**Time Series Analysis   
First Assignment**

**(data : Amazone Stock Price)**

**학과 : 경영학과**

**학번 : 201143184**

**이름 : 박동규**

options(digits=4)

library(astsa)

library(TSA)

library(forecast)

**##1. simulation data**

x0 <-

read.csv("C:\\Users\\dgpar\\Desktop\\data\_science\\lecture\_note\\data\\AMZN.csv")

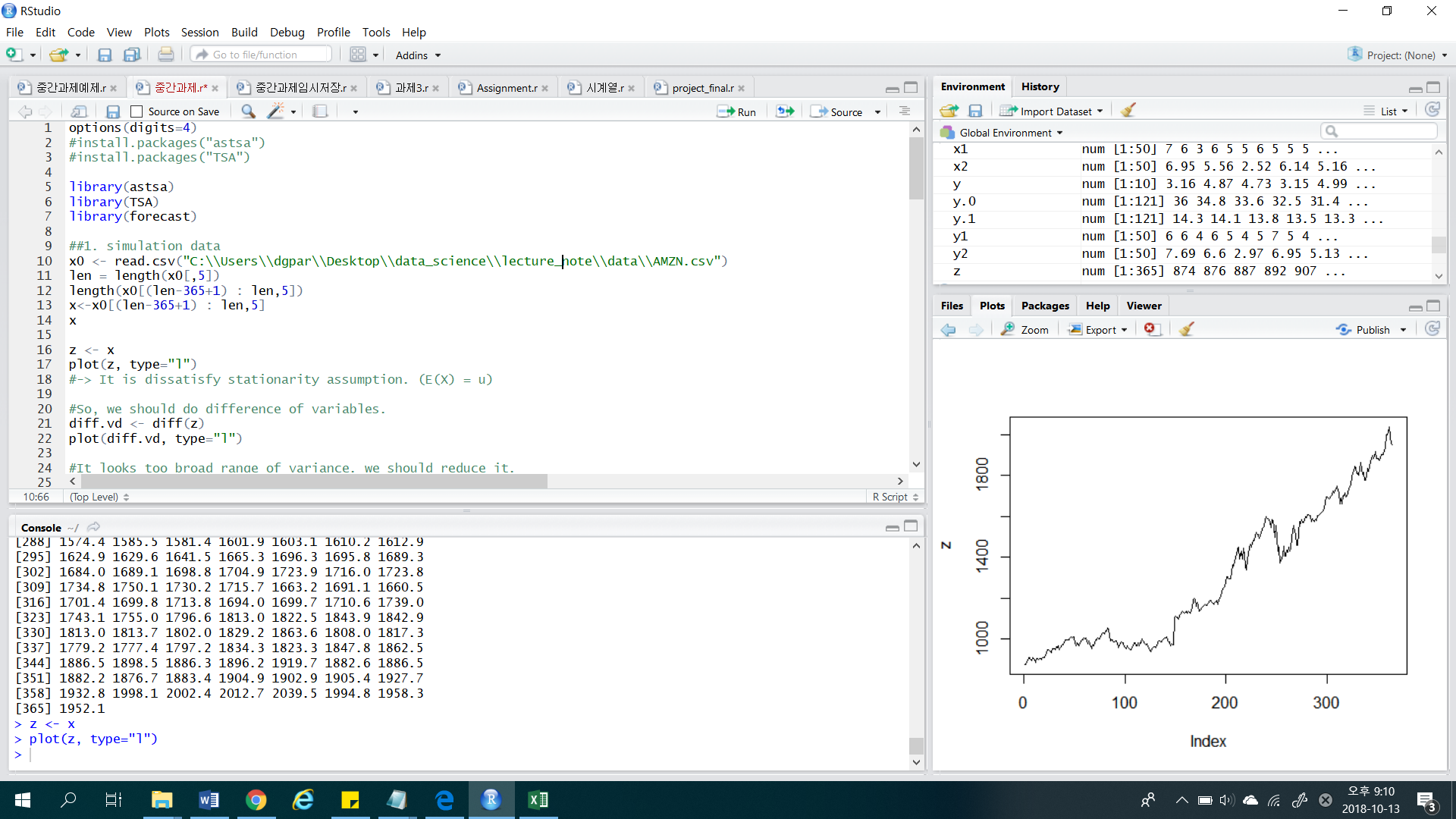
# 최근 1년치(365일) 데이터를 추출한다.

len = length(x0[,5])

length(x0[(len-365+1) : len,5])

z<-x0[(len-365+1) : len,5]

plot(z, type="l")



-> 그래프 해석 : 해당 그래프는 정상시계열 가정( E(X) = u, 등분산성 )을 위반한다. 따라서, 변수변환과 차분을 시행하도록 한다.

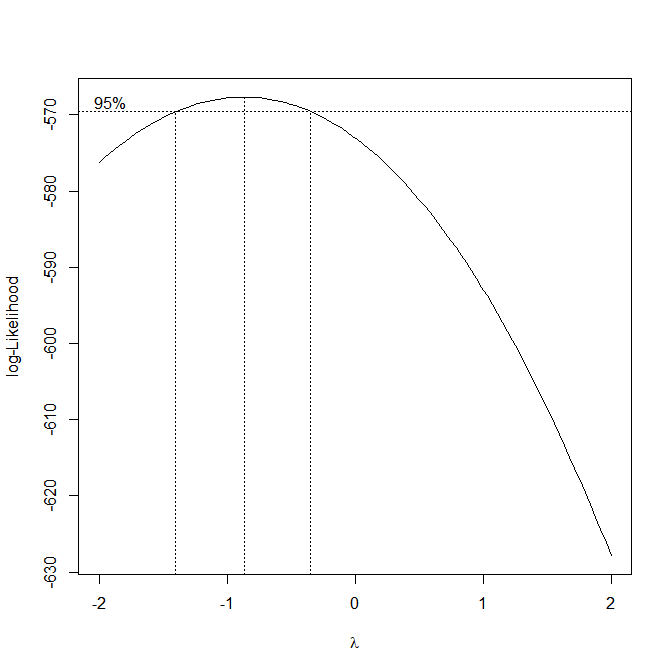
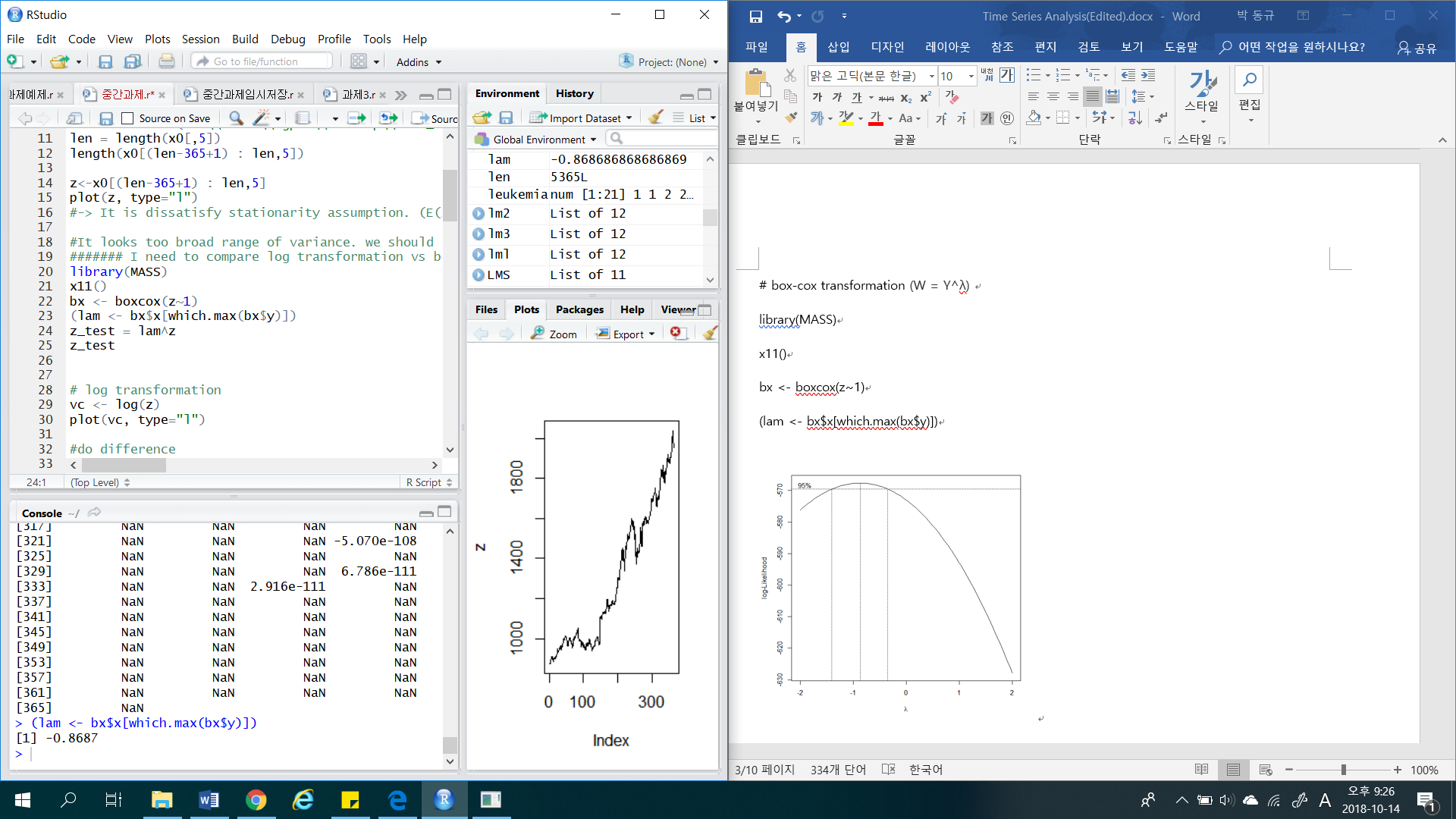
# box-cox transformation (W = Y^λ)

library(MASS)

x11()

bx <- boxcox(z~1)

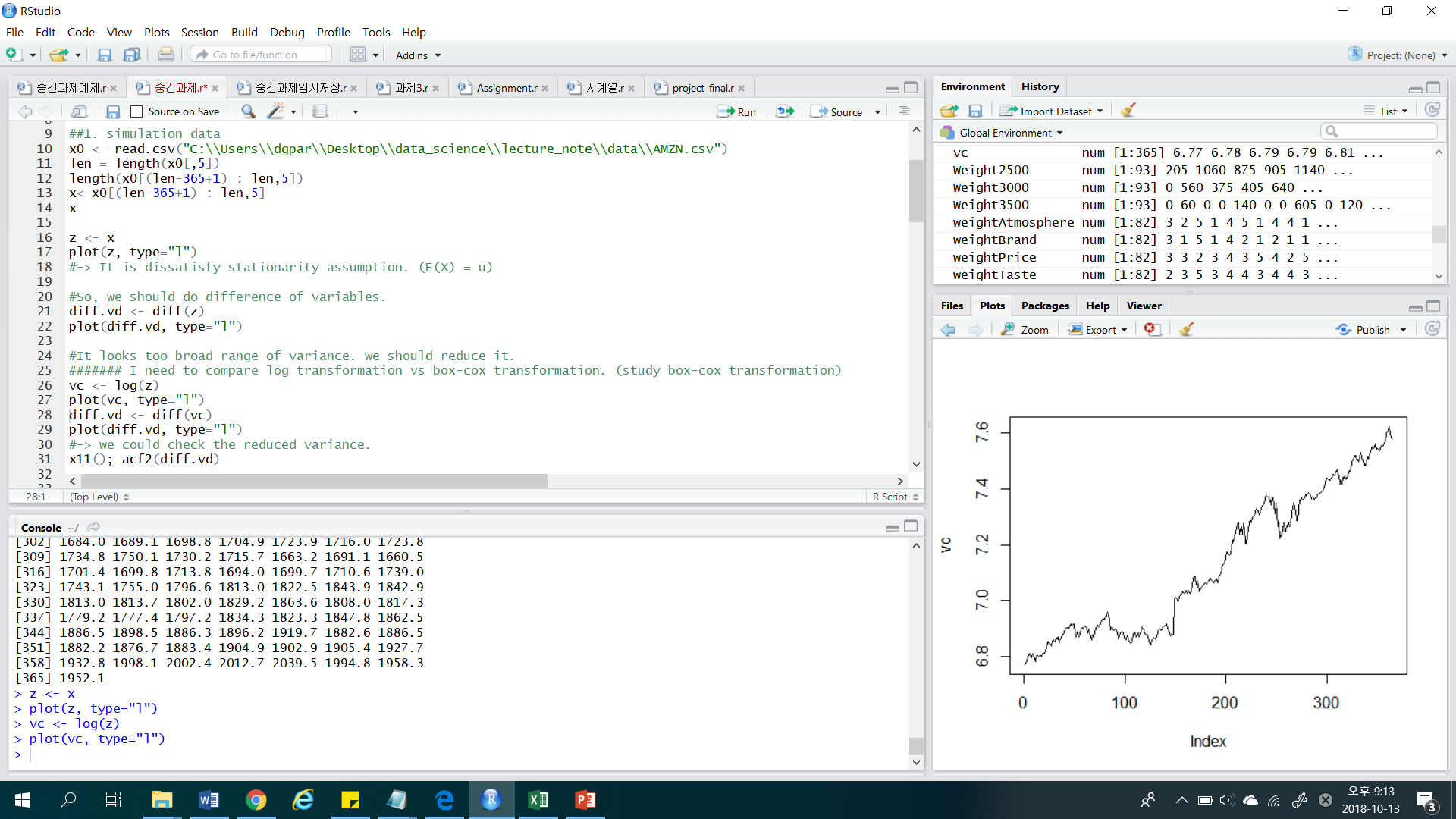
(lam <- bx$x[which.max(bx$y)])



# 로그변환을 사용하여 변수변환을 실시.

vc <- log(z)

plot(vc, type="l")

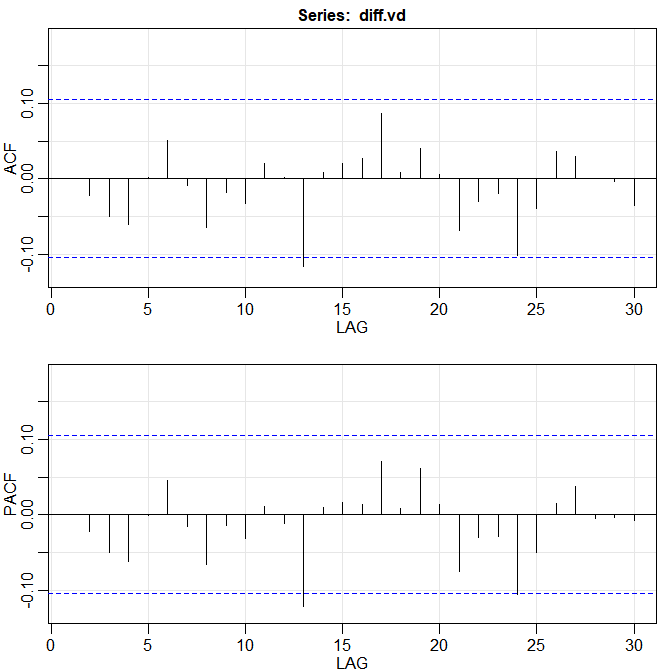
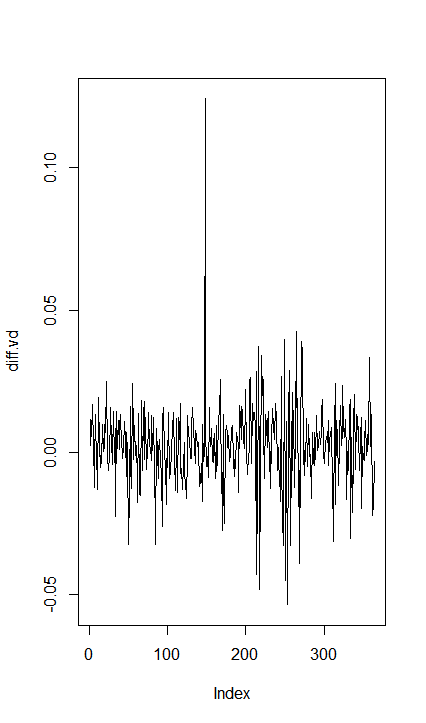


-> 그래프 해석 : 값들의 분산이 많이 줄어든 것으로 보인다. 하지만, E(X) = u의 조건을 만족하지 못하므로 차분을 통해 정상 시계열로 변환을 시도한다.

#차분 실시

diff.vd <- diff(vc)

plot(diff.vd, type="l"); x11(); acf2(diff.vd)

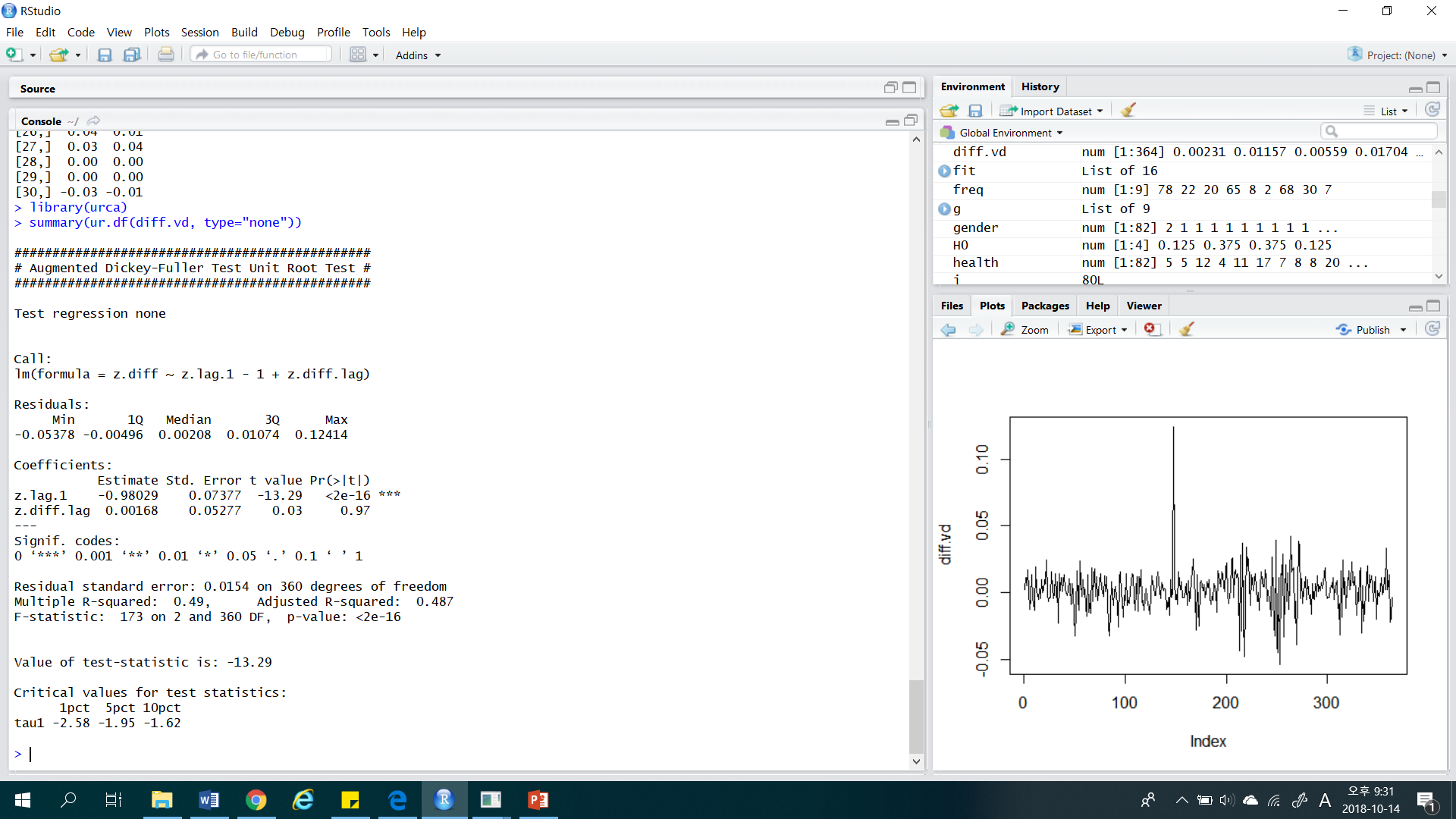


# 1차 차분의 단위근 검정 실시

# check unit root test.

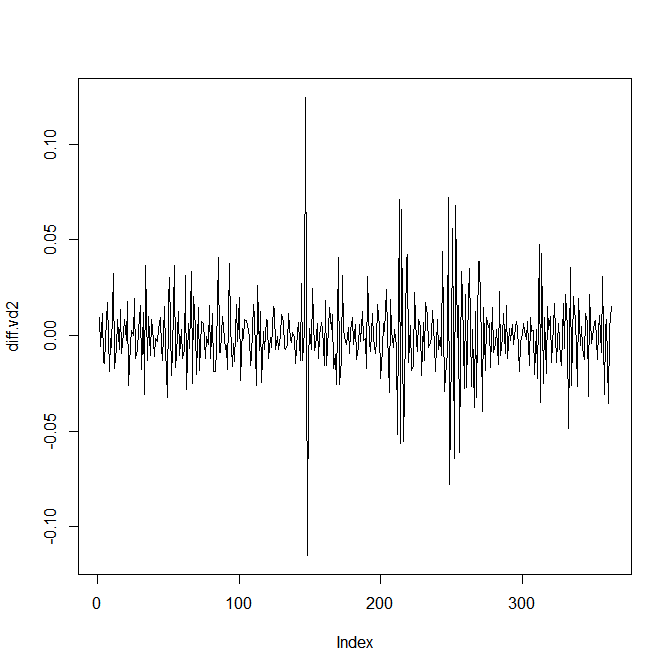
library(urca)

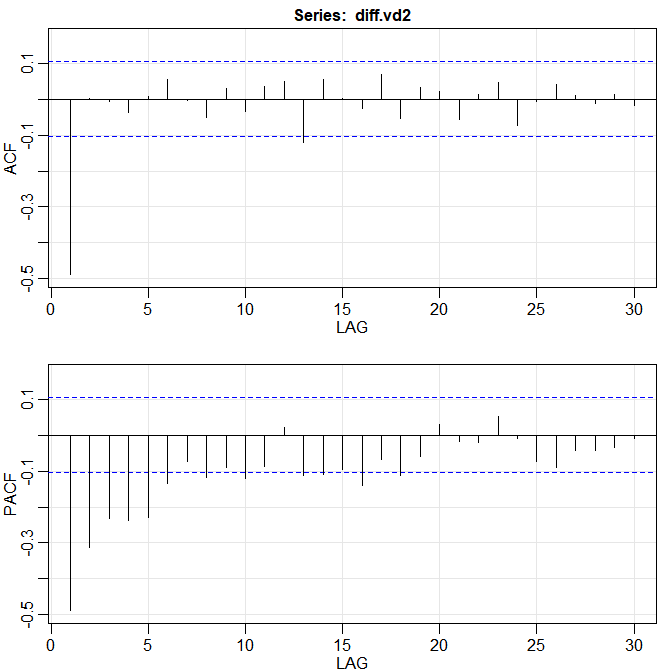
summary(ur.df(diff.vd, type="none"))



#-> 단위근 검정 결과, p-value가 0.05이하로 차분을 하면 안되는 것으로 파악된다. 하지만, 원 데이터는 누가봐도 E(X) = u가 아닌 것 같아 차분을 해야 한다고 생각했다. 또한 1차 차분결과, ARIMA 모형의 모델을 추측하기 어려우므로 차분을 한번 더 진행해 보았다.

diff.vd2 <- diff(diff.vd)

plot(diff.vd2, type="l"); x11(); acf2(diff.vd2)



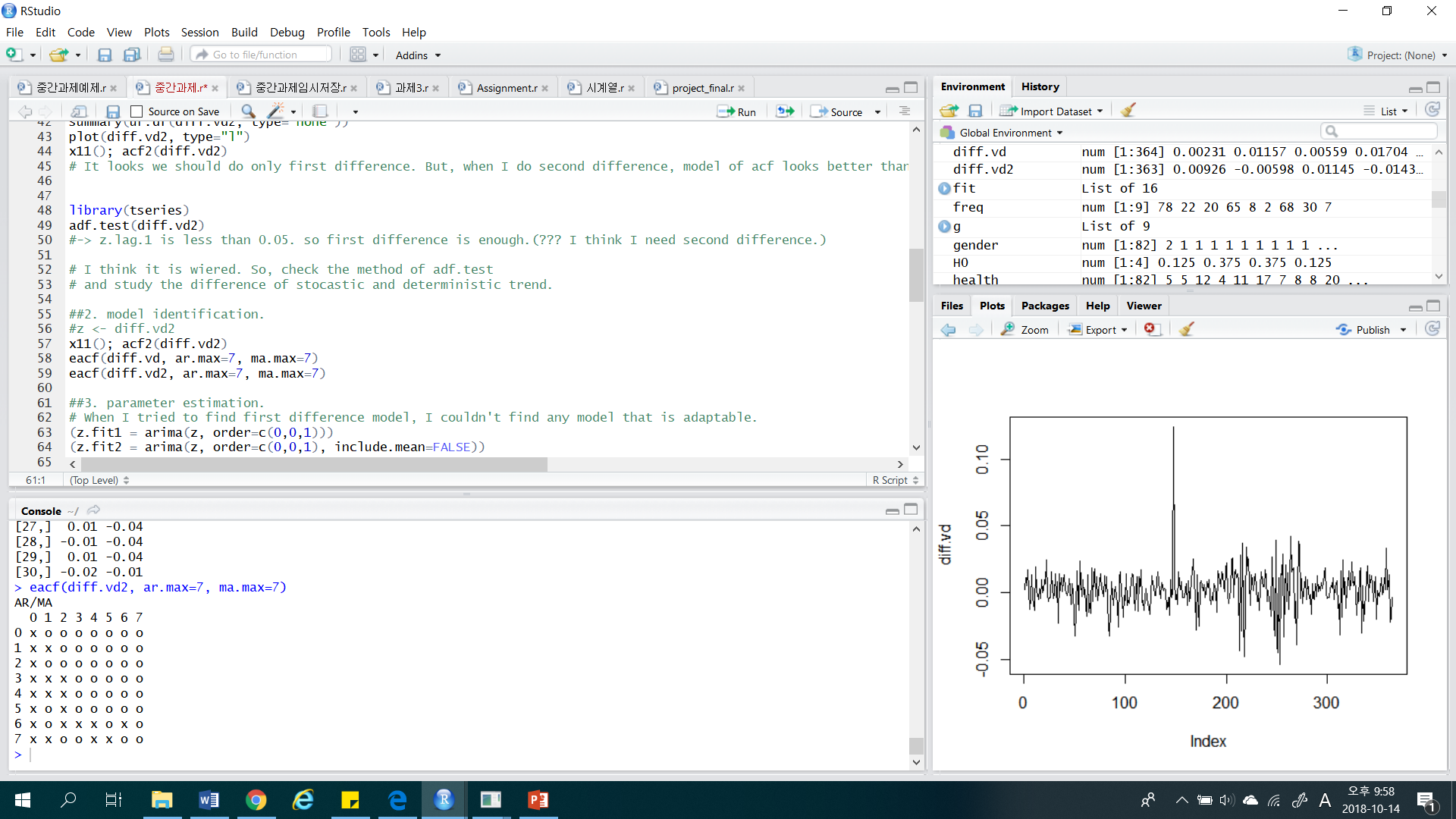
-> 2차 차분결과, 평균과 분산이 정상화 되었으며 acf와 pacf값이 안정적으로 나옴을 확인. 따라서, 2차 차분 데이터로 분석을 진행하였습니다.

**##2. model identification.**

z <- diff.vd2

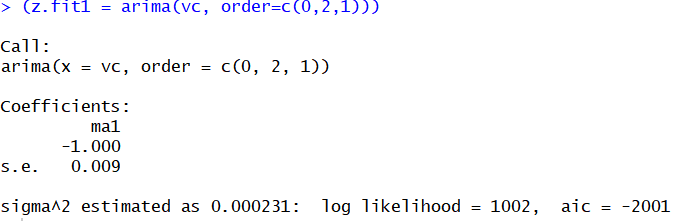
x11(); acf2(z)

eacf(z, ar.max=7, ma.max=7)

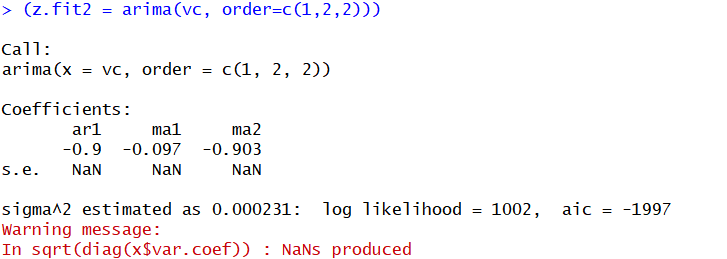


**##3. parameter estimation.**

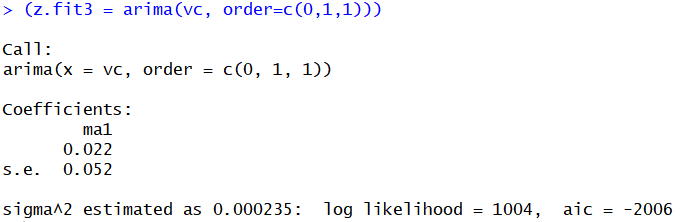
(z.fit1 = arima(vc, order=c(0,2,1)))



(z.fit2 = arima(vc, order=c(0,2,1), include.mean=FALSE))



(z.fit3 = arima(vc, order=c(1,2,2)))



#-> z.fit2, z.fit3의 경우 유효한 모델이 아니며 z.fit1이 유의하게 나왔으므로 해당 모델을 선택한다. 해당 모델의 경우 ma모델이 2 sigma 범위를 벗어나 유의한 모형이며, AIC값이 가장 작음을 확인 할 수 있다.

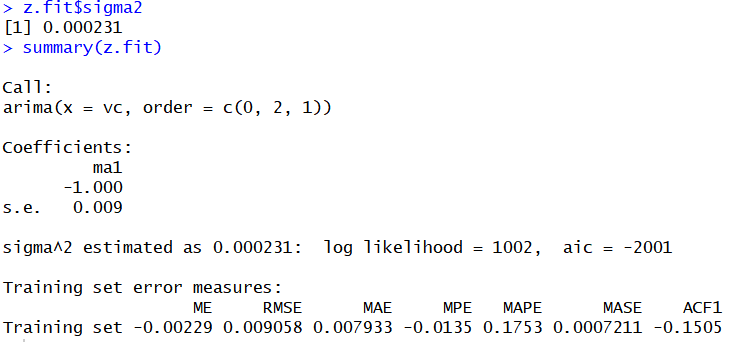
#final model

z.fit <- z.fit1

names(z.fit)

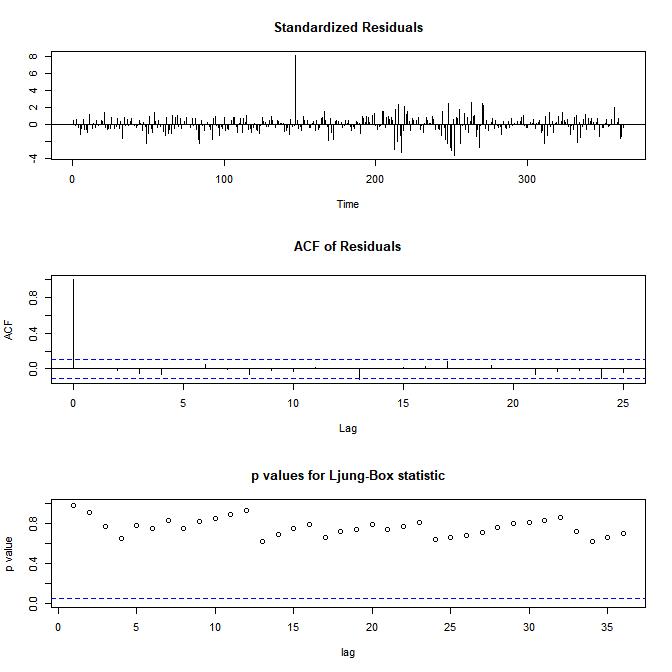
z.fit$sigma2

summary(z.fit)



##portmanteau test of autocorrelation within residuals

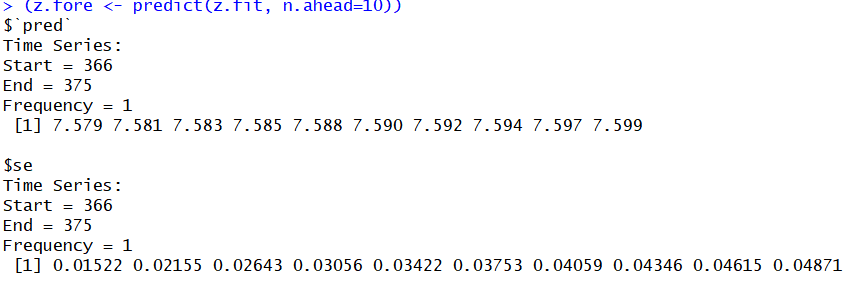
tsdiag(z.fit, gof.lag=36)



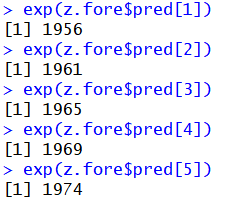
-> 해석 : 1) Standardized Residuals를 보면 대체로 안정된 것 같지만, 주가의 급등과 급락시 굉장히 큰 차이를 나타낸다. 이는 주가의 급격한 변화를 감지하기엔 한계점을 지니는 것으로 인식된다. 2) ACF의 경우 자기상관성을 나타내지 않으며 안정화 되어 있다. 3) p-value들도 상대적으로 큰값을 나타내며 유의한 모델임을 확인할 수 있다.

**##4. forecast**

(z.fore <- predict(z.fit, n.ahead=10))



예측값) 실제값)

1일후 : 1939.01

2일후 : 1987.15

3일후 : 1990.00

4일후 : 1989.87

5일후 : 1970.19

-> 예측값과 관측값을 살펴보면, 전체적인 추세는 맞음을 확인 할 수 있다. 하지만, 대략적인 추세 이상의 의미는 해당 분석만으로는 파악하기 어려워 보인다.

**##5. results**

#residual analysis

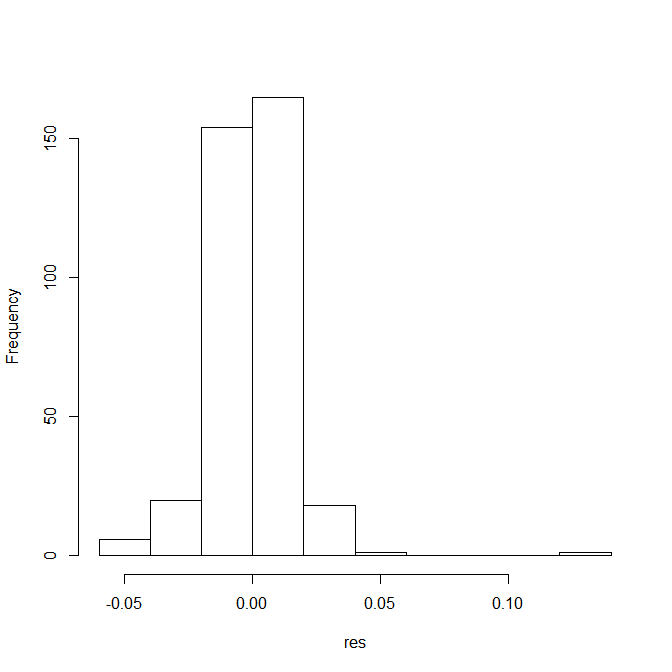
res = z.fit$res

summary(res)

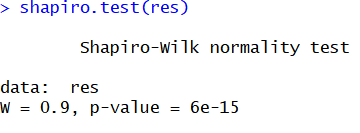


x11()

hist(res, main='')



shapiro.test(res)

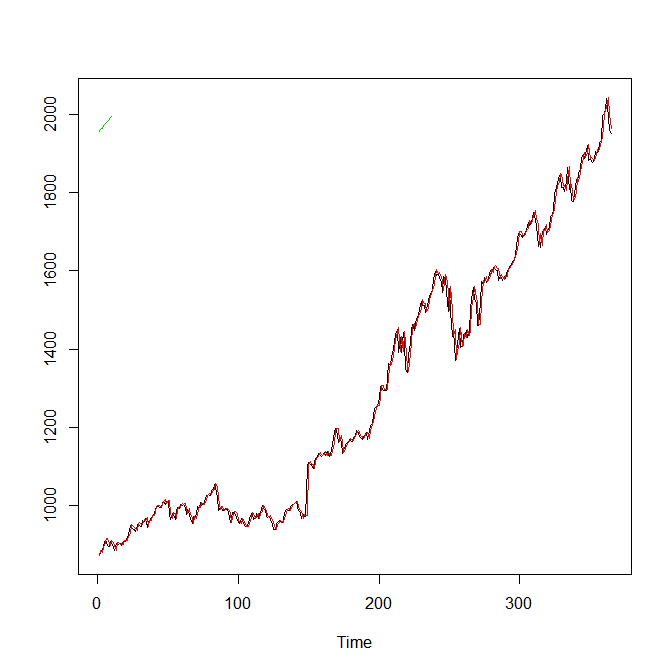


-> shapiro test 결과, 해당 모델이 정규분포를 나타내지는 않는다. 하지만 히스토그램을 살펴보면 대체로 중앙에 값들이 밀집되어 있음을 파악 할 수 있다.

# tsplots

est = z-res; x11();

ts.plot(ts(z), ts(est), ts(z.fore$pred), col=1:3, xlim=c(400,510), ylim=c(-5,5))



\* 최종 결과 해석 : 주가 데이터의 경우 2차 차분을 한 MA1 모델이 유효한 것으로 보인다. 하지만, 해당 분석만으로는 대략적인 추세 이외에 다른 정보는 발견하기 어려운 것으로 판단된다. 또한, 주가의 급등이나 급락을 발견하지 못하는 한계점을 파악 할 수 있다.