STAT3032S19_HW6

Mingming Xu

5/01/2019

Problem1 (a)

```r
soi=read.table("SOIvalues.txt")
soits = ts(soi, frequency=12,start=c(1876,1),end = c(2019,3))
mean(soits)
```

```
## [1] 0.0779523
```

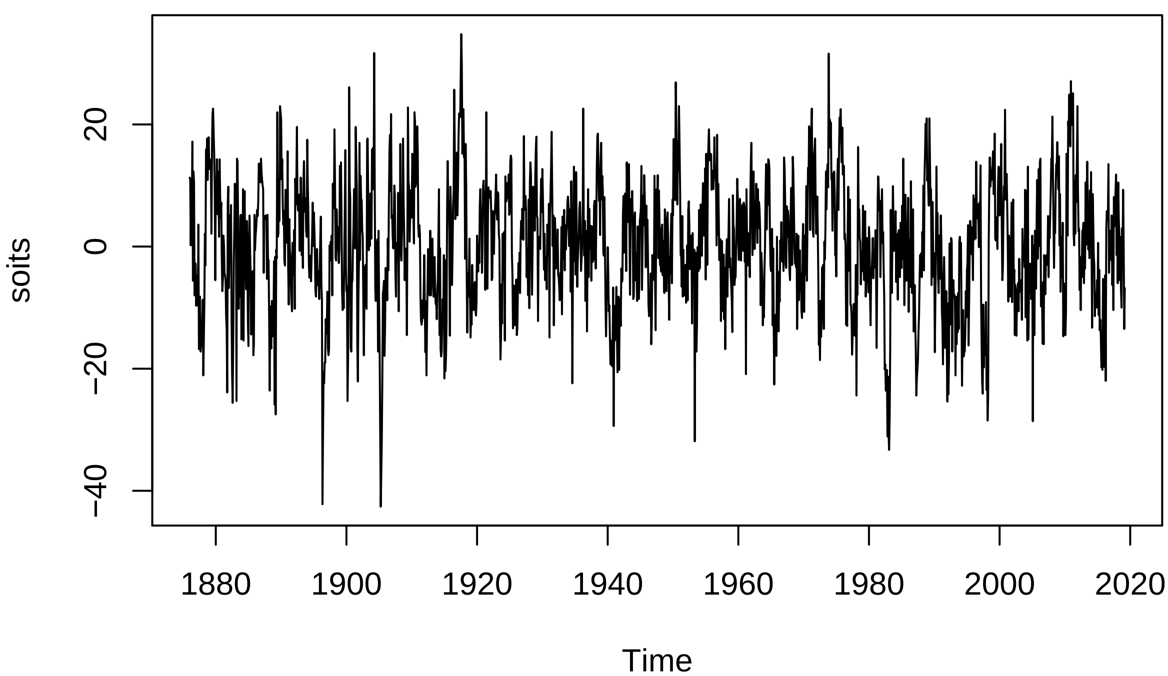
```r
var(soits)
```

```
##          y
## y 109.5344
```

The estimated mean of this time series is 0.0779523 and the variance of this time series is 109.5344.

(b)

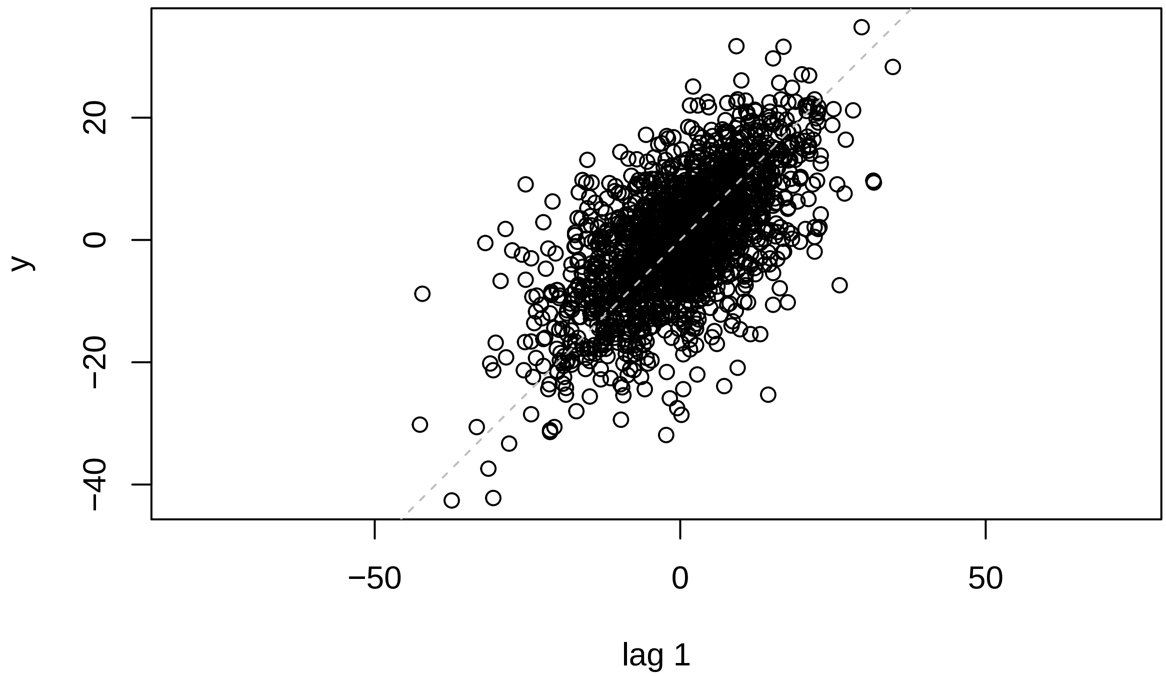
```r
ts.plot(soits)
```



(c)This time series looks weakly stationary.Because this time series plot does not have obvious trend (up or down over time),seasonality or cycles.This time series fluctuates around 0 and the range of fluctuation is limited in this plot.

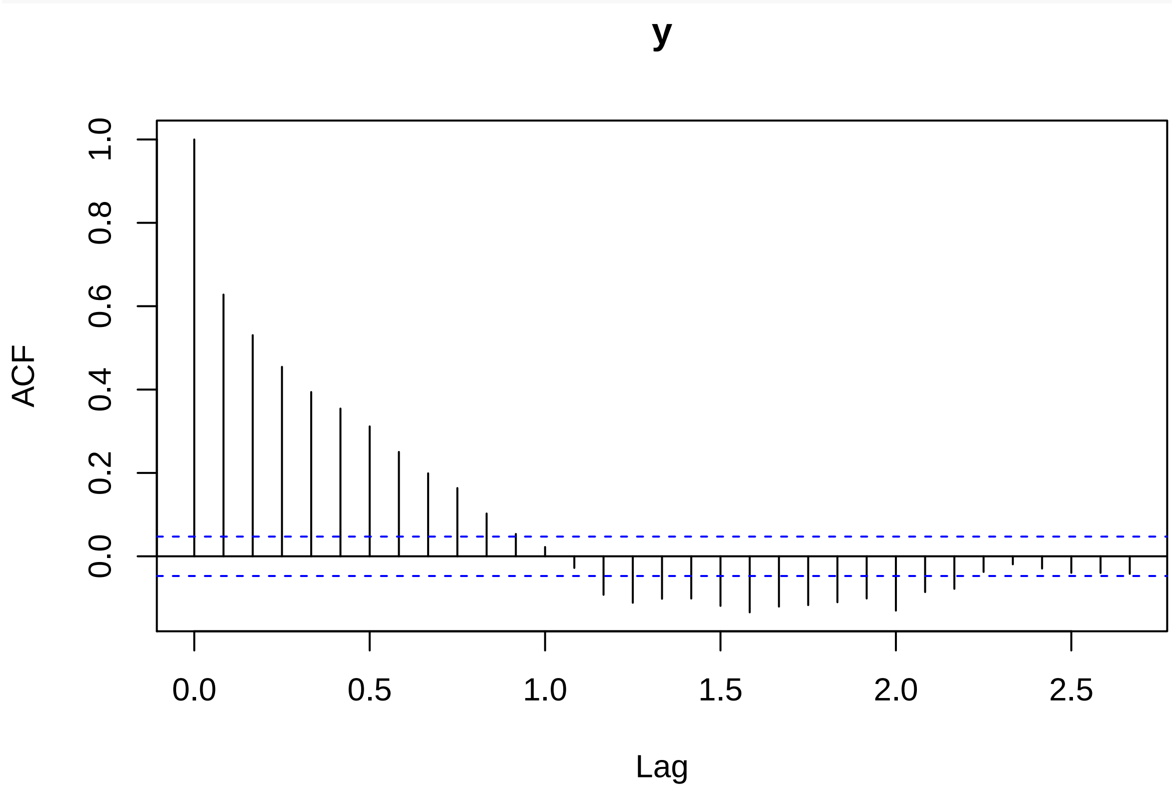
(d)

```r
lag.plot(soits,lags = 1)
```



(e)

```
acf(soits)
```




```
acf(soits,plot = FALSE)
```

```
##
## Autocorrelations of series 'soits', by lag
##
```

```
## 0.0000 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500
##  1.000  0.628  0.531  0.455  0.394  0.354  0.312  0.250  0.199  0.164
## 0.8333 0.9167 1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833
##  0.103  0.054  0.022 -0.028 -0.092 -0.112 -0.102 -0.101 -0.119 -0.135
## 1.6667 1.7500 1.8333 1.9167 2.0000 2.0833 2.1667 2.2500 2.3333 2.4167
## -0.121 -0.117 -0.110 -0.101 -0.130 -0.086 -0.078 -0.038 -0.019 -0.029
## 2.5000 2.5833 2.6667
## -0.040 -0.040 -0.042
```

The estimated autocorrelation at lag 1 is 0.628.

(f)

```r
#estimate of delta
#rho_1=phi_1=0.628
mean(soits)*(1-0.628)
```

```
## [1] 0.02899825
```

My estimate for δ is 0.02899825 .

(g)

```r
#estimate of sigma^2
#variance*(1-phi_1^2)
var(soits)*(1-0.628^2)
```

```
##          y
## y 66.33577
```

My estimate for σ^2 is 66.33577.

(h)

```r
arima(soits,order = c(1,0,0))
```

```
##
## Call:
## arima(x = soits, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.6282     0.0762
## s.e.  0.0187     0.5275
##
## sigma^2 estimated as 66.26:  log likelihood = -6043.86,  aic = 12093.71
```

Based on the output ,the estimated μ is 0.0762 , the estimated ϕ_1 is 0.6282 and the estimated variance σ^2 is 66.26. So,my results is almost close to the what I answered for parts(f) and (g).

(i) The fitted modle is $\hat{y_t} = 0.0762 + 0.6282 * (y_{t-1} - 0.0762)$

```r
len=length(soits)
y_tmi1=soits[len]
0.0762+0.6282*(y_tmi1-0.0762)
```

```
## [1] -4.243429
```

The predicted value of the SOL level in April 2019 is -4.243429.

Problem 2

(a)

To prove $Cov(a+V, b+W) = Cov(V, W)$: $Cov(a+V, b+W)$ $=E[(a+V)(b+W)] - E(a+V)E(b+W)$ $=E[ab+aW+bV+VW] - (a+E(V))(b+E(W))$ $=ab+aE(W)+bE(v)+E(VW)-ab-aE(W)-bE(v)-E(V)E(W)$ $=E(VW)-E(V)E(W) = Cov(V, W)$

(b)

If $y_t = \sigma+\phi_1 y_{t-1}+w_t$ where $\phi \neq 0$, $\rho_h = Corr(y_t-\sigma, y_{t_h}-\sigma)$ $=Cov(y_t-\sigma, y_{t-h}-\sigma)/(\sqrt{Var(y_t-\sigma)}\sqrt{Var(y_{t-h}-\sigma)})$

As in part(a),we let $y_t-\sigma$as $a+V$,where $-\sigma$as a and y_t as V,let $y_{t-h}-\sigma$ as $b+W$,whre $-\sigma$ as b and y_t as W.Then,$Cov(y_t-\sigma, y_{t-h}-\sigma) = Cov(-\sigma+y_t, -\sigma+y_{t_h}) = Cov(a+V, b+W) = Cov(V, W) = Cov(y_t, y_{t-h})$.

Also, $\sqrt{Var(y_t-\sigma)}\sqrt{Var(y_{t-h}-\sigma)} = \sqrt{Var(y_t)}\sqrt{Var(y_{t-h})}$

Hence,$\rho_h = Corr(y_t-\sigma, y_{t-h}-\sigma) = Cov(y_t, y_{t-h}/(\sqrt{Var(y_t)}\sqrt{Var(y_{t-h})}) = Corr(y_t-\sigma, y_{t-h}-\sigma)$

(c)No.$Corr(aV, bW) = E(aV*bW) - E(aV)E(bW) = abE(VW) - abE(V)E(W) = ab(E(VW) - E(V)E(W)) = abCov(V, W)$

(d)

$\gamma_h = \phi_1\gamma_{h-1} = \phi_1\phi_1\gamma_{h-2} = \phi_1\phi_1\phi_1\gamma_{h-3} = ... = \phi_1 * ... * \gamma_0 = \phi_1^n * \gamma_0$

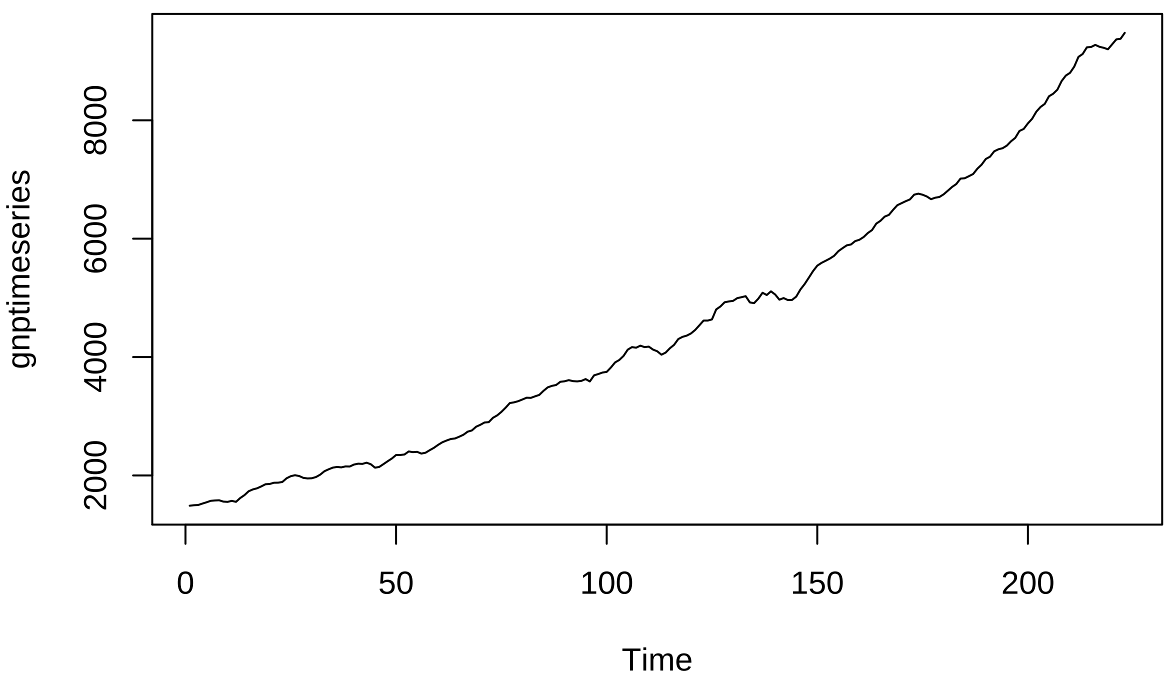
(e)

$\lim h \to \infty \rho_h = 0$ So, when h is so large, there is a so weak correlationship between y_t and y_{t-h},which means that past information becomes negligible in Layman terms

Problem 3

(a)

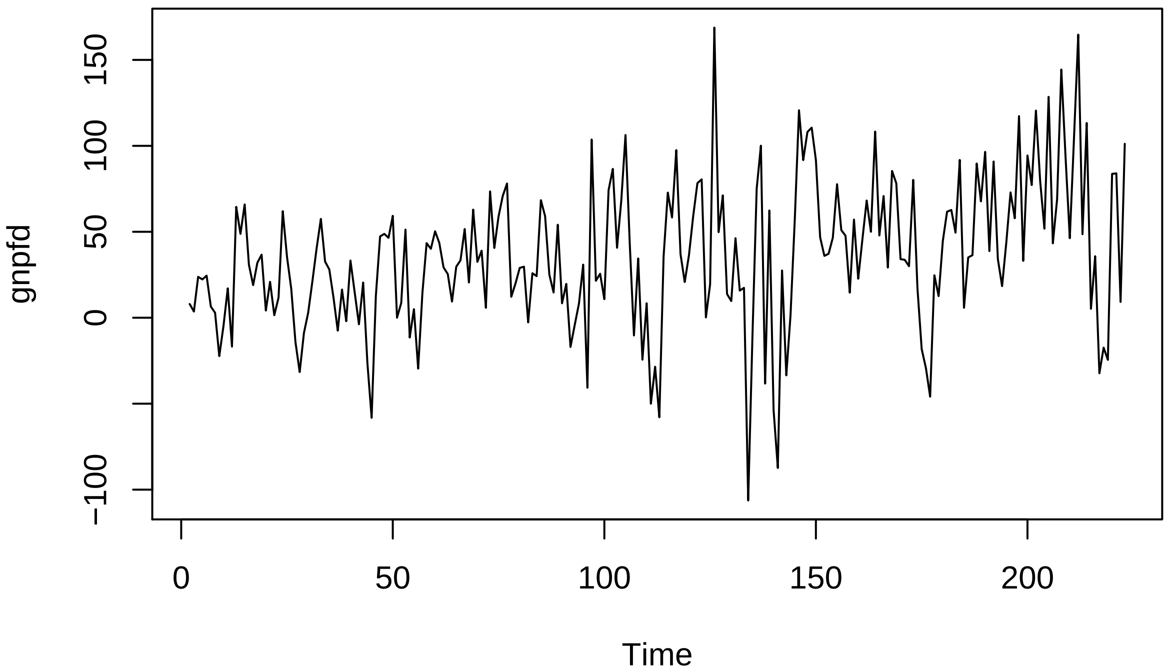
```
gnp=scan(file = "gnp.csv",skip=1)
gnptimeseries =ts(gnp)
ts.plot(gnptimeseries)
```



Based on the plot, thers is an increasing tendency implying this time series doesn't have stationarity.

(b)

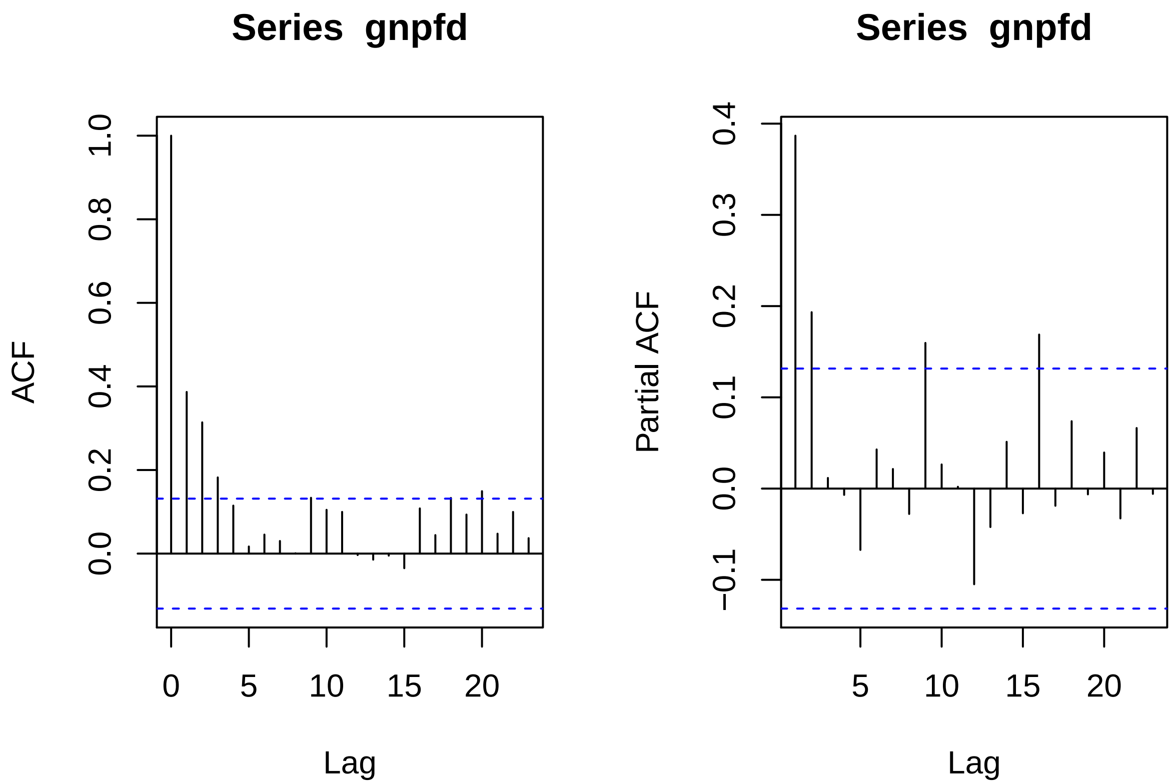
```r
gnpfd=diff(gnptimeseries,lag=1)
ts.plot(gnpfd)
```



Because this time series fluctuates around 0 and the range of fluctuation is limited in this plot,there is the weak stationarity of the first difference of the quarterly GNP.

(c)

```r
par(mfrow=c(1,2))
acf(gnpfd)
pacf(gnpfd)
```




```r
par(mfrow=c(1,1))
```

Based on the ACF and PACF pots, we need to compare the AR(2) and MA(3)

```r
arima(gnpfd,order=c(0,0,3))
```

```
##
## Call:
## arima(x = gnpfd, order = c(0, 0, 3))
##
## Coefficients:
##          ma1     ma2     ma3  intercept
##       0.2978  0.2661  0.1370    35.9984
## s.e.  0.0660  0.0633  0.0619     4.3644
##
## sigma^2 estimated as 1471:  log likelihood = -1124.72,  aic = 2259.45
```

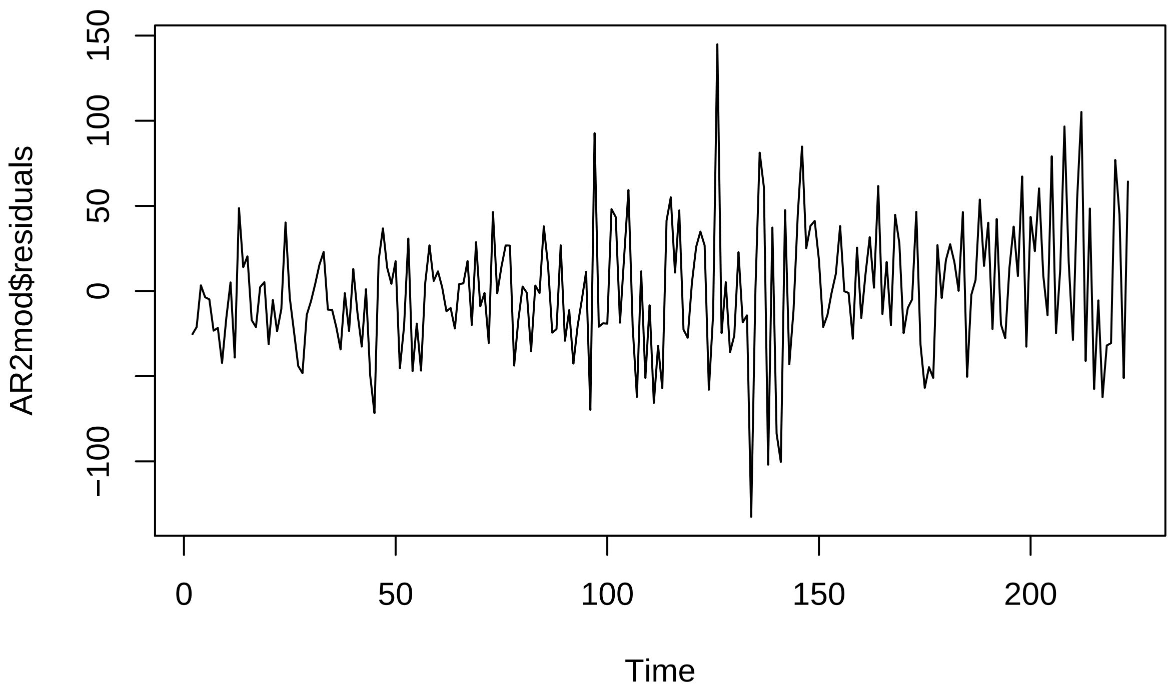
```r
arima(gnpfd,order=c(2,0,0))
```

```
##
## Call:
## arima(x = gnpfd, order = c(2, 0, 0))
##
## Coefficients:
##          ar1     ar2  intercept
##       0.3136  0.1931    36.0519
## s.e.  0.0662  0.0663     5.1613
##
## sigma^2 estimated as 1457:  log likelihood = -1123.67,  aic = 2255.34
```

```
AR2mod=arima(gnpfd,order=c(2,0,0))
```

Because AR(2) model has a smaller AIC,AR(2) is the appropriate time series model for the first diffrence of the quarterly GNP data. And the fitted modle is $

```
plot(AR2mod$residuals)
```



The residuals look like white noise

(d)

```
n=length(gnptimeseries)
muhat=AR2mod$coef[3]
phihat1=AR2mod$coef[1]
phihat2=AR2mod$coef[2]
y_tminus1= gnptimeseries[n]
y_tminus2= gnptimeseries[n-1]
y_tminus3= gnptimeseries[n-2]
muhat+phihat1*(y_tminus1-y_tminus2-muhat)+phihat2*(y_tminus2-y_tminus3-muhat)+y_tminus1
```

```
## intercept
##  9529.198
```

The predict value of the GNP of the 4th quarter of 2002 is 9529.198 .

Problem 4

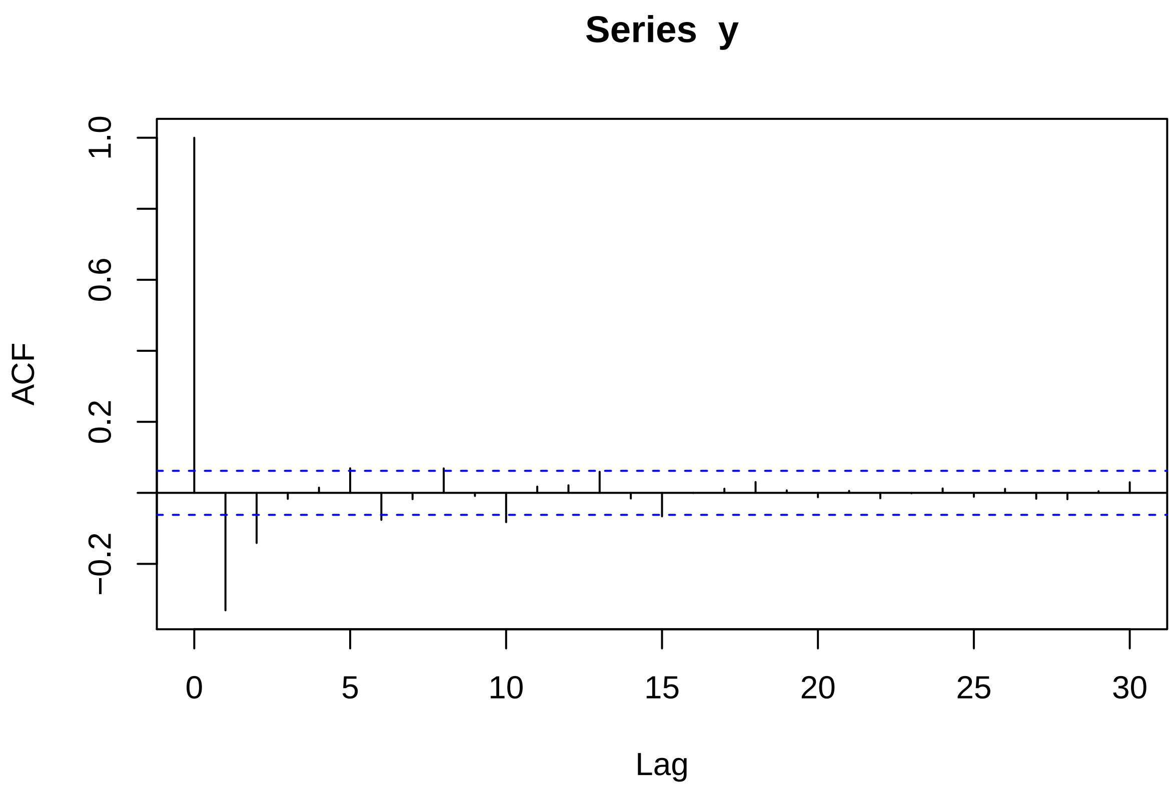
(a) C. y_t has a MA(2) model.

(b) Mean:$E(y_t) = \mu$, and $\mu = 0$. Variance : $Var(y_t) = \sigma_w^2(1 + \phi_1^2 + \phi_2^2) = 0.01(1 + 0.025 + 0.04) = 0.01065$

(c)$\rho_5 = Corr(y_t, y_{t_5}) = 0$.It is MA(2) model.Hence,for a lag 5 of ${y_t}, autocorrelation term is 0 because of 5>=2+1.

(d)

```r
set.seed(3032)
#MA(2) model
y=arima.sim(n=1000, list(ma=c(-0.5, -0.2)), sd=sqrt(0.01))
acf(y)
```




```r
acf(y,plot = F)
```

```
##
## Autocorrelations of series 'y', by lag
##
##      0      1      2      3      4      5      6      7      8      9
##  1.000 -0.331 -0.141 -0.017  0.015  0.069 -0.076 -0.018  0.069 -0.009
##     10     11     12     13     14     15     16     17     18     19
## -0.083  0.018  0.021  0.059 -0.016 -0.066  0.000  0.012  0.031  0.007
##     20     21     22     23     24     25     26     27     28     29
## -0.012  0.006 -0.015 -0.001  0.012 -0.011  0.011 -0.017 -0.018  0.005
##     30
##  0.030
```

```r
pacf(y,plot = F)
```

```
##
## Partial autocorrelations of series 'y', by lag
##
##      1      2      3      4      5      6      7      8      9     10
## -0.331 -0.281 -0.207 -0.145 -0.023 -0.090 -0.085  0.004 -0.002 -0.094
##     11     12     13     14     15     16     17     18     19     20
## -0.062 -0.050  0.026  0.033 -0.028 -0.051 -0.053 -0.009  0.015  0.011
##     21     22     23     24     25     26     27     28     29     30
##  0.012 -0.009  0.000  0.012 -0.014 -0.007 -0.023 -0.032 -0.027  0.008
```

```r
mean(y)
```

```
## [1] -0.0007444642
```

```r
var(y)
```

```
## [1] 0.01314148
```

(e)

Based on the generated data in Part (d), the estimated mean is -0.0007444642, estimated variance is 0.01314148, and estimated autocorrelation at lag 5 of {yt} is 0.069. These estimated values are close to the theoretical values in Part (b) and (c).

(f)

```r
arima(y,order = c(0,0,2))
```

```
##
## Call:
## arima(x = y, order = c(0, 0, 2))
##
## Coefficients:
##           ma1      ma2  intercept
##       -0.5390  -0.1862     -5e-04
## s.e.   0.0311   0.0320      9e-04
##
## sigma^2 estimated as 0.009921:  log likelihood = 887.29,  aic = -1766.58
```

The fitted model is $\hat{y_t} = -5e - 04 + w_t - 0.5390 w_{t-1} - 0.1862 * w_{t-2}$,and $hat\sigma_w^2 = 0.009921$.So, the estimated coefficient values are close to the theoretical values in the population model.