problem1

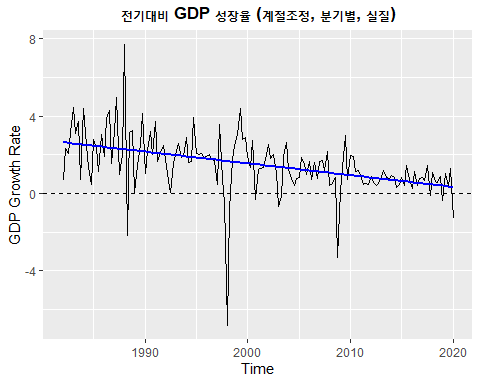
Harry Woo

2020 6 17

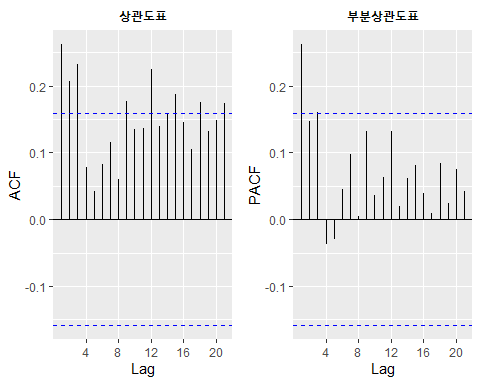
# Package Loading  
rm(list = ls())  
library(tseries)  
library(fabletools)  
library(forecast)  
library(grid)  
library(gridExtra)  
library(ggfortify)  
library(zoo)  
library(dplyr)  
theme\_t <- theme(plot.title = element\_text(size = 13, hjust = 0.5, vjust = 1.5, face = "bold"))  
  
# Data Loading  
gdp <- read.csv("gdpq.csv", header = TRUE)  
gdp\_gr <- ts(gdp[,2], start=1982, frequency=4) # Growth Rate

# Check stationarity of original data  
autoplot(gdp\_gr) + xlab("Time") + ylab("GDP Growth Rate") + theme\_t +   
 ggtitle(label = "전기대비 GDP 성장율 (계절조정, 분기별, 실질)") +  
 geom\_hline(yintercept = 0, lty = 2, col = "black") +  
 geom\_smooth(aes(x = as.Date(time(gdp\_gr)), y = gdp\_gr),  
 se = FALSE, method = "lm", lty = 1, col = "blue")

## `geom\_smooth()` using formula 'y ~ x'



acf1 <- ggAcf(gdp\_gr) + ggtitle("상관도표") + theme\_t  
pacf1 <- ggPacf(gdp\_gr) + ggtitle("부분상관도표") + theme\_t  
grid.arrange(acf1, pacf1, ncol = 2)



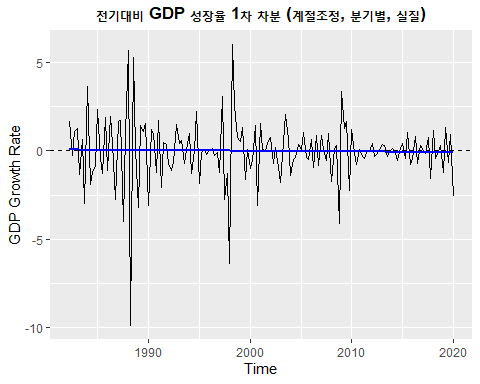
adf.test(gdp\_gr, alternative = "stationary", k = 0)

## Warning in adf.test(gdp\_gr, alternative = "stationary", k = 0): p-value smaller  
## than printed p-value

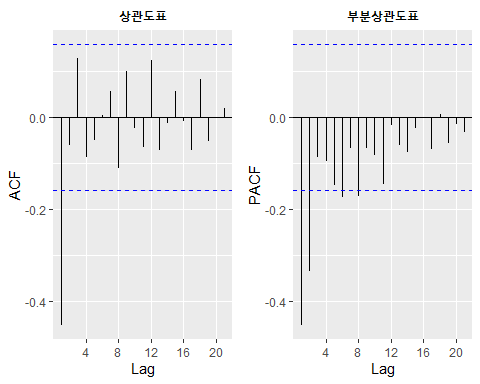
##   
## Augmented Dickey-Fuller Test  
##   
## data: gdp\_gr  
## Dickey-Fuller = -11.27, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary

# Check stationarity of differenced data  
autoplot(diff(gdp\_gr)) + xlab("Time") + ylab("GDP Growth Rate") + theme\_t +  
 ggtitle(label = "전기대비 GDP 성장율 1차 차분 (계절조정, 분기별, 실질)") +  
 geom\_hline(yintercept = 0, lty = 2, col = "black") +  
 geom\_smooth(aes(x = as.Date(time(diff(gdp\_gr))), y = diff(gdp\_gr)),  
 se = FALSE, method = "lm", lty = 1, col = "blue")

## `geom\_smooth()` using formula 'y ~ x'



acf2 <- ggAcf(diff(gdp\_gr)) + ggtitle("상관도표") + theme\_t  
pacf2 <- ggPacf(diff(gdp\_gr)) + ggtitle("부분상관도표") + theme\_t  
grid.arrange(acf2, pacf2, ncol = 2)



adf.test(diff(gdp\_gr), alternative = "stationary", k = 0)

## Warning in adf.test(diff(gdp\_gr), alternative = "stationary", k = 0): p-value  
## smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff(gdp\_gr)  
## Dickey-Fuller = -19.882, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary

data.frame("(0,1,1)" = Arima(gdp\_gr, order = c(0, 1, 1))$aic,  
 "(2,1,0)" = Arima(gdp\_gr, order = c(2, 1, 0))$aic)

## X.0.1.1. X.2.1.0.  
## 1 539.1278 564.2766

# ARIMA fitting  
gdpp\_fit = Arima(gdp\_gr, order=c(0,1,1))  
gdpp\_fit

## Series: gdp\_gr   
## ARIMA(0,1,1)   
##   
## Coefficients:  
## ma1  
## -0.9340  
## s.e. 0.0242  
##   
## sigma^2 estimated as 1.966: log likelihood=-267.56  
## AIC=539.13 AICc=539.21 BIC=545.18

auto.arima(gdp\_gr, seasonal = FALSE)

## Series: gdp\_gr   
## ARIMA(1,1,1)   
##   
## Coefficients:  
## ar1 ma1  
## 0.1037 -0.9414  
## s.e. 0.0849 0.0224  
##   
## sigma^2 estimated as 1.96: log likelihood=-266.82  
## AIC=539.63 AICc=539.8 BIC=548.7

# Empty data frame  
df <- data.frame(p = double(), d = double(), q = double(),   
 aic = double(), aicc = double(), bic = double())  
  
# 과대적합 체크  
for(p in 0:4){  
 for(d in 0:1){  
 for(q in 0:2){  
 df <- rbind(df, data.frame(p = p, d = d, q = q,   
 Arima(gdp\_gr, order = c(p, d, q))[c("aic", "aicc", "bic")]))  
 }  
 }  
}  
  
df[df$aic == min(df$aic),]

## p d q aic aicc bic  
## 5 0 1 1 539.1278 539.2084 545.1756

df[df$aicc == min(df$aicc),]

## p d q aic aicc bic  
## 5 0 1 1 539.1278 539.2084 545.1756

df[df$bic == min(df$bic),]

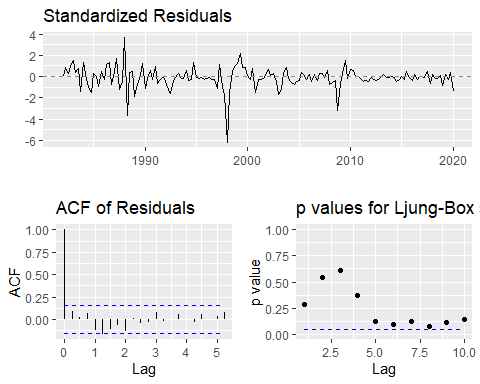
## p d q aic aicc bic  
## 5 0 1 1 539.1278 539.2084 545.1756

# Final Model  
gdpp\_fit = Arima(gdp\_gr, order=c(0,1,1))

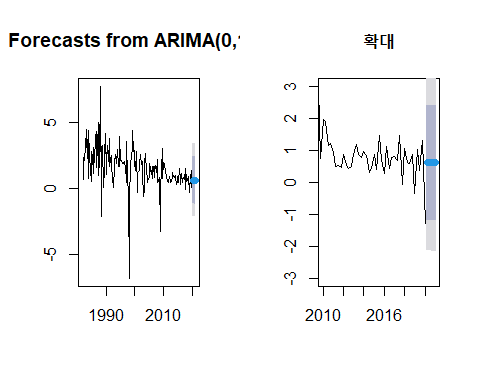
# Check residuals  
resdiag <- attributes(ggtsdiag(gdpp\_fit))

## Warning: `mutate\_()` is deprecated as of dplyr 0.7.0.  
## Please use `mutate()` instead.  
## See vignette('programming') for more help  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_warnings()` to see where this warning was generated.

# Grid  
grid.newpage()  
pushViewport(viewport(layout = grid.layout(2, 2)))  
define\_region <- function(row, col){  
 viewport(layout.pos.row = row, layout.pos.col = col)  
}   
print(resdiag$plots[[1]], vp = define\_region(1, 1:2))  
print(resdiag$plots[[2]], vp = define\_region(2, 1))  
print(resdiag$plots[[3]], vp = define\_region(2, 2))



# Forecasting  
  
xlim <- c(as.yearqtr("2010-01-01", format = "%Y-%m-%d"),  
 as.yearqtr("2021-01-01", format = "%Y-%m-%d"))  
  
par(mfrow = c(1,2))  
plot(forecast(gdpp\_fit, h = 4))  
plot(forecast(gdpp\_fit, h = 4), xlim = xlim, ylim = c(-3, 3), main = "확대")



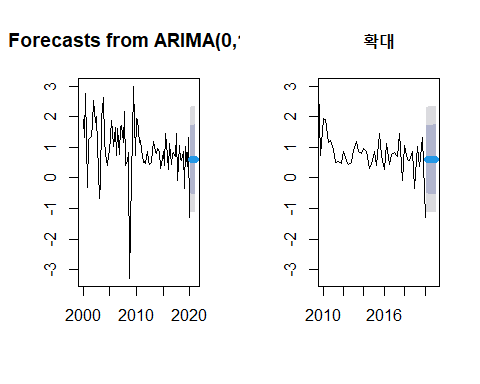
forecast(gdpp\_fit, h = 4)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2020 Q2 0.6101871 -1.186514 2.406888 -2.137630 3.358004  
## 2020 Q3 0.6101871 -1.190426 2.410801 -2.143614 3.363988  
## 2020 Q4 0.6101871 -1.194331 2.414705 -2.149584 3.369959  
## 2021 Q1 0.6101871 -1.198226 2.418600 -2.155542 3.375916

# Subsetting Time Series for further analysis  
gdp\_gr2 <- window(gdp\_gr, 2000, 2020)  
auto.arima(gdp\_gr2)

## Series: gdp\_gr2   
## ARIMA(0,1,1)   
##   
## Coefficients:  
## ma1  
## -0.9315  
## s.e. 0.0377  
##   
## sigma^2 estimated as 0.7686: log likelihood=-103.5  
## AIC=210.99 AICc=211.15 BIC=215.76

gdpp\_fit2 = Arima(gdp\_gr2, order = c(0, 1, 1))  
  
par(mfrow = c(1,2))  
plot(forecast(gdpp\_fit2, h = 4))  
plot(forecast(gdpp\_fit2, h = 4), xlim = xlim, main = "확대")



forecast(gdpp\_fit2, h=4)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2020 Q2 0.60006 -0.5234940 1.723614 -1.118268 2.318388  
## 2020 Q3 0.60006 -0.5261293 1.726249 -1.122298 2.322418  
## 2020 Q4 0.60006 -0.5287584 1.728878 -1.126319 2.326439  
## 2021 Q1 0.60006 -0.5313814 1.731501 -1.130330 2.330450

# 주석처리 모음  
  
# subsetting?  
# gdp\_gr2 <- window(gdp\_gr, 2000, 2020)  
  
# library(zoo)  
# gr <- as.zoo(gdp\_gr)  
# index(gr)  
  
# 성장률은 로그변환 후 차분과 근사함. 일종의 차분을 통하여 추세가 제거됨.  
# gghistogram(gdp\_gr, add.kde = TRUE)  
# shapiro.test(gdp\_gr)  
  
# ggseasonplot(gdp\_gr, year.labels = TRUE, year.labels.left = TRUE)  
# ggseasonplot(diff(gdp\_gr), year.labels = TRUE, year.labels.left = TRUE)  
  
# 모든 자기상관도표는 시차 0에서 acf=1  
# 점선이 유의수준에 따른 기각역  
# 륭-박스 검정은 m차까지 자기상관관계가 존재하지 않는다는 귀무가설  
  
# spectrum(gdp\_sa)  
# spectrum(gdp\_gr)