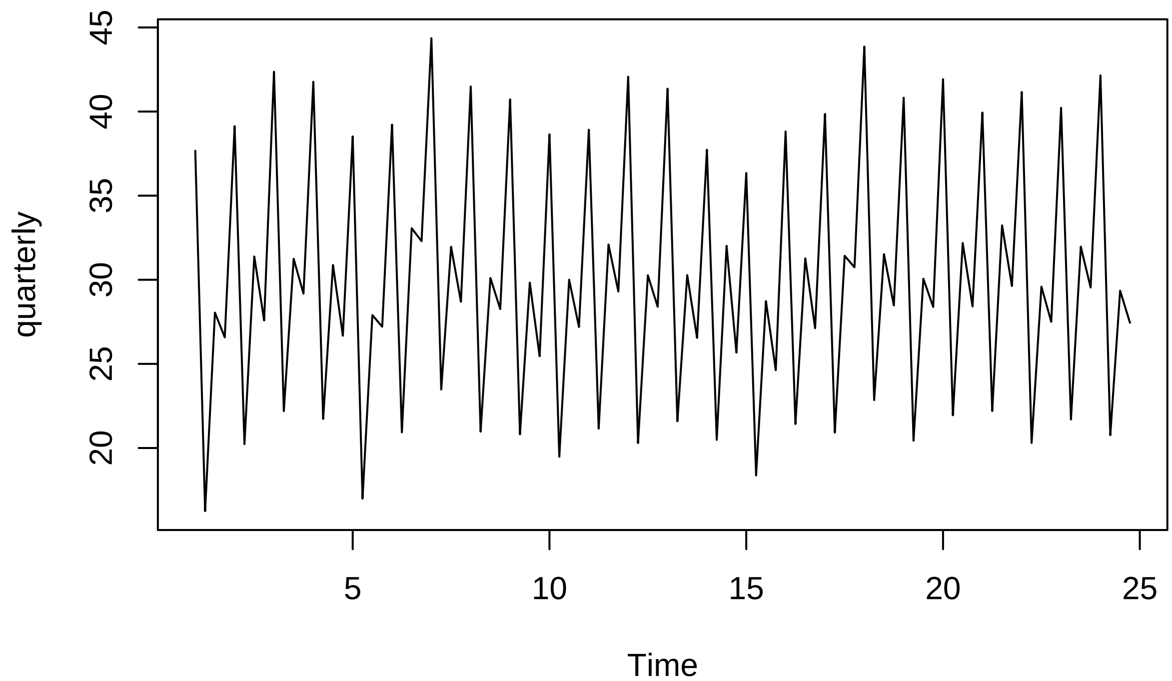
a4q2

Question 2

Part a)

```
quarterly<-read.table("quarterly.txt")
quarterly<-ts(quarterly[,1], frequency=4)
plot(quarterly)
```



Comment: The plot has some seasonal trend, which suggests that we should use seasonal indicators to model the data. It has a period of 4, representing 4 quarters in a year, and we can use 3 parameters to estimate that.

Part b)

We fit the regression model by $Y_t = \beta_0 + \sum_{j=1}^{3} \beta_j \ X_{t,j} + \epsilon_t$, where $X_{i,1}, X_{i,2}$,and $X_{i,3}$ are indicator parameters corresponding to the first three quarters.

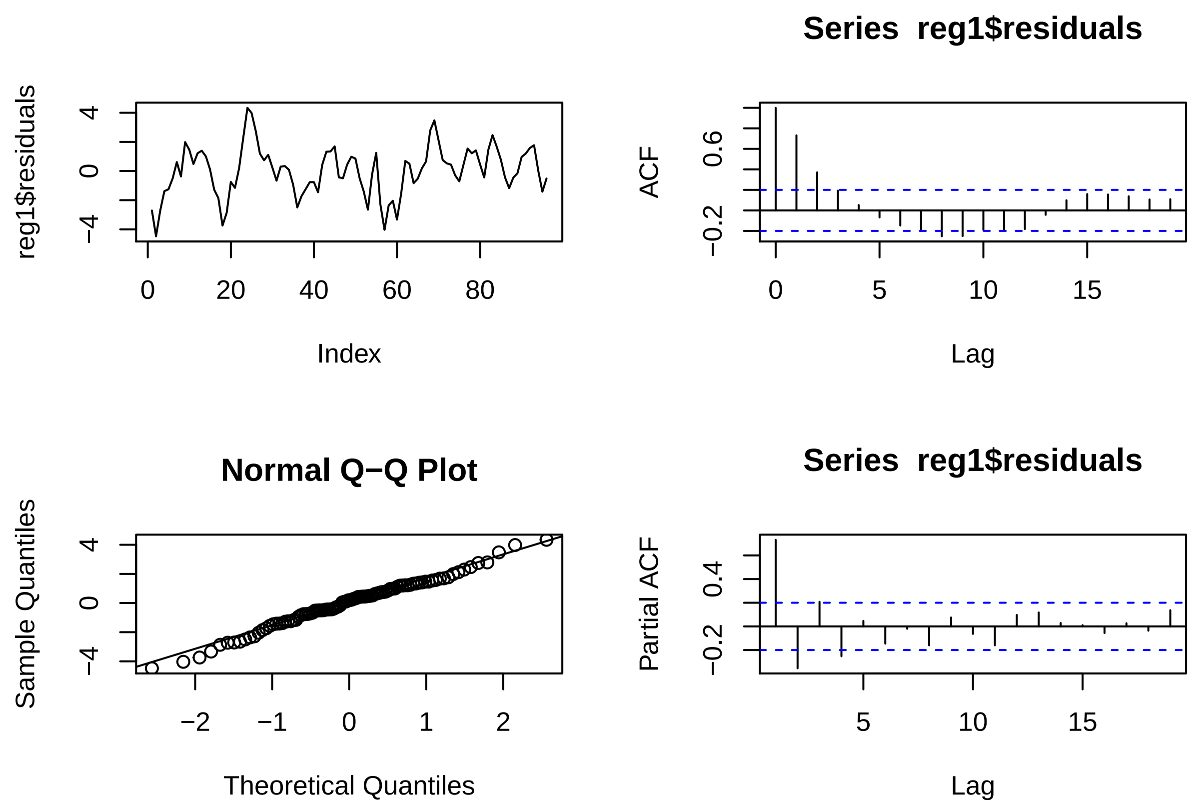
```
quar=factor(cycle(quarterly))
contrasts(quar) <- contr.treatment(4, base = 4)
reg1=lm(quarterly~quar)
summary(reg1)
```

```
##
## Call:
## lm(formula = quarterly ~ quar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4793 -0.9912  0.1957  1.1946  4.3401
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.9554     0.3487  80.181  < 2e-16 ***
## quar1        12.4240     0.4931  25.197  < 2e-16 ***
## quar2        -7.2268     0.4931 -14.657  < 2e-16 ***
## quar3         2.8124     0.4931   5.704 1.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.708 on 92 degrees of freedom
## Multiple R-squared:  0.9467, Adjusted R-squared:  0.9449
## F-statistic: 544.2 on 3 and 92 DF,  p-value: < 2.2e-16
```

Part c)

```r
par(mfcol=c(2,2))
plot(reg1$residuals,type="l")
qqnorm(reg1$residuals)
qqline(reg1$residuals)
acf(reg1$residuals)
acf(reg1$residuals,type="partial")
library(lawstat,quietly=T)
```

```
## Loading required package: stats4
## Loading required package: splines
## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```




```r
runs.test(reg1$residuals)
```

```
##
##  Runs Test - Two sided
##
## data:  reg1$residuals
## Standardized Runs Statistic = -4.925, p-value = 8.437e-07
```

```r
library(randtests,quietly=T)
```

```
##
## Attaching package: 'randtests'
##
## The following object is masked from 'package:lawstat':
##
##     runs.test
```

```r
difference.sign.test(reg1$residuals)
```

```
##
##  Difference Sign Test
##
## data:  reg1$residuals
## statistic = 0.5276, n = 96, p-value = 0.5978
## alternative hypothesis: nonrandomness
```

```r
turning.point.test(reg1$residuals)
```

```
##
##  Turning Point Test
##
## data:  reg1$residuals
## statistic = -6.2724, n = 96, p-value = 3.555e-10
## alternative hypothesis: non randomness
```

```r
shapiro.test(reg1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  reg1$residuals
## W = 0.989, p-value = 0.6146
```

From the residual plots, we see some extremely large and small residuals, and some expanding variance as the fitted value increasees. The qq-plot looks not quite straight, with periodic movements along the qqline. The runs test has a p-value far less than 0.05, which is near 0. So we would like to reject the null hypothesis and conclude that the residuals are not independent. The Shapiro-Wilk test disagrees with that, giving a p-value of 0.6146, which tells us that there is no evidence to reject Normality. Similarly, the difference sign test has p-value 0.5978, with lack of evidence to reject the null hypothesis, but turning point test has a p-value near 0 showing no randomness in the residual. Even though some tests are showing contradicting results, overall we can conclude that the residual is close to IID normal distribution by looking at the graphs of residuals.

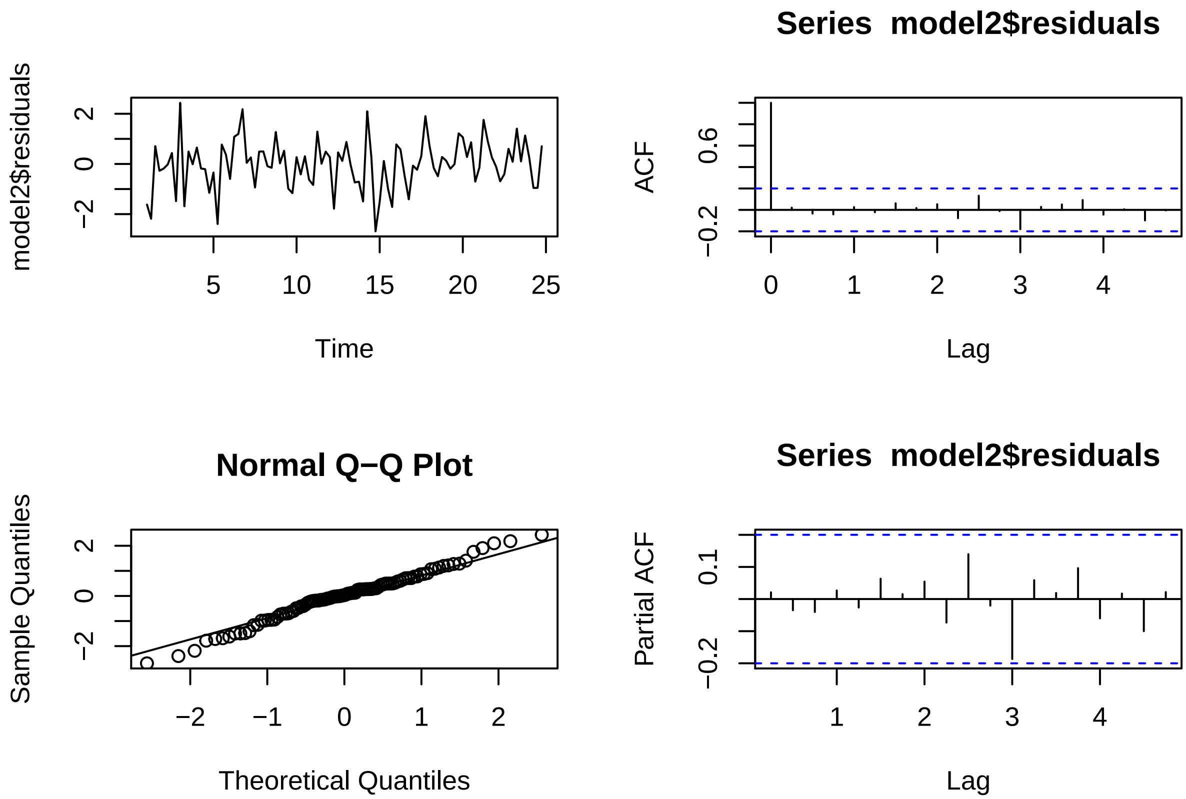
Part d)

We propose the model $ARMA(2,3)$

```r
resi<-ts(reg1$residuals,frequency=4)
model2<-arima(resi,order=c(2,0,3),method="ML")
model2
```

```
##
## Call:
## arima(x = resi, order = c(2, 0, 3), method = "ML")
##
## Coefficients:
##          ar1      ar2      ma1      ma2     ma3  intercept
##       1.6959  -0.8104  -0.6138  -0.5528  0.4031     0.0245
## s.e.  0.1640   0.1731   0.2444   0.1191  0.2536     0.2113
##
## sigma^2 estimated as 0.9632:  log likelihood = -135.28,  aic = 284.56
```

```r
par(mfcol=c(2,2))
plot(model2$residuals,type="l")
qqnorm(model2$residuals)
qqline(model2$residuals)
acf(model2$residuals)
acf(model2$residuals,type="partial")
```




```r
model2
```

```
##
## Call:
## arima(x = resi, order = c(2, 0, 3), method = "ML")
##
## Coefficients:
##          ar1      ar2      ma1      ma2     ma3  intercept
##       1.6959  -0.8104  -0.6138  -0.5528  0.4031     0.0245
## s.e.  0.1640   0.1731   0.2444   0.1191  0.2536     0.2113
##
## sigma^2 estimated as 0.9632:  log likelihood = -135.28,  aic = 284.56
```

Comment: The new residual looks random and there is no pattern suggesting in the plots. We have covered the negative spike at lag=1 in the previous model.

Part e)

As we have done the work in part d, the new model is the combination of linear regression model and a ARMA(3,2) model.

Part f)

```r
resipre<-predict(model2,n.ahead=8)$pred
lmpre<-predict.lm(reg1,newdata=data.frame(quar=factor(c(1,2,3,4,1,2,3,4))))
lmpre
```

```
##        1        2        3        4        5        6        7        8
## 40.37936 20.72855 30.76775 27.95537 40.37936 20.72855 30.76775 27.95537
```

```r
lmpre+resipre
```

```
##        Qtr1     Qtr2     Qtr3     Qtr4
## 25 40.37236 20.35785 30.43242 27.68990
## 26 40.20370 20.64858 30.77728 28.03916
```

Comment: The new model is better as the predicted values are much closer to the expected values.