proj_arima

lbq

2020/11/28

```
getwd()
```

```
## [1] "F:/FA2020/VE406/proj/Stock_Price_Prediction-main/Stock_Price_Prediction-main/ARIMA"
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.0.3
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
TSLA.full = read.csv("11_28.csv", header = T)
```

for simple multilinear regression model

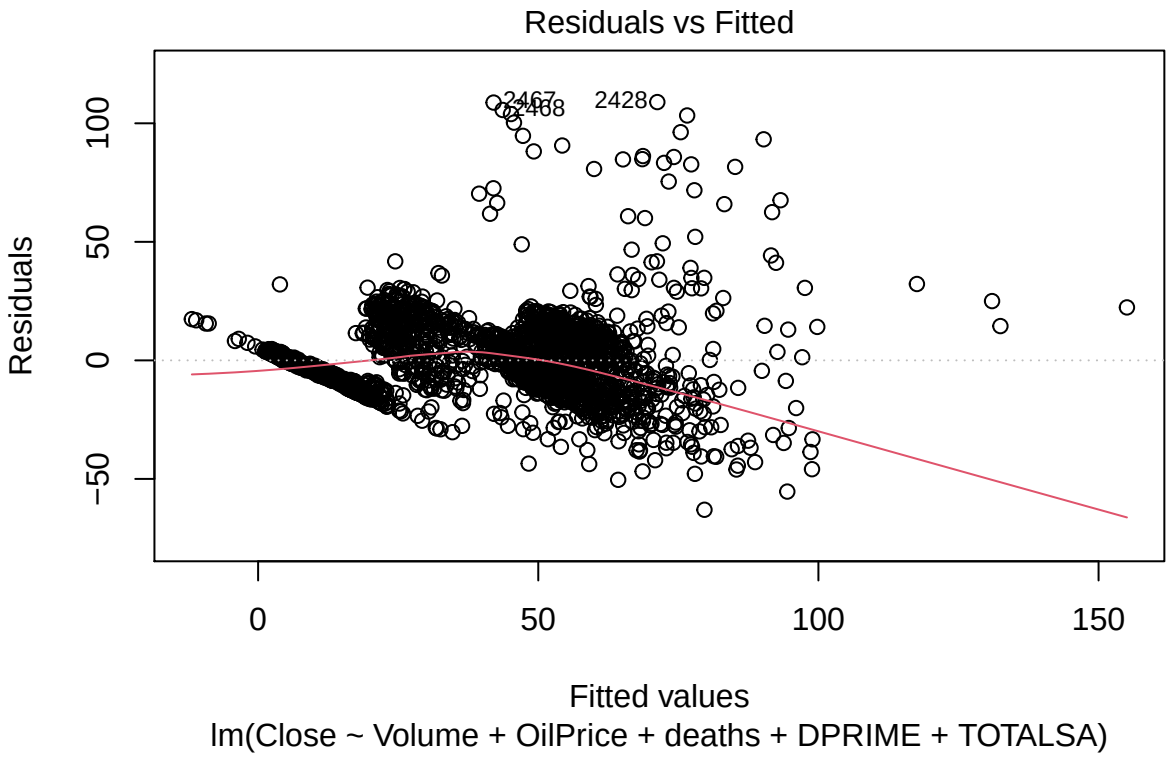
$Y_i = \beta X_i + \varepsilon_i$

```
TSLA.full$Date = as.Date(TSLA.full$Date, format = "%Y-%m-%d")
TSLA.full.lm = lm(Close ~ Volume + OilPrice + deaths + DPRIME + TOTALSA, data = TSLA.full)
summary(TSLA.full.lm)
```

```
##
## Call:
## lm(formula = Close ~ Volume + OilPrice + deaths + DPRIME + TOTALSA,
##     data = TSLA.full)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.013  -8.752  -2.171   8.254 108.953
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.347e+00  4.035e+00   1.821   0.0688 .
## Volume       3.507e-07  1.297e-08  27.034   <2e-16 ***
## OilPrice    -4.412e-01  1.434e-02 -30.773   <2e-16 ***
## deaths       3.740e+00  1.726e+00   2.167   0.0304 *
## DPRIME       6.350e+00  4.571e-01  13.892   <2e-16 ***
## TOTALSA      6.012e+01  6.156e+00   9.765   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.1 on 2447 degrees of freedom
##   (161 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.6587, Adjusted R-squared:  0.658
## F-statistic: 944.7 on 5 and 2447 DF,  p-value: < 2.2e-16
```

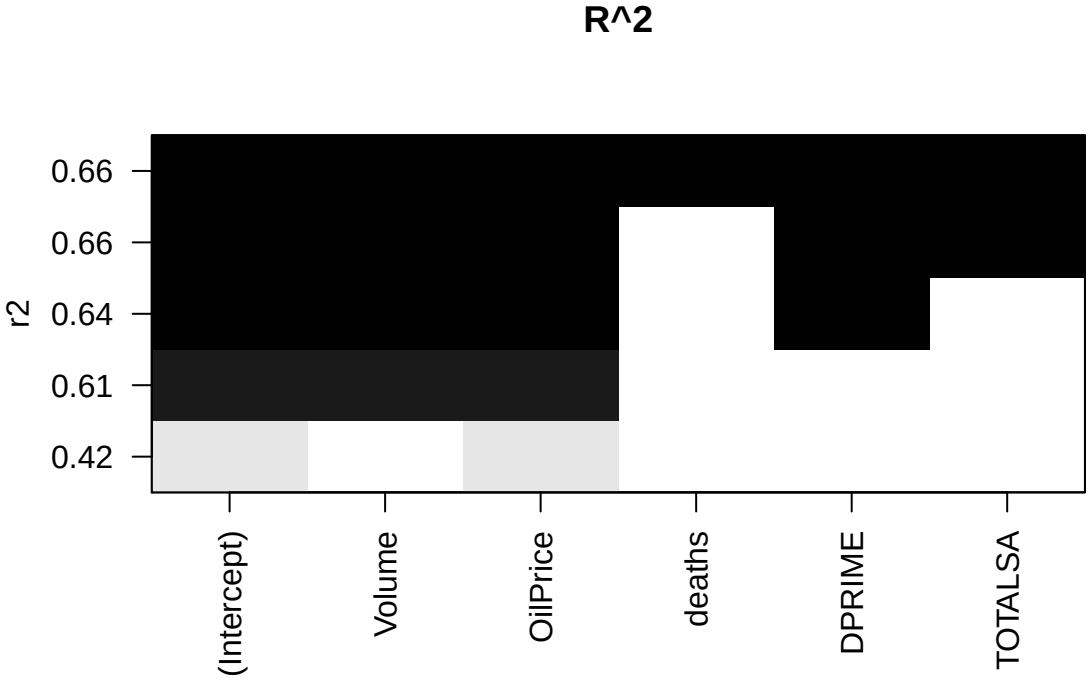
```
plot(TSLA.full.lm, which = 1)
```



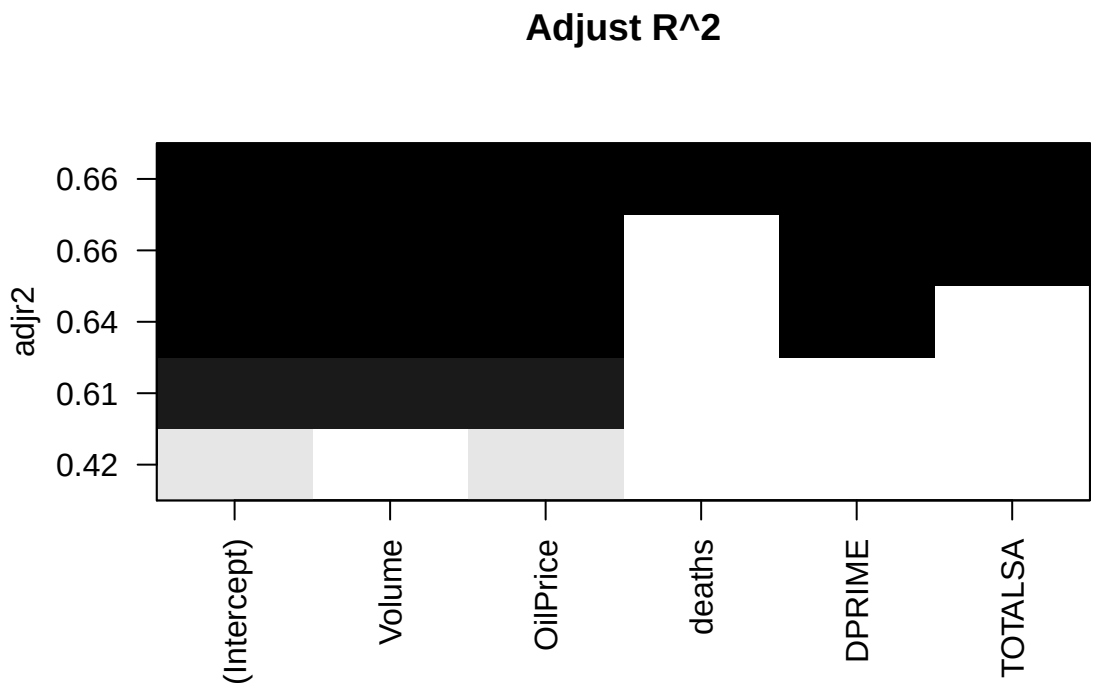

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.0.3
```

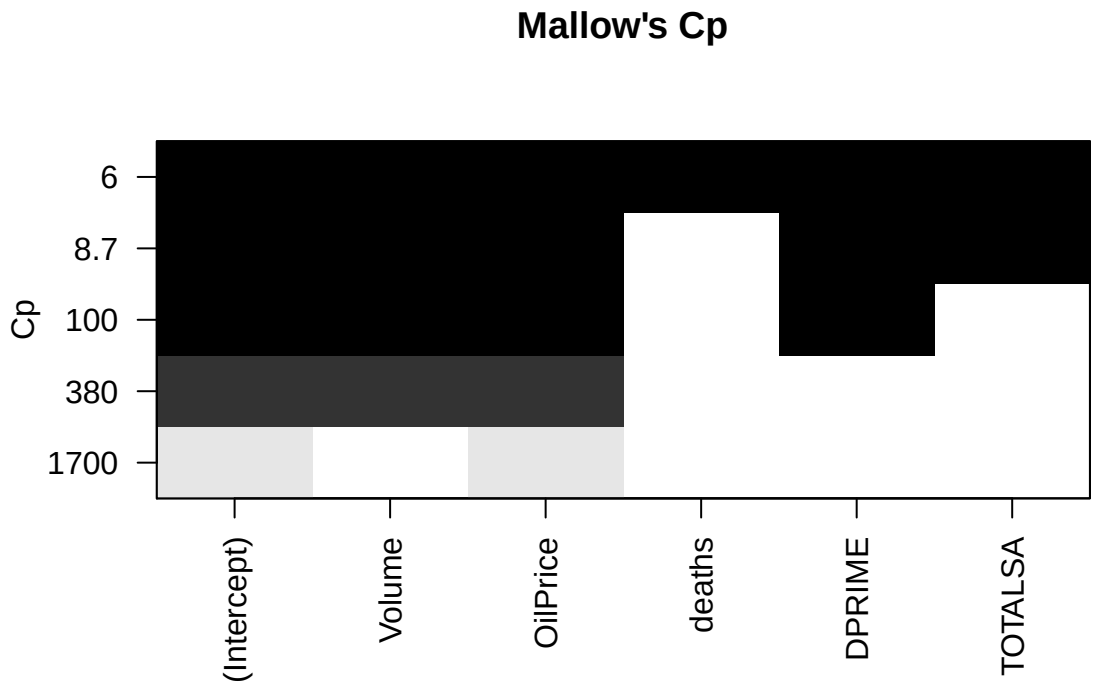
```
regsubsets.out = regsubsets(
  Close ~ Volume + OilPrice + deaths + DPRIME + TOTALSA, data = TSLA.full,
  nbest = 1,
  nvmax = NULL,
  method = "exhaustive"
)
plot(regsubsets.out , scale = "r2", main = "R^2")
```



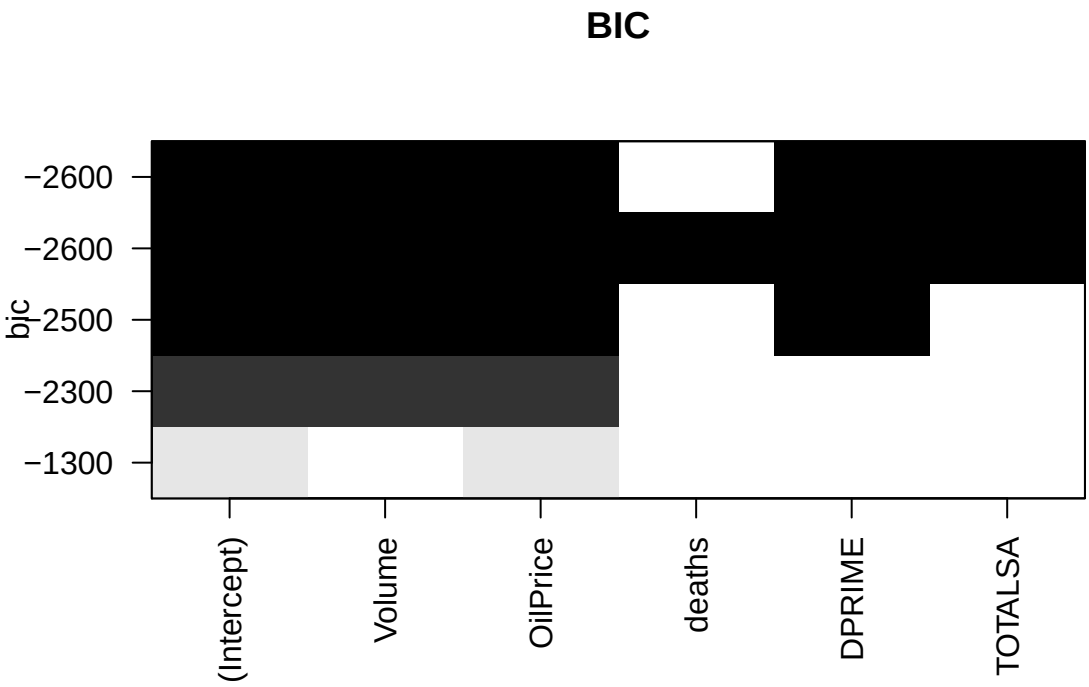

```
plot(regsubsets.out , scale = "adjr2", main = "Adjust R^2")
```




```
plot(regsubsets.out , scale = "Cp", main = " Mallow's Cp ")
```




```
plot(regsubsets.out , main = "BIC")
```

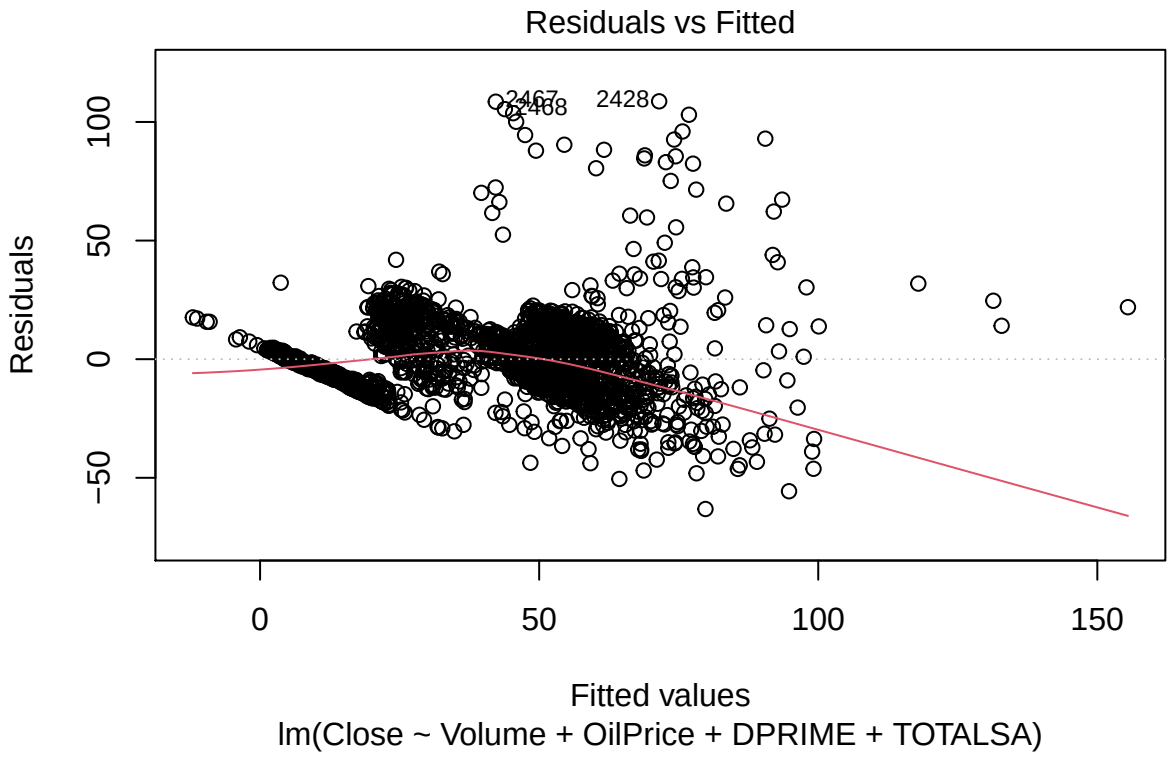


There is no evidence that the deaths is significant

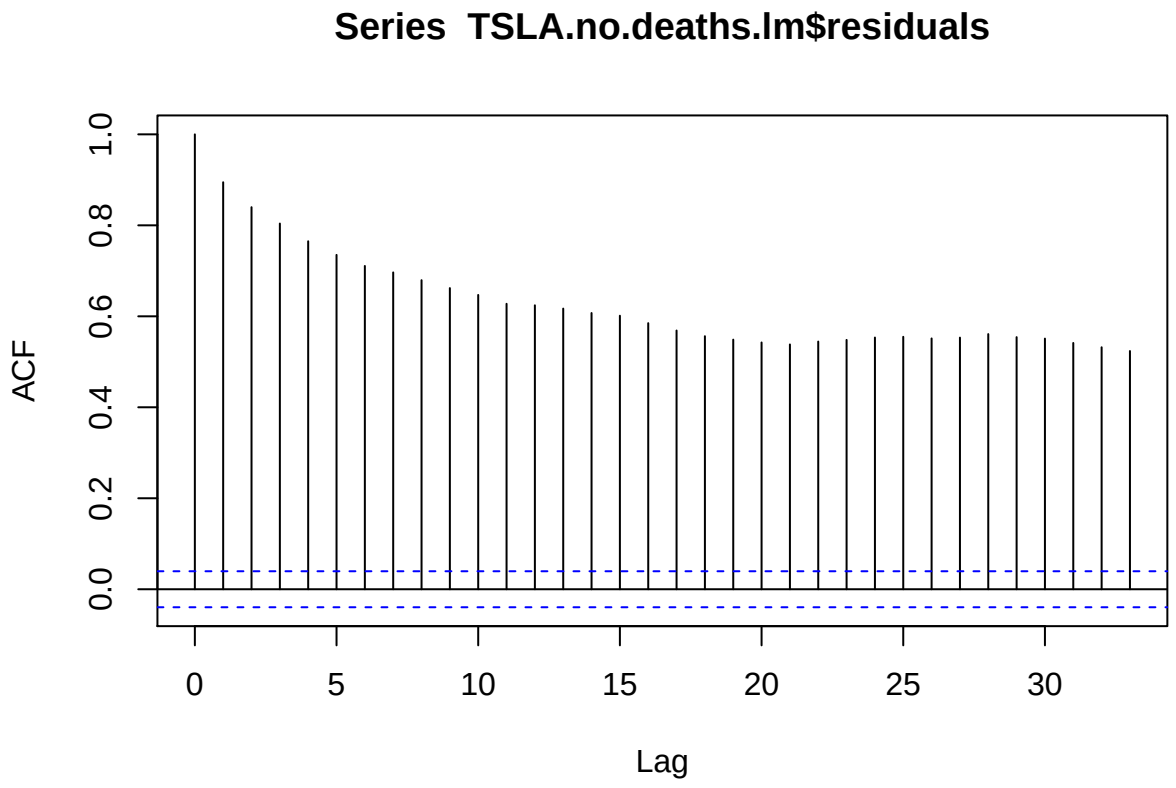
```
TSLA.no.deaths.lm = lm(Close ~ Volume + OilPrice + DPRIME + TOTALSA, data = TSLA.full)
summary(TSLA.no.deaths.lm)
```

```
##
## Call:
## lm(formula = Close ~ Volume + OilPrice + DPRIME + TOTALSA, data = TSLA.full)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.148  -8.745  -2.163   8.366 108.701
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.374e+00  4.038e+00   1.826   0.0679 .
## Volume       3.513e-07  1.298e-08  27.073   <2e-16 ***
## OilPrice    -4.434e-01  1.431e-02 -30.974   <2e-16 ***
## DPRIME       6.416e+00  4.564e-01  14.059   <2e-16 ***
## TOTALSA      6.009e+01  6.161e+00   9.754   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.11 on 2448 degrees of freedom
##   (161 observations deleted due to missingness)
## Multiple R-squared:  0.6581, Adjusted R-squared:  0.6575
## F-statistic:  1178 on 4 and 2448 DF,  p-value: < 2.2e-16
```

```r
plot(TSLA.no.deaths.lm, which = 1)
```

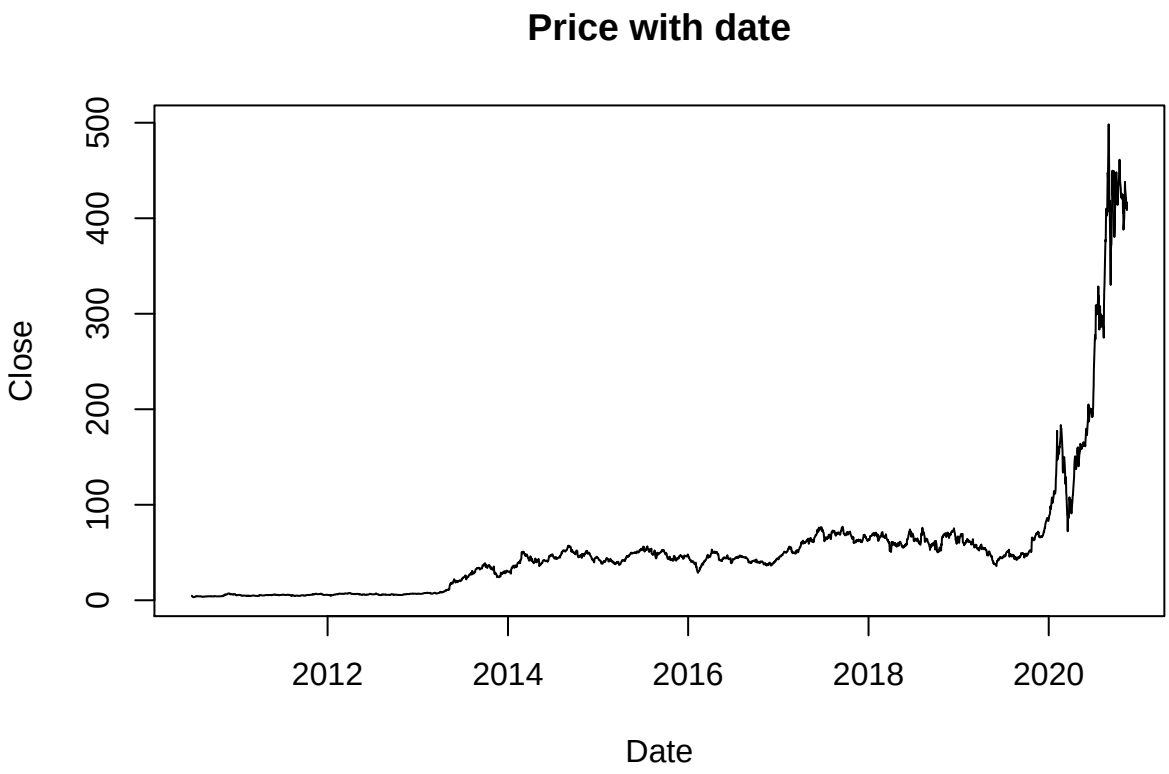



```r
acf(TSLA.no.deaths.lm$residuals)
```



From acf plot the errors are highly correlated.

```
plot(TSLA.full$Date, TSLA.full$Close, type = 'l',
     xlab = "Date", ylab = "Close", main = "Price with date")
```



the time series is not stationary so we log the Price and diff it

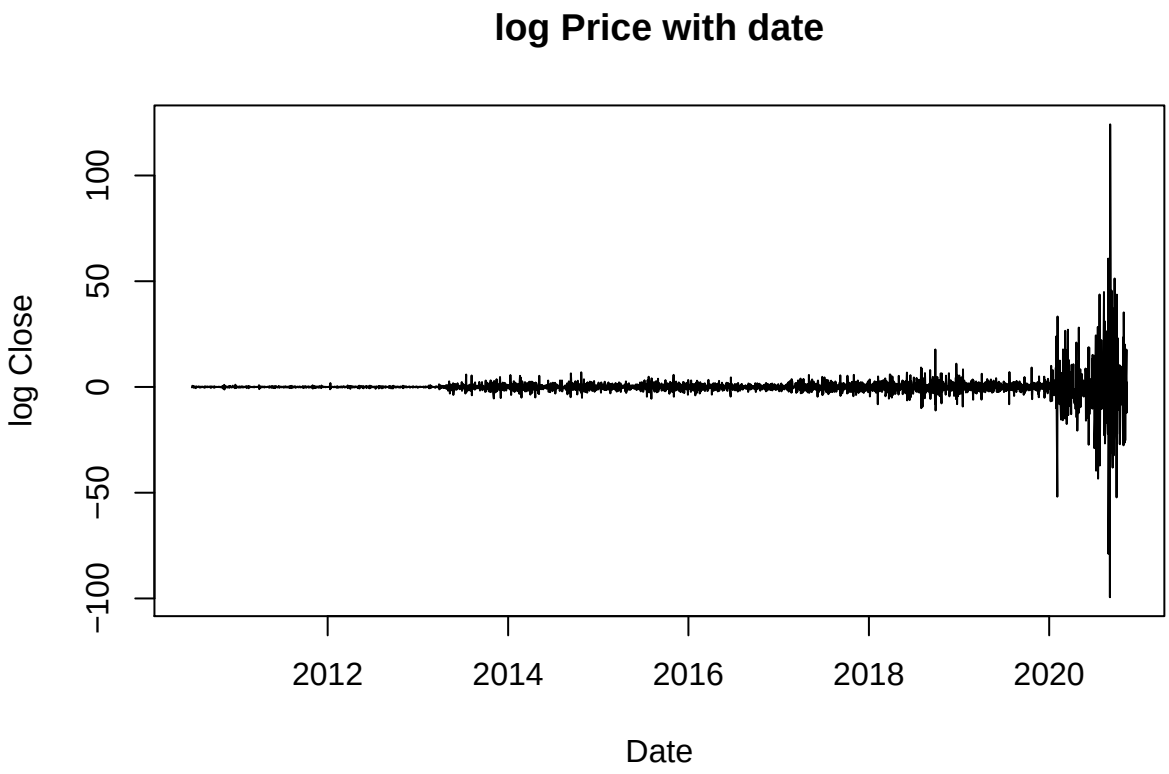
```
Price = log(TSLA.full$Close)
ndiffs(Price)
```

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

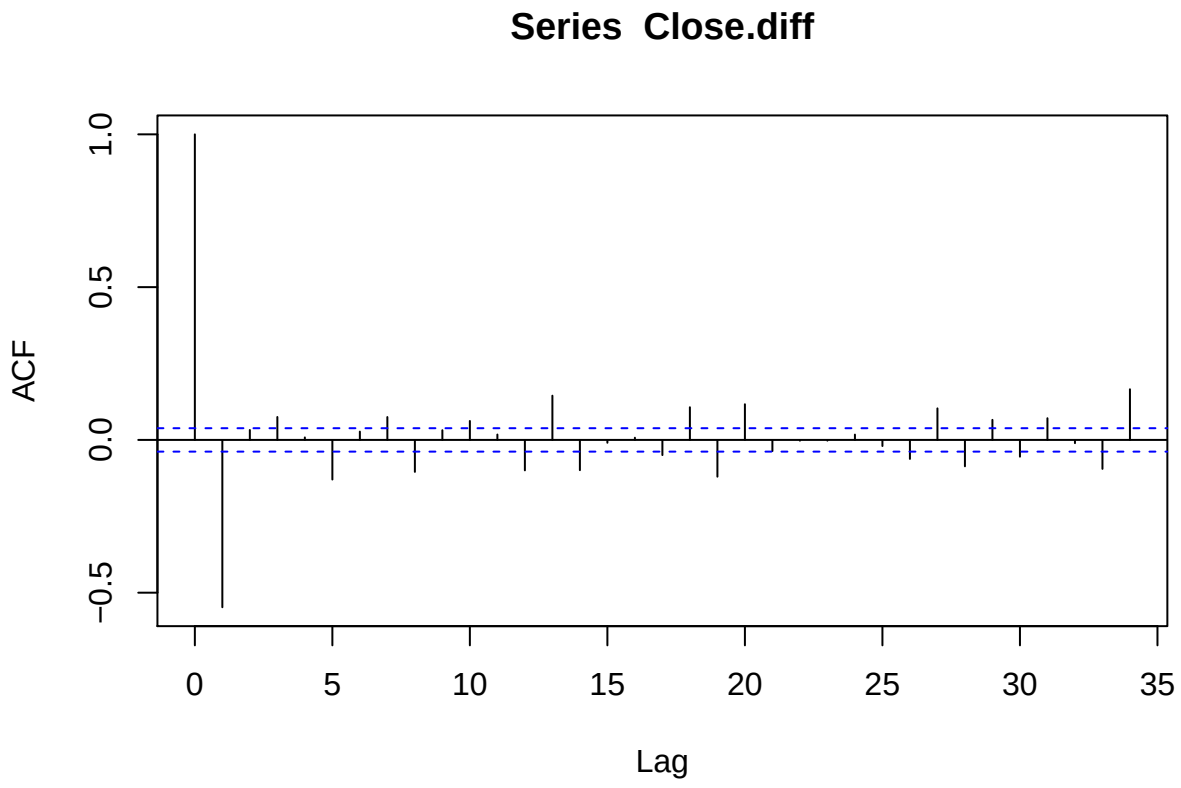
```
Close.diff = diff(TSLA.full$Close, 1)
ndiffs(Close.diff)
```

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

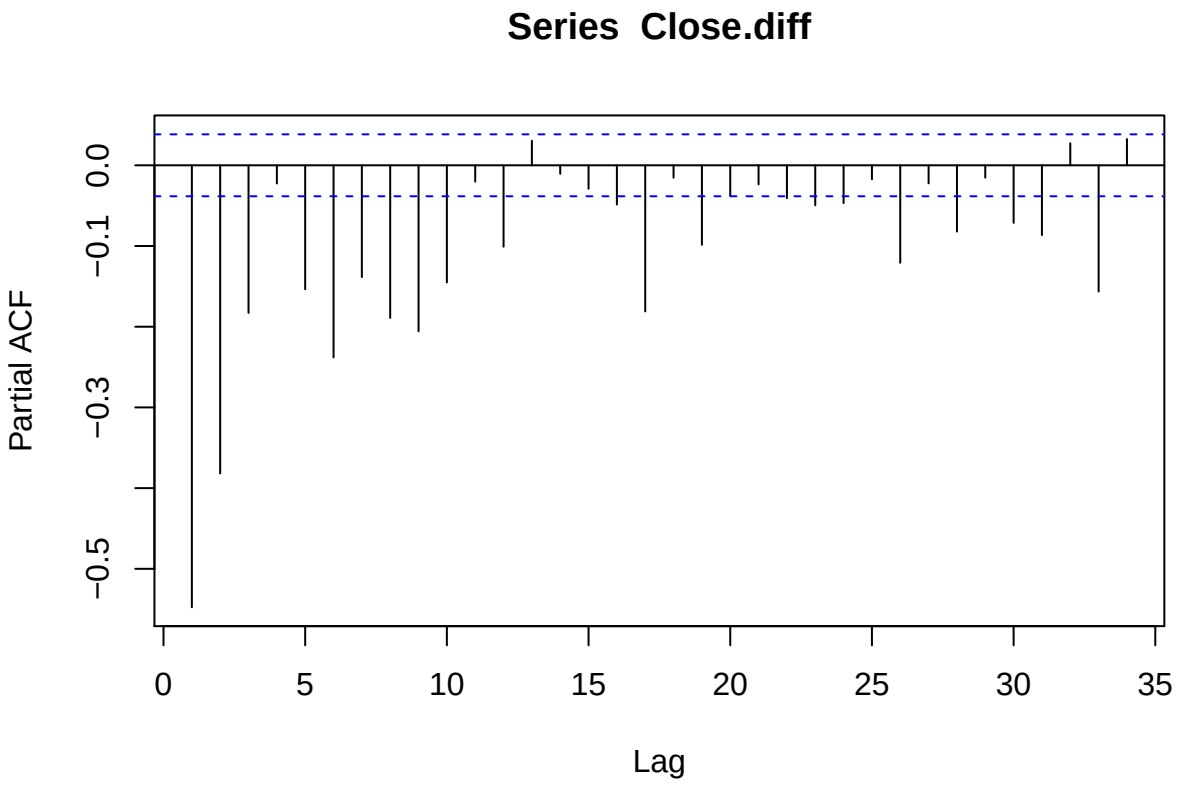
```
Close.diff = diff(Close.diff, 1)
plot(TSLA.full$Date[1:length(Close.diff)] ,Close.diff, type = 'l',
     xlab = "Date", ylab = "log Close", main = "log Price with date")
```




```
acf(Close.diff)
```




```
pacf(Close.diff)
```




```
auto.arima(TSLA.full$Close, trace = T, xreg = cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPR
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,2,2)                    : -1471.996
##  ARIMA(0,2,0)                    : 11810.58
##  ARIMA(1,2,0)                    : 11113.72
##  ARIMA(0,2,1)                    : Inf
##  ARIMA(1,2,2)                    : -1645.739
##  ARIMA(0,2,2)                    : Inf
##  ARIMA(1,2,1)                    : -1523.86
##  ARIMA(1,2,3)                    : -1651.51
##  ARIMA(0,2,3)                    : Inf
##  ARIMA(2,2,3)                    : -1761.045
##  ARIMA(3,2,3)                    : Inf
##  ARIMA(2,2,4)                    : -1763.744
##  ARIMA(1,2,4)                    : -1673.424
##  ARIMA(3,2,4)                    : Inf
##  ARIMA(2,2,5)                    : -1762.126
##  ARIMA(1,2,5)                    : -1674.423
##  ARIMA(3,2,5)                    : Inf
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(2,2,4)                    : Inf
```

```
##  ARIMA(2,2,5)                    : Inf
##  ARIMA(2,2,3)                    : Inf
##  ARIMA(1,2,5)                    : Inf
##  ARIMA(1,2,4)                    : Inf
##  ARIMA(1,2,3)                    : Inf
##  ARIMA(1,2,2)                    : Inf
##  ARIMA(1,2,1)                    : Inf
##  ARIMA(2,2,2)                    : Inf
##  ARIMA(1,2,0)                    : 11178.25
##
##  Best model: Regression with ARIMA(1,2,0) errors
```

```
## Series: TSLA.full$Close
## Regression with ARIMA(1,2,0) errors
##
## Coefficients:
##           ar1  xreg1   xreg2   xreg3     xreg4
##       -0.4844      0  0.1188  0.0946  -60.3090
## s.e.   0.0178      0  0.0257  0.0755    3.4079
##
## sigma^2 estimated as 5.196:  log likelihood=-5583.11
## AIC=11178.22   AICc=11178.25   BIC=11213.08
```

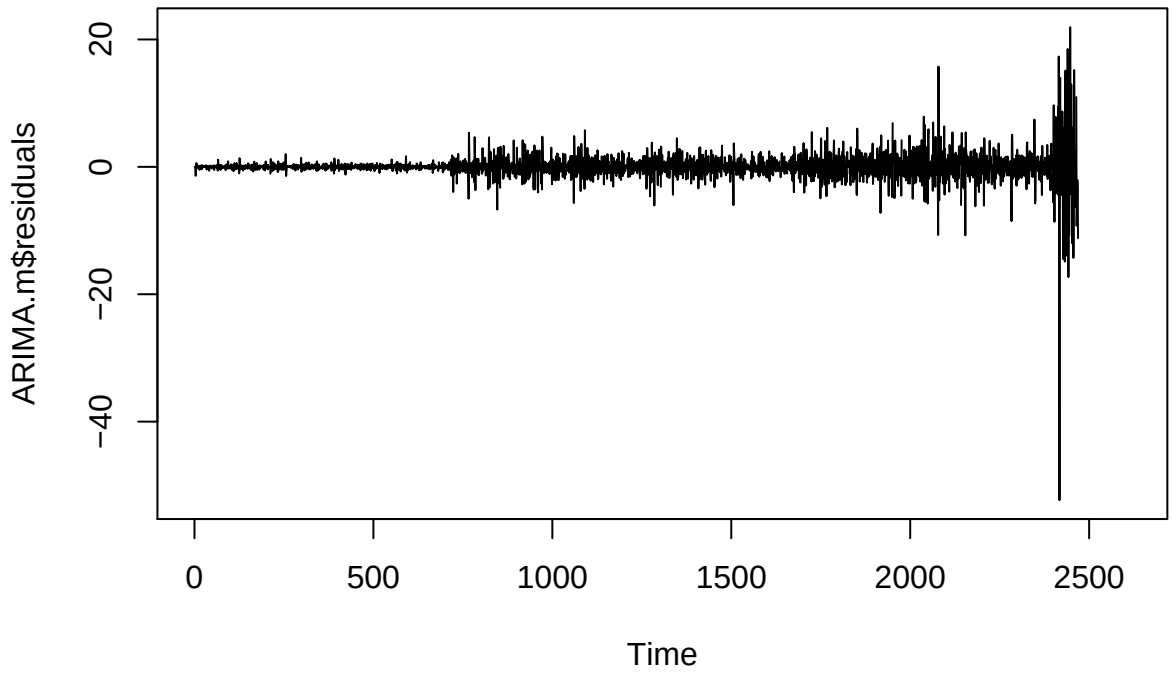
The best model is ARIMA(1,2,0) SO

```
ARIMA.m = arima(TSLA.full$Close,
                order = c(1,2,0),
                xreg = cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)
                method="ML")
summary(ARIMA.m)
```

```
##
## Call:
## arima(x = TSLA.full$Close, order = c(1, 2, 0), xreg = cbind(TSLA.full$Volume,
##     TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA), method = "ML")
##
## Coefficients:
##           ar1
##       -0.4844
## s.e.   0.0178
##       cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)1
##                                                                                       0
## s.e.                                                                                  0
##       cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)2
##                                                                                  0.1188
## s.e.                                                                               0.0257
##       cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)3
##                                                                                  0.0946
## s.e.                                                                               0.0755
##       cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)4
##                                                                                -60.3090
## s.e.                                                                             3.4079
##
## sigma^2 estimated as 5.527:  log likelihood = -5583.11,  aic = 11178.22
##
## Training set error measures:
```

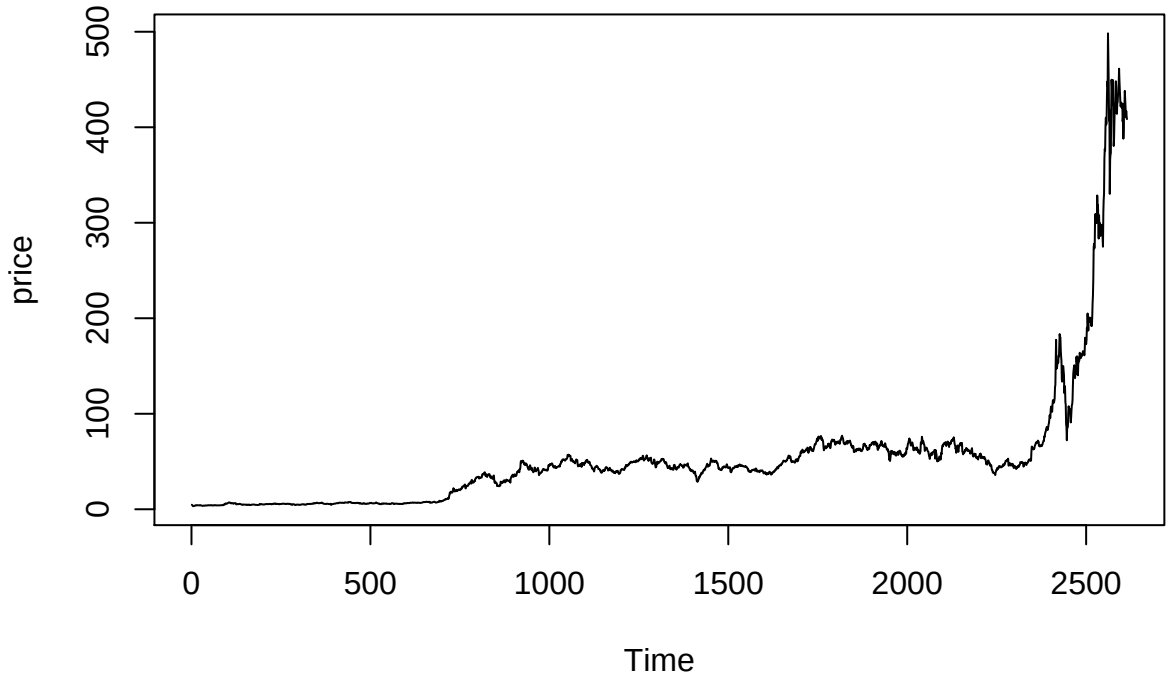
```
##                         ME     RMSE      MAE         MPE     MAPE      MASE
## Training set -0.005657444 2.350017 1.171349 -0.01540201 3.400764 0.7808799
##                     ACF1
## Training set -0.1690465
```

```r
plot(ARIMA.m$residuals)
```

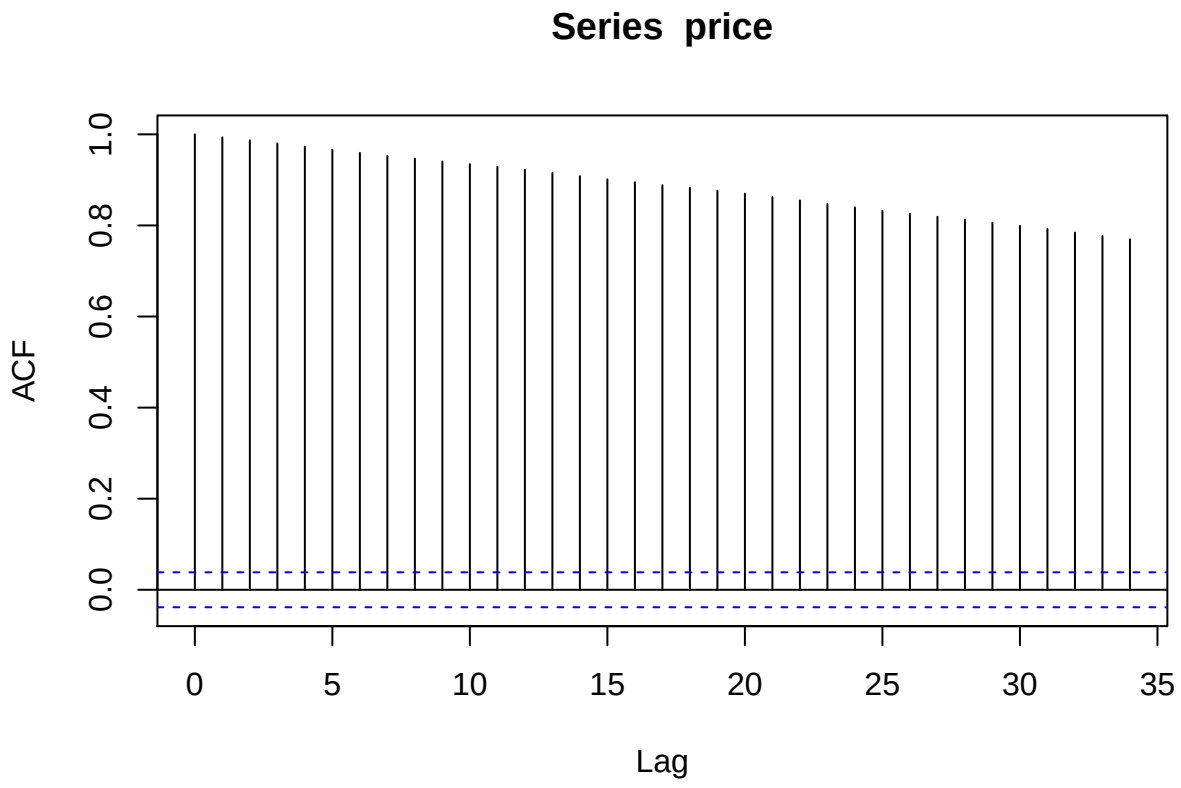



```r
P = predict(ARIMA.m,
        newxreg = cbind(TSLA.full$Volume, TSLA.full$OilPrice, TSLA.full$DPRIME, TSLA.full$TOTALSA)  )
```

```r
price = ts(data = TSLA.full$Close)
plot(price)
```




```
acf(price)
```




```
ndiffs(price)
```

```
## [1] 2
```

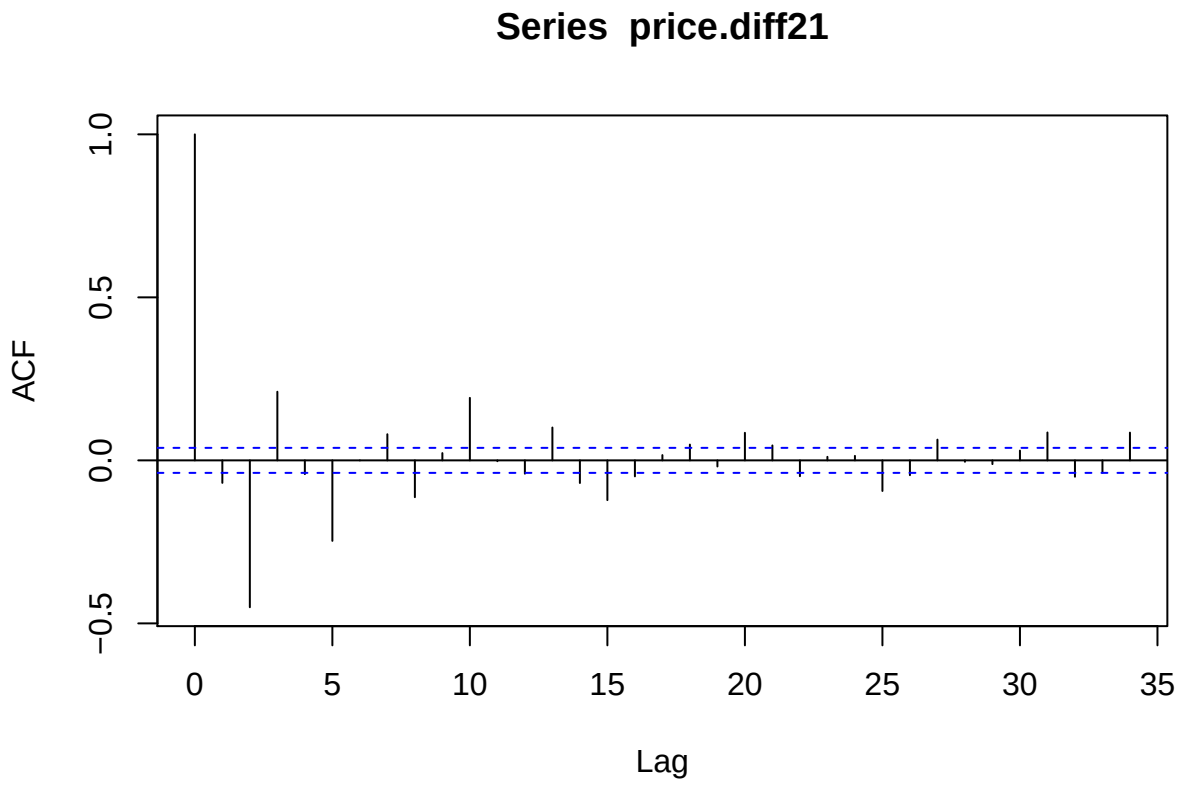
```
price.diff2 = diff(price, 2)
ndiffs(price.diff2)
```

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

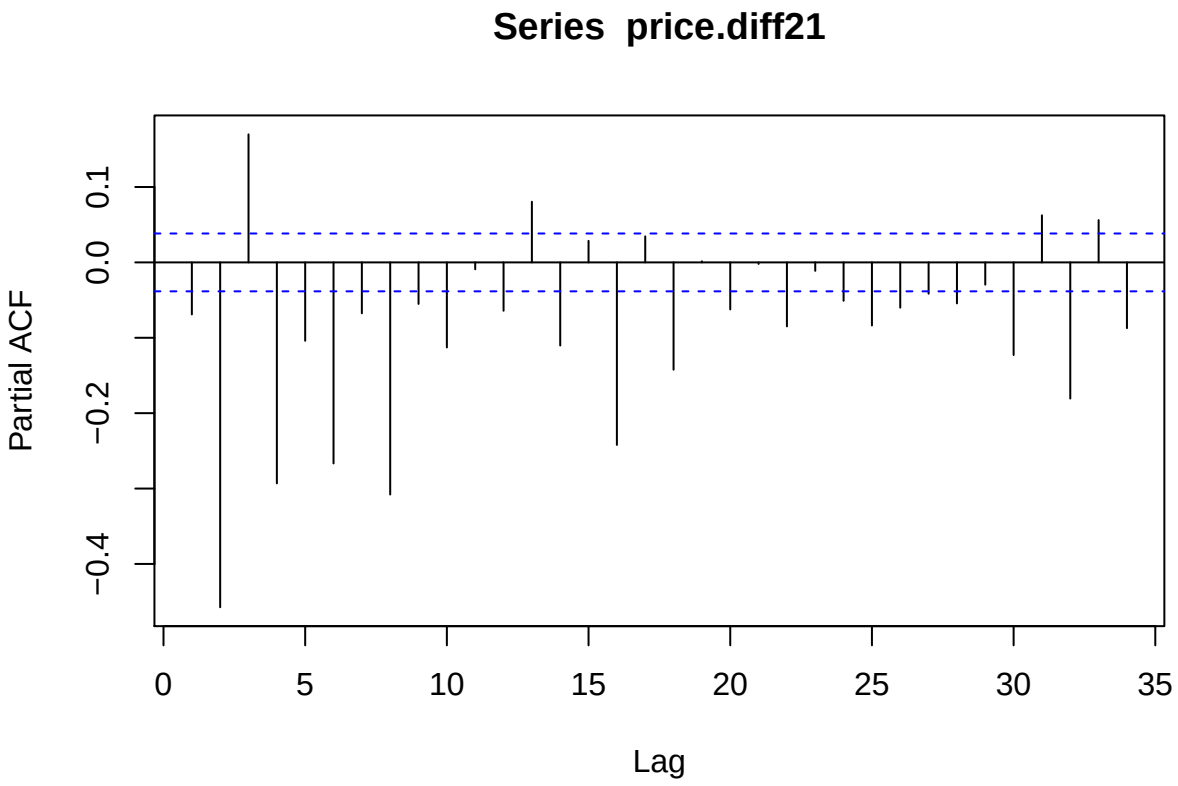
```
price.diff21 = diff(price.diff2, 1)
ndiffs(price.diff21)
```

```
## [1] 0
```

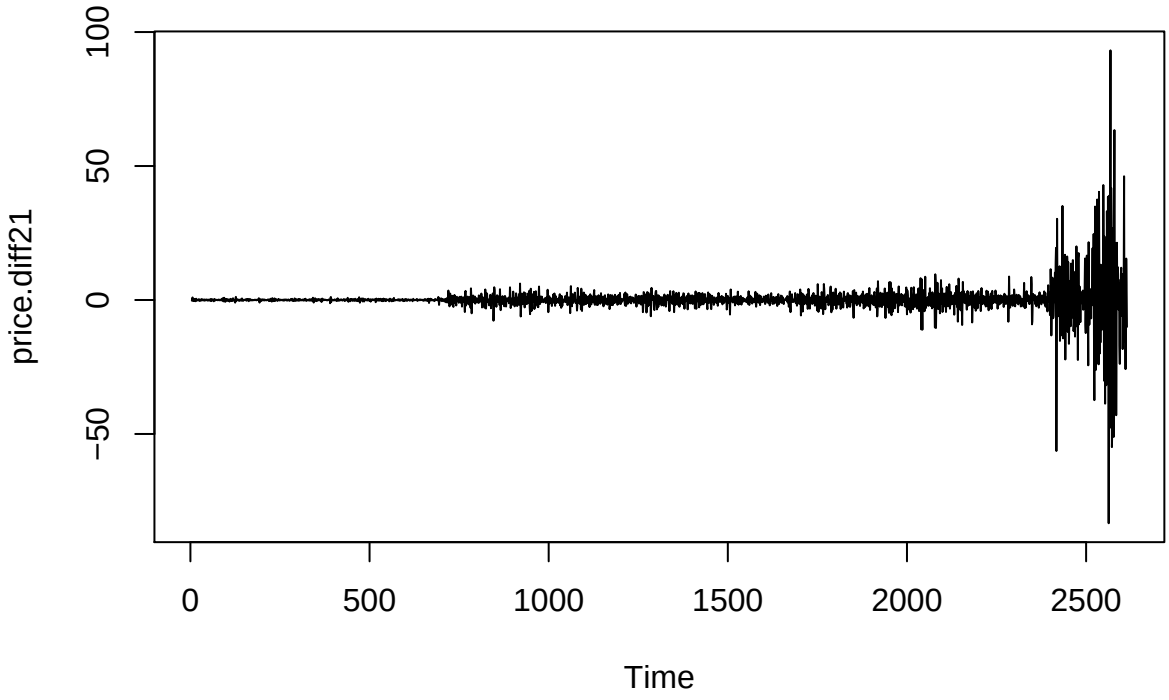
```
acf(price.diff21)
```




```
pacf(price.diff21)
```




```
plot(price.diff21)
```




```
auto.arima(price, trace = T)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,2,2)                    : Inf
##  ARIMA(0,2,0)                    : 17044.8
##  ARIMA(1,2,0)                    : 16116.68
##  ARIMA(0,2,1)                    : Inf
##  ARIMA(2,2,0)                    : 15707.75
##  ARIMA(3,2,0)                    : 15622.1
##  ARIMA(4,2,0)                    : 15623.8
##  ARIMA(3,2,1)                    : Inf
##  ARIMA(2,2,1)                    : Inf
##  ARIMA(4,2,1)                    : Inf
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(3,2,0)                    : 15627.26
##
##  Best model: ARIMA(3,2,0)
```

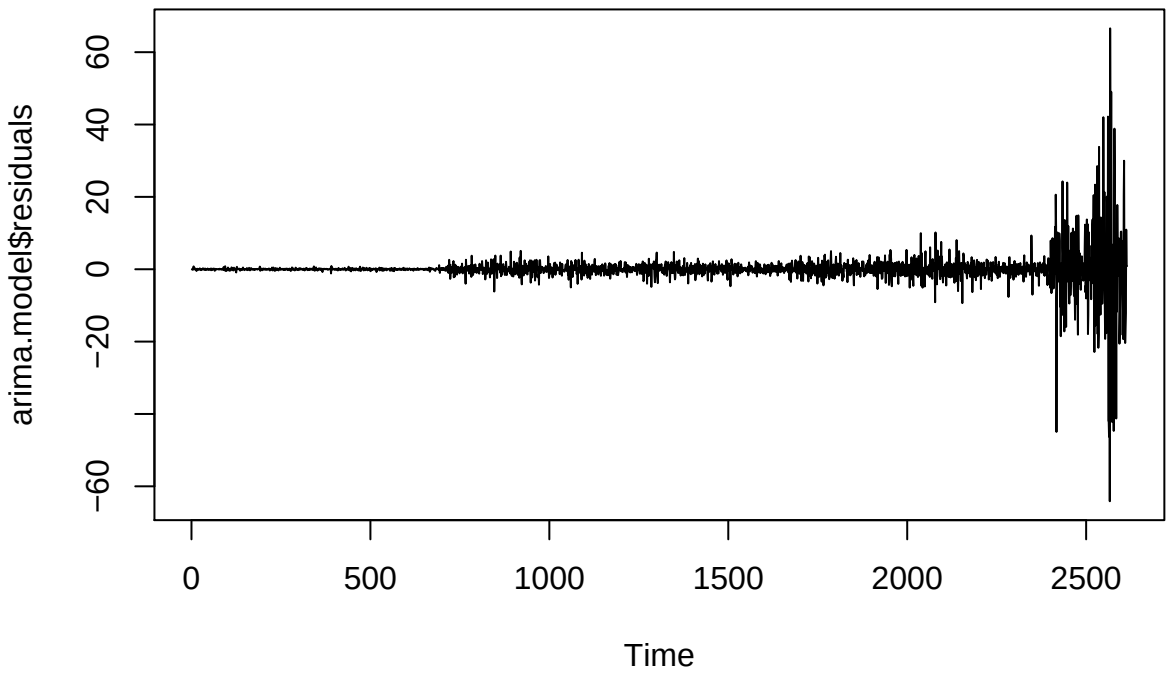
```
## Series: price
## ARIMA(3,2,0)
##
## Coefficients:
##           ar1     ar2      ar3
```

```
##       -0.8263  -0.520  -0.1827
## s.e.   0.0192   0.023   0.0193
##
## sigma^2 estimated as 23.17:  log likelihood=-7809.62
## AIC=15627.25   AICc=15627.26   BIC=15650.72
```

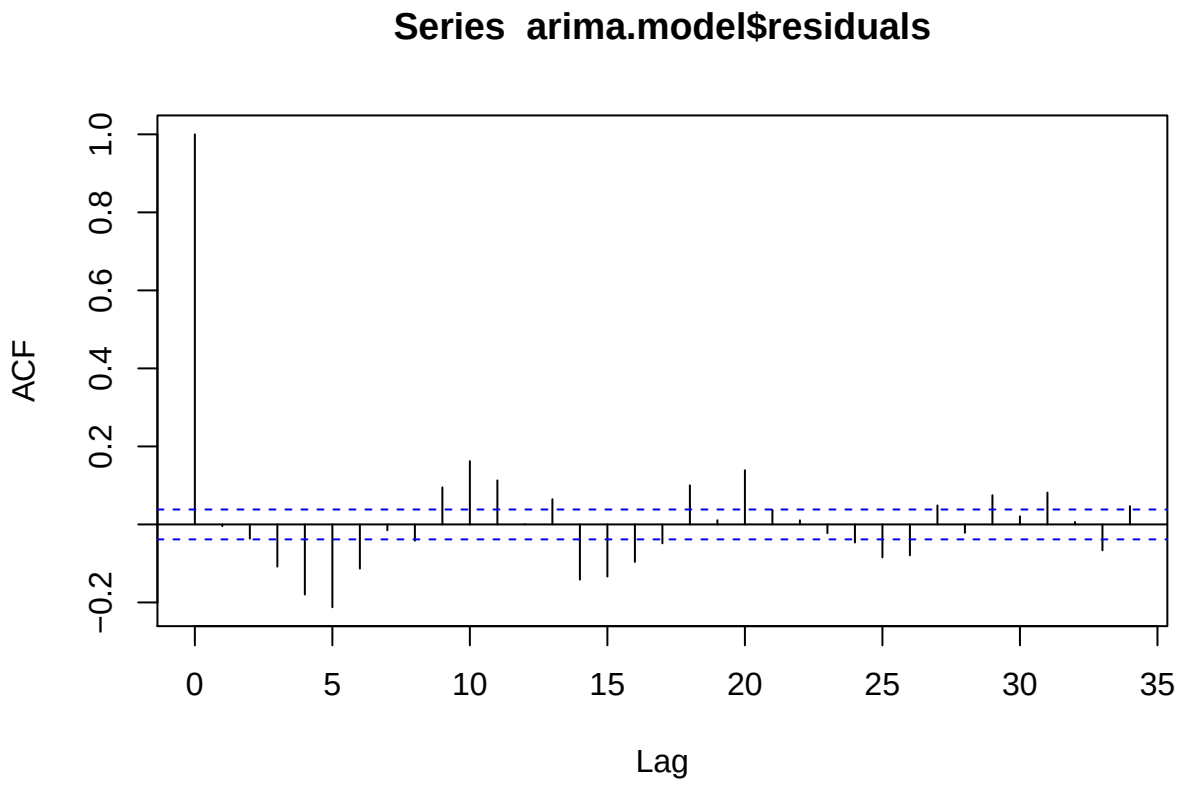
```
arima.model = arima(price, order = c(3,2,0), method = "ML")

summary(arima.model)
```

```
##
## Call:
## arima(x = price, order = c(3, 2, 0), method = "ML")
##
## Coefficients:
##           ar1     ar2      ar3
##       -0.8263  -0.520  -0.1827
## s.e.   0.0192   0.023   0.0193
##
## sigma^2 estimated as 23.14:  log likelihood = -7809.62,  aic = 15627.25
##
## Training set error measures:
##                        ME     RMSE      MAE       MPE     MAPE     MASE
## Training set -0.001950509 4.808729 1.743037 0.0121328 2.780508 1.161996
##                      ACF1
## Training set -0.004363582
```

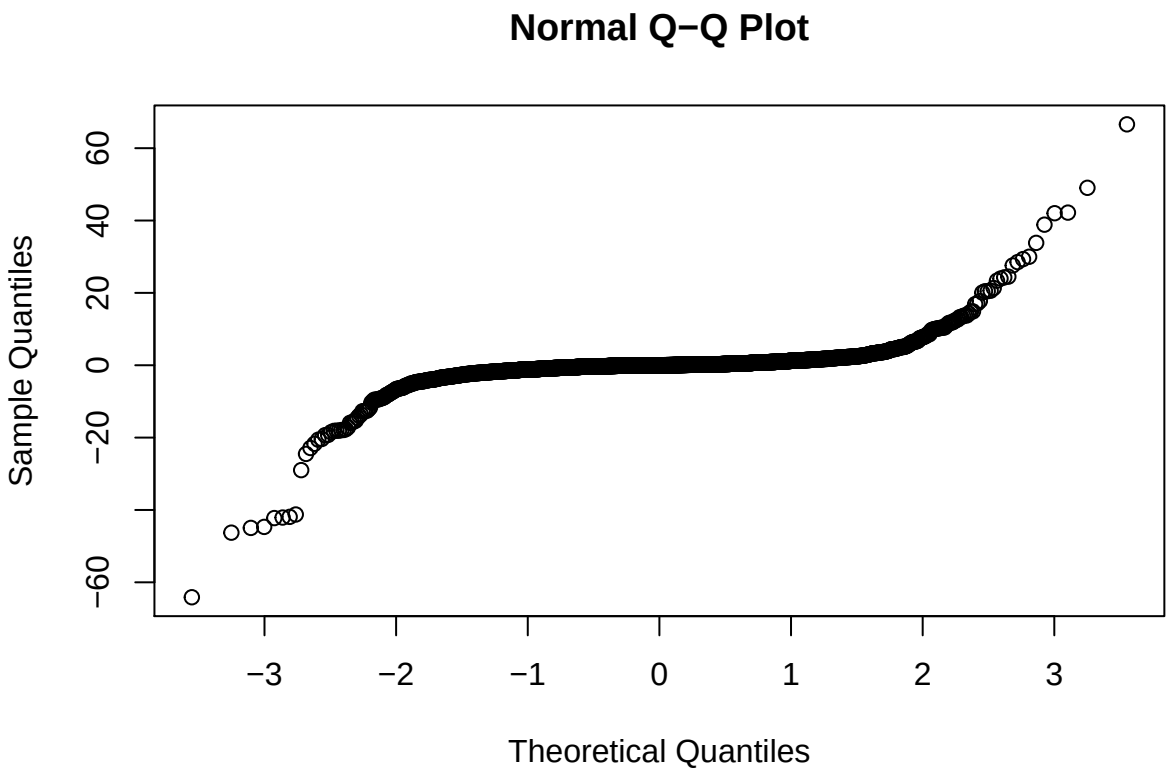
```
plot(arima.model$residuals)
```




```
acf(arima.model$residuals)
```




```
qqnorm(arima.model$residuals)
```




```
Box.test(arima.model$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.model$residuals
## X-squared = 0.04983, df = 1, p-value = 0.8234
```

```
pred = predict(arima.model, 10)
pred
```

```
## $pred
## Time Series:
## Start = 2615
## End = 2624
## Frequency = 1
##  [1] 406.5806 404.6747 401.6749 399.3270 397.0068 394.5246 392.0428 389.6398
##  [9] 387.2011 384.7508
##
## $se
## Time Series:
## Start = 2615
## End = 2624
## Frequency = 1
##  [1]  4.810570  7.417724 10.382657 14.022083 18.106999 22.389519 27.012926
##  [8] 31.946606 37.137646 42.579304
```

```r
plot(c(2605:2615),price[(length(price)-10) : length(price)-1], type = 'l')
lines(pred$pred, col = "red")
```