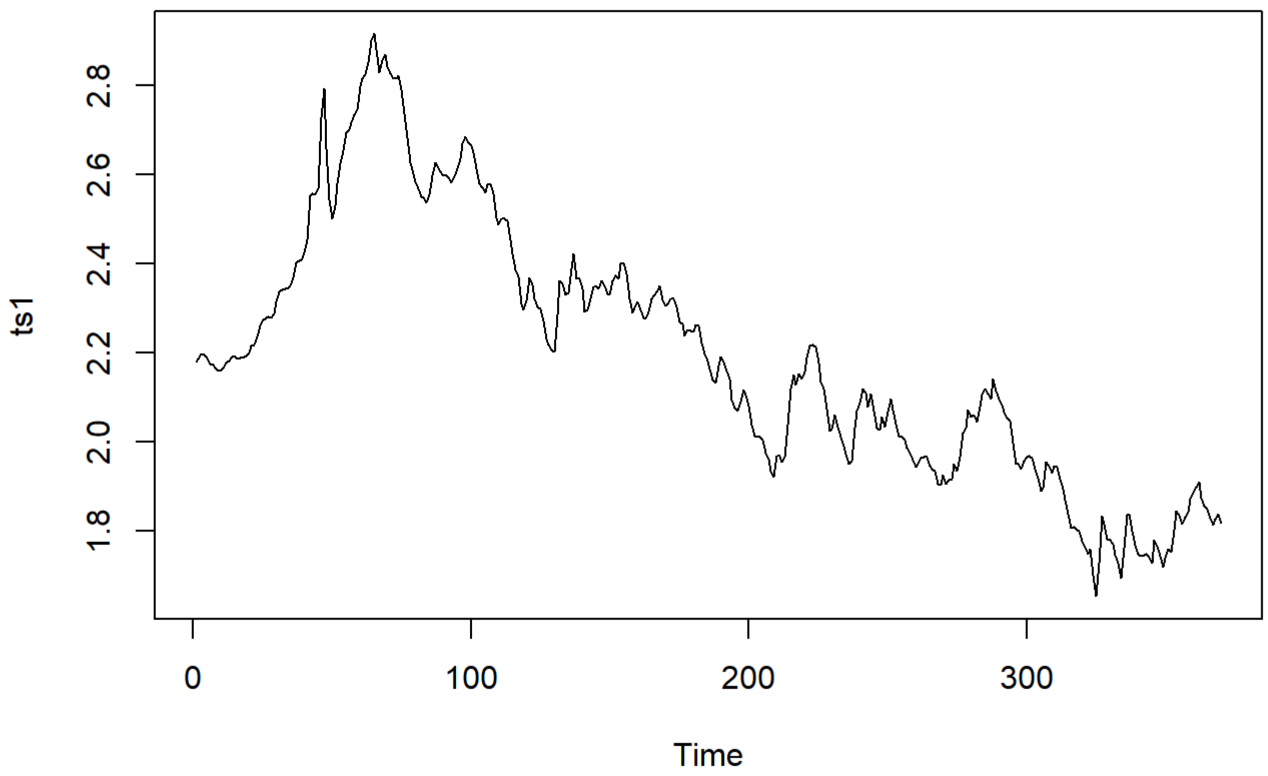
hw2

20989977 Zhang Mingtao

2024/3/28

1.

```
########################## 1
df = read.table("C://Users//张铭韬//Desktop//学业//港科大//MSDM5053时间序列//作业//assignment2//m-mortg.txt",header=F)
# log transformation
ts1 = log(ts(df$V4))
plot(ts1)
```



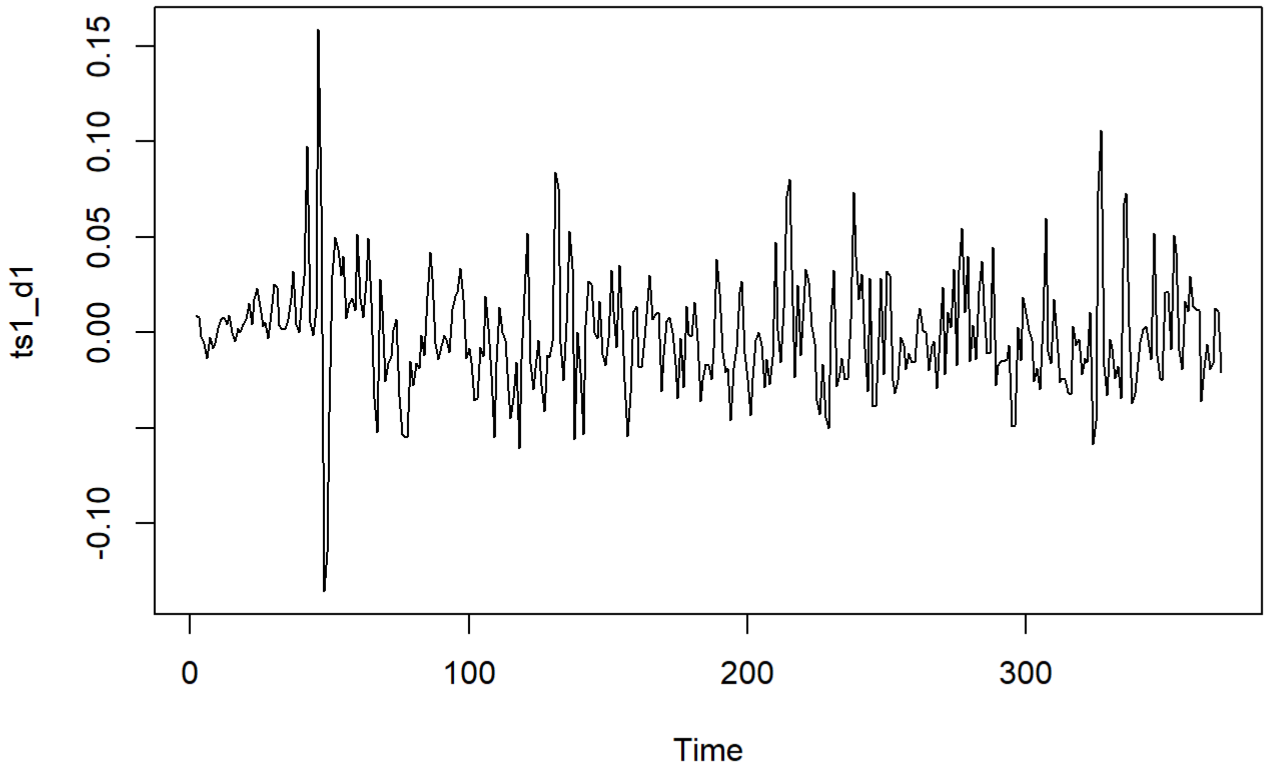

```
# 差分+平稳性检验
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
ndiffs(ts1) # 差分一次即平稳
```

```
## [1] 1
```

```
ts1_d1 = diff(ts1)
plot(ts1_d1)
```




```
library(aTSA)
```

```
##
## 载入程辑包：'aTSA'
```

```
## The following object is masked from 'package:forecast':
##
##     forecast
```

```
## The following object is masked from 'package:graphics':
##
##     identify
```

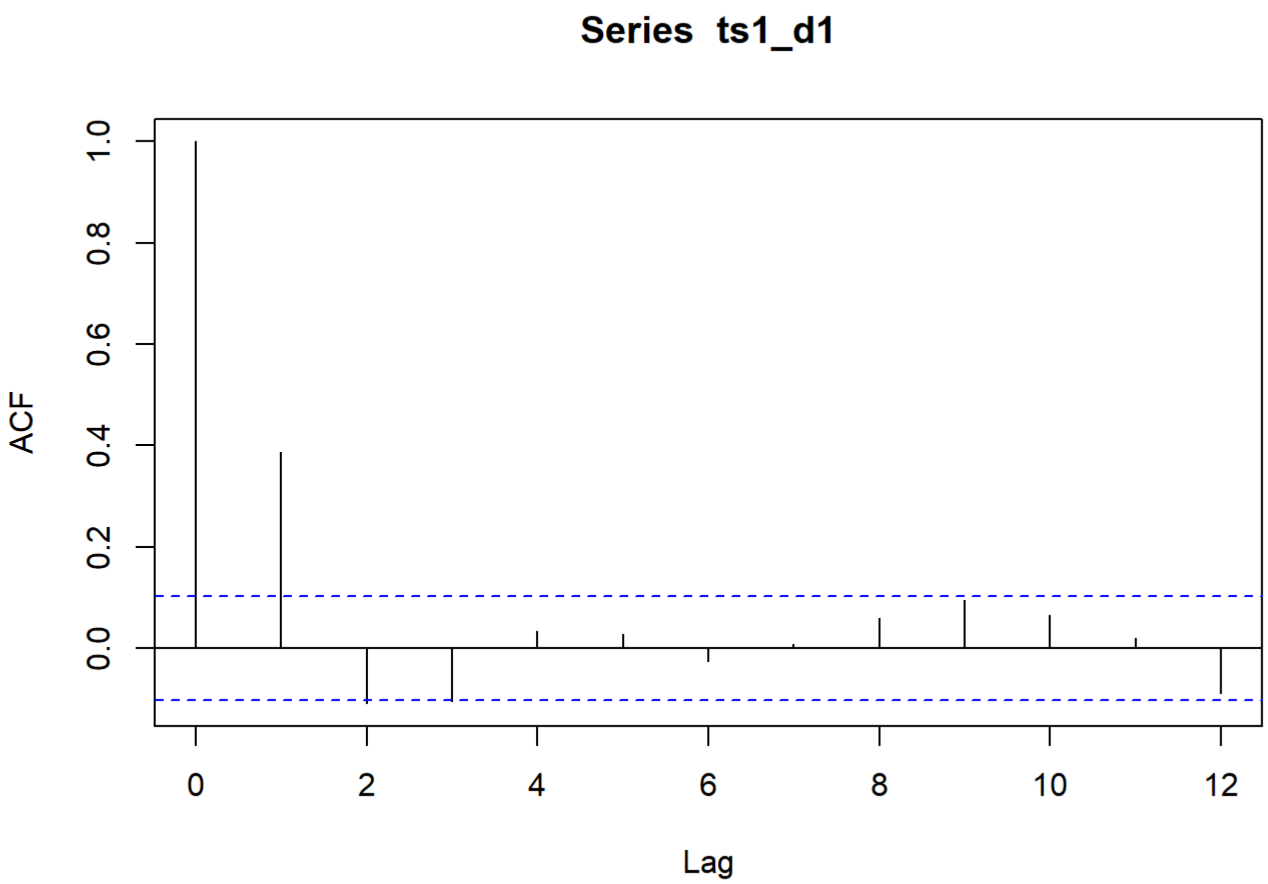
```
adf.test(ts1_d1) # p值 < 0.05，拒绝H0，表示平稳
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag    ADF p.value
## [1,]   0 -12.72    0.01
## [2,]   1 -14.46    0.01
## [3,]   2 -10.51    0.01
## [4,]   3  -8.95    0.01
## [5,]   4  -8.37    0.01
## [6,]   5  -7.66    0.01
## Type 2: with drift no trend
##      lag    ADF p.value
## [1,]   0 -12.72    0.01
## [2,]   1 -14.46    0.01
## [3,]   2 -10.52    0.01
## [4,]   3  -8.96    0.01
## [5,]   4  -8.37    0.01
## [6,]   5  -7.67    0.01
## Type 3: with drift and trend
##      lag    ADF p.value
## [1,]   0 -12.74    0.01
## [2,]   1 -14.51    0.01
## [3,]   2 -10.56    0.01
## [4,]   3  -9.01    0.01
## [5,]   4  -8.44    0.01
## [6,]   5  -7.75    0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

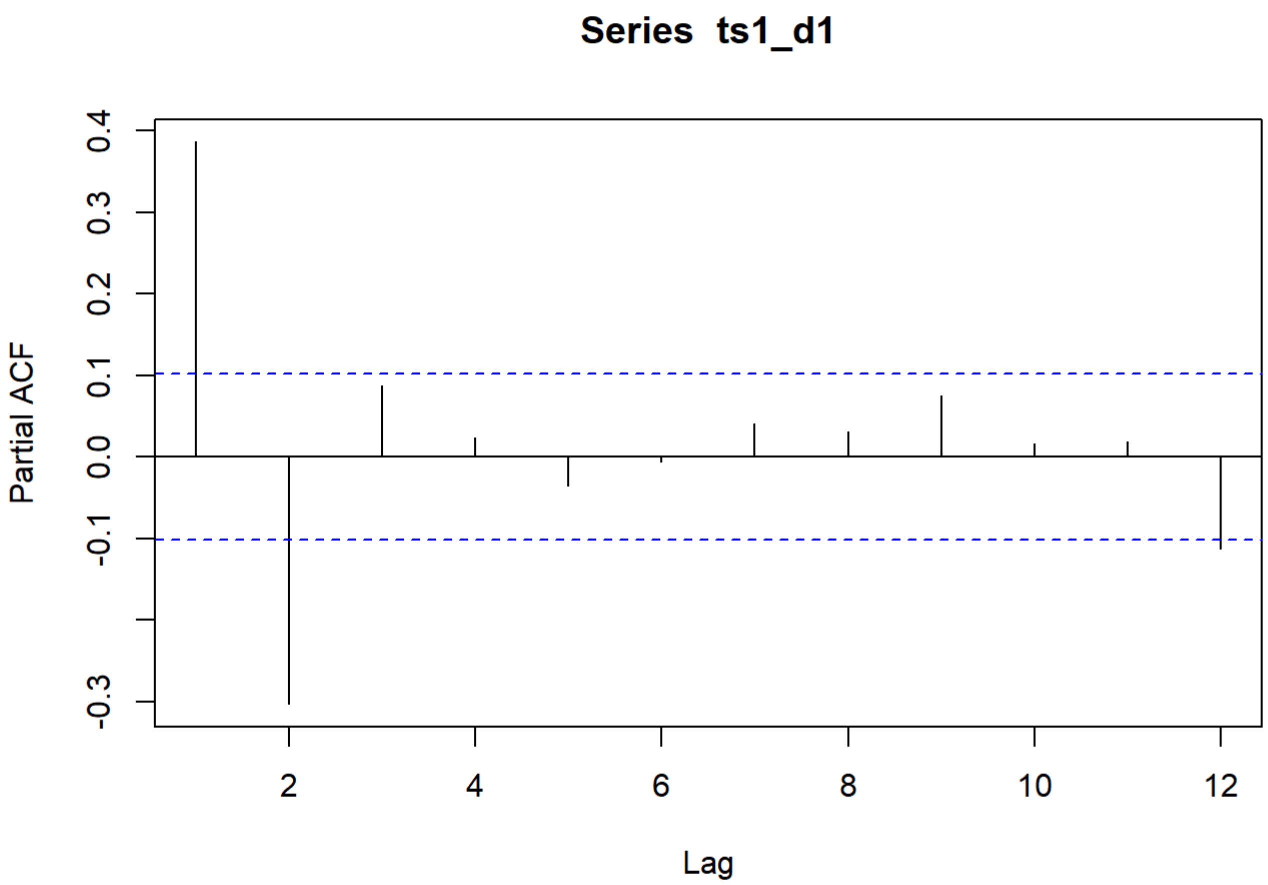
```
# 白噪声检验
for( i in c(5,9,12) ){
  print(Box.test(ts1_d1,lag=i,type="Ljung-Box"))
} # < 0.05 则非白噪声，有继续分析的意义
```

```
##
##  Box-Ljung test
##
## data:  ts1_d1
## X-squared = 64.662, df = 5, p-value = 1.317e-12
##
##
##  Box-Ljung test
##
## data:  ts1_d1
## X-squared = 69.768, df = 9, p-value = 1.691e-11
##
##
##  Box-Ljung test
##
## data:  ts1_d1
## X-squared = 74.503, df = 12, p-value = 4.561e-11
```

```
# 观察
acf(ts1_d1, lag.max = 12)
```




```
pacf(ts1_d1, lag.max = 12)
```




```
# 模型识别
# 1.
library(TSA)
```

```
## Registered S3 methods overwritten by 'TSA':
##   method       from
##   fitted.Arima forecast
##   plot.Arima   forecast
```

```
##
## 载入程辑包：'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
eacf(ts1_d1) # 可粗略选取最接近左上角的参数，ARMA(2,0)，但无法详细评估
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x o o o o o o o o o  o  o  o
## 1 x x o x o o o o o o o  o  o  o
## 2 x x o o o o o o o o o  o  o  o
## 3 x x o o o o o o o o o  o  o  o
## 4 x x o o o o o o o o o  o  o  o
## 5 x x x o o o o o o o o  o  o  o
## 6 x x x o o o o o o o o  o  o  o
## 7 x x x x o o o o o o o  o  o  o
```

```
# 2.
# library(forecast) 默认结合AIC指标与复杂度
auto.arima(ts1_d1) # ARMA(2,1)
```

```
## Series: ts1_d1
## ARIMA(2,0,1) with zero mean
##
## Coefficients:
##          ar1      ar2     ma1
##       0.3058  -0.2278  0.2206
## s.e.  0.1372   0.0760  0.1375
##
## sigma^2 = 0.0006909:  log likelihood = 820.42
## AIC=-1632.85   AICc=-1632.74   BIC=-1617.2
```

```
auto.arima(ts1) # ARIMA(2,1,1)
```

```
## Series: ts1
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1      ar2     ma1
##       0.3058  -0.2278  0.2206
## s.e.  0.1372   0.0760  0.1375
##
## sigma^2 = 0.000691:  log likelihood = 820.42
## AIC=-1632.85   AICc=-1632.74   BIC=-1617.2
```

```
# 参数估计
md = Arima(ts1, order = c(2,1,1), include.drift = T, method = 'ML')
md
```

```
## Series: ts1
## ARIMA(2,1,1) with drift
##
## Coefficients:
##          ar1      ar2     ma1    drift
##       0.3077  -0.2296  0.2179  -0.0010
## s.e.  0.1372   0.0759  0.1377   0.0018
##
## sigma^2 = 0.0006923:  log likelihood = 820.58
## AIC=-1631.16   AICc=-1630.99   BIC=-1611.6
```

```
# (1 - φ_1*B - φ_2*B^2) * (1 - B)^d * p_t = θ_0 + (1 - θ_1*B) * a_t, where d = 1, φ_1 = 0.3077, φ_2 = -0.2296, θ_0 = -0.001, θ_1 = 0.2179
```

```
# 参数显著性检验
# t统计量
t = abs(md$coef)/sqrt(diag(md$var.coef))
# 自由度
df_t = length(ts1)-length(md$coef)
# pt()
pt(t,df_t,lower.tail = F)
```

```
##         ar1         ar2         ma1       drift
## 0.012786165 0.001325663 0.057244604 0.286018526
```

```
# p<0.05，则显著，可见ma1系数不显著，进而考虑ARIMA(2,1,0)模型
```

```
md2 = Arima(ts1, order = c(2,1,0), include.drift = T, method = 'ML')
md2
```

```
## Series: ts1
## ARIMA(2,1,0) with drift
##
## Coefficients:
##          ar1      ar2    drift
##       0.5034  -0.3031  -0.0010
## s.e.  0.0496   0.0495   0.0017
##
## sigma^2 = 0.0006947:  log likelihood = 819.44
## AIC=-1630.87   AICc=-1630.76   BIC=-1615.23
```

```
# (1 - φ_1*B - φ_2*B^2) * (1 - B)^d * p_t = θ_0 + a_t, where d = 1, φ_1 = 0.5034, φ_2 = -0.3031, θ_0 = -0.001
```

```
t = abs(md2$coef)/sqrt(diag(md2$var.coef))
df_t = length(ts1)-length(md2$coef)
pt(t,df_t,lower.tail = F)
```

```
##          ar1          ar2        drift
## 8.245244e-22 1.190366e-09 2.788068e-01
```

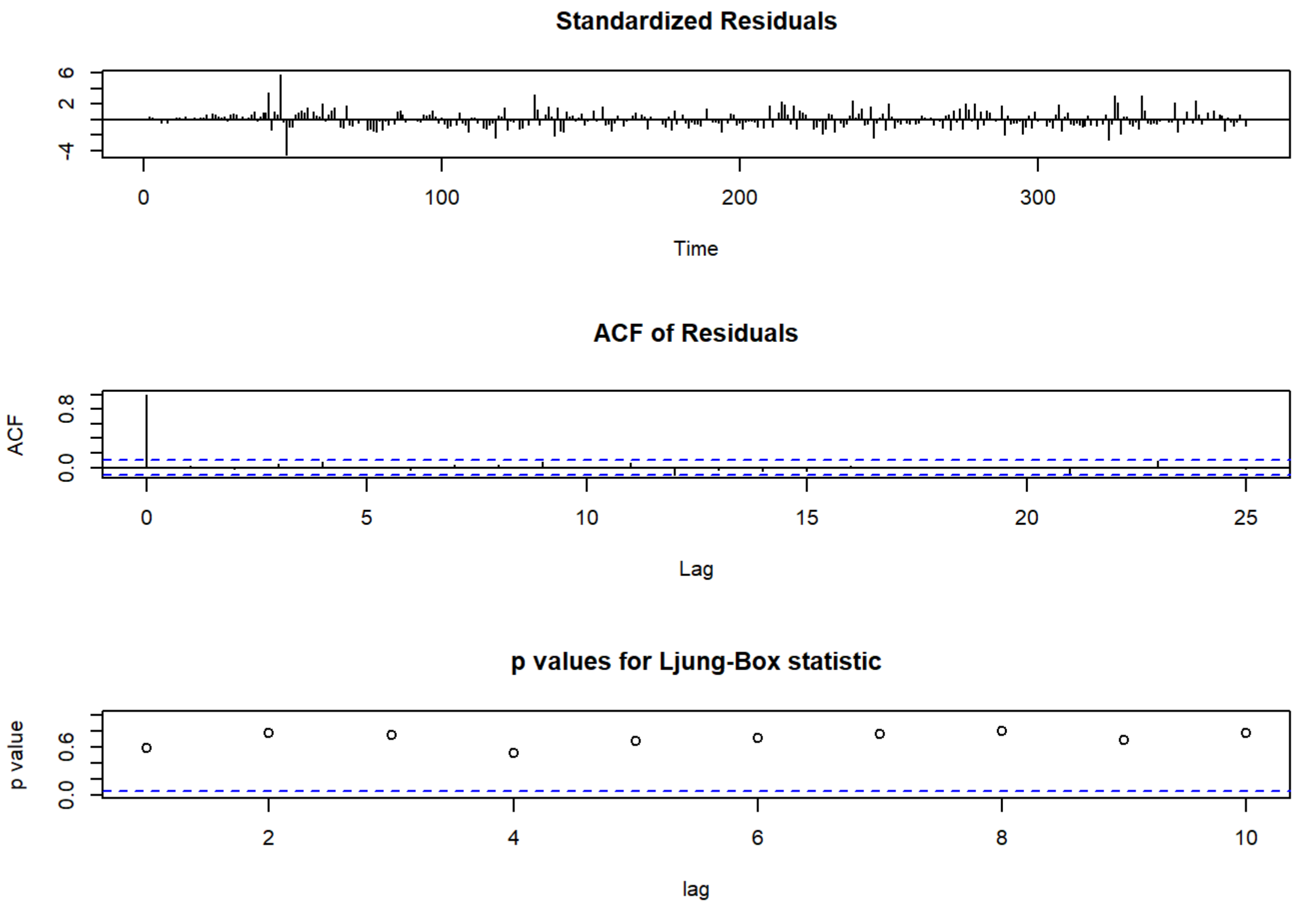
```
# 均显著，drift可忽略，选择ARIMA(2,1,0)模型：(1 - 0.5034*B - -0.3031*B^2) * (1 - B) * p_t = a_t
```

```
# 残差检验
library(stats)
Box.test(md2$residuals,lag=12,type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  md2$residuals
## X-squared = 11.468, df = 12, p-value = 0.4893
```

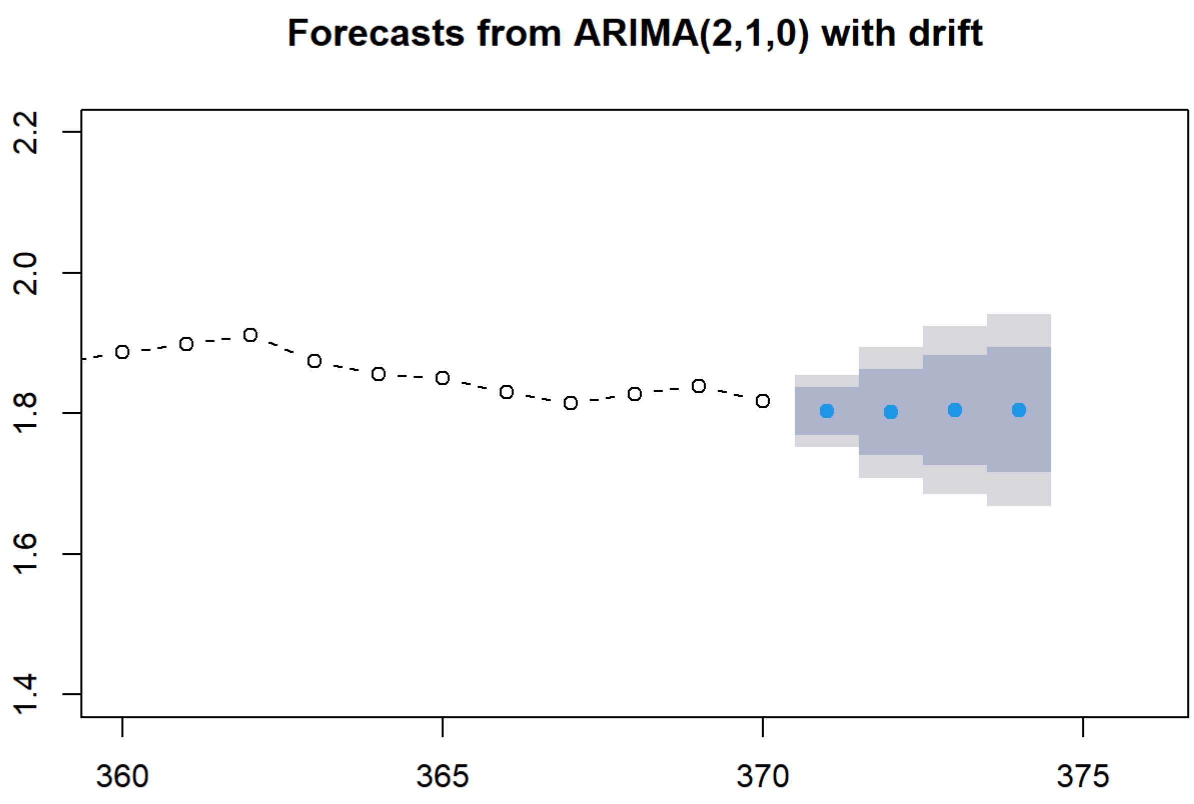
```
# p value is larger than 0.05 so we cannot refuse H0: the first 12 lags of residuals'ACF are al
l zero, thus is white noise.
```

```
tsdiag(md2)
```




```
# The standardized residuals are basically distributed near the zero horizontal line;
# the autocorrelation function quickly drops to within the two dotted lines;
# the P values of the Ljung-Box statistics are all greater than 0.05
# therefore, the model passes the test.
```

```
# 模型预测
# (1) Arima函数 对应 forecast::forecast
fore.gnp = forecast::forecast(md2,4) # 后四列为置信区间
plot(fore.gnp, lty=2, pch=1, type='b',xlim=c(360,376),ylim=c(1.4,2.2))
```



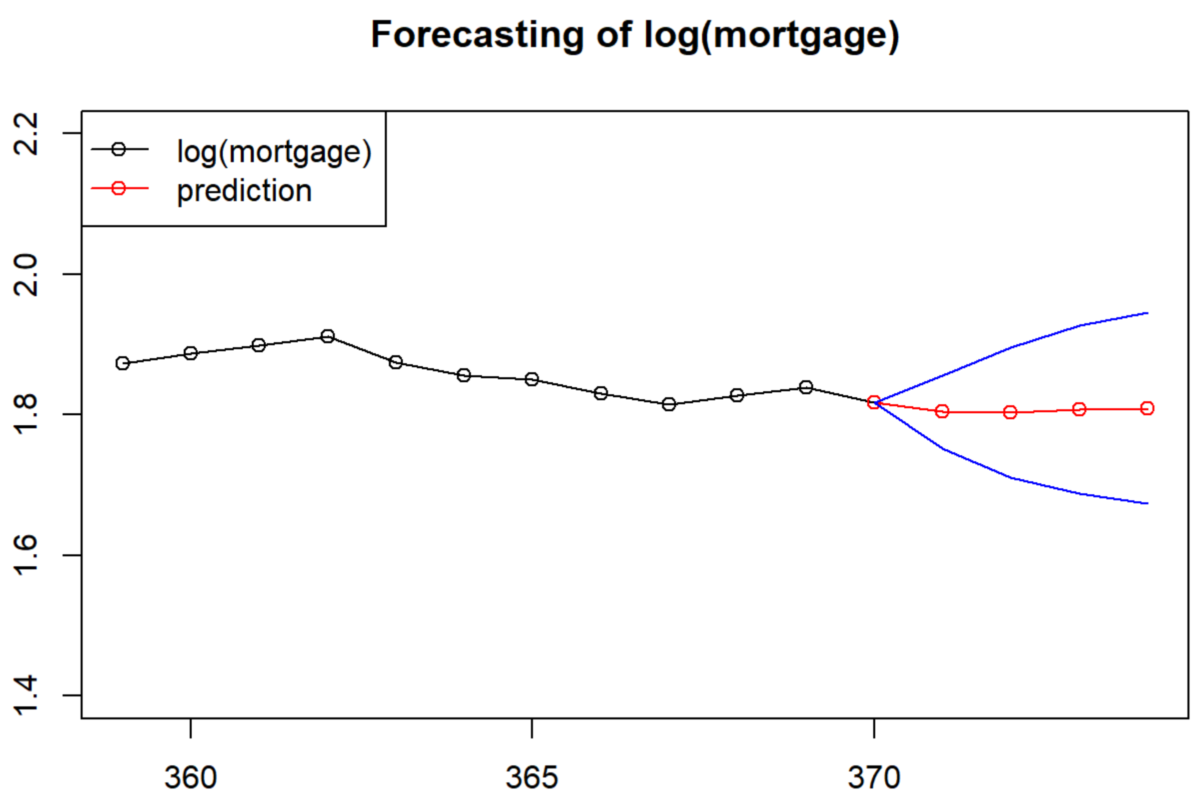

```
# lines(fore.gnp$fitted, col=2, pch=2, type='b')
```

```
# (2) arima函数 对应 predict, 实际上arima函数未考虑drift, 准确度不如Forecast包的Arima函数, 此处仅作画图演示
md2_2 = arima(ts1, order = c(2,1,0), method = 'ML')
fore=predict(md2_2, 4)
fore
```

```
## $pred
## Time Series:
## Start = 371
## End = 374
## Frequency = 1
## [1] 1.804167 1.803473 1.807330 1.809484
##
## $se
## Time Series:
## Start = 371
## End = 374
## Frequency = 1
## [1] 0.02626175 0.04743085 0.06091632 0.06955892
```

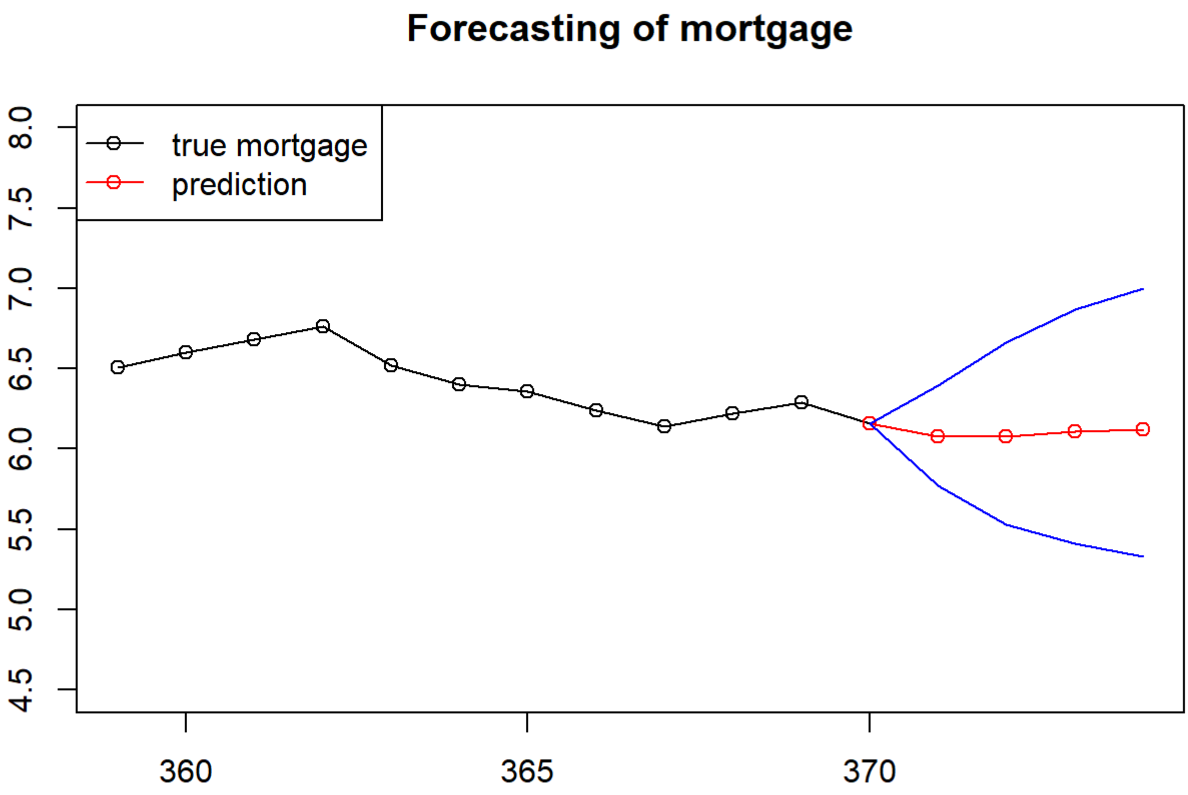
```
U=append(ts1[370],fore$pred+1.96*fore$se)
L=append(ts1[370],fore$pred-1.96*fore$se)

plot(359:370,ts1[359:370],xlim=c(359,374),ylim=c(1.4,2.2),type="o",ylab="",xlab="",main="Forecasting of log(mortgage)")
lines(370:374,append(ts1[370],fore$pred),type="o",col="red")
lines(370:374, U,type="l",col="blue")
lines(370:374, L,type="l",col="blue")
legend(x="topleft",c("log(mortgage)","prediction"),lty=c(1,1),pch=c(1,1),col=c("black","red"))
```



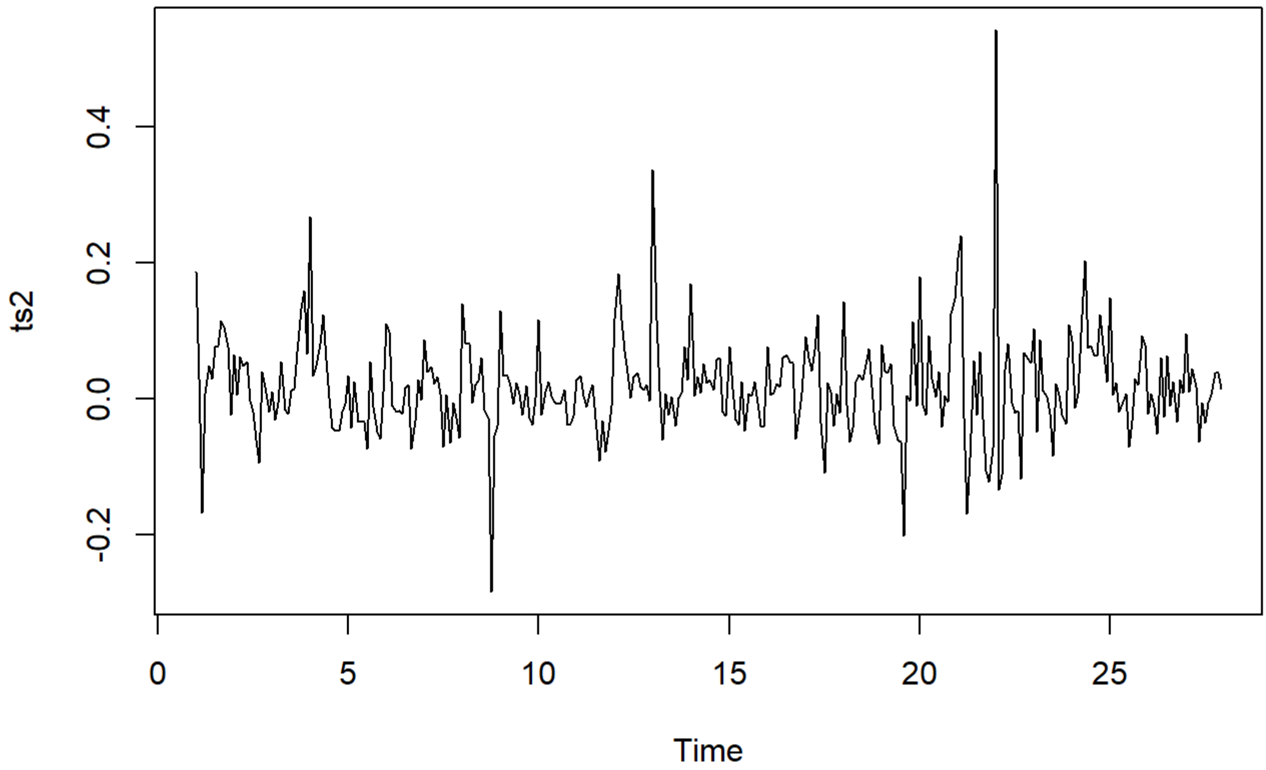

```
# 因为原数据进行过对数化,需要还原
U1=fore$pred + 1.96 * fore$se
L1=fore$pred - 1.96 * fore$se
U2=exp(U1)
L2=exp(L1)
E1 = exp(fore$pred+fore$se*fore$se/2)

plot(359:370,df$V4[359:370],xlim=c(359,374),ylim=c(4.5,8),type="o",ylab="",xlab="",main="Forecasting of mortgage")
lines(370:374,append(df$V4[370],E1),type="o",col="red")
lines(370:374,append(df$V4[370],U2),type="l",col="blue")
lines(370:374,append(df$V4[370],L2),type="l",col="blue")
#points(temp.date[2:7],p,type="o")
legend(x="topleft",c("true mortgage","prediction"),lty=c(1,1),pch=c(1,1),col=c("black","red"))
```

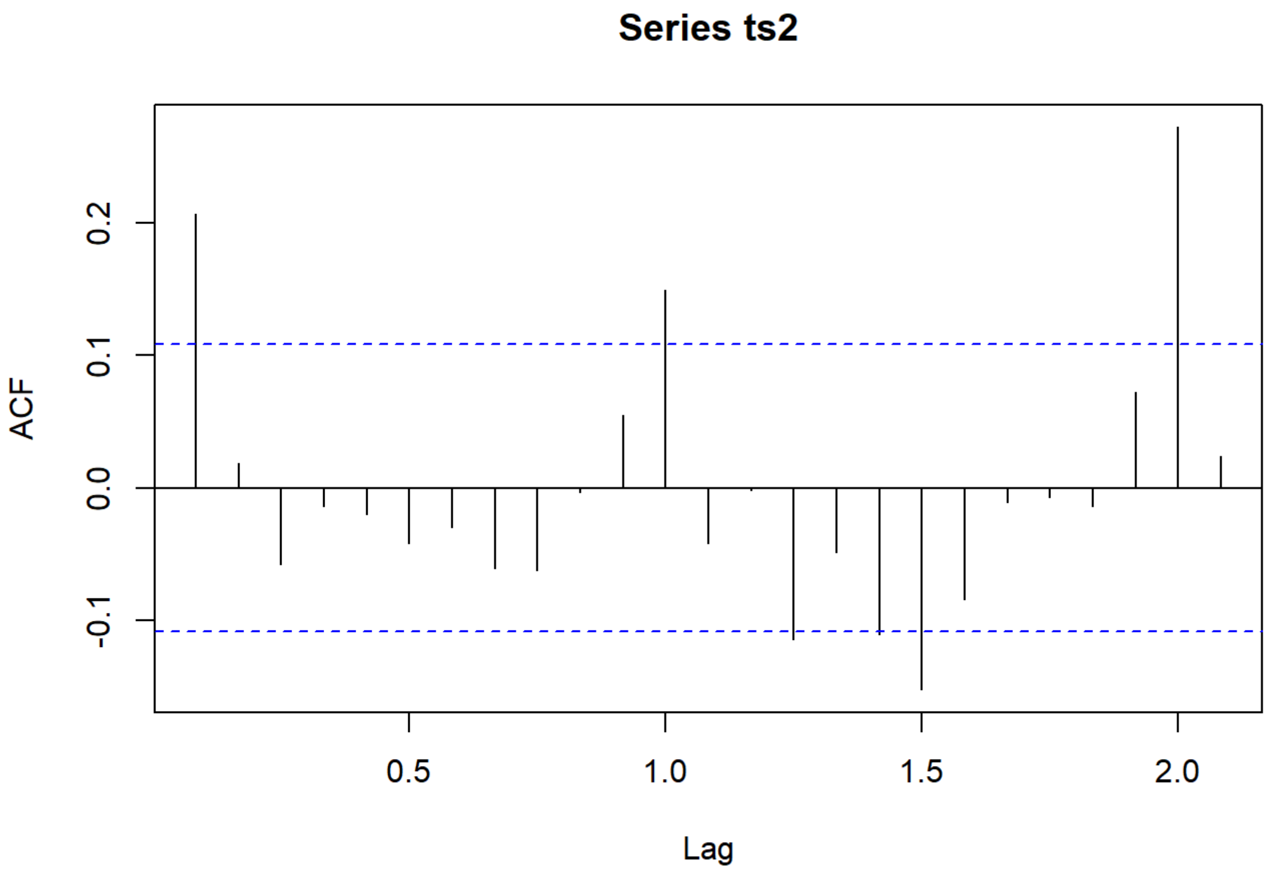


2.

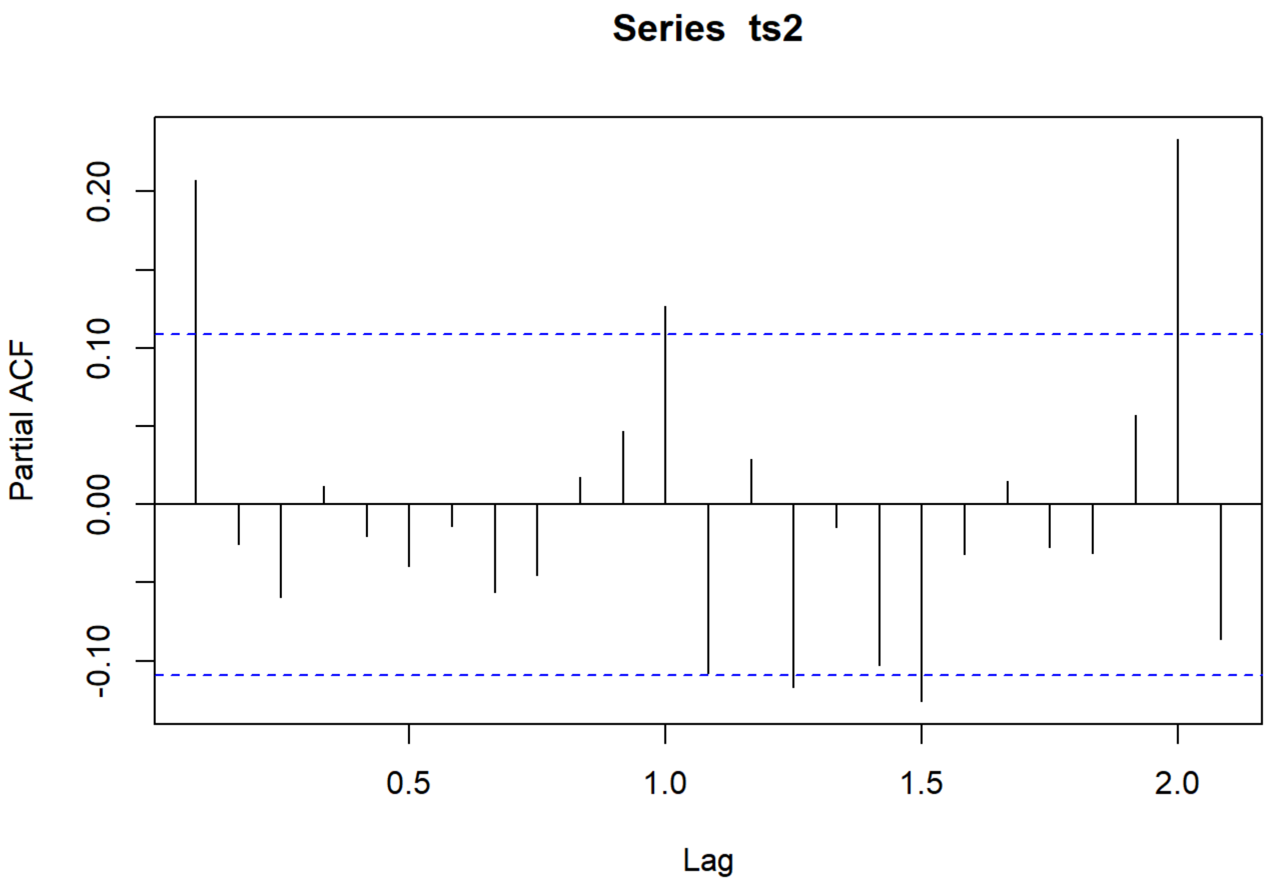
```
########################## 2
df2 = read.table("C://Users//张铭韬//Desktop//学业//港科大//MSDM5053时间序列//作业//assignment
2//m-dec1-8006.txt",header=F)
ts2 = ts(df2$V2,frequency=12) # frequency=12 is very impoortant for use in Arima() and auto.ari
ma()
plot(ts2)
```




```
# 观察季节性
acf(ts2) # 明显季节性
```



```
pacf(ts2) # 明显季节性
```




```
# 参数估计
arima(ts2, order = c(0, 0, 1), seasonal = list(order = c(1,0,1), period = 12))
```

```
##
## Call:
## arima(x = ts2, order = c(0, 0, 1), seasonal = list(order = c(1, 0, 1), period = 12))
##
## Coefficients:
##          ma1    sar1     sma1  intercept
##       0.2409  0.9995  -0.9830     0.0179
## s.e.  0.0517  0.0014   0.0241     0.0131
##
## sigma^2 estimated as 0.00409:  log likelihood = 420.42,  aic = -832.83
```

```
# or this function, actually their results are the same
est=Arima(ts2, order = c(0, 0, 1), seasonal = c(1,0,1))
est
```

```
## Series: ts2
## ARIMA(0,0,1)(1,0,1)[12] with non-zero mean
##
## Coefficients:
##          ma1    sar1     sma1    mean
##       0.2409  0.9995  -0.9830  0.0179
## s.e.  0.0517  0.0014   0.0241  0.0131
##
## sigma^2 = 0.004141:  log likelihood = 420.42
## AIC=-830.83   AICc=-830.64   BIC=-811.93
```

```
# ARIMA(p, d, q)×(P, D, Q)_s model : ΦP(B^s) * φp(B) * (1 - B)^d * (1 - B^s)^D * pt = θ0 + θq(B) * ΘQ(B^s) * at
# In this case, the model is actually: (1-0.9995*B^12) * pt = 0.0179 + (1-0.2409*B) * (1+0.983*B^12) * at
```

```
# 参数显著性检验
# t统计量
t = abs(est$coef)/sqrt(diag(est$var.coef))
# 自由度
df_t = length(ts2)-length(est$coef)
# pt()
pt(t,df_t,lower.tail = F)
```

```
##          ma1          sar1          sma1     intercept
##  2.353853e-06  0.000000e+00 3.273583e-129  8.648045e-02
```

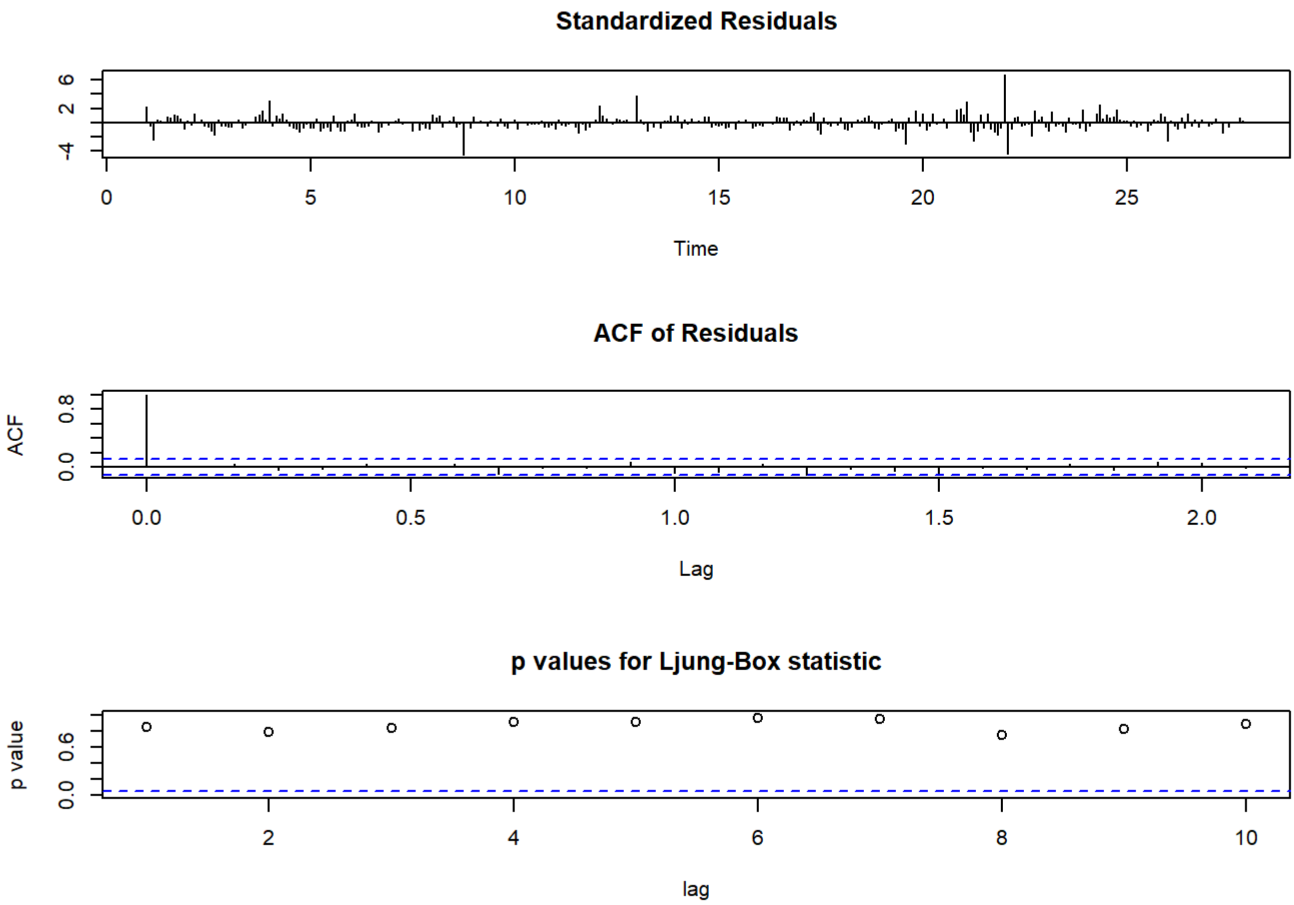
```
# p<0.05, 均显著
```

```
# 残差检验
Box.test(est$residuals,lag=24,type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  est$residuals
## X-squared = 23.922, df = 24, p-value = 0.466
```

```
# p value is larger than 0.05 so we cannot refuse H0: the first 12 lags of residuals'ACF are all zero, thus is white noise.
```

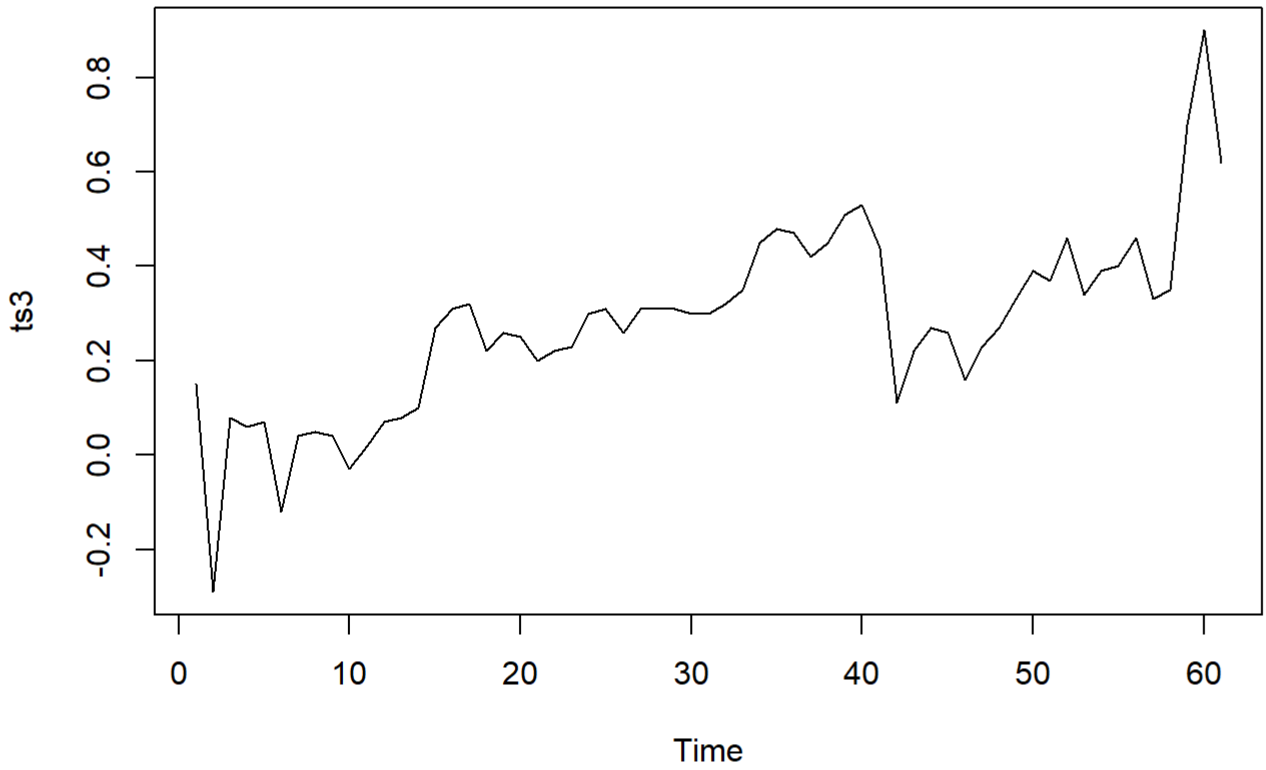
```
tsdiag(est)
```



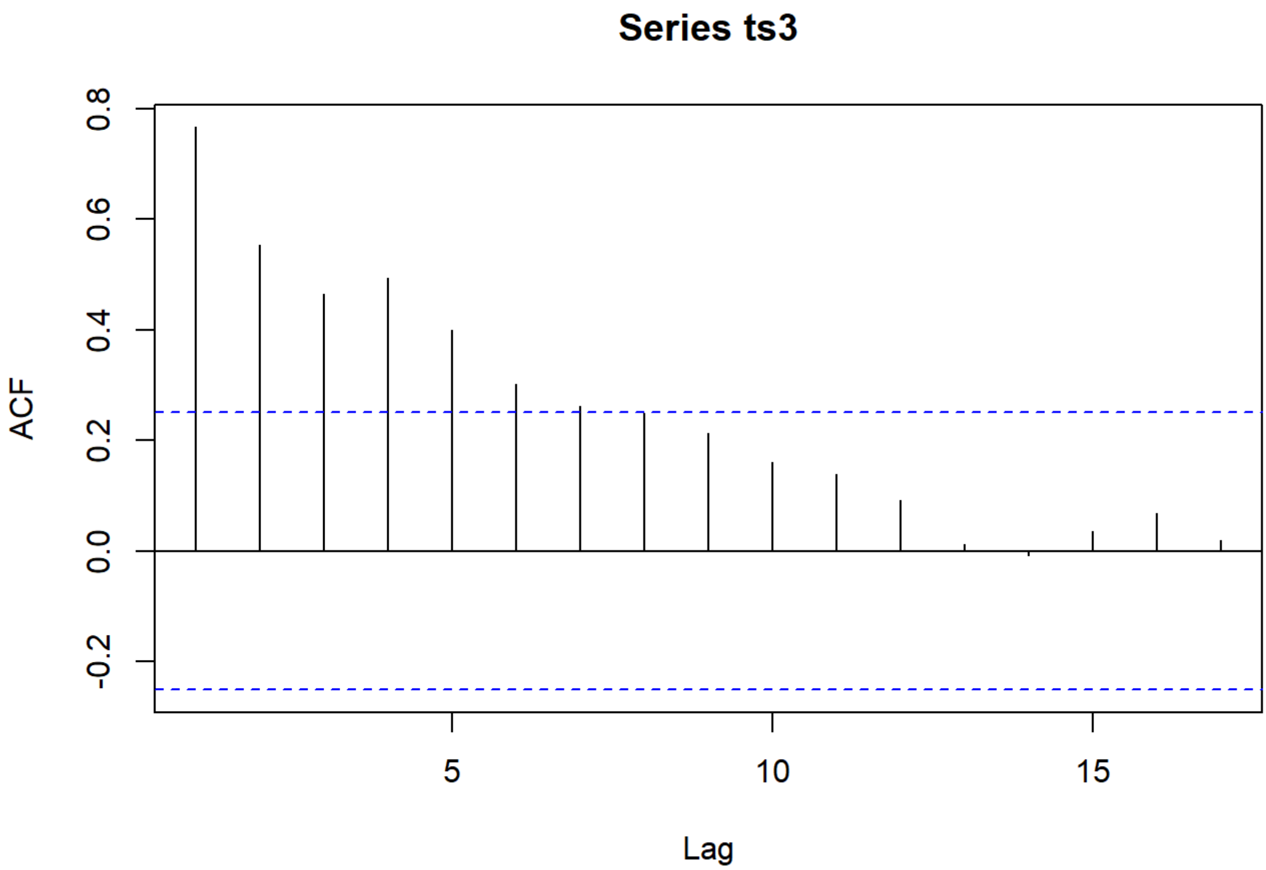

```
# The standardized residuals are basically distributed near the zero horizontal line;
# the autocorrelation function quickly drops to within the two dotted lines;
# the P values of the Ljung-Box statistics are all greater than 0.05
# therefore, the model passes the test.
```

3.

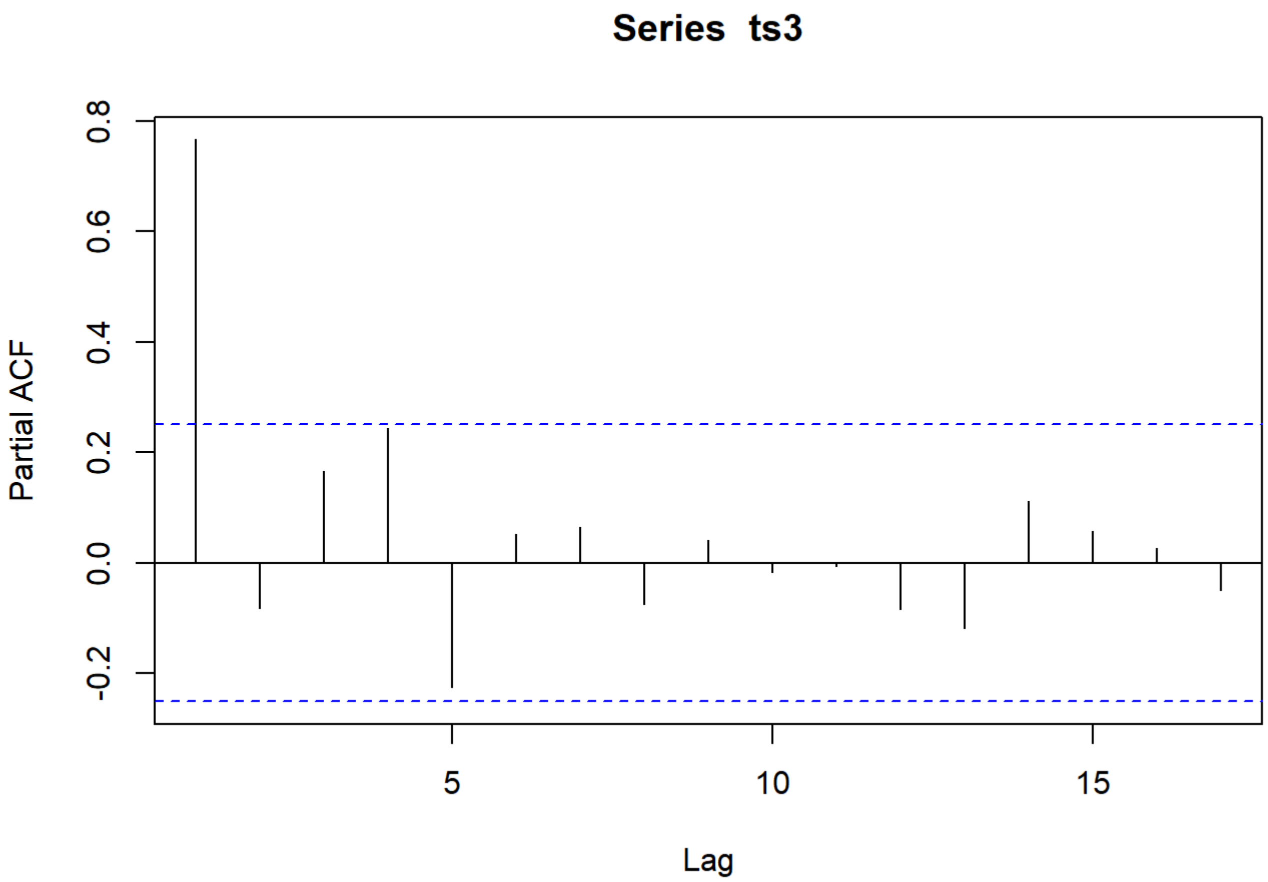
```
######################### 3
df3 = read.table("C://Users//张铭韬//Desktop//学业//港科大//MSDM5053时间序列//作业//assignment 2//q-aa-earn.txt",header=F)
ts3 = ts(df3$V4)
plot(ts3)
```




```
# (p)acf图 无季节性
acf(ts3)
```



```
pacf(ts3)
```




```
# 判断差分+平稳性检验
ndiffs(ts3) # d=1, 差分一次即平稳
```

```
## [1] 1
```

```
ts3_d1 = diff(ts3)
library(tseries)
```

```
##
## 载入程辑包：'tseries'
```

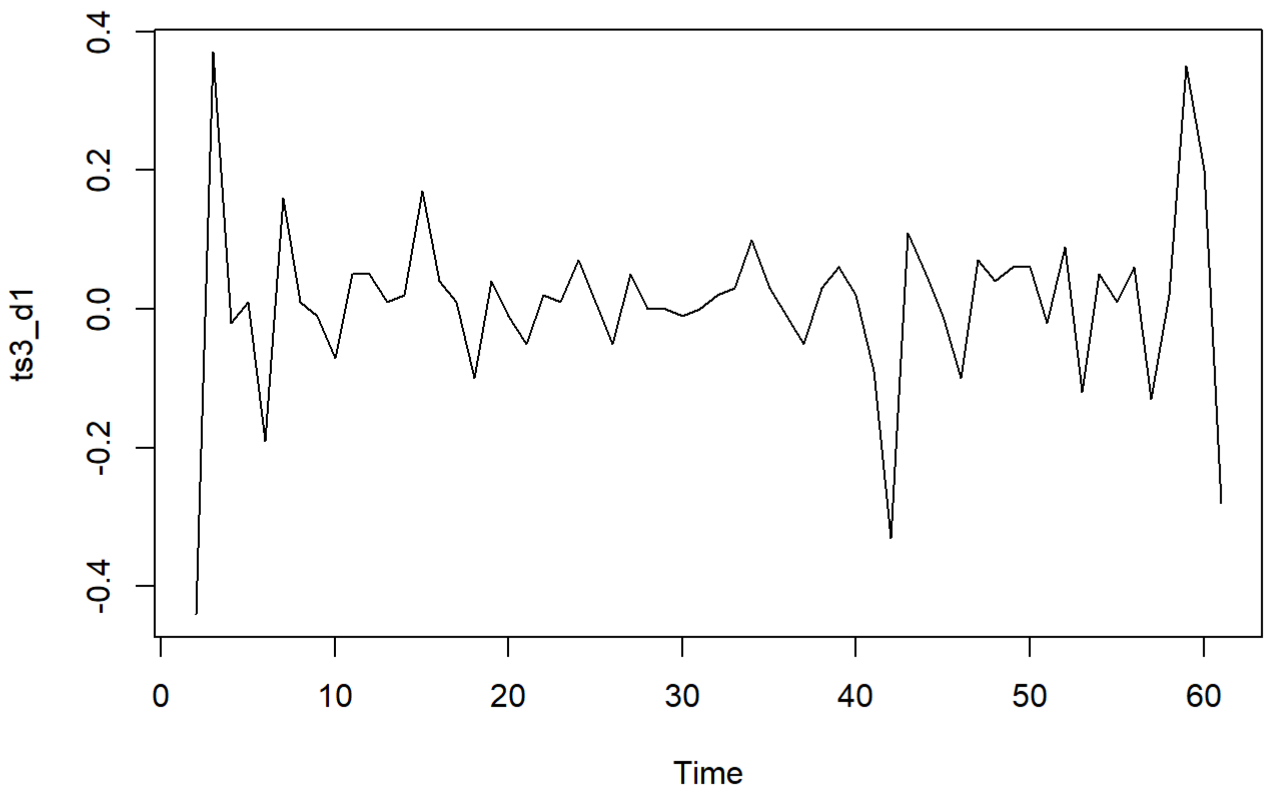
```
## The following objects are masked from 'package:aTSA':
##
##     adf.test, kpss.test, pp.test
```

```
pp.test(ts3_d1) # p值 < 0.05, 拒绝H0, 表示平稳
```

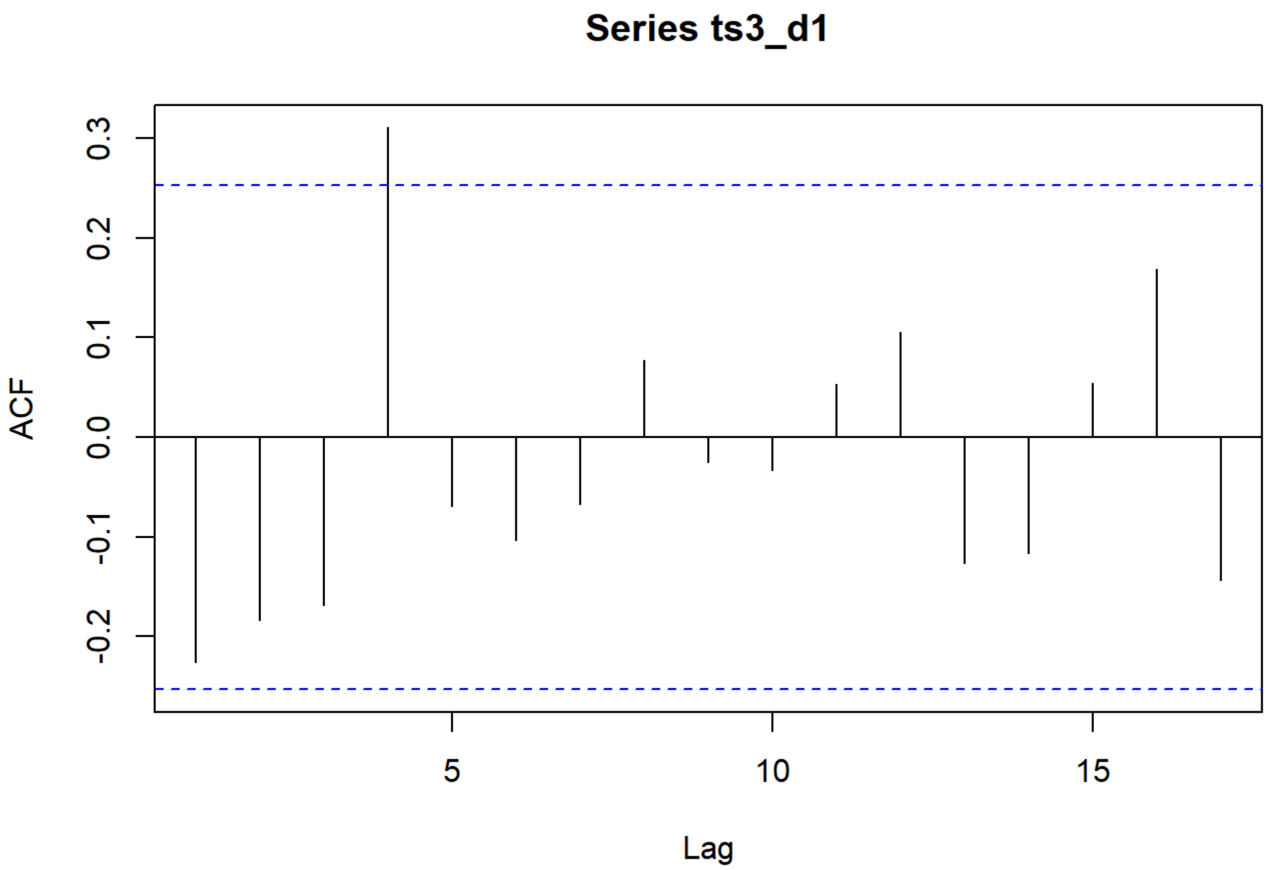
```
## Warning in pp.test(ts3_d1): p-value smaller than printed p-value
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  ts3_d1
## Dickey-Fuller Z(alpha) = -66.815, Truncation lag parameter = 3, p-value
## = 0.01
## alternative hypothesis: stationary
```

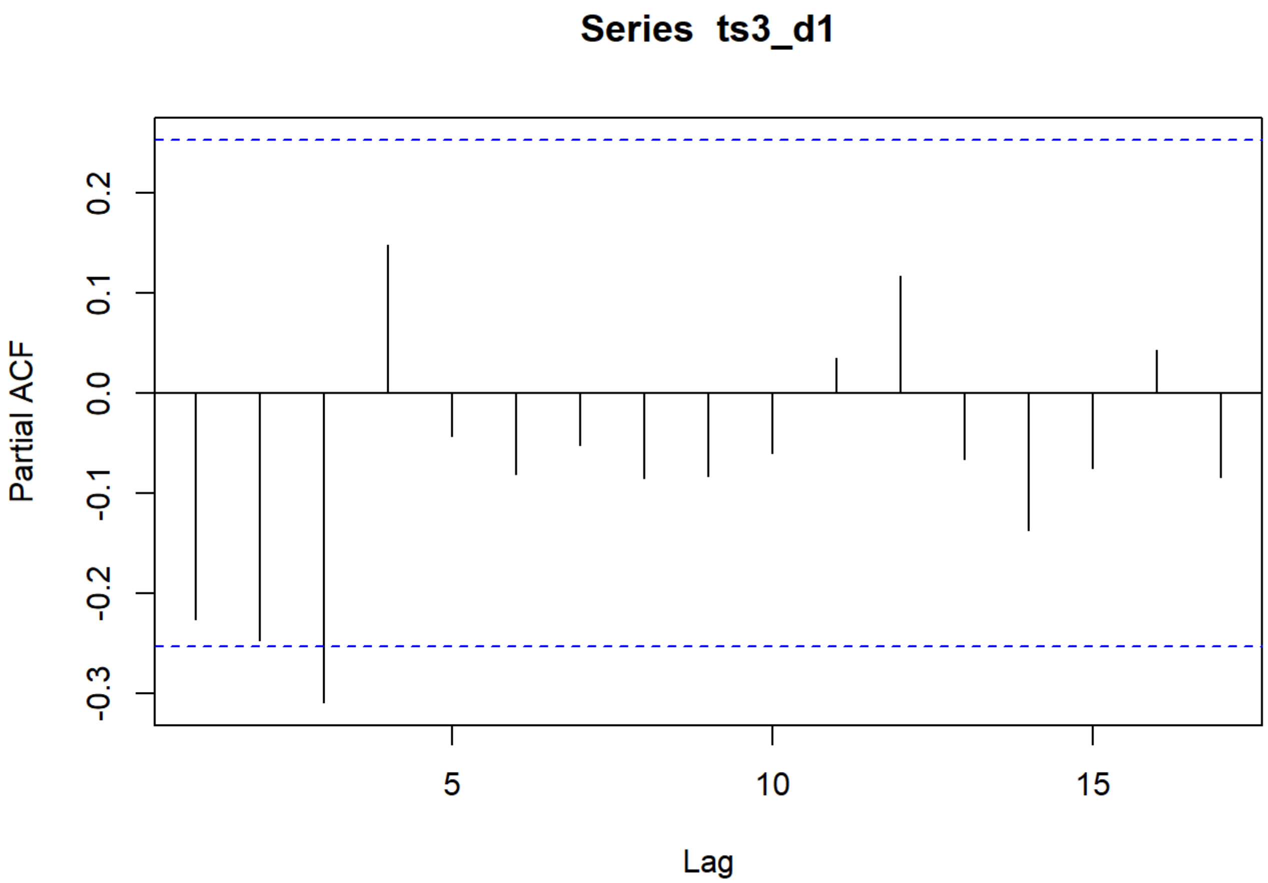
```
plot(ts3_d1)
```




```
acf(ts3_d1)
```


Series ts3_d1
ACF
0.3
0.2
0.1
0.0
-0.1
-0.2
5
10
15
Lag

```
pacf(ts3_d1)
```


Series ts3_d1
Partial ACF
0.2
0.1
0.0
-0.1
-0.2
-0.3
5
10
15
Lag

```
# 白噪声检验
for( i in c(2,4,6,8) ){
  print(Box.test(ts3_d1,lag=i,type="Ljung-Box"))
} # < 0.05 则非白噪声, 有继续分析的意义, 但在8阶以后滞后可认为无相关性
```

```
##
##  Box-Ljung test
##
## data:  ts3_d1
## X-squared = 5.3741, df = 2, p-value = 0.06808
##
##
##  Box-Ljung test
##
## data:  ts3_d1
## X-squared = 13.636, df = 4, p-value = 0.008554
##
##
##  Box-Ljung test
##
## data:  ts3_d1
## X-squared = 14.702, df = 6, p-value = 0.0227
##
##
##  Box-Ljung test
##
## data:  ts3_d1
## X-squared = 15.452, df = 8, p-value = 0.05092
```

```
# 模型识别
# 1.
eacf(ts3_d1) # 可粗略选取ARMA((0,1,2),(1,2))
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o o x o o o o o o o  o  o  o
## 1 x o o x o o o o o o o  o  o  o
## 2 x o o o o o o o o o o  o  o  o
## 3 x o o o o o o o o o o  o  o  o
## 4 x x o o o o o o o o o  o  o  o
## 5 x o o o o o o o o o o  o  o  o
## 6 o o o o o o o o o o o  o  o  o
## 7 o x x o o o o o o o o  o  o  o
```

```
# 2.
auto.arima(ts3_d1) # ARMA(0,0,1)
```

```
## Series: ts3_d1
## ARIMA(0,0,1) with zero mean
##
## Coefficients:
##           ma1
##       -0.4426
## s.e.   0.1290
##
## sigma^2 = 0.01381:  log likelihood = 43.73
## AIC=-83.45   AICc=-83.24   BIC=-79.26
```

```
auto.arima(ts3) # ARIMA(0,1,1)
```

```
## Series: ts3
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.4426
## s.e.   0.1290
##
## sigma^2 = 0.01381:  log likelihood = 43.73
## AIC=-83.45   AICc=-83.24   BIC=-79.26
```

```
# 参数估计
md3 = Arima(ts3, order = c(0,1,1), include.drift = T, method = 'ML')
md3
```

```
## Series: ts3
## ARIMA(0,1,1) with drift
##
## Coefficients:
##           ma1   drift
##       -0.5079  0.0112
## s.e.   0.1312  0.0074
##
## sigma^2 = 0.01358:  log likelihood = 44.71
## AIC=-83.43   AICc=-83   BIC=-77.15
```

```
# (1 - B)^d * p_t = θ_0 + (1 - θ_1*B) * a_t, where d = 1, θ_0 = 0.0112, θ_1 = -0.5079
```

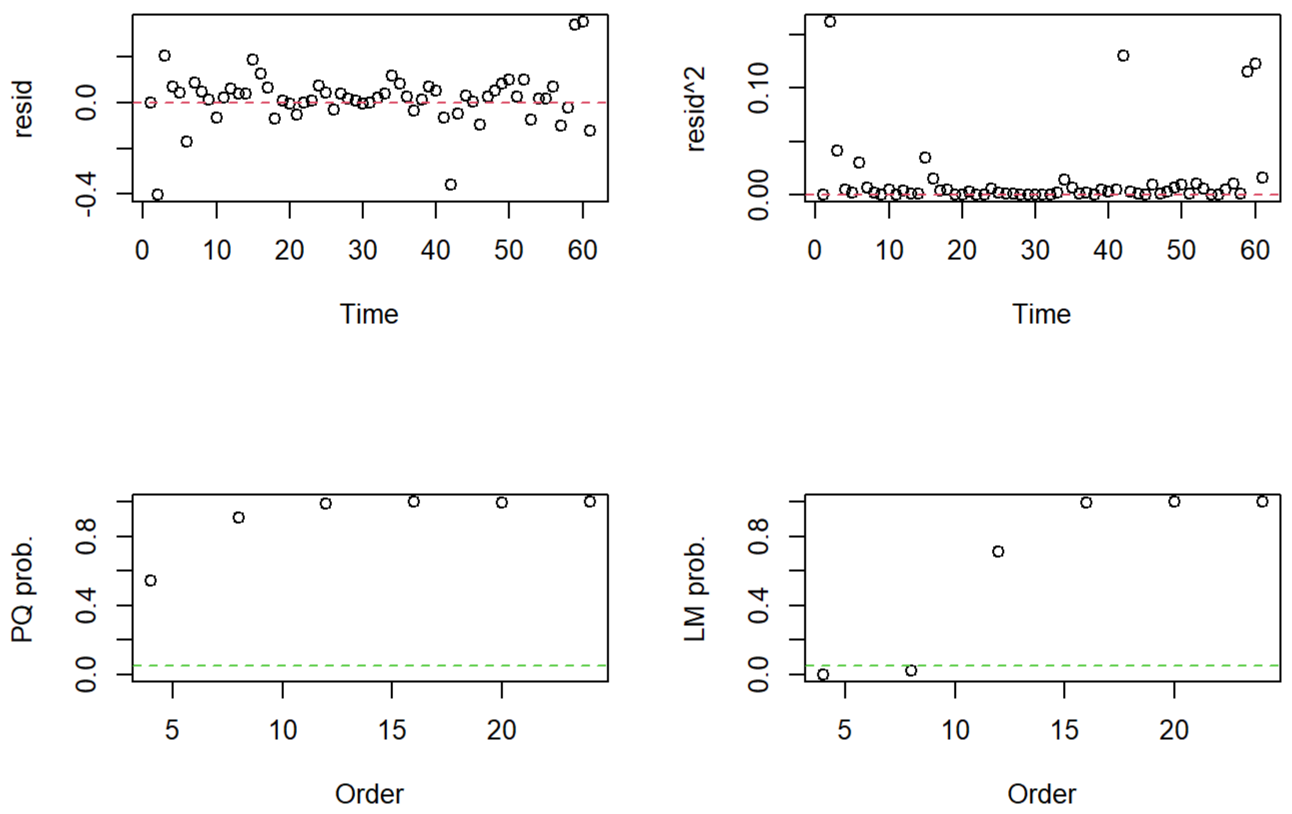
```
# 参数显著性检验
# t统计量
t = abs(md3$coef)/sqrt(diag(md3$var.coef))
# 自由度
df_t = length(ts3)-length(md3$coef)
# pt()
pt(t,df_t,lower.tail = F)
```

```
##          ma1       drift
## 0.000136693 0.067542097
```

```
# p<0.05, 显著, drift可忽略, 选择ARIMA(0,1,1)模型: (1 - B) * p_t = (1 + 0.5079 * B) * a_t
```

```
# 异方差检验 library(aTSA)
arch.test(arima(ts3, order = c(0,1,1), method = 'ML'), output = T)
```

```
## ARCH heteroscedasticity test for residuals
## alternative: heteroscedastic
##
## Portmanteau-Q test:
##      order   PQ p.value
## [1,]     4 3.08   0.545
## [2,]     8 3.37   0.909
## [3,]    12 3.66   0.989
## [4,]    16 4.02   0.999
## [5,]    20 7.86   0.993
## [6,]    24 8.37   0.999
## Lagrange-Multiplier test:
##      order     LM  p.value
## [1,]     4 39.031 1.71e-08
## [2,]     8 15.884 2.62e-02
## [3,]    12  8.017 7.12e-01
## [4,]    16  4.473 9.96e-01
## [5,]    20  0.928 1.00e+00
## [6,]    24  0.178 1.00e+00
```



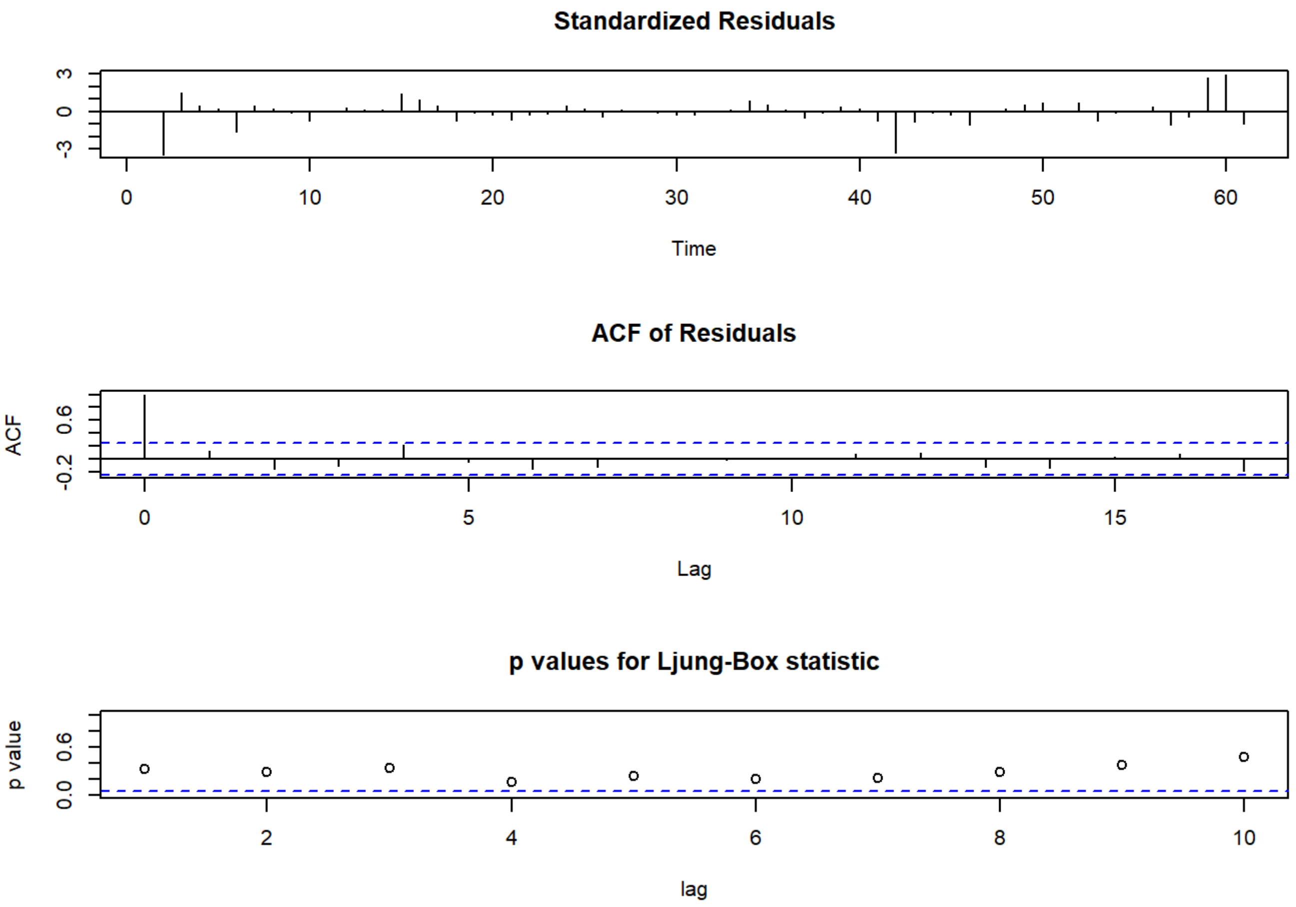

```
# 上半残差序列及平方序列的散点图，下半PQ检验和LM检验的P值，p>0.05，所以不拒绝原假设，不具备异方差性，不考虑GARCH
```

```
# 残差检验
Box.test(md3$residuals,lag=12,type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  md3$residuals
## X-squared = 10.737, df = 12, p-value = 0.5516
```

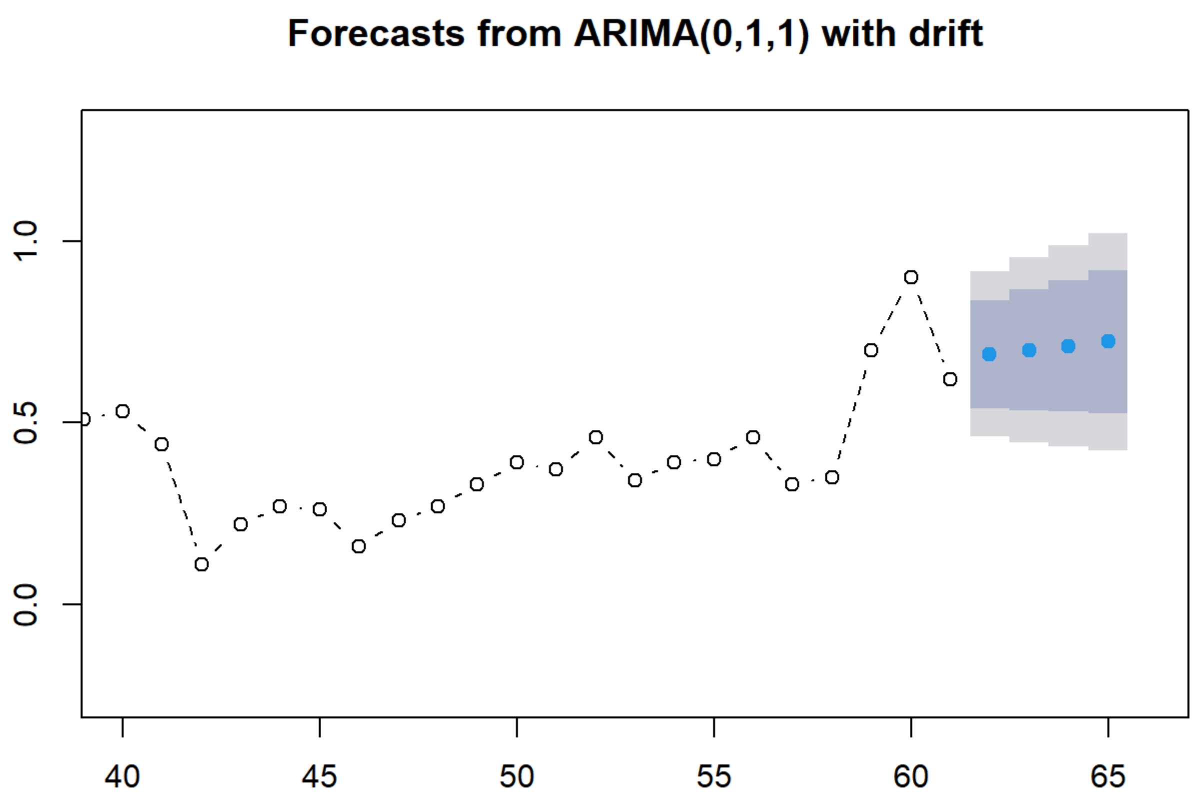
```
# p value is larger than 0.05 so we cannot refuse H0: the first 12 lags of residuals'ACF are all zero, thus is white noise.
```

```
tsdiag(md3)
```




```
# The standardized residuals are basically distributed near the zero horizontal line;
# the autocorrelation function quickly drops to within the two dotted lines;
# the P values of the Ljung-Box statistics are all greater than 0.05
# therefore, the model passes the test.
```

```
# 模型预测
# (1) Arima函数 对应 forecast::forecast
fore.gnp = forecast::forecast(md3,4) # 后四列为置信区间
plot(fore.gnp, lty=2, pch=1, type='b',xlim=c(40,66),ylim=c(-0.25,1.3))
```




```
# lines(fore.gnp$fitted, col=2, pch=2, type='b')
```

```
# (2) arima函数 对应 predict, 实际上arima函数未考虑drift, 准确度不如Forecast包的Arima函数, 此处仅作画图演示
md3_2 = arima(ts3, order = c(0,1,1), method = 'ML')
fore=predict(md3_2, 4)
fore
```

```
## $pred
## Time Series:
## Start = 62
## End = 65
## Frequency = 1
## [1] 0.6753347 0.6753347 0.6753347 0.6753347
##
## $se
## Time Series:
## Start = 62
## End = 65
## Frequency = 1
## [1] 0.1165388 0.1334200 0.1483930 0.1619879
```

```
U=append(ts3[61],fore$pred+1.96*fore$se)
L=append(ts3[61],fore$pred-1.96*fore$se)

plot(40:61,ts3[40:61],xlim=c(40,66),ylim=c(-0.25,1.3),type="o",ylab="",xlab="",main="Forecastin
g of Earnings")
lines(61:65,append(ts3[61],fore$pred),type="o",col="red")
lines(61:65, U,type="l",col="blue")
lines(61:65, L,type="l",col="blue")
legend(x="topleft",c("Earnings","prediction"),lty=c(1,1),pch=c(1,1),col=c("black","red"))
```