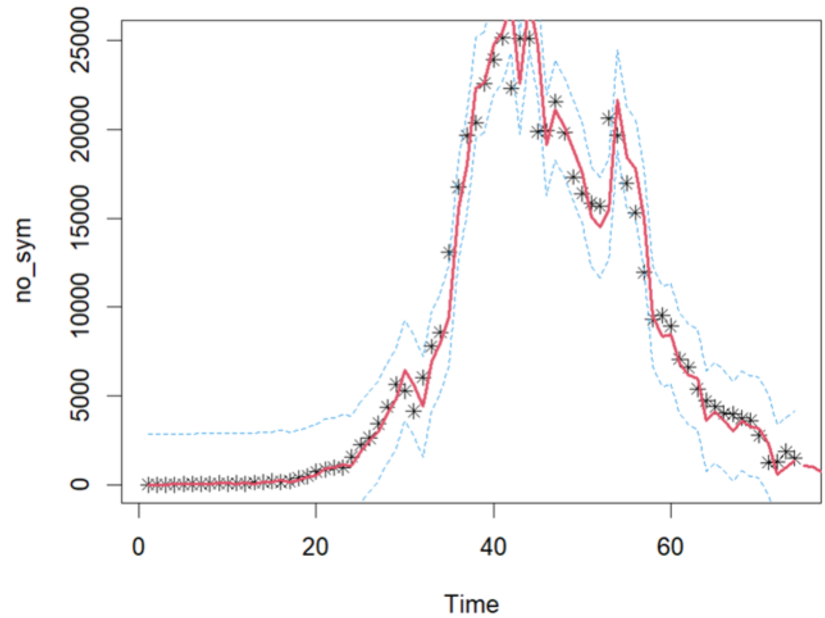

图 5.1.3：上下 95%置信区间

六. 对残差建立条件异方差模型

6.1 观察集群效应

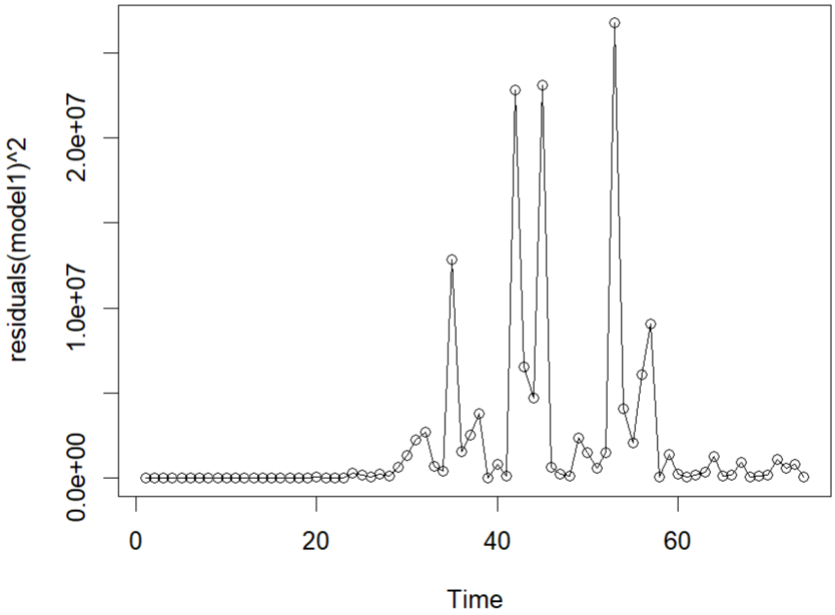

图 6.1.1：残差平方时序图

显然 30-60 时期的波动显著大于其它时期，因此猜测存在集群效应。

6.2. ARCH 效应检验

```
> for(i in 1:5)
+   print(Box.test(residuals(model1)^2,lag=i))

        Box-Pierce test

data:  residuals(model1)^2
X-squared = 1.6212, df = 1, p-value = 0.2029


        Box-Pierce test

data:  residuals(model1)^2
X-squared = 2.2397, df = 2, p-value = 0.3263


        Box-Pierce test

data:  residuals(model1)^2
X-squared = 9.2138, df = 3, p-value = 0.02658


        Box-Pierce test

data:  residuals(model1)^2
X-squared = 10.175, df = 4, p-value = 0.03758


        Box-Pierce test

data:  residuals(model1)^2
X-squared = 10.313, df = 5, p-value = 0.06683
```

图 6.2.1：Portmanteau Q 检验

延迟阶数为 3，4 时的 Portmanteau Q 检验的 P 值小于 0.05，因此残差序列存在短期的自相关性，存在 ARCH 效应

6.3. 确定条件异方差模型

```
Model:
GARCH(0,1)

Residuals:
     Min       1Q   Median       3Q      Max 
-3.38461 -0.09363  0.04481  0.31416  3.20665 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 1.947e+06   3.063e+05    6.357 2.05e-10 ***
a1 4.513e-01   2.735e-01    1.650    0.099 .  
```

图 6.3.1：ARCH(1)模型

```
GARCH(0,2)

Residuals:
     Min       1Q   Median       3Q      Max
-3.47959 -0.12307  0.04877  0.33075  3.31730

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.845e+06   2.862e+05    6.445 1.15e-10 ***
a1 4.043e-01   2.666e-01    1.517    0.129
a2 1.260e-08   1.694e-01    0.000    1.000
```

图 6.3.2：ARCH(2)模型

```
Model:
GARCH(0,3)

Residuals:
     Min       1Q   Median       3Q      Max
-3.60248 -0.09272  0.05300  0.31498  3.22208

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.742e+06   3.034e+05    5.744 9.27e-09 ***
a1 1.635e-01   3.797e-01    0.431    0.667
a2 1.324e-08   2.353e-01    0.000    1.000
a3 4.083e-01   3.571e-01    1.143    0.253
```

图 6.3.3：ARCH(3)模型

```
Model:
GARCH(0,4)

Residuals:
     Min       1Q   Median       3Q      Max
-2.86455 -0.09696  0.05488  0.33730  2.94802

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.640e+06   5.981e+05    2.742  0.00611 **
a1 1.418e-01   4.694e-01    0.302  0.76262
a2 2.611e-10   2.825e-01    0.000  1.00000
a3 3.685e-01   4.125e-01    0.893  0.37163
a4 2.997e-01   2.421e-01    1.238  0.21560
```

图 6.3.4：ARCH(4)模型

除了 ARCH(1),其他模型都有参数不显著且接近 0，所以选择 ARCH(1)模型。

```
Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 97.136, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 0.010776, df = 1, p-value = 0.9173
```

图 6.3.5：ARCH(1)模型统计量

Jarque Bera 检验表明 ARCH(1)模型的残差非正态，而 Box-Ljung 检验结果表明残差的平方是白噪声序列。

七． 对序列建立条件异方差模型

7.1 检验集群效应

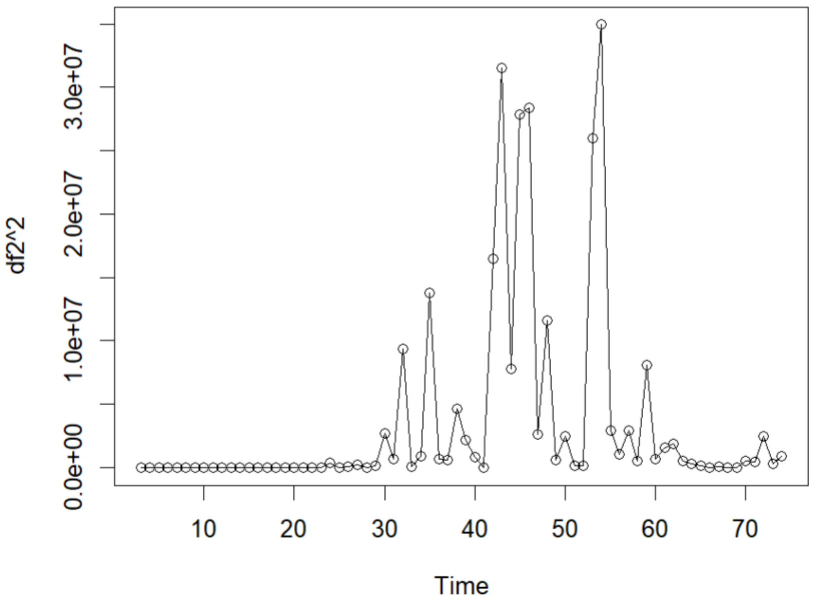

图 7.1.1：平方时序图

观察平方时序图可知，序列在中间部分的波动很大，显著的大于两侧，因此猜测存在集群效应

7.2 ARCH 检验

```
> for(i in 1:5)
+   print(Box.test(df2^2,lag=i))

        Box-Pierce test

data:  df2^2
X-squared = 15.39, df = 1, p-value = 8.745e-05


        Box-Pierce test

data:  df2^2
X-squared = 18.888, df = 2, p-value = 7.915e-05


        Box-Pierce test

data:  df2^2
X-squared = 24.917, df = 3, p-value = 1.607e-05


        Box-Pierce test

data:  df2^2
X-squared = 25.122, df = 4, p-value = 4.754e-05


        Box-Pierce test

data:  df2^2
X-squared = 25.836, df = 5, p-value = 9.603e-05
```

图 7.2.1：LM 检验


```
> for(i in 1:5)
+   print(Box.test(df2^2,lag=i))

        Box-Pierce test

data:  df2^2
X-squared = 15.39, df = 1, p-value = 8.745e-05


        Box-Pierce test

data:  df2^2
X-squared = 18.888, df = 2, p-value = 7.915e-05


        Box-Pierce test

data:  df2^2
X-squared = 24.917, df = 3, p-value = 1.607e-05


        Box-Pierce test

data:  df2^2
X-squared = 25.122, df = 4, p-value = 4.754e-05


        Box-Pierce test

data:  df2^2
X-squared = 25.836, df = 5, p-value = 9.603e-05
```

图 7.2.2：Portmanteau Q 检验

LM 检验与 Portmanteau Q 检验显示 1 阶至 5 阶 ARCH 模型均显著成立。

7.3 拟合 ARCH 模型

```
Model:
GARCH(0,1)

Residuals:
      Min        1Q    Median        3Q       Max
-2.205468 -0.261547 -0.003806  0.195383  2.715504

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 3.382e+06   7.164e+05    4.721 2.35e-06 ***
a1 9.795e-01   7.552e-01    1.297    0.195
```

图 7.3.1：ARCH(1)模型

```
Model:
GARCH(0,2)

Residuals:
      Min        1Q    Median        3Q       Max
-2.266420 -0.301858 -0.002235  0.205798  2.800585

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 3.204e+06   6.480e+05    4.944 7.64e-07 ***
a1 7.547e-01   6.532e-01    1.155    0.248
a2 4.790e-09   1.937e-01    0.000    1.000
```

图 7.3.2：ARCH(2)模型

```
Model:
GARCH(0,3)

Residuals:
      Min        1Q    Median        3Q       Max
-2.190327 -0.367226 -0.006323  0.214279  2.703823

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 3.026e+06   8.254e+05    3.666 0.000246 ***
a1 4.418e-01   3.839e-01    1.151 0.249792
a2 1.585e-08   2.696e-01    0.000 1.000000
a3 1.879e-01   3.313e-01    0.567 0.570682
```

图 7.3.3：ARCH(3)模型

```
Model:
GARCH(0,4)

Residuals:
      Min        1Q    Median        3Q       Max 
-2.294299 -0.394908 -0.008885  0.225259  2.869105 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 2.848e+06   7.160e+05    3.978 6.95e-05 ***
a1 1.965e-01   1.269e-01    1.549    0.121    
a2 5.612e-02   1.975e-01    0.284    0.776    
a3 1.104e-01   2.366e-01    0.466    0.641    
a4 8.076e-09   1.796e-01    0.000    1.000    
```

图 7.3.4：ARCH(4)模型

```
GARCH(0,5)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.34263 -0.38867 -0.01162  0.23539  2.55088 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)   
a0 2.670e+06   8.420e+05    3.171  0.00152 **
a1 2.131e-01   1.421e-01    1.499  0.13375   
a2 5.289e-02   2.084e-01    0.254  0.79968   
a3 1.097e-01   2.663e-01    0.412  0.68042   
a4 5.771e-08   2.089e-01    0.000  1.00000   
a5 8.728e-02   1.352e-01    0.646  0.51854   
```

图 7.3.5：ARCH(5)模型

虽然以上模型都有参数不显著，但考虑到除了 ARCH(1)模型外的其它 ARCH 模型都有接近 0 的参数，所以选择 ARCH1 模型。

7.4. ARCH 模型预测

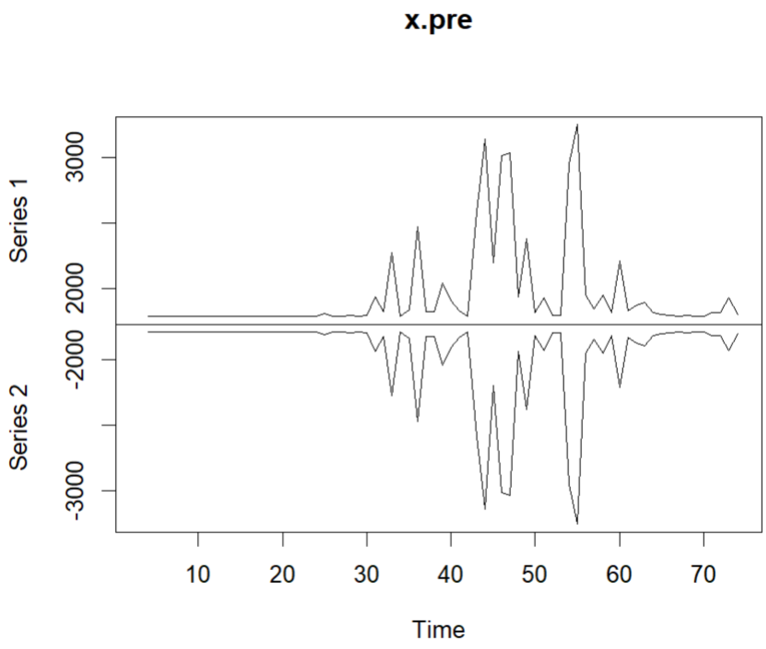

图 7.4.1：ARCH(1)模型 95%置信区间

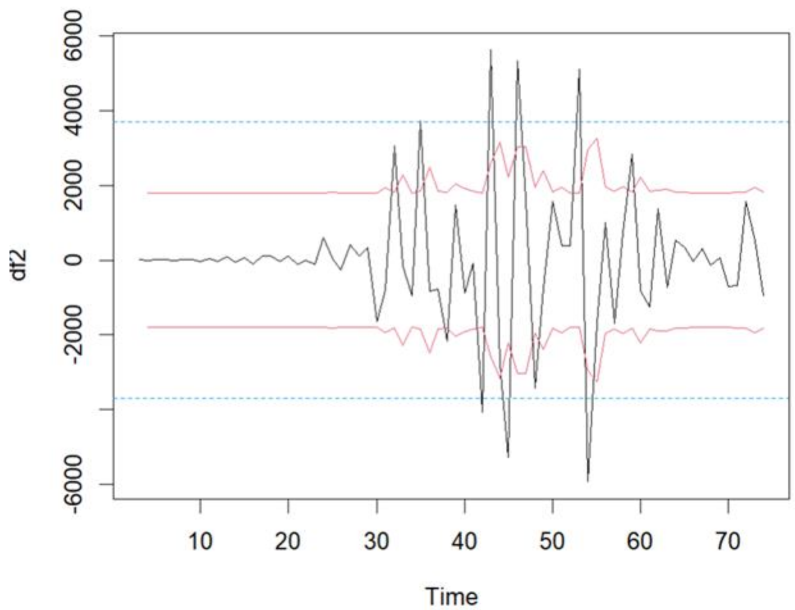

图 7.4.2：条件异方差与方差齐性置信区间比较

据上图可知，条件异方差模型拟合的置信区间比无条件方差两条平行线给出的 95%的置信区间更加符合原序列的真实波动情况

附录：

```r
library(tseries)
library(forecast)
library(lmtest)
library(zoo)
library(FinTS)
library(xlsxjars)
library(rJava)
library(xlsx)
#导入数据
mydata <- read.xlsx("D:\\MathModering\\RFunction\\上海新增 3.1-5.13.xlsx",1)
no_sym <- mydata[,3]
no_sym=ts(no_sym,start=1)
#画图观察,显然非平稳
plot(no_sym,type='o',ylab=expression(无症状感染者),xlab=expression(时间),main=expression(无症状感染者时序图))
#一阶差分作图观察，单位根检验(p=0.1724>0.05),不平稳
plot(diff(no_sym),type='o',ylab=expression(一阶差分无症状感染者),xlab=expression(时间),main=expression(一阶差分无症状感染者时序图))
adf.test(diff(no_sym))
#二阶差分，单位根检验(p=0.01<0.05),平稳
plot(diff(diff(no_sym)),type='o',ylab=expression(二阶差分无症状感染者),xlab=expression(时间),main=expression(二阶差分无症状感染者时序图))
adf.test(diff(diff(no_sym)))
df2=diff(diff(no_sym))
#白噪声检验(p=0.005646<0.05),不是白噪声序列
for(i in 1:3)
  print(Box.test(df2),lag=6*i,type='Ljung-Box')
#拟合模型
acf(df2)
pacf(df2)
model1=arima(no_sym,order=c(2,2,3),transform.pars=F,
             fixed=c(NA,NA,0,0,NA))

#画出残差图
plot(model1$residual,type='o')
abline(h=0)
#对残差进行白噪声检验(p=0.95),通过检验，是白噪声
for(i in 1:3)
  print(Box.test(model1$residual,lag=6*i))
#正态 QQ 图
hist(rstandard(model1),xlab='Standardlized~~Resiuals')
qqnorm(residuals(model1))
```

```
qqline(residuals(model1))
#残差性质图
tsdiag(model1)
#Portmanteau Q 检验
for(i in 1:3)
  print(Box.test(residuals(model1),lag=i))
#对残差建立 GARCH 模型
res.fit <- garch(model1$residuals,oreder=c(0,1))
#预测并画出预测图
x.fore=forecast(model1,h=5)
x.fore
plot(x.fore)
no_sym.fore=forecast(model1,h=5)
L1=no_sym.fore$fitted-1.96*sqrt(model1$sigma2)
U1=no_sym.fore$fitted+1.96*sqrt(model1$sigma2)
L2=ts(no_sym.fore$lower[,2],start=1)
U2=ts(no_sym.fore$upper[,2],start=1)
c1=min(no_sym,L1,L2)
c2=max(no_sym,L1,L2)
plot(no_sym,type='p',pch=8)
lines(no_sym.fore$fitted,col=2,lwd=2)
lines(no_sym.fore$mean,col=2,lwd=2)
lines(L1,col=4,lty=2)
lines(U1,col=4,lty=2)
lines(L2,col=4,lty=2)
lines(U2,col=4,lty=2)
#对参数进行显著性检验
t1=-0.6551/0.1084
pt(t1,df=69,lower.tail = T)
t2=-0.7242/0.1099
pt(t2,df=69,lower.tail = T)
t3=-0.3708/0.1655
pt(t3,df=69,lower.tail = T)
#残差分析

#残差相关性质图
win.graph(width=6.5,height=6)
tsdiag(model1)
#残差分布
hist(residuals(model1))
#正态检验
qqnorm(residuals(model1))
qqline(residuals(model1))
#聚集效应
```

```
plot(df2^2,type='o')
#LM 检验
for (i in 1:5)
  print(ArchTest(df2),lag=i)
#Portmanteau Q 检验
for(i in 1:5)
  print(Box.test(df2^2,lag=i))


#建立条件异方差模型
model2=auto.arima(no_sym)
for(i in 1:5)
  print(Box.test(residuals(model1)^2,lag=i))
for (i in 1:3)
  print(ArchTest(residuals(model1),lag=i))
eacf(abs(residuals(model2)))
eacf(abs(residuals(model1)^2))
plot(residuals(model1)^2,type='o')
r.fit1 <- garch(model1$residuals,order=c(0,1))
summary(r.fit1)

r.fit2 <- garch(model1$residuals,order=c(0,2))
summary(r.fit2)


r.fit3 <- garch(model1$residuals,order=c(0,3))
summary(r.fit3)

r.fit4 <- garch(model1$residuals,order=c(0,4))
summary(r.fit4)

r.fit5 <- garch(model1$residuals,order=c(0,2))
summary(r.fit5)

plot(residuals(r.fit5),type='o')
acf(residuals(r.fit5)^2,na.action=na.omit)
gBox(r.fit5,method='squared')

eacf(abs(diff(diff(no_sym))))
model2=garch(x=diff(no_sym),order=c(0,3))
qqnorm(residuals(model2))
qqline(residuals(model2))
shapiro.test(na.omit(residuals(model2)))
#Arch 模型
```

```
df2.fit1 <- garch(df2,order=c(0,1))
summary(df2.fit1)

df2.fit2 <- garch(df2,order=c(0,2))
summary(df2.fit2)

sumdf2.fit3 <- garch(df2,order=c(0,3))
summary(df2.fit3)

df2.fit4 <- garch(df2,order=c(0,4))
summary(df2.fit4)

df2.fit5 <- garch(df2,order=c(0,5))
summary(df2.fit5)

df2.fit1 <- garch(df2,oreder=c(1,1))
summary(df2.fit1)
x.fit <- garch(df2,oreder=c(0,1))
#绘制波动置信区间
x.pre <- predict(x.fit)
plot(x.pre)
#对比
plot(df2)
lines(x.pre[,1],col=2)
lines(x.pre[,2],col=2)
abline(h=1.96*sd(df2),col=4,lty=2)
abline(h=-1.96*sd(df2),col=4,lty=2)
```