STAT153 HW5

Ziyuan Li

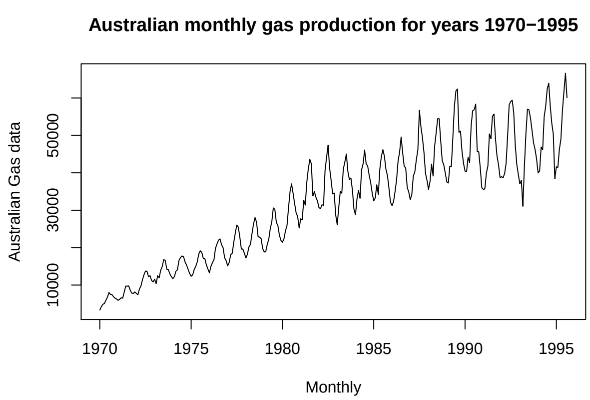
11/10/2021

1.Exploratory data analysis

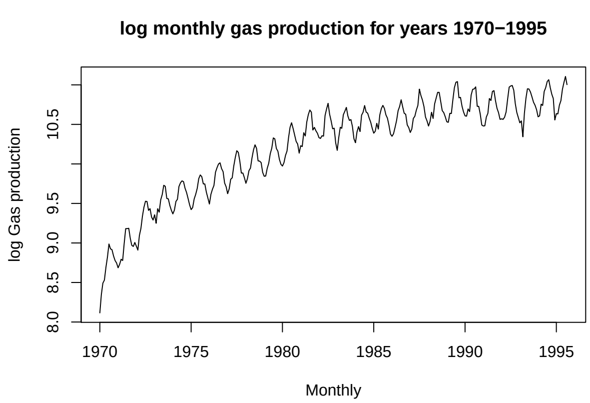
1a

```
rm(list = ls())
load("/Users/li/Desktop/STAT153/Hw5/gas_data.Rdata")

plot.ts(gas_data,type = "l", xlab = "Monthly", ylab = "Australian Gas data", main = "Australian monthly
```



The plot is not homoskedastic, so try $ V_t = f(Y_t)=log(Y_t)$ as variance stabilizing transform.It looks more homoskedastic.

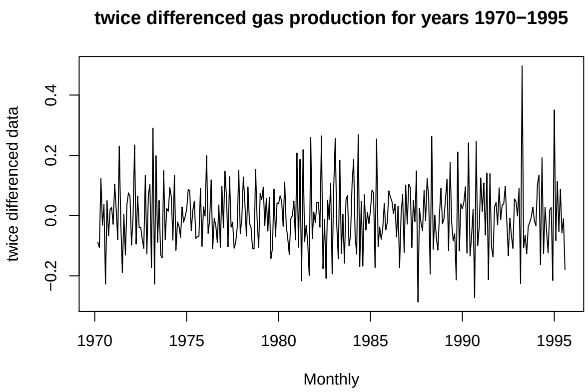


1b

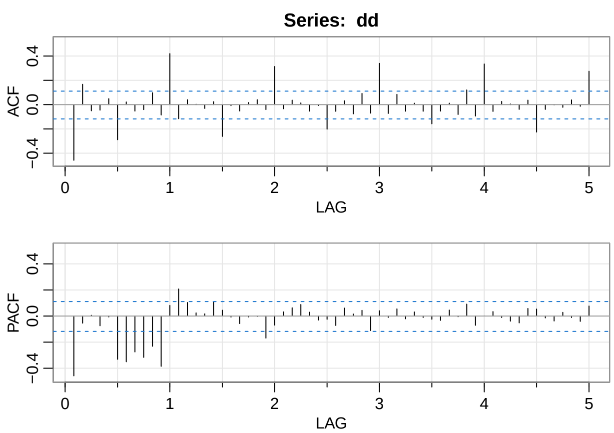
V_t is not stationarity because it's mean is related with t.

1c

```
dd = diff(diff(lggas))
plot.ts(dd,type = "l", xlab = "Monthly", ylab = "twice differenced data", main = "twice differenced gas
```




```
library(astsa)
acf2(dd,max.lag = 60)
```



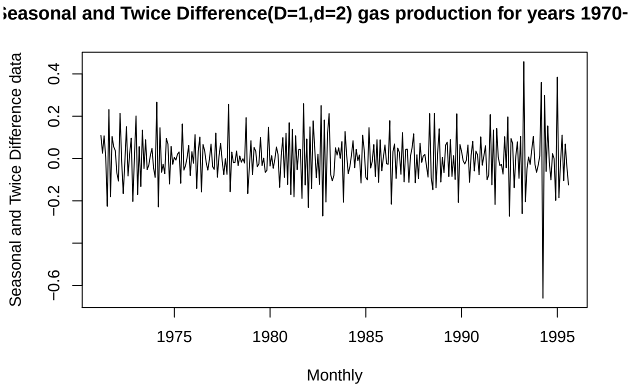
1)From the ACF we could see seasonal period h = 12 months.

2)Neither ACF and PACF has a reasonable cutoff. We probably have ARIMA(p,q) p > 0 and q > 0.

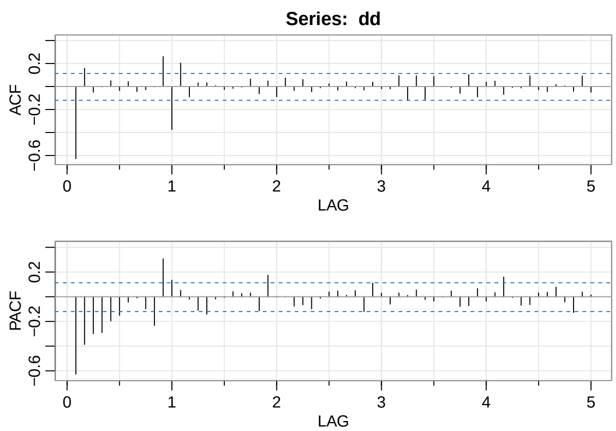
1d

Implement D=1,S=12,d=2:

```
dd = diff(diff(diff(lggas, lag=12)))
plot.ts(dd,type = "l", xlab = "Monthly", ylab = "Seasonal and Twice Difference data", main = "Seasonal a
```




```
acf2(dd,max.lag = 60)
```



ACF:

Large negative autocorrelation at lag 1 and then a bunch of small autocorrelations followed by large autocorrelations at lags 11, 12, 13 (the one at lag 12 is quite large), so it could be a $ARMA(0,1) \times (0,1)_{12}$

PACF:

Not very interpretable.

2. Model fitting and diagnostics

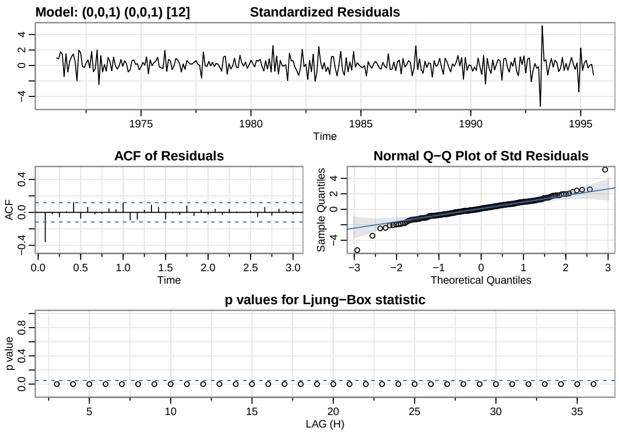
1

$ARMA(0,1) \times (0,1)_{12}$

Comment: The Ljung-Box-Pierce test shows the model fit well in-sample, which means the hypothesis that data x1, … ,xn was generated from a causal and invertible ARMA(p,q) model could not be rejected.

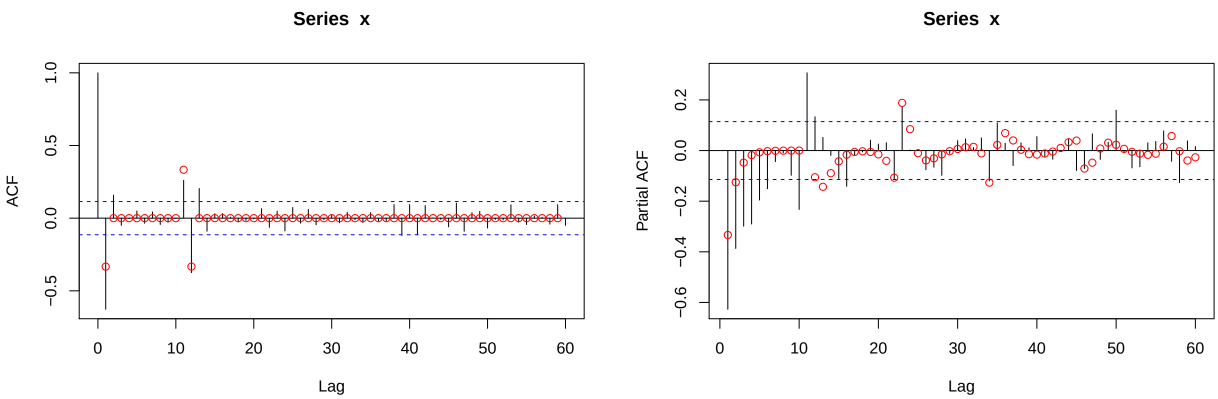
However, the sample PACF is not fitted with theoretical pacf.

```
model1 = sarima(dd,p=0,d=0,q=1,S=12,P=0,D=0,Q=1)
```

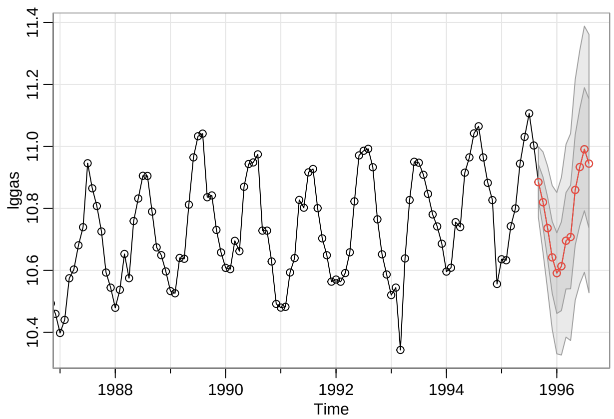



```
res = as.vector(dd)

#question 6
#cal acf pacf
th = c(model1$fit$coef[1],rep(0, 10), model1$fit$coef[2])
T = 60
corrs = ARMAacf(ma = th, lag.max = T)
par.corrs = ARMAacf(ma = th, lag.max = T, pacf = T)
all.plot = function(x,corr,pcorr){
  acf(x,lag.max = T)
  points(corr[2:60],col='red')
  pacf(x,lag.max = T)
  points(pcorr[1:60],col='red')
}
all.plot(res,corrs,par.corrs)
```




```
#Forecast
modelf1 = sarima.for(lggas,n.ahead = 12,p=0,d=2,q=1,S=12,P=0,D=1,Q=1)
```



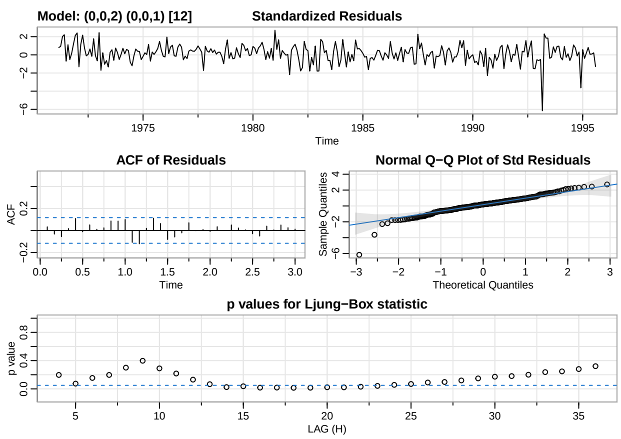
2

$ARMA(0, 2) \times (0, 1)_{12}$

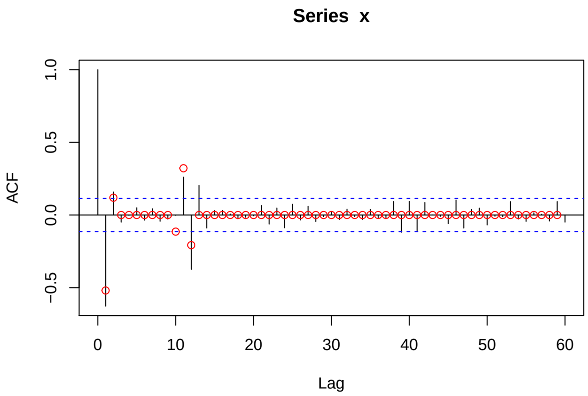
Comment: The Ljung-Box-Pierce test shows the model fit well in-sample, which means the hypothesis that data was generated from a causal and invertib ARMA(p,q) model could not be rejected.

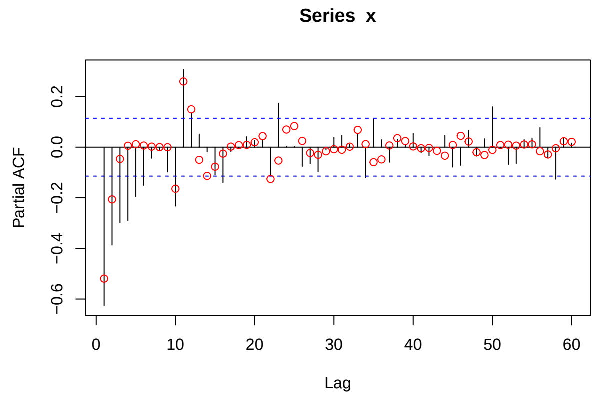
The sample ACF and PACF is fitted well with theoretical acf pacf.

```
model2 = sarima(dd,p=0,d=0,q=2,S=12,P=0,D=0,Q=1)
```

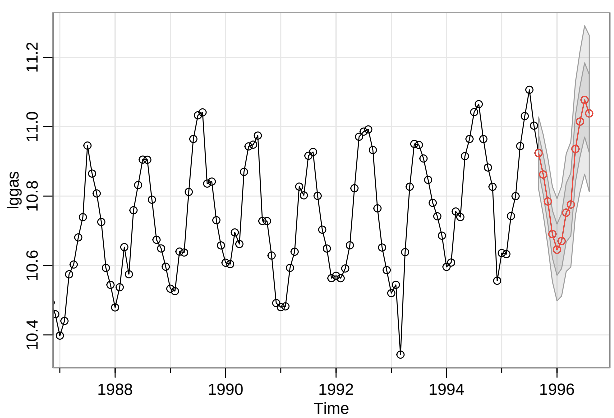



```
th = c(model2$fit$coef[1],model2$fit$coef[2],rep(0, 9), model2$fit$coef[3])
T = 60
corrs = ARMAacf(ma = th, lag.max = T)
par.corrs = ARMAacf(ma = th, lag.max = T, pacf = T)
all.plot(res,corrs,par.corrs)
```






```
modelf2 = sarima.for(lggas,n.ahead = 12,p=0,d=2,q=2,S=12,P=0,D=1,Q=1)
```

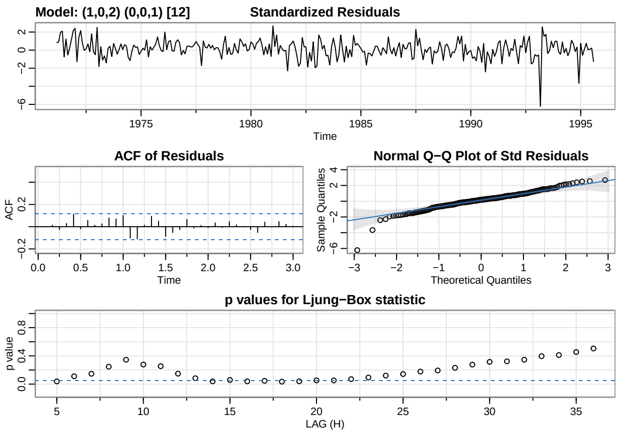


3

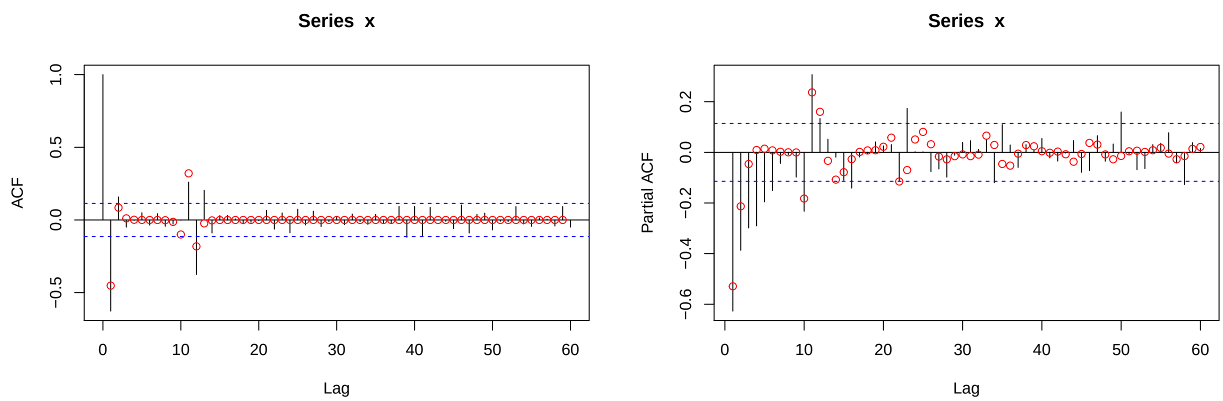
$ARMA(1,2)\times(0,1)_{12}$ Comment: The Ljung-Box-Pierce test shows the model fit worse than model1 and moel in-sample, which means the hypothesis that data was generated from a causal and invertible ARMA(p,q) model could not be rejected.

However, the sample PACF is not fitted with theoretical pacf.

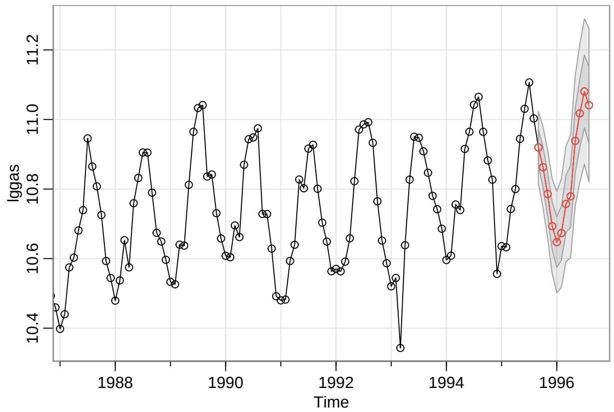
```
model3 = sarima(dd,p=1,d=0,q=2,S=12,P=0,D=0,Q=1)
```




```
#question 6
#cal acf pacf
ar1 = model3$fit$coef[1]
th = c(model3$fit$coef[2],model3$fit$coef[3],rep(0, 9), model3$fit$coef[4])
T = 60
corrs = ARMAacf(ar = ar1, ma = th, lag.max = T)
par.corrs = ARMAacf(ma = th, lag.max = T, pacf = T)
all.plot(res,corrs,par.corrs)
```




```
#Forecast
modelf1 = sarima.for(lggas,n.ahead = 12,p=1,d=2,q=2,S=12,P=0,D=1,Q=1)
```



3. Model selection

3a

```r
model1$AIC
model2$AIC
model3$AIC
print('AIC:Model 2 fit best')

model1$AICc
model2$AICc
model3$AICc
print('AICc:Model 2 fit best')
model1$BIC
model2$BIC
model3$BIC
print('BIC:Model 1 fit best')
```

```
## [1] -2.750871
## [1] -2.959146
## [1] -2.956979
## [1] "AIC:Model 2 fit best"
## [1] -2.750589
## [1] -2.958676
## [1] -2.95627
## [1] "AICc:Model 2 fit best"
## [1] -2.700754
## [1] -2.896501
## [1] -2.881804
## [1] "BIC:Model 1 fit best"
```

3b

cross-validation scores:

```
Model 1   Model 2   Model 3    Model 4
```

1 1.0236025 0.8169198 0.8612357 1.3123145

Model 2 yields the smallest cross-validation score.

```
sse = matrix(NA, nrow=11,ncol=3) # forecasting out 11 different times, with 3 models
for(i in 1:11){

  ## Split train/test

  train = window(lggas,start=1970,end=1984+i-.0001)
  test = window(lggas, start=1984+i, end=1984+i+.999)
  # if using a standard vector
  #train.test.split.point = 124+12*(i-1) # last point of train
  #train = l1[5:train.test.split.point]
  #test = l1[(train.test.split.point+1):(train.test.split.point+12)]

  ## Fit
  par(mfrow=c(3,1))
  model1 = sarima.for(train,n.ahead = 12,p=0,d=2,q=1,S=12,P=0,D=1,Q=1,plot = FALSE)
  model2 = sarima.for(train,n.ahead = 12,p=0,d=2,q=2,S=12,P=0,D=1,Q=1,plot = FALSE)
  model3 = sarima.for(train,n.ahead = 12,p=1,d=2,q=2,S=12,P=0,D=1,Q=1,plot = FALSE)

  ## Test
  sse[i,1] = sum((test - model1$pred)^2/12)
  sse[i,2] = sum((test - model2$pred)^2/12)
  sse[i,3] = sum((test - model3$pred)^2/12)
}
apply(sse,2,sum)
```

```
## [1] 0.08530021 0.06807665 0.07176964
```

```
#for lm
for(i in 1:11){
  #from 1970 to 1984 is 15years = 180 month
  train.test.split.point = 180 +12*(i-1) # last point of train
  train = dd[1:train.test.split.point]
  test = dd[(train.test.split.point+1):(train.test.split.point+12)]


  t = 1:180 +12*(i-1)
  lm1=lm(train ~ 1+t+I(t.^2)+I(t.^3)+I(t.%12))
  t = [length(train)+1:length(train)+12]

  predict(lm1,newdata= data.frame(1,t,I(t^2),I(t^3),I(t%12)))
  sse[i] = sum((test - predict)^2/12)
}
apply(sse,2,sum)
```