STAT 621 HW 3

Yingying Xu

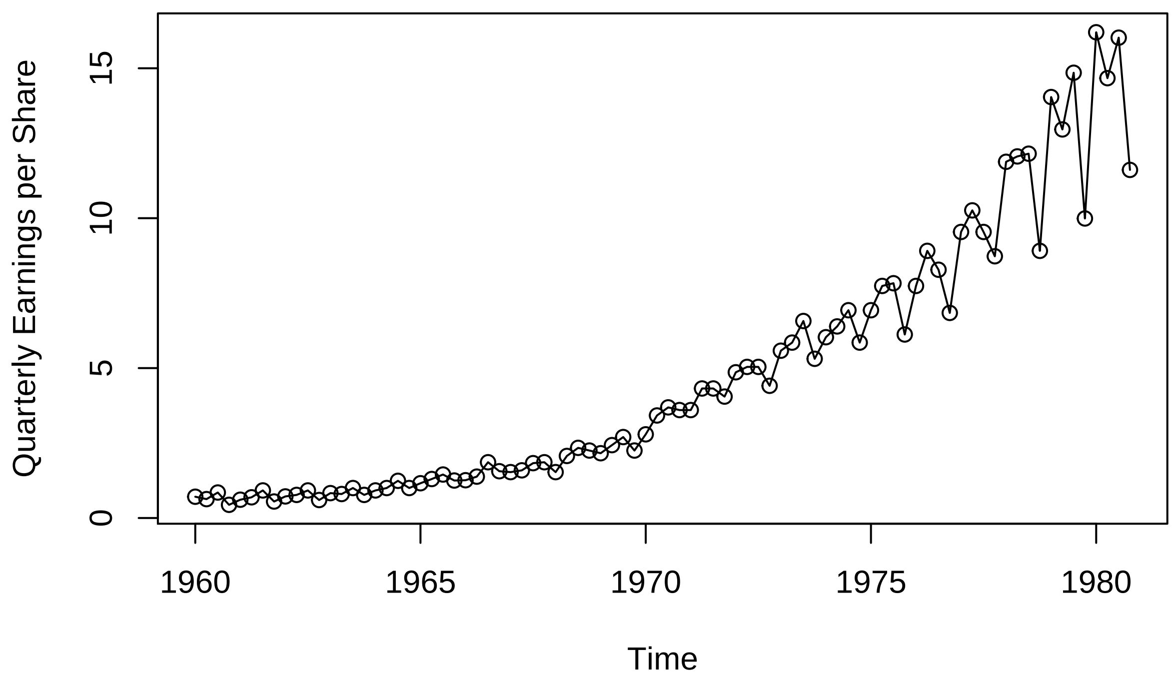
2/9/2018 (Due)

```
library(astsa)
library(ggplot2)
library(stats)
```

Q.1 Problem 2.1 Structural Model

a) Fit the regression model

```
plot(jj,type="o",ylab="Quarterly Earnings per Share")
```




```
trend = time(jj)-1970        #center time
Q     = factor(cycle(jj))    #factor(quarter)
reg   = lm(log(jj)~0+trend+Q,na.action = NULL)
#model.matrix(reg)
summary(reg)
```

```
##
```

```
## Call:
## lm(formula = log(jj) ~ 0 + trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## trend 0.167172   0.002259   74.00   <2e-16 ***
## Q1    1.052793   0.027359   38.48   <2e-16 ***
## Q2    1.080916   0.027365   39.50   <2e-16 ***
## Q3    1.151024   0.027383   42.03   <2e-16 ***
## Q4    0.882266   0.027412   32.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

b) The estimated average annual increase in the logged earnings per share

The average annual increase in the logged earning per share is the coefficient of treand, which means every year the earning increases $0.167 * 100\% = 16.7\%$. In addition, we could find the annual increase is equal to $5.3\% + 8.1\% + 15.1\% - 11.8\% = 16.7\%$

c) 3rd to 4th quarter

Decreased from third quarter to fourth quarter by $115.1\% - 88.2\% = 26.9\%$

d) Include the intercept term

suppose to include an intercept term, then Q1 is omitted due to the collinearity. The intercept is the base line for Q1 and the coefficients for Q2, Q3, and Q4 become the increased percentage relative to Q1 (baseline). For example, $Q2\% = 8.1\% = 5.3\% + 2.8\%$. The coefficient of the trend 0.167 does not change.

```r
reg2 = lm(log(jj)~trend+Q,na.action = NULL)
summary(reg2)
```

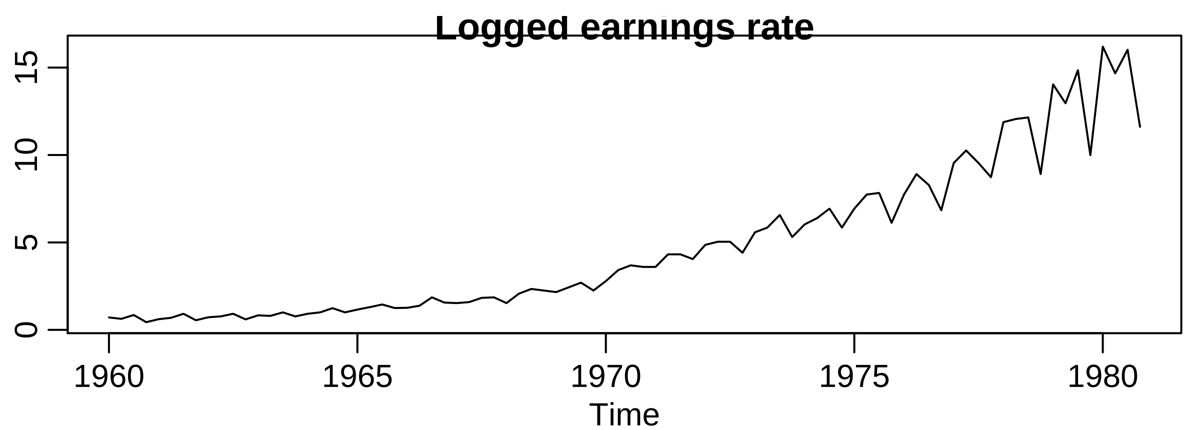
```
##
## Call:
## lm(formula = log(jj) ~ trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.052793   0.027359  38.480  < 2e-16 ***
## trend        0.167172   0.002259  73.999  < 2e-16 ***
```

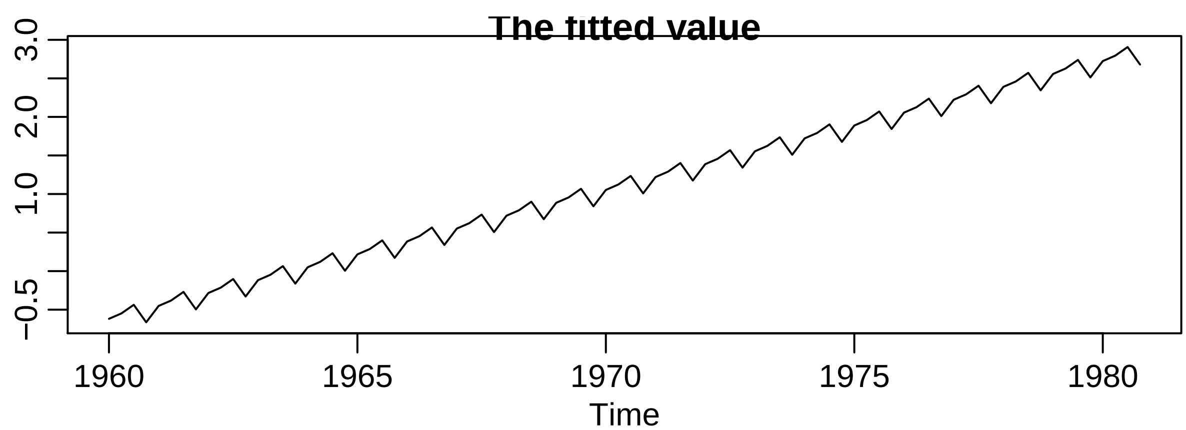
```
## Q2           0.028123   0.038696   0.727   0.4695
## Q3           0.098231   0.038708   2.538   0.0131 *
## Q4          -0.170527   0.038729  -4.403 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

e)

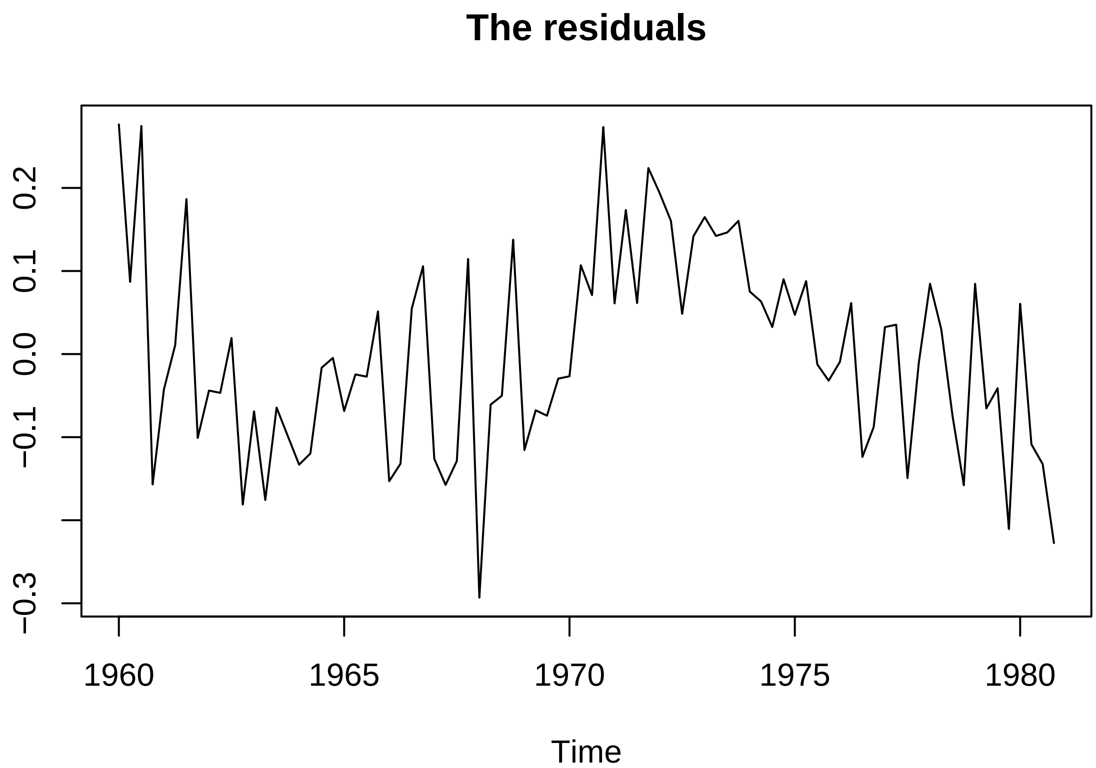
```
trend = time(jj)-1970       #center time
Q     = factor(cycle(jj))   #factor(quarter)
fit = lm(log(jj)~0+trend+Q,na.action=NULL)
resid_x = resid(fit)
x_fit = fitted(fit)

par(mfrow=c(2,1),mar=c(2.5,2.5,0,0)+.5,mgp=c(1.6,.6,0))
plot.ts(jj,ylab='',main="Logged earnings rate")
plot.ts(x_fit,ylab='',main="The fitted value")
```






```
plot.ts(resid_x,ylab='',main="The residuals")
```



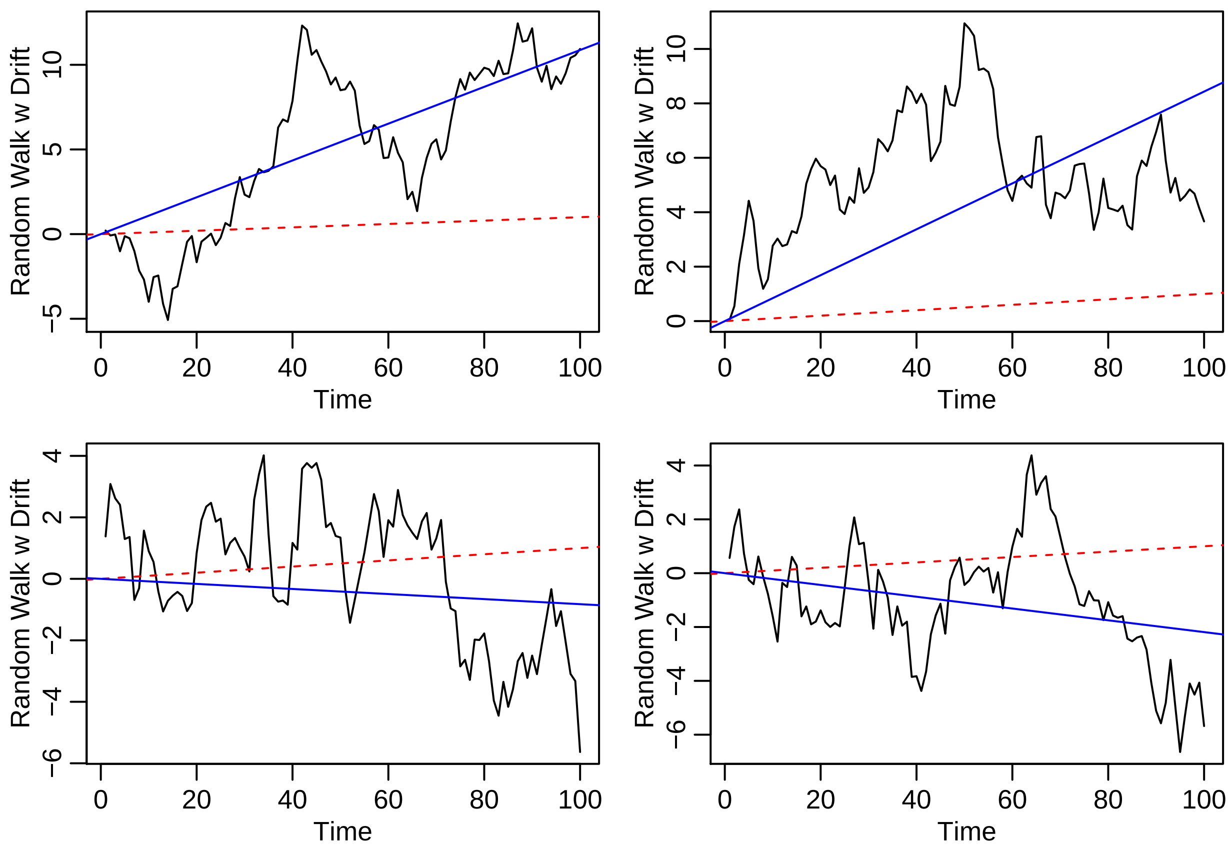
From the residual graph, the residuals look white noise. Therefore, it appears the model fits the data well. We may also use 1-2 lags autoregression model to test.

Q.2 Problem 2.3

a) random walk with drift

$x_t = \delta + x_{t-1} + w_t = 0.01t + \sum w_t$

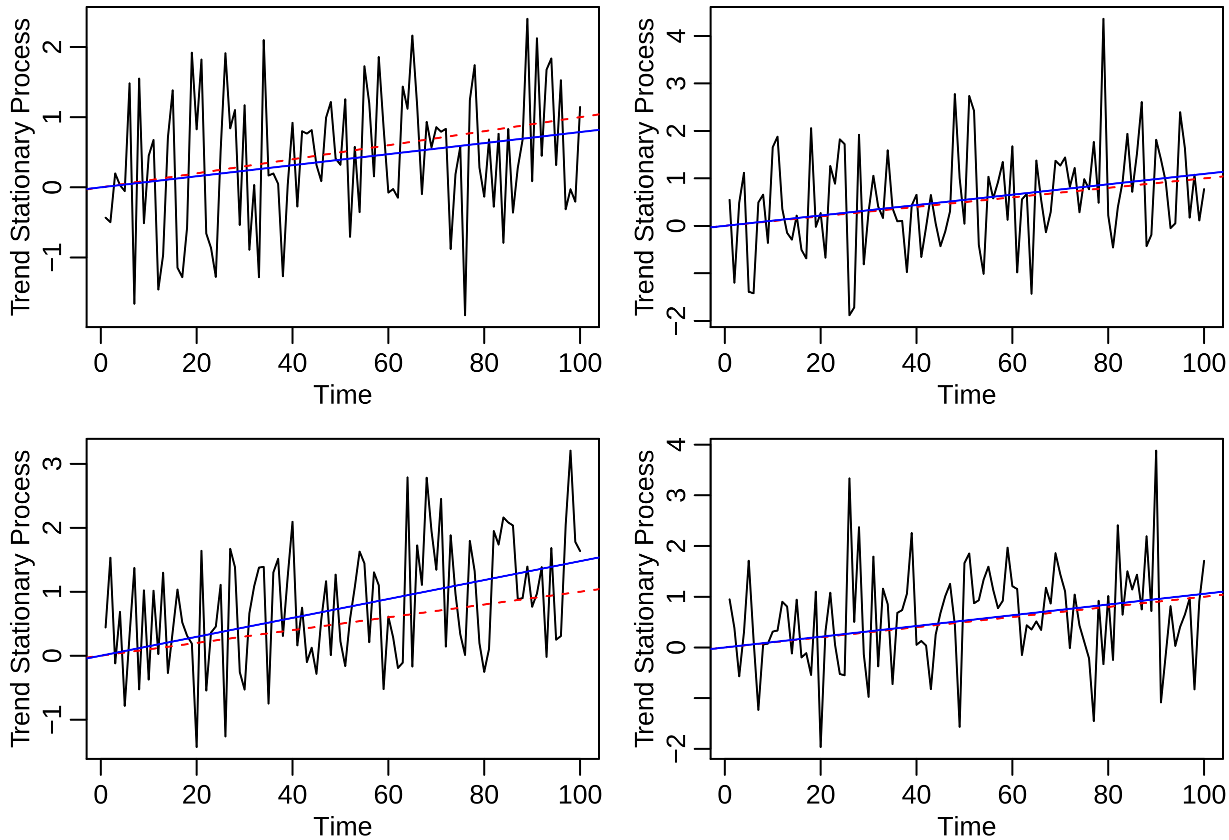
```
par(mfrow=c(2,2),mar=c(2.5,2.5,0,0)+.5,mgp=c(1.6,.6,0)) #set up
 for (i in 1:4){
   x=ts(cumsum(rnorm(100,.01,1)))       #data
   regx=lm(x~0+time(x),na.action=NULL) #regression
   plot(x,ylab='Random Walk w Drift')  #plots
    abline(a=0,b=.01,col=2,lty=2)      #true mean (red - dashed)
    abline(regx,col=4)                 #fitted line(blue - solid)
 }
```



b) trend stationary process

$y_t = \delta t + w_t = 0.01t + w_t$

```
par(mfrow=c(2,2),mar=c(2.5,2.5,0,0)+.5,mgp=c(1.6,.6,0)) #set up
 for (i in 1:4){
   w=rnorm(100,0,1)
   y=ts(0.01*(1:100)+w)                       #data
   regy=lm(y~0+time(y),na.action=NULL)        #regression
   plot(y,ylab='Trend Stationary Process')    #plots
    abline(a=0,b=.01,col=2,lty=2)             #true mean (red - dashed)
    abline(regy,col=4)                        #fitted line(blue - solid)
 }
```



c)

Compare Random walk with Drift and a trend stationary process, we can find the true mean and fitted line is almost the same for trend stationary process, but very different in the random walk case (non-stationary). Because of the non-stationary, the fitted line in random walk with drift is highly influenced by the $\sum w_t$. With the increase of t, the variance is explosion.

Q.3 Problem 2.10

```
head(oil)
```

```
## [1] 26.20 26.07 26.34 24.95 26.27 29.37
```

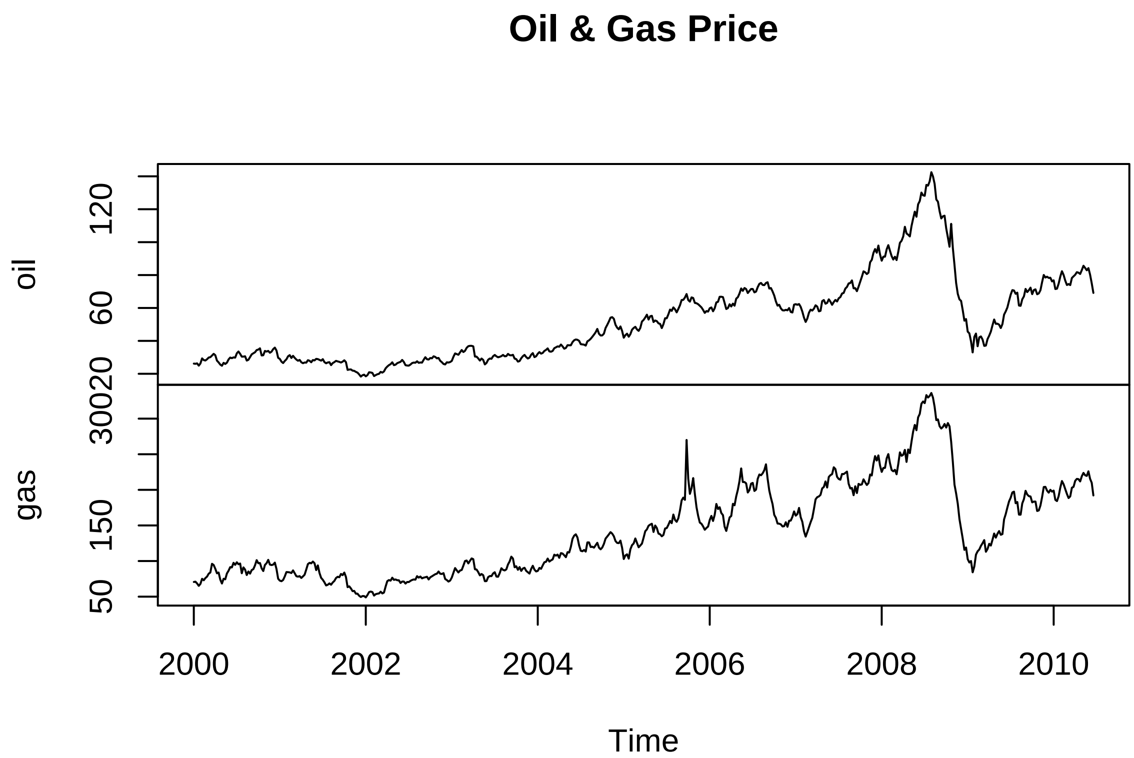
```
head(gas)
```

```
## [1] 70.636 71.040 68.490 65.137 67.918 75.117
```

a) Plot the data

The graph looks like autoregressions or moving average.

```
par(mfrow=c(2,1))
plot.ts(cbind(oil,gas),main="Oil & Gas Price")
```



b)

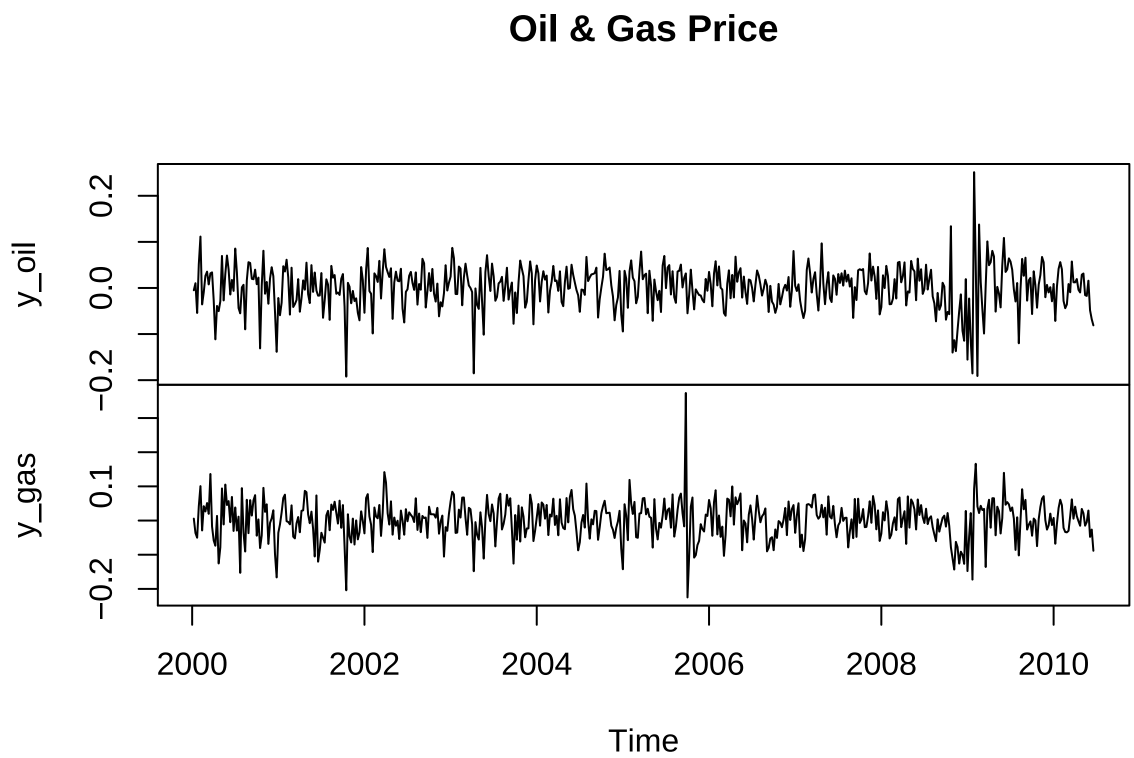
When the return is near zero, $log(1+r) = r - r^2/2 + r^3/3 - ... \approx r$. Therefore, the return can be defined as

$$r_t = (x_t - x_{t-1})/x_{t-1} \approx log(1+r_t) = log(x_t/x_{t-1}) = log(x_t) - log(x_{t-1}) = \nabla log(x_t) = y_t$$

c) ACFs

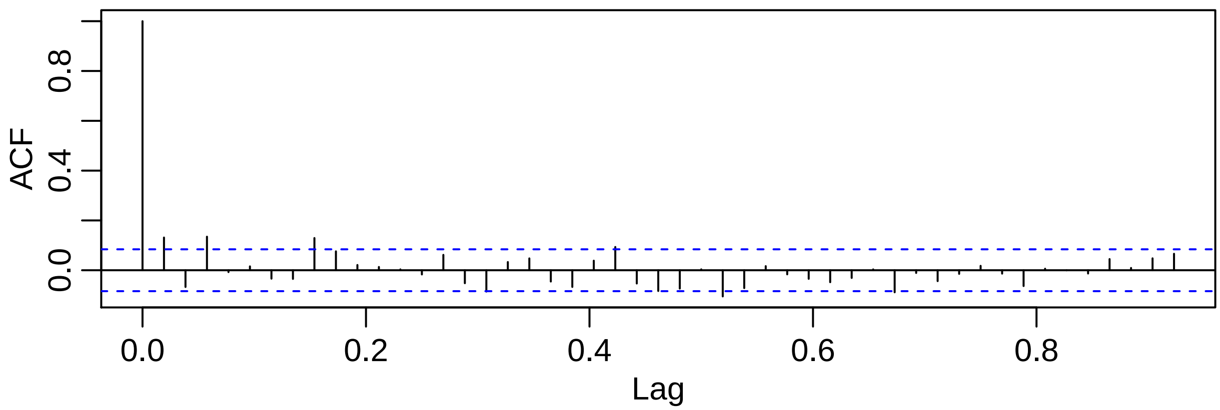
Transform the data

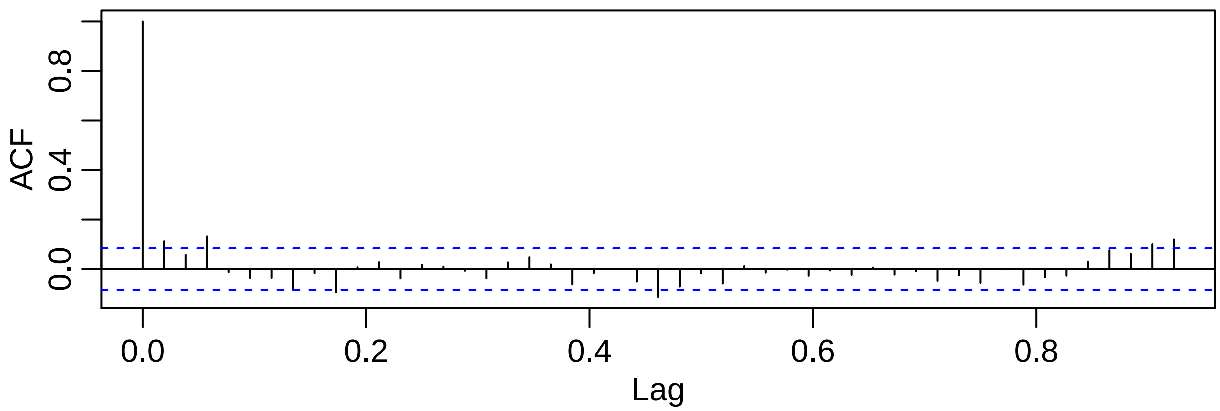
```
y_oil=diff(log(oil))
y_gas=diff(log(gas))
plot.ts(cbind(y_oil,y_gas),main="Oil & Gas Price")
```



The sample ACFs of the transformed data

```
par(mfrow=c(2,1),mar=c(2.5,2.5,0,0)+.5,mgp=c(1.6,.6,0))
acf(y_oil,48,main="Oil")
acf(y_gas,48,main="Gas")
```



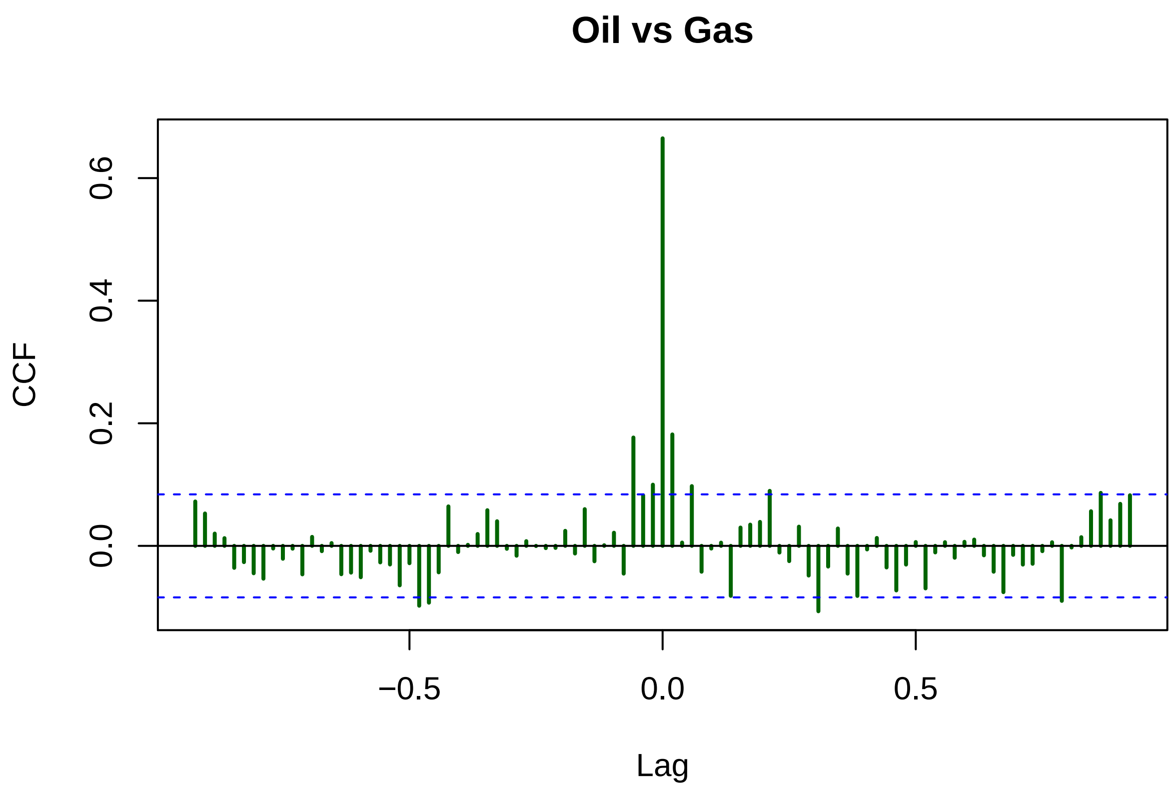


Comments: The ACF graph showing the growth rate of gas and oil might be MA(3) process.

d)

Plot the CCF

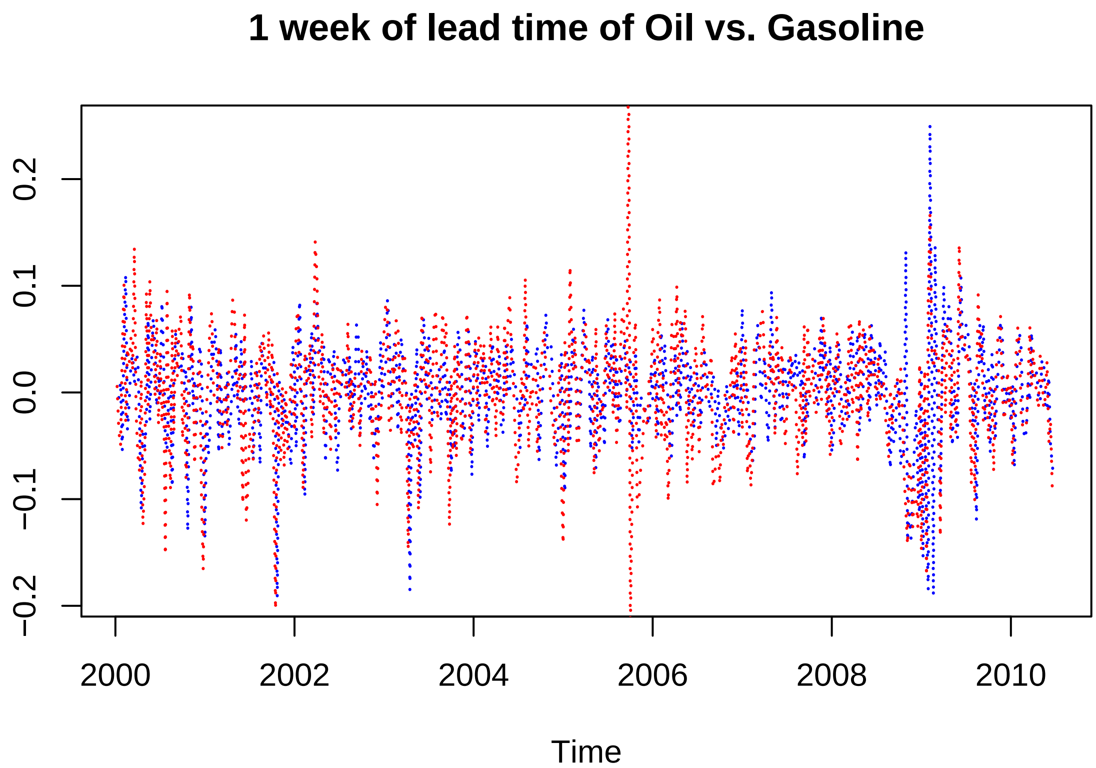
```
ccf(y_gas,y_oil,48,main="Oil vs Gas",ylab="CCF",lwd=2,col="darkgreen")
```



The left side is Gas leads Oil.

e) Scatterplots of the oil and gas growth rate serie

```
plot.ts(lag(y_oil,-1),lty=3,lwd=1.5,col="blue",ylab='', main="1 week of lead time of Oil vs. Gasoline")
lines(y_gas,lty=3,lwd=1.5,col="red",ylab='')
```



By trying different number (1,2,3 weeks) of lead of oil price, I find 1 week of lead time of oil price (growth rate) is almost the same rate as gasoline price. As shown in the graph, the blue (1 week lead time of oil) is of the same magnitude as the growth rate gasoline price.

f)

i) Fit the regression

```
poil=diff(log(oil))
pgas=diff(log(gas))
indi=ifelse(poil<0,0,1)
mess=ts.intersect(pgas,poil,poilL=lag(poil,-1),indi)
summary(fit<-lm(pgas~poil+poilL+indi,data=mess))
```

```
##
## Call:
## lm(formula = pgas ~ poil + poilL + indi, data = mess)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18451 -0.02161 -0.00038  0.02176  0.34342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006445   0.003464  -1.860  0.06338 .
```

```
## poil         0.683127   0.058369  11.704  < 2e-16 ***
## poilL        0.111927   0.038554   2.903  0.00385 **
## indi         0.012368   0.005516   2.242  0.02534 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04169 on 539 degrees of freedom
## Multiple R-squared:  0.4563, Adjusted R-squared:  0.4532
## F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

$$G_t = \alpha_1 + \alpha_2 I_t + \beta_1 O_t + \beta_2 O_{t-1} + w_t$$

The results shows the coefficients α_2,β_1 and β_2 are significant, which means the gasoline prices respond oil price in current period and the period before. The growth in oil price will have positive effects on gasoline price.

ii)

when there is negative growth in oil price at time t, the indiator $I_t = 0$.The model becomes

$$G_t = \alpha_1 + \beta_1 O_t + \beta_2 O_{t-1} + w_t = -0.006445 + 0.683127 O_t + 0.111927 O_{t-1}$$

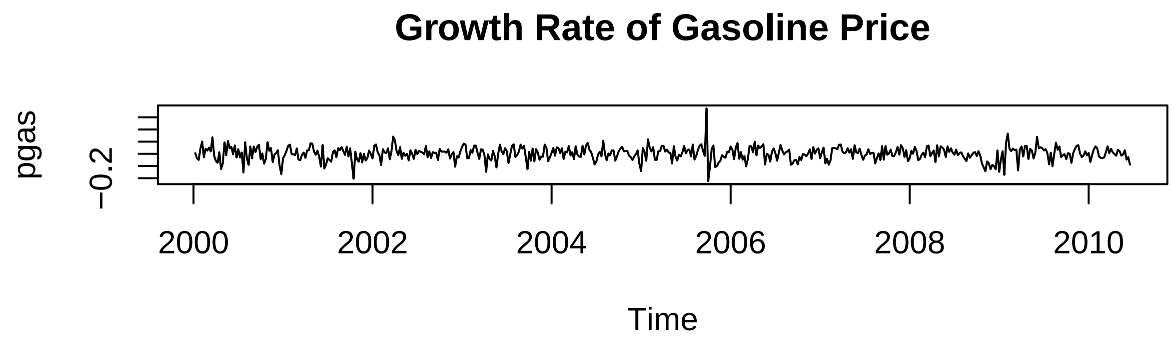
If there is positive growth in oil price, the model becomes

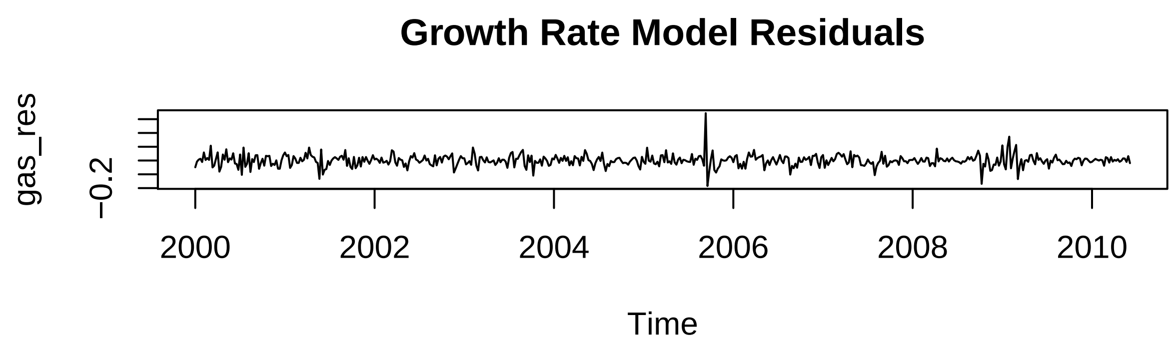
$$G_t = \alpha_1 + \alpha_2 + \beta_1 O_t + \beta_2 O_{t-1} + w_t = 0.005923 + 0.683127 O_t + 0.111927 O_{t-1}$$

Therefore, the fitted model is asymmetry.

iii) Analyze the residuals

```
#head(time(pgas))
gas_res = ts(resid(fit),start=c(2000,1),frequency=52)
par(mfrow=c(2,1))
plot.ts(pgas,main="Growth Rate of Gasoline Price")
plot.ts(gas_res,main="Growth Rate Model Residuals")
```





When there is great changes in the growth rate, the absolute value of residual also becomes big, suggesting this model is good for constant growth rate, but not a good fit to catch up big changes.

Q.4 Problem 2.6 Deterministic Trend

$$x_t = \beta_0 + \beta_1 t + w_t, w_t \sim (0, \sigma_w^2)$$

a) x_t is nonstationary

$E[x_t] = E[\beta_0 + \beta_1 t] = \beta_0 + \beta_1 t$, which is time-varying. Therefore, it is non-stationary. $Var[x_t] = Var[\beta_0 + \beta_1 t + w_t] = \sigma_w^2$, which is time-invariant. The trend may be considered as having stationary behavior around a linear trend.

b) Prove $\nabla x_t = x_t - x_{t-1}$ is stationary

$$z_t = \nabla x_t = x_t - x_{t-1} = (\beta_0 + \beta_1 t + w_t) - (\beta_0 + \beta_1(t-1) + w_{t-1}) = \beta_1 + w_t - w_{t-1}$$

mean $E[z_t] = E[\beta_1 + w_t - w_{t-1}] = \beta_1$ varaince $Var[z_t] = Var[\beta_1 + w_t - w_{t-1}] = Var[w_t] + Var[w_{t-1}] = 2\sigma_w^2$ autocovariance function:

$$\gamma_z(h) = cov(z_{t+h}, z_t) = E[(z_{t+h} - \beta_1)(z_t - \beta_1)] = E[(w_{t+h} - w_{t+h-1})(w_t - w_{t-1})]$$

If $h = 0$, $\gamma_z(h) = 2\sigma_w^2$; If $h = \pm 1$, $\gamma_z(h) = -\sigma_w^2$; If $h = else$, $\gamma_z(h) = 0$. The autocovariance function only depends on h. $z_t = \nabla x_t$ is stationary.

c) Test $x_t = \beta_0 + \beta_1 t + y_t$ when y_t is stationary

$$z_t = \nabla x_t = x_t - x_{t-1} = (\beta_0 + \beta_1 t + y_t) - (\beta_0 + \beta_1(t-1) + y_{t-1}) = \beta_1 + y_t - y_{t-1}$$

mean $E[\nabla x_t] = E[\beta_1 + y_t - y_{t-1}] = \beta_1 + \mu_y - \mu_y = \beta_1$, independent of time. varaince $Var[\nabla x_t] = Var[\beta_1 + y_t - y_{t-1}] = Var[y_t - y_{t-1}]$ autocovariance function

$$\gamma_z(h) = cov(z_{t+h}, z_t) = E[(z_{t+h} - \beta_1)(z_t - \beta_1)] = E[(y_{t+h} - y_{t+h-1})(y_t - y_{t-1})]$$

Since y_t is a stationary process, the autocovariance only depends $|h|$.

$$E[(y_{t+h} - y_{t+h-1})(y_t - y_{t-1})] = E[((y_{t+h} - \mu_y) - (y_{t+h-1} - \mu_y))((y_t - \mu_y) - (y_{t-1} - \mu_y))]$$

$$= E[(y_{t+h} - \mu)(y_t - \mu)] - E[(y_{t+h-1} - \mu)(y_t - \mu)] - E[(y_{t+h} - \mu)(y_{t-1} - \mu)] + E[(y_{t+h-1} - \mu)(y_{t-1} - \mu)]$$

$$= \gamma_y(h) - \gamma_y(h-1) - \gamma_y(h+1) + \gamma_y(h) = 2\gamma_y(h) - \gamma_y(h-1) - \gamma_y(h+1)$$

Therefore, $\gamma_z(h) = \gamma_{\nabla x}(h)$ has autocovariance only depends on $|h|$ and constant mean, It is stationary.