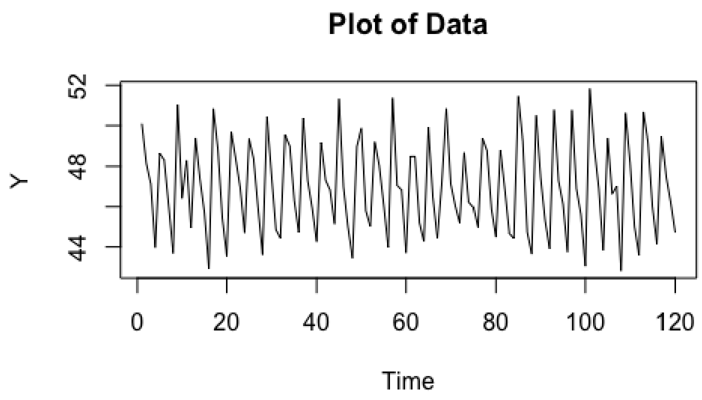
**Question 3**

First, we plot the data to see the general pattern:




We notice that there is a seasonal component in the data, so the assumed simple causal model is applicable:

$$Y_t = \beta_0 + \sum_{k=1}^{3} \beta_k X_{t,k} + \epsilon_t$$

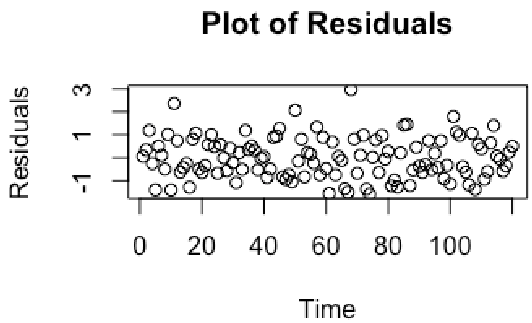
We fit the model and the following estimates:

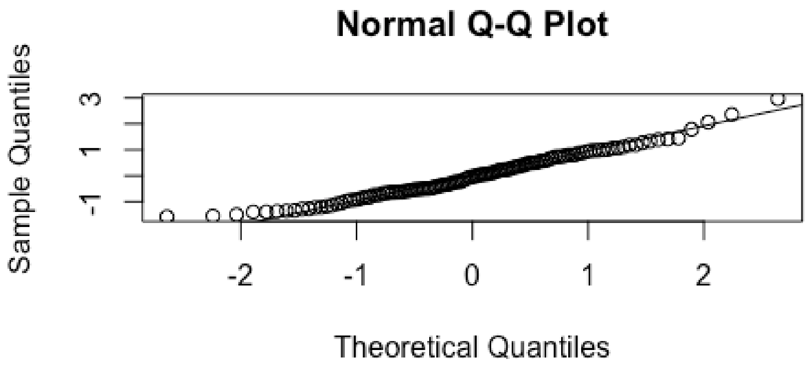
```
Residual Standard Error=0.9076
R-Square=0.855
F-statistic (df=3, 116)=228.0701
p-value=0
```

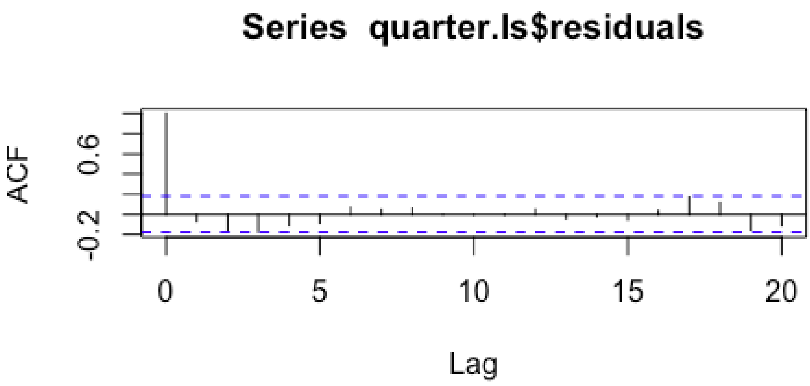
| | Estimate | Std.Err | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| Intercept | 44.2095 | 0.1657 | 266.8058 | 0 |
| X1 | 5.8279 | 0.2343 | 24.8700 | 0 |
| X2 | 3.5992 | 0.2343 | 15.3595 | 0 |
| X3 | 1.7326 | 0.2343 | 7.3938 | 0 |

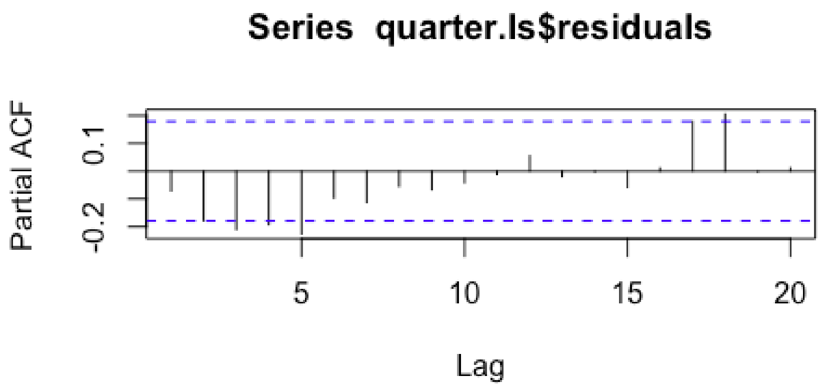
Here, $\widehat{\beta_0} = 44.2095$, $\widehat{\beta_1} = 5.8279$, $\widehat{\beta_2} = 3.5992$, $\widehat{\beta_3} = 1.7326$

Then we perform Residual Analysis:





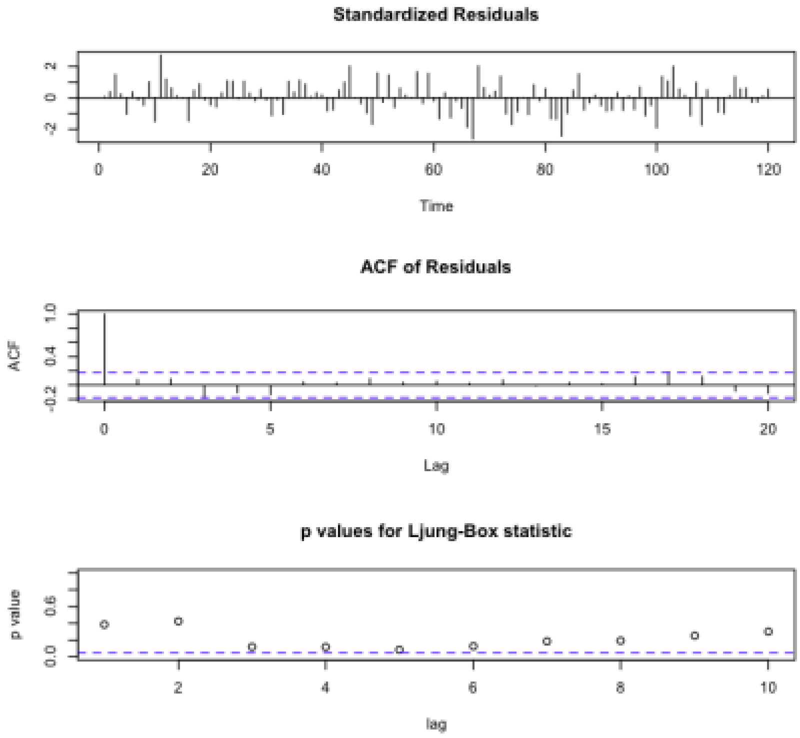

- From the first plot, we see no particular evidence against stationarity i.e. no changing mean or variance.
- The q-q plot doesn't show any evidence against normality, it fits the line perfectly in the middle part of the graph.
- Sample ACF plot shows no significant spikes, hence, correlation is 0 (no exponential decay)
- Sample PACF plot shows spikes at lags 3,4,5,18, which suggest that there may be some correlation.
- Since there is still some correlation left and there is spikes only in SPACF graph let's use MA (2) and ARMA (1,1) models i.e.

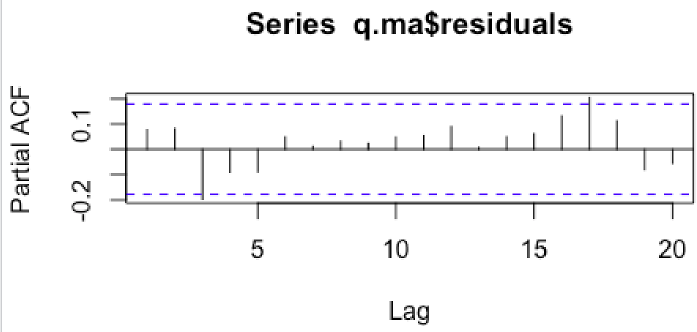
$$Y_t = \beta_0 + \sum_{k=1}^{3} \beta_k X_{t,k} + \epsilon_t$$

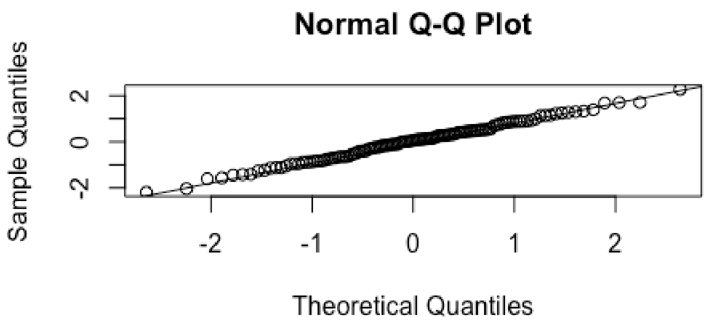
$$\epsilon_t \sim MA(2)\ or\ ARMA(1,1)$$

MA (2):

```
Call:
arima(x = quarter.y, order = c(0, 0, 2), xreg = x, method = "ML")

Coefficients:
          ma1      ma2  intercept      x1      x2      x3
      -0.3147  -0.3656    44.1987  5.8326  3.6044  1.7375
s.e.   0.0778   0.0718     0.1691  0.2572  0.3013  0.2583

sigma^2 estimated as 0.6997:  log likelihood = -149.13,  aic = 312.27
```

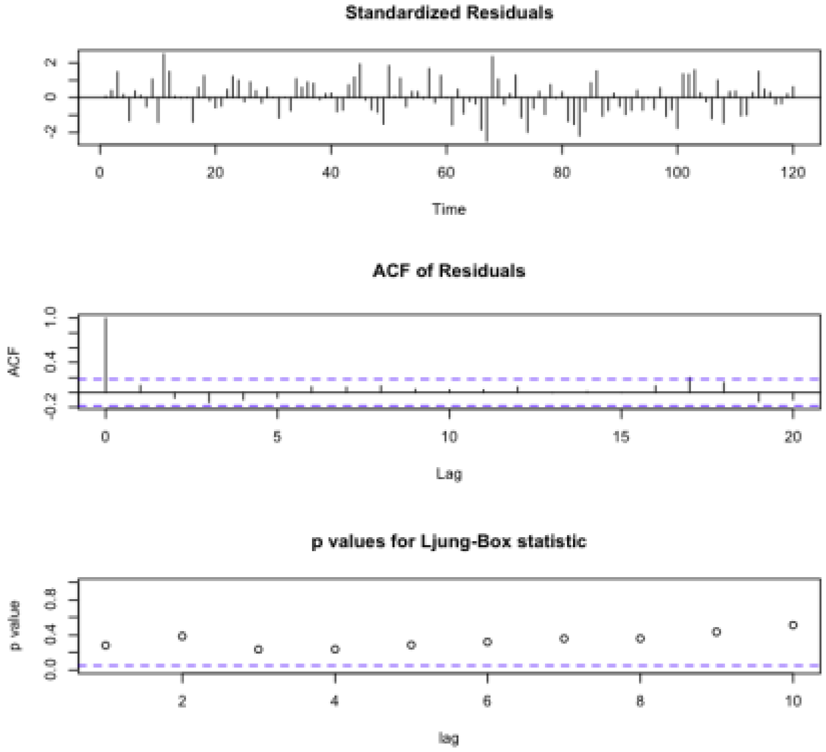


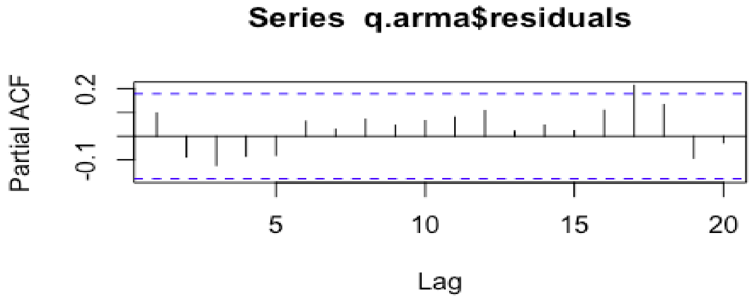



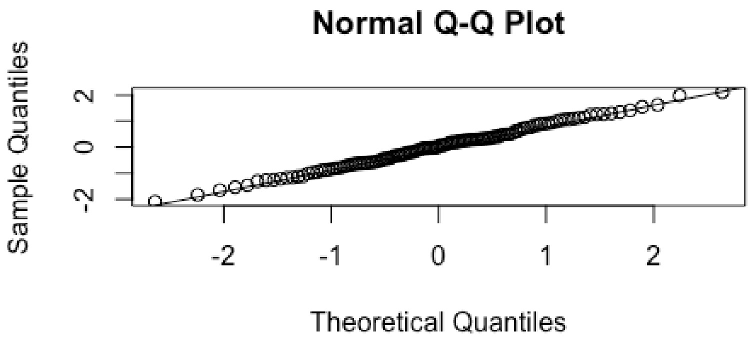



ARMA (1, 1):

```
Call:
arima(x = quarter.y, order = c(1, 0, 1), xreg = x, method = "ML")

Coefficients:
         ar1      ma1  intercept      x1      x2      x3
      0.5496  -0.8642    44.1980  5.8312  3.6041  1.7364
s.e.  0.1021   0.0522     0.1564  0.2535  0.2492  0.2544

sigma^2 estimated as 0.6965:  log likelihood = -148.8,  aic = 311.59
```



- The estimated variance for MA (2) model of WN process is 0.6997 which is lower than the variance of the original residuals 0.9076.
- The estimated variance for ARMA (1,1) model of WN process is 0.6965 which is lower than the variance of the original residuals 0.9076 and the variance of MA (2).
- Also, AIC for MA (2) is 312.27 and is higher than AIC for ARMA (1,1) 311.59.
- So, we can conclude that ARMA (1,1) is a better model based on AIC statistic and rule of parsimony.
- *For MA (2) Model:*
- We see no obvious patterns of violation of stationarity and symmetry from the first plot.
- From ACF plot there are no spikes. So, we have no evidence against WN.
- From third plot, we see that at lag 5 dot is a bit below the blue dashed line, which suggests that we should reject $H_0$: all correlations are zero.
- We notice spikes at lags 3, 17 in the PACF plot. From plot 5 we see no evidence against normality since the plot fits the line in the middle. However, spikes in PACF plot and rejection of $H_0$ in Ljung-Box test suggests that the process $Z_t$ is not WN and MA (2) is not a good model.
- *For ARMA (1,1):*
- We see no obvious patterns of violation of stationarity and symmetry from the first plot.
- From ACF plot there are no spikes.
- From third plot, we see that all dots are above blue dashed line, which suggests that there is no evidence to reject $H_0$: all correlations are zero. So, we have no evidence against WN.
- We notice spike at lag 17 in the PACF plot too, however, behavior of Ljung-Box statistic suggests that spike at lag 17 is not significant.
- From plot 5 we see no evidence against normality since the plot fits the line in the middle.
- Hence, there are no evidences against normality or WN of $Z_t$. Moreover, we can use ARMA (1,1) model to test regression parameters and to construct CI and PI.

*Forecasting and PI:*

Forecasted values for $Y_{121}$, …, $Y_{123}$ correspondingly:

49.85434
47.70594
45.88161

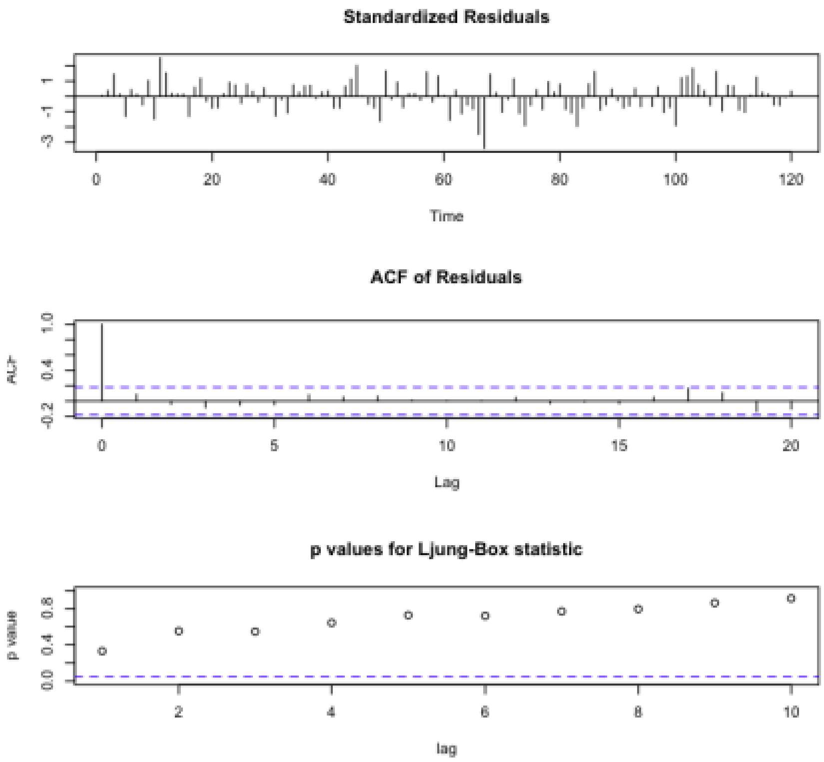
PI's for $Y_{121}$, …, $Y_{123}$ correspondingly:

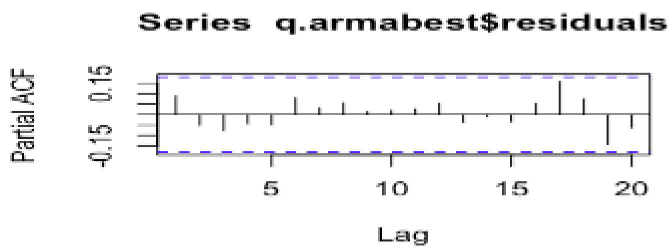
(48.18518, 51.52350)
(45.95615, 49.45573)
(44.10818, 47.65504)

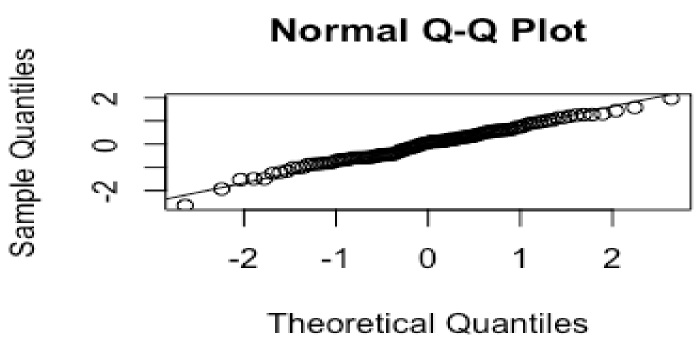
We could consider the ARMA (2,2) model which has more parameters, but fits perfectly.

```
Call:
arima(x = quarter.y, order = c(2, 0, 2), xreg = x, method = "ML")

Coefficients:
         ar1      ar2      ma1     ma2  intercept      x1      x2      x3
      1.5230  -0.6104  -1.9618  0.9999    44.2078  5.8358  3.6058  1.7367
s.e.  0.0739   0.0791   0.0575  0.0581     0.1562  0.2503  0.2493  0.2513

sigma^2 estimated as 0.6012:  log likelihood = -142.42,  aic = 302.85
```

As we see both AIC and variance of the ARMA (2,2) is lower than the ones of the models indicated above. Now, let's take a look at the plots:



From the plots:

- We see no obvious patterns of violation of stationarity and symmetry from the first plot.
- From ACF plot there are no spikes. So, we have no evidence against WN.
- From third plot, we see that all dots are above blue dashed line, which suggests that there is no evidence to reject $H_0$: all correlations are zero. So, we have no evidence against WN.
- We notice no spikes in the PACF plot. So, we have no evidence against WN.
- From plot 5 we see no evidence against normality since the plot fits the line in the middle.
- Hence, there are no evidences against normality or WN of $Z_t$. Moreover, we can use ARMA (2,2) model to test regression parameters and to construct CI and PI.

  *Forecasting and PI:*

  Forecasted values for $Y_{121}, \ldots, Y_{123}$ correspondingly:
  50.06729
  47.81915
  45.93855
  PI's for $Y_{121}, \ldots, Y_{123}$ correspondingly:
  (48.50486, 51.62972)
  (46.12214, 49.51616)
  (44.18941, 47.68769)

  We see that there is no big difference between forecasted values and PI's of ARMA (1,1) and ARMA (2,2) models. Since, there are less parameters in ARMA(1,1) by rule of parsimony I chose ARMA(1,1) Model.

**R code**

```
## Question 3
mypath = "/Users/safurasuleymanovs/Desktop/4B/Stat/A4/"
data.set ="quarter_y.txt"
quarter.y = scan(paste(mypath,data.set,sep=""))
quarter.ts = ts(scan(paste(mypath,data.set,sep="")))
plot(quarter.ts, type="l",xlab="Time",ylab="Y", main="Plot of Data")
b1 <-diag(3)
b <- rbind(b1,c(0,0,0))
x <- rbind(b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b,b)
quarter.ls <-lsfit(x,quarter.y)
ls.print(quarter.ls)

## Residual Analysis
plot(quarter.ls$residuals,main="Plot of Residuals", xlab="Time", ylab="Residuals")
qqnorm(quarter.ls$residuals)
qqline(quarter.ls$residuals)
acf(quarter.ls$residuals)
pacf(quarter.ls$residuals)

## Arima
q.ma <- arima(quarter.y,order=c(0,0,2),xreg=x,method="ML")
pacf(q.ma$residuals)
qqnorm(q.ma$residuals)
qqline(q.ma$residuals)
tsdiag(q.ma)
q.arma <- arima(quarter.y,order=c(1,0,1),xreg=x,method="ML")
pacf(q.arma$residuals)
qqnorm(q.arma$residuals)
qqline(q.arma$residuals)
tsdiag(q.arma)
q.armabest <- arima(quarter.y,order=c(2,0,2),xreg=x,method="ML")
pacf(q.armabest$residuals)
qqnorm(q.armabest$residuals)
qqline(q.armabest$residuals)
tsdiag(q.armabest)

## Forecating and PI
new.x <- b1
q.pr <- predict(q.arma,n.ahead=3,newxreg=new.x,se.fit=TRUE)
## Predicted value
y.pred <- q.pr$pred
y.sd <- q.pr$se
## Upper bounds of PIs
U <- y.pred + 2*y.sd
## Lower bounds of PIs
L <- y.pred - 2*y.sd
```