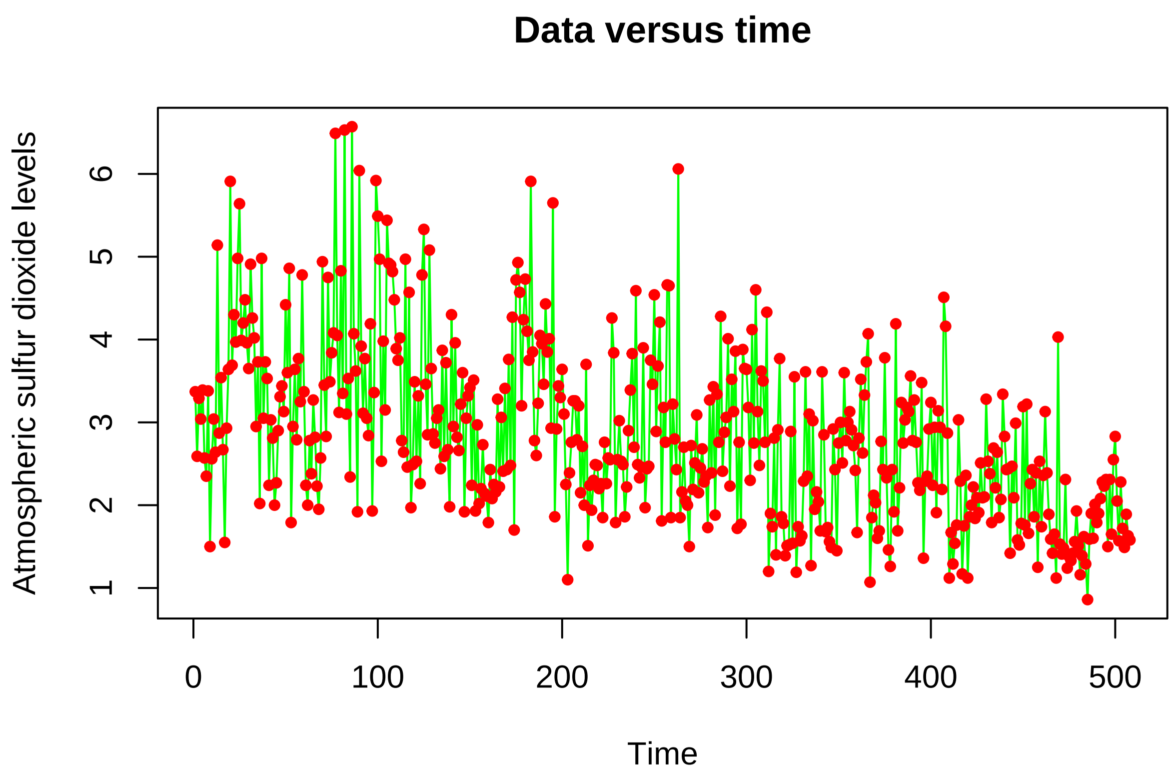
2. Part 1: Determination of d.

(a) Present a plot of the original data.



(b) Make observations about possible trends.

There seems to have a downward trend.

(c) Report the results of the least squares polynomial fit and the relative sizes of coefficients.

```
##   (Intercept)          time     I(time^2)
##  3.600651e+00 -1.761490e-03 -3.569062e-06
```

a least squares estimate for a polynomial model

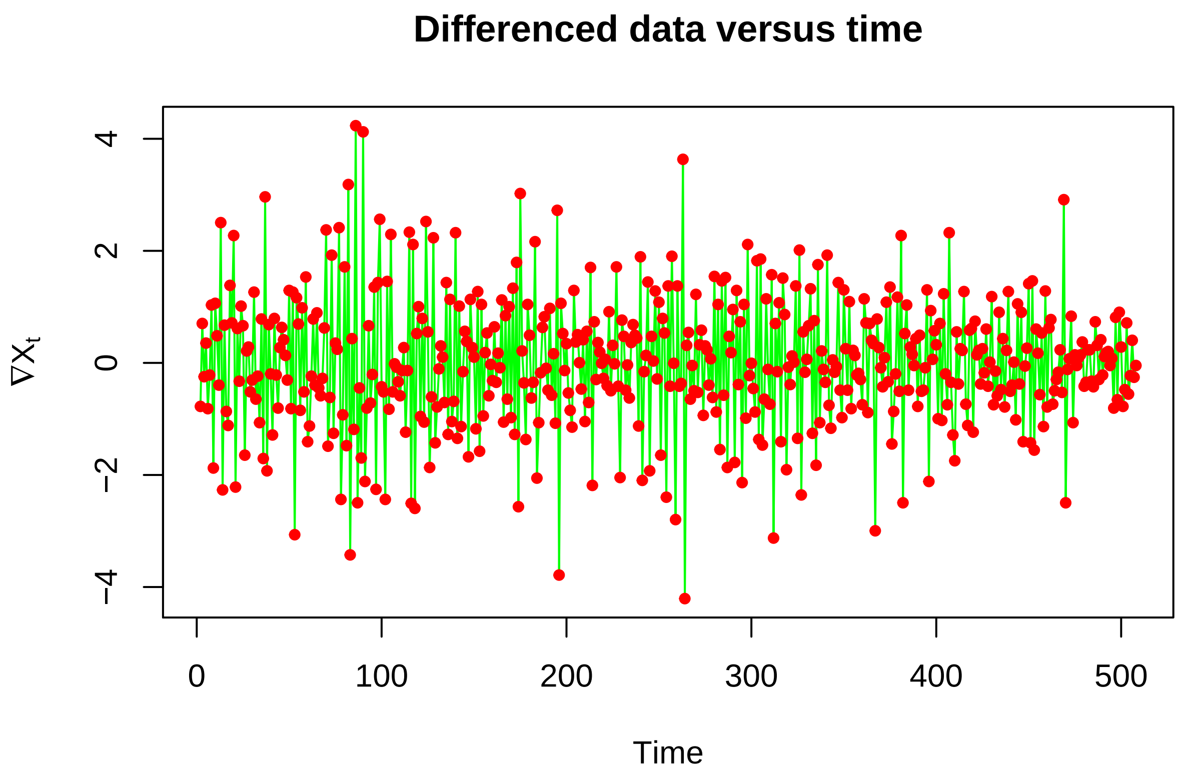
$$a_0 + a_1 t + a_2 t^2$$

$a_0 = 3.600651$, $a_1 = -1.761490 \times 10^{-3}$, $a_2 = -3.569062 \times 10^{-6}$.

(d) Specify d and explain your choice.

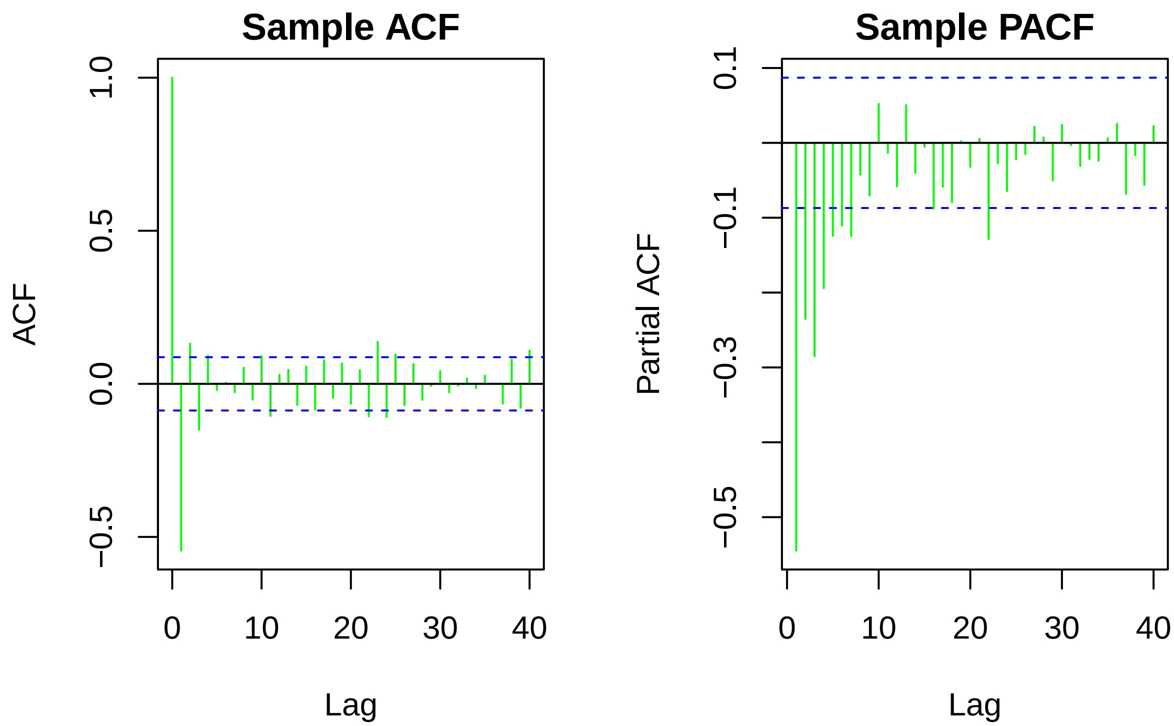
The coefficient a_2 is -3.569062×10^{-6} which is relatively small, so i think a linear model is more appropriate. In other words, we can differencing the data once to remove the linear trend, so d = 1.

(e) For the chosen d, display a plot of the mean-centered differenced data.



3. Part 2: Determination of p and q for the mean-centered differenced data.

(a) Show the plot of the sample acf/pacf for the mean-centered differenced data.



(b) Give observations on possible orders of dependency.

The acf is significantly nonzero at lag 4 and almost insignificant after lag 4 while there seems an exponential decay in pacf values for small increasing lags, so we might try an MA(4) model as an initial guess.

(c) For the ARMA(p, q) estimated using MLE, show plots of the model acf/pacf values together with the sample acf/pacf and plots of the model residuals for four choices of p and q.

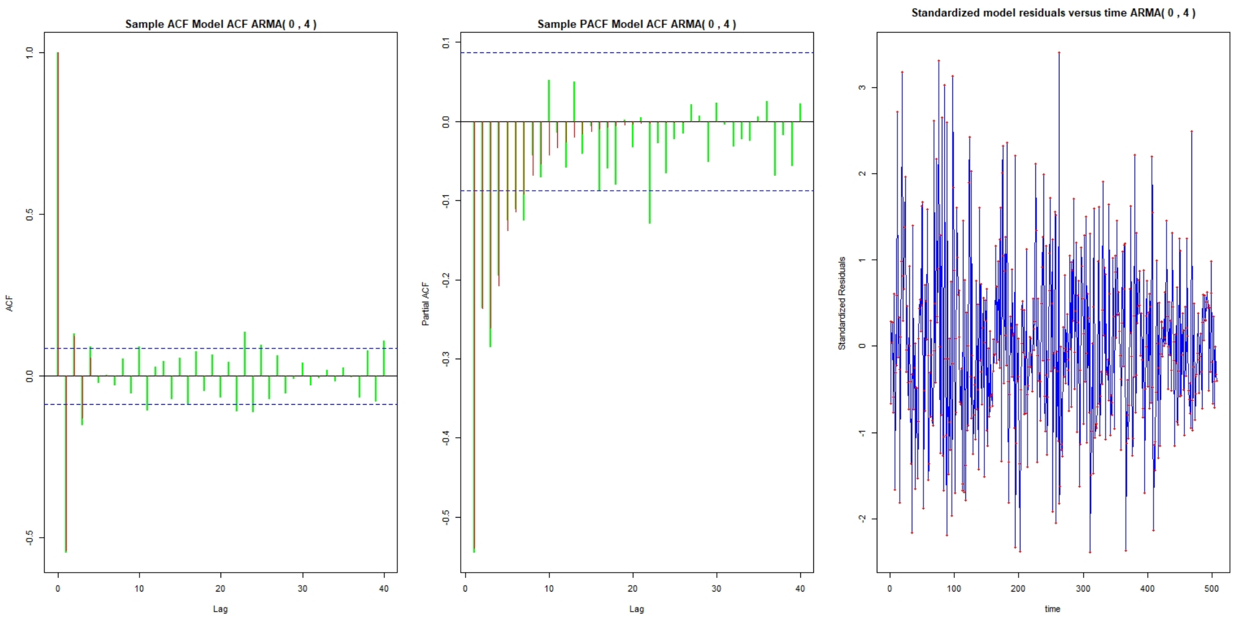


Figure 1: MA(4) model

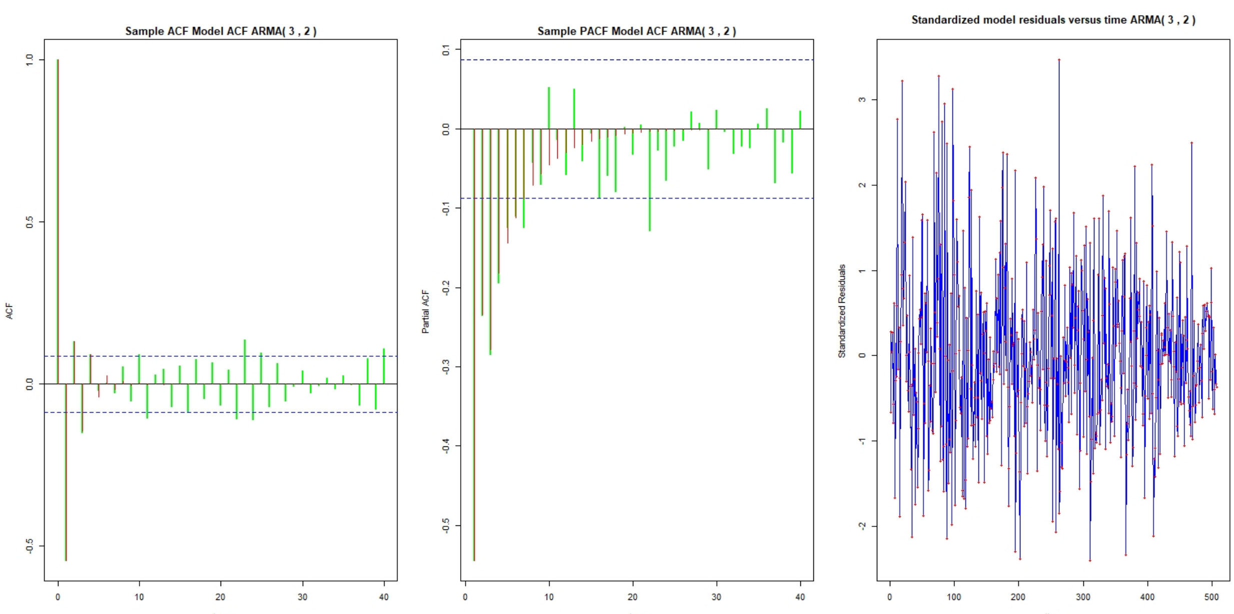


Figure 2: ARMA(3,2) model

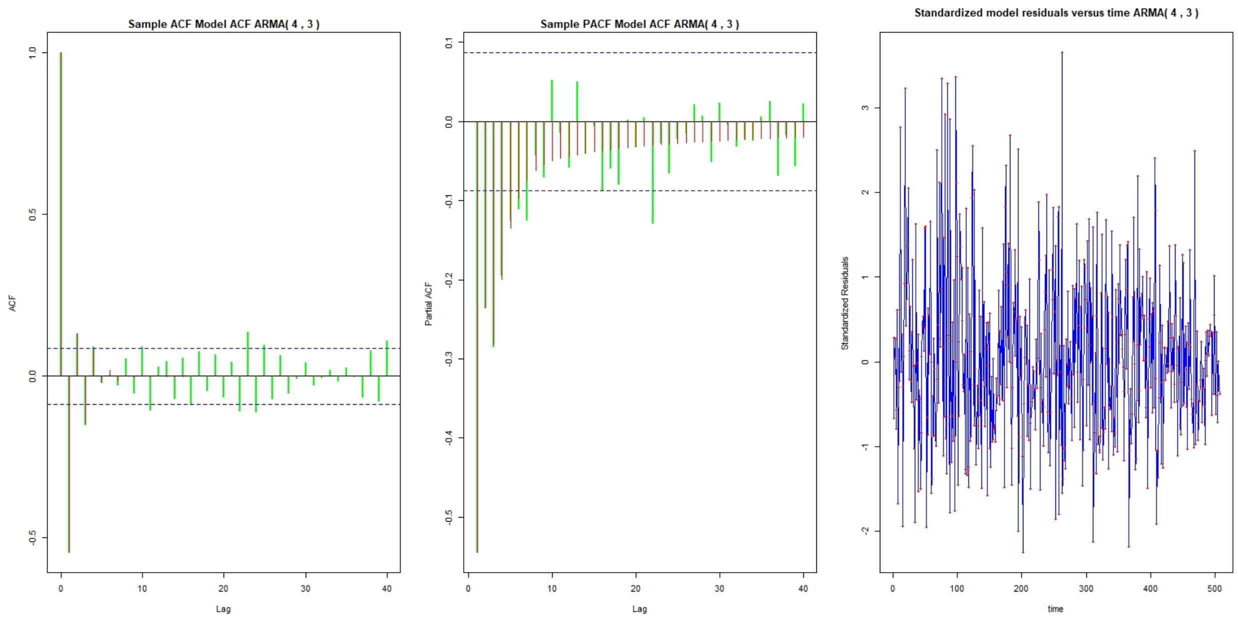


Figure 3: ARMA(4,3) model

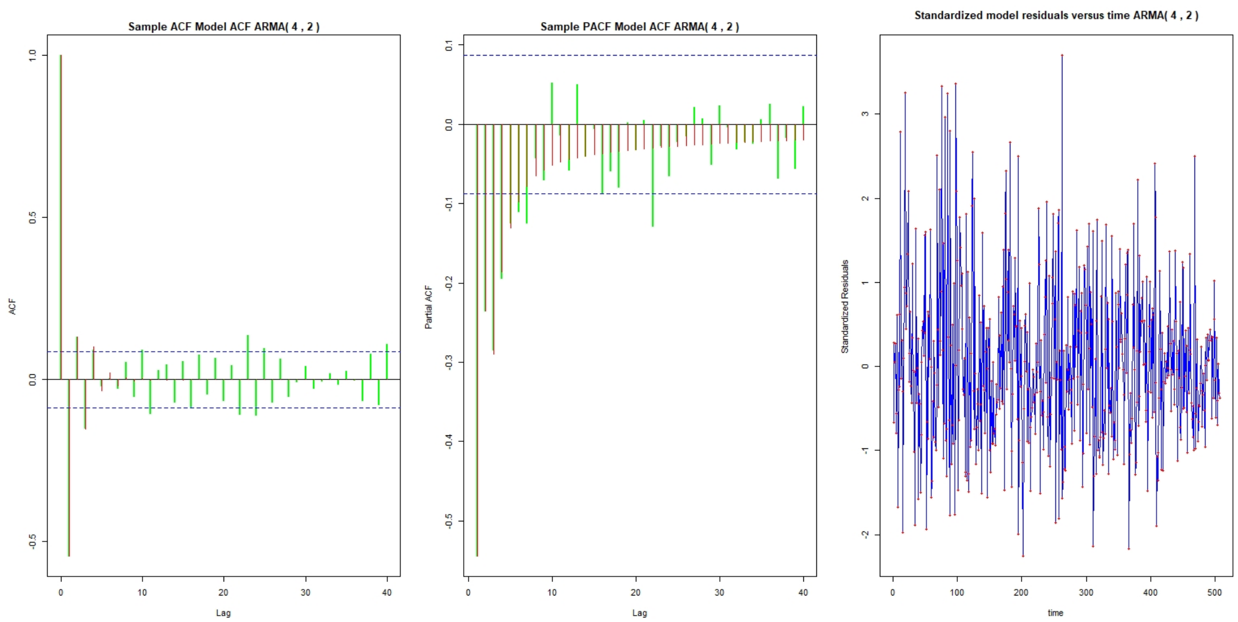


Figure 4: ARMA(4,2) model

(d) Give the aic or aicc values for each of the estimated models in (c).

```
##                    AIC
## ARMA( 0 , 4 ) 1321.771
## ARMA( 3 , 2 ) 1323.238
## ARMA( 4 , 3 ) 1309.032
## ARMA( 4 , 2 ) 1307.177
```

(e) Specify the p and q values you choose and give the reason.

The standardized residuals appear stationary for these 4 models, we observed a good agreement in significant acf values in ARMA(3,2), ARMA(4,3), and ARMA(4,2) models. we observed a good agreement in significant pacf values up to lag 4 in ARMA(3,2) model, up to lag 5 in ARMA(4,2) and ARMA(4,3) models. Moreover, ARMA(4,2) has the minimum AIC values. In conclusion, i would choose p = 4, q = 2. The ARMA(4,2) model.

4. Part 3: Use MLE to fit the ARMA(p, q) model for the chosen p and q and analyze the model.

(a) Specify p, d, and q.

ARIMA(p, d, q) model with p = 4, d = 1, q = 2.

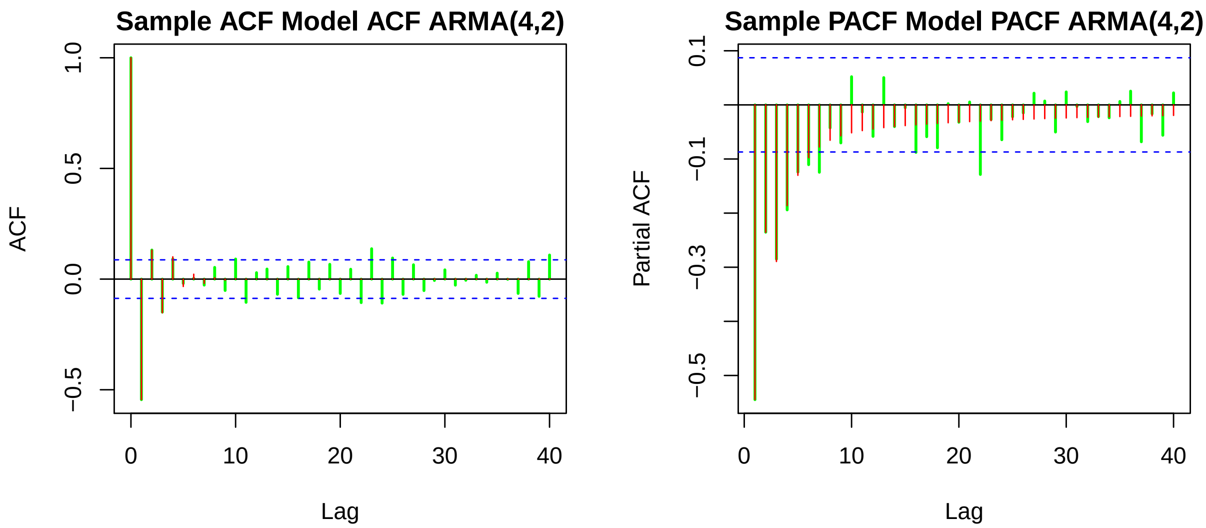
(b) Give the estimated coefficients for the MLE fit.

```
##           ar1           ar2           ar3           ar4           ma1
##  6.953413e-01  1.130086e-01 -1.096243e-01  1.319973e-01 -1.593950e+00
##           ma2     intercept
##  5.939507e-01  6.582580e-06
```

(c) Give the value of the AIC or AICC.

```
## [1] 1307.177
```

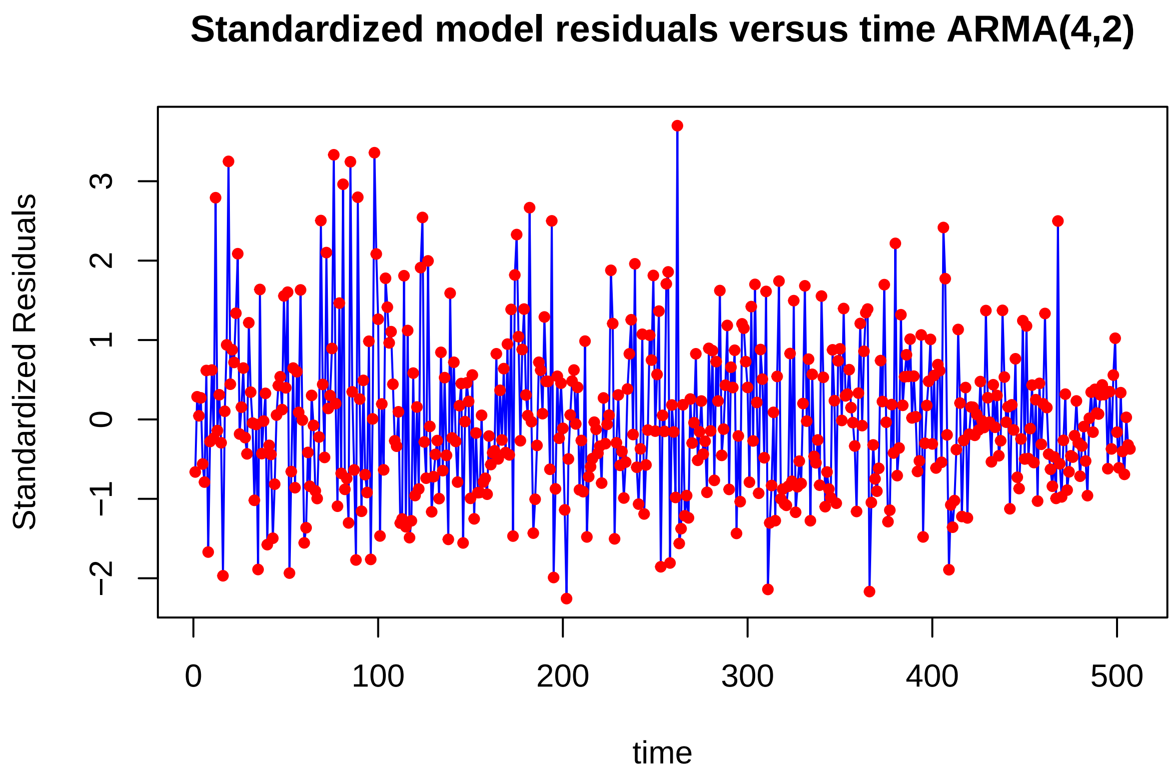
(d) Plot the model and sample acf/pacf values together.



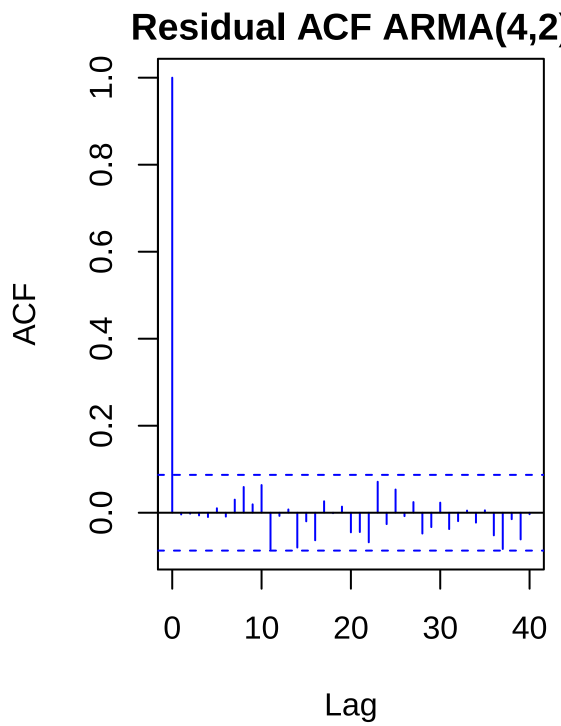
(e) Use the plot from (d) to assess the quality of the model fit.

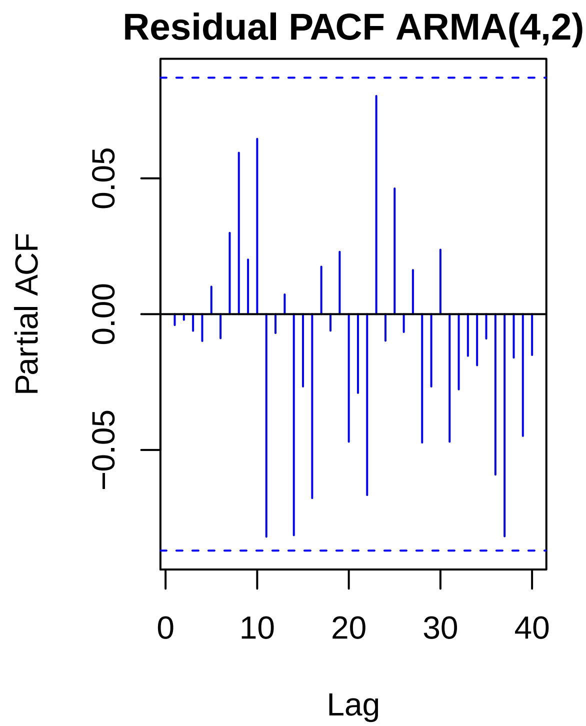
We observe a good agreement in significant acf values, and a good agreement in significant pacf values up to lag 5.

(f) Plot the standardized model residuals.



(g) Plot the sample acf/pacf for the standardized model residuals.



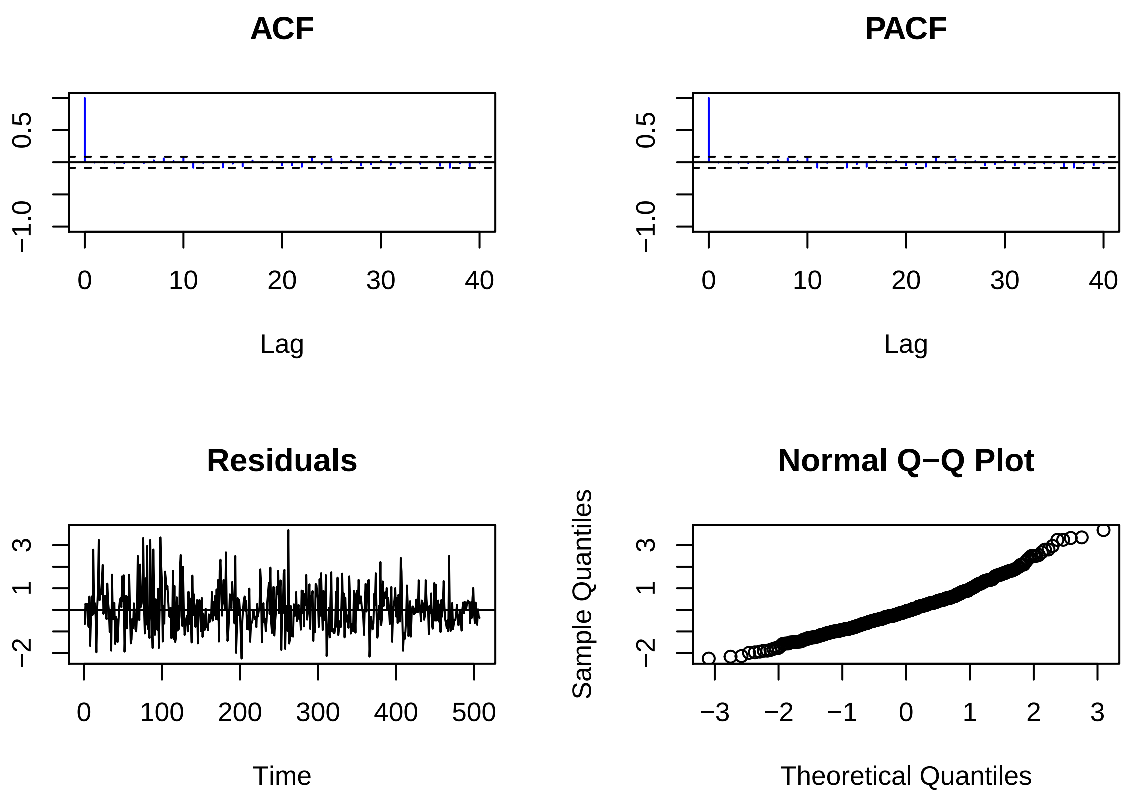


(h) Assess the plots from (f) and (g) with respect to the hypothesis that the model residuals behave like iid noise.

The standardized residual plot appear stationary and the residual acf and pacf plots do not support the rejection of iid noise hypothesis.

(i) Evaluate the Ljung-Box and McLeod-Li statistics and indicate if they support rejection of the hypothesis that the model residuals behave like iid noise.

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     15.65    0.7384
## McLeod-Li Q                Q ~ chisq(20)     53.26     1e-04 *
## Turning points T  (T-336.7)/9.5 ~ N(0,1)       345    0.3792
## Diff signs S        (S-253)/6.5 ~ N(0,1)       260     0.282
## Rank P       (P-64135.5)/1905.5 ~ N(0,1)     63045    0.5671
```



The Ljung-Box statistic is 15.65 with p-value 0.7384. This fails to support the rejection of iid noise hypothesis.

The McLeod-Li statistic is 53.26 with p-value almost 0. This support the rejection of iid noise hypothesis.

(j) Using (h) and (i), give a final assessment on the validity of the hypothesis that the model residuals behave like iid noise.

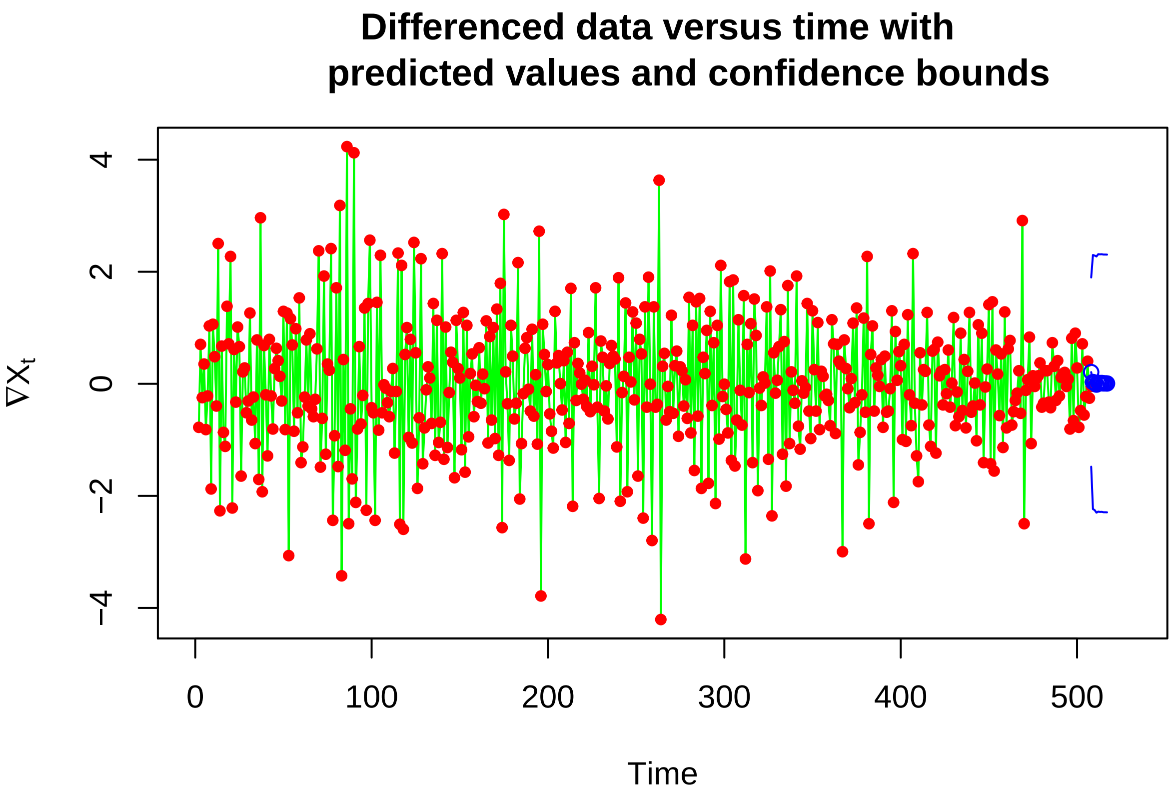
The residual acf and pacf plots do not support the rejection of iid noise hypothesis, and The Ljung-Box statistic test fails to support the rejection of iid noise hypothesis. In balance, i would conclude that residuals behave like iid noise.

(k) Use the results from (e) and (j) to give a summary evaluation about the quality of the fitted model.

We observe a good agreement in significant acf values and good agreement in significant pacf values up to lag 5. The standardized residual plot appear stationary, and the residual acf and pacf plots do not support the rejection of iid noise hypothesis. The Ljung-Box statistic test also fails to support the rejection of iid noise hypothesis. Overall, I would conclude that ARMA(4,2) model is a good fit.

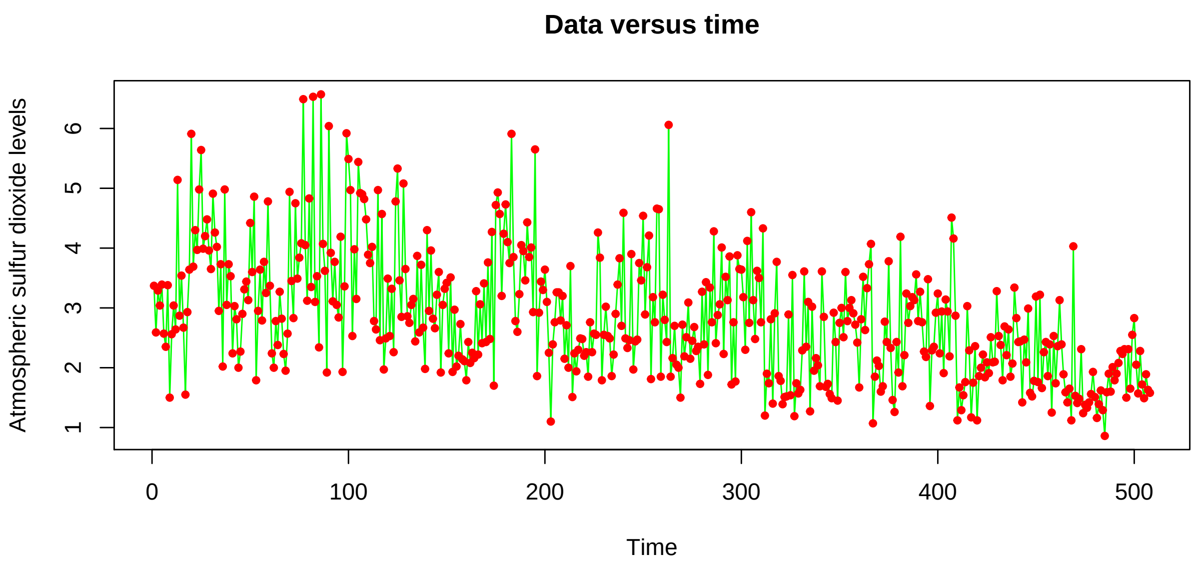
5. Part 4: Use the estimated model to make a forecast.

(a) Plot the data together with prediction of values for 10 time steps past the last time of the data together with the confidence bounds.



Appendix (code)

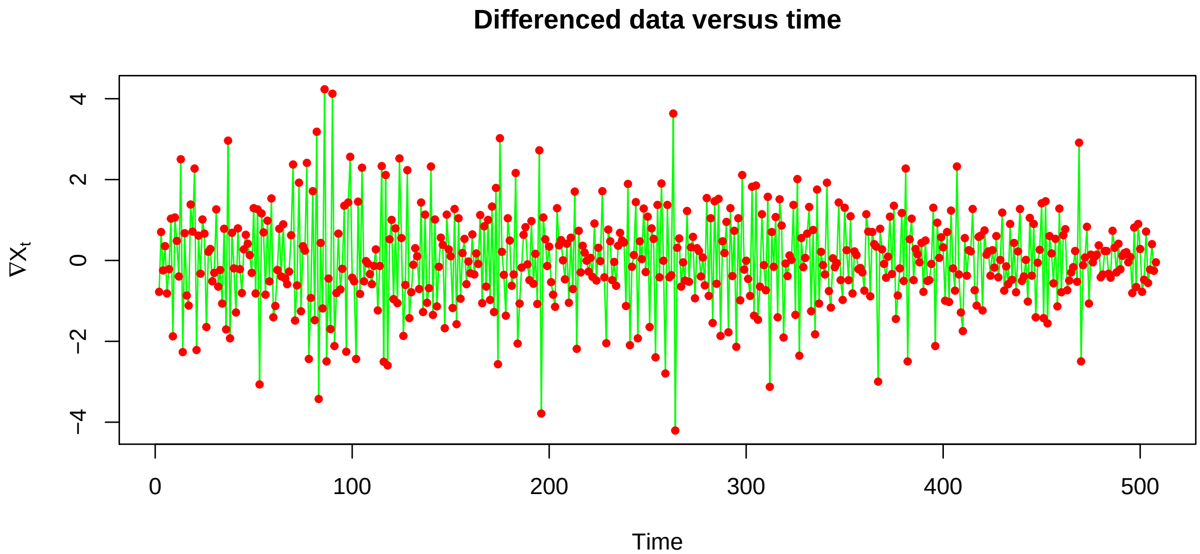
```
# load packages
library(TSA)
library(itsmr)

# read and plot data
so2 = read.table('so2.txt')
plot(so2$V1, type = 'l', col = 'green',
     xlab = 'Time', ylab = 'Atmospheric sulfur dioxide levels',
     main = 'Data versus time')
points(so2$V1, pch = 20, col = 'red')
```



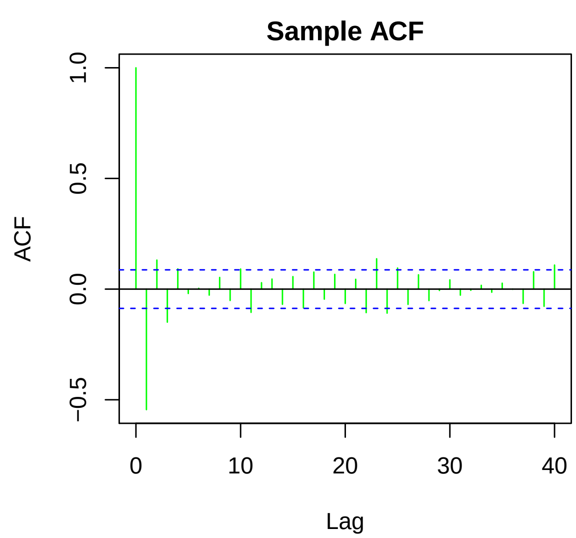

```
# polynomial fit
time = 1:508
polyfit = lm(so2$V1 ~ time + I(time^2))
coef(polyfit)
```

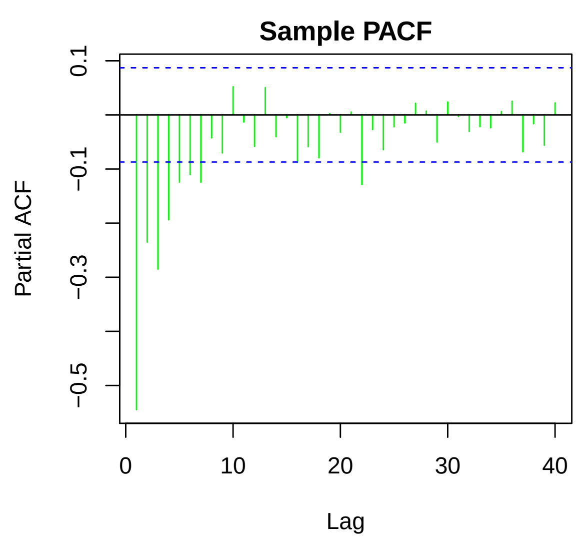
```
##   (Intercept)          time     I(time^2)
##  3.600651e+00 -1.761490e-03 -3.569062e-06
```

```
# differencing
diffso2 = diff(so2$V1)
# mean - centered
diffso2 = diffso2 - mean(diffso2)

# plot differenced data
plot(x = 2:508, y = diffso2,
     type = 'l', col = 'green',
     xlab = 'Time', ylab = expression(paste(nabla,X[t])),
     main = 'Differenced data versus time')
points(x = 2:508, y = diffso2,
       pch = 20, col = 'red')
```




```
# Sample ACF and Sample PACF
par(mfrow = c(1,2))
sampleacf = stats::acf(diffso2, lag.max = 40, plot = F)
plot(sampleacf, main = "", col = 'green')
title("Sample ACF",line=0.5)

samplepacf = stats::pacf(diffso2, lag.max = 40, plot = F)
plot(samplepacf, main = "", col = 'green')
title("Sample PACF", line=0.5)
```






```
# Choose p and q
p = 0:4          # adjust different p values
q = 0:4          # adjust differnt q values
# container
```

```r
order = matrix(NA, nrow = length(p)*length(q), ncol = 1)
parms = expand.grid(p,q)
row.names(order) = paste('ARMA(', parms[,2], ',', parms[,1],')')
colnames(order) = 'AIC'

# for loop
counter = 1
for (i in p){
  for (j in q){
    # fit model
    thisfit = stats::arima(diffso2, order = c(i,0,j), method = 'ML',
                           optim.control = list(maxit = 1000))
    # coef
    arcoef = thisfit$coef[0:i]
    macoef = thisfit$coef[i+1:j]

    # plot
    x11()
    par(mfrow = c(1,3))

    # acf
    plot(sampleacf, main = "", col = 'green', lwd = 2)
    lines(x = 0:40, ARMAacf(ar = arcoef,
                            ma = macoef,
                            lag.max = 40),
          type = 'h', col = 'red')
    title(paste("Sample ACF Model ACF ARMA(",i ,",", j, ")"),
          line=0.5)

    # pacf
    plot(samplepacf, main = "", col = 'green', lwd = 2)
    lines(ARMAacf(ar = arcoef,
                  ma = macoef,
                  lag.max = 40, pacf = TRUE),
          type = 'h', col = 'red')
    title(paste("Sample PACF Model ACF ARMA(",i ,",", j, ")"),
          line=0.5)

    # residual
    fitresid = rstandard(thisfit)
    plot(fitresid, type = 'l', col = 'blue',
         xlab = 'time', ylab = 'Standardized Residuals',
         main = '')
    title(paste("Standardized model residuals versus time ARMA(",i ,",", j, ")"))
    points(fitresid, pch = 20, col = 'red')

    # aic
    order[counter,1] = thisfit$aic
    counter = counter + 1
  }
}
```

```r
# AIC values for choosed 4 models
df = as.data.frame(order[c(5,18,24,23),])
```

```r
colnames(df) = 'AIC'
df
```

```
##                      AIC
## ARMA( 0 , 4 ) 1321.771
## ARMA( 3 , 2 ) 1323.238
## ARMA( 4 , 3 ) 1309.032
## ARMA( 4 , 2 ) 1307.177
```

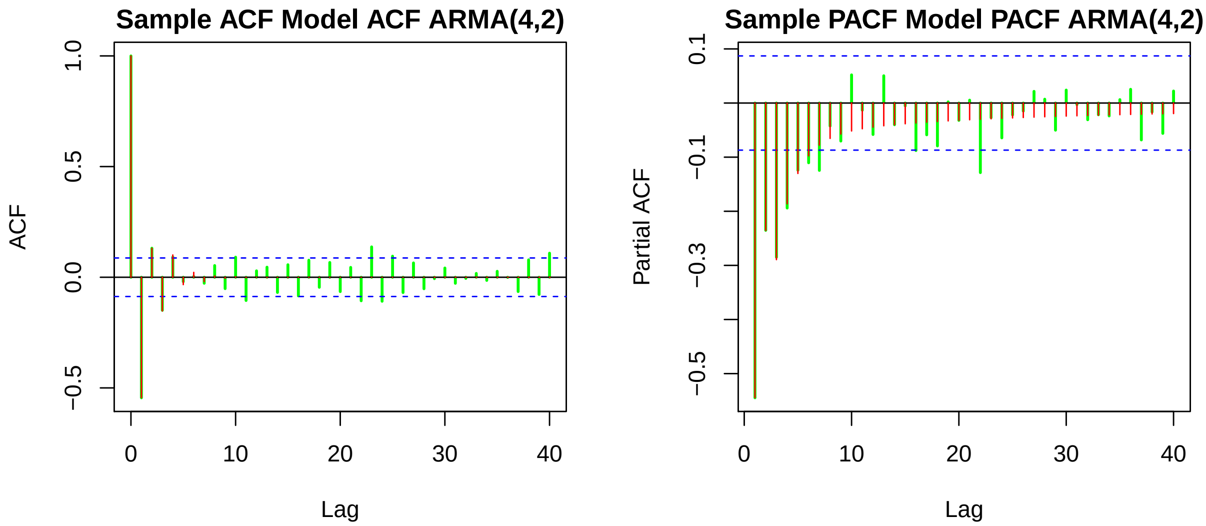
```r
# ARMA(4,2) model
arma42fit = stats::arima(diffso2, order = c(4,0,2),
                         method = 'ML',
                         optim.control = list(maxit = 1000))
# coef
arma42fit$coef
```

```
##           ar1           ar2           ar3           ar4           ma1
##  6.953413e-01  1.130086e-01 -1.096243e-01  1.319973e-01 -1.593950e+00
##           ma2     intercept
##  5.939507e-01  6.582580e-06
```

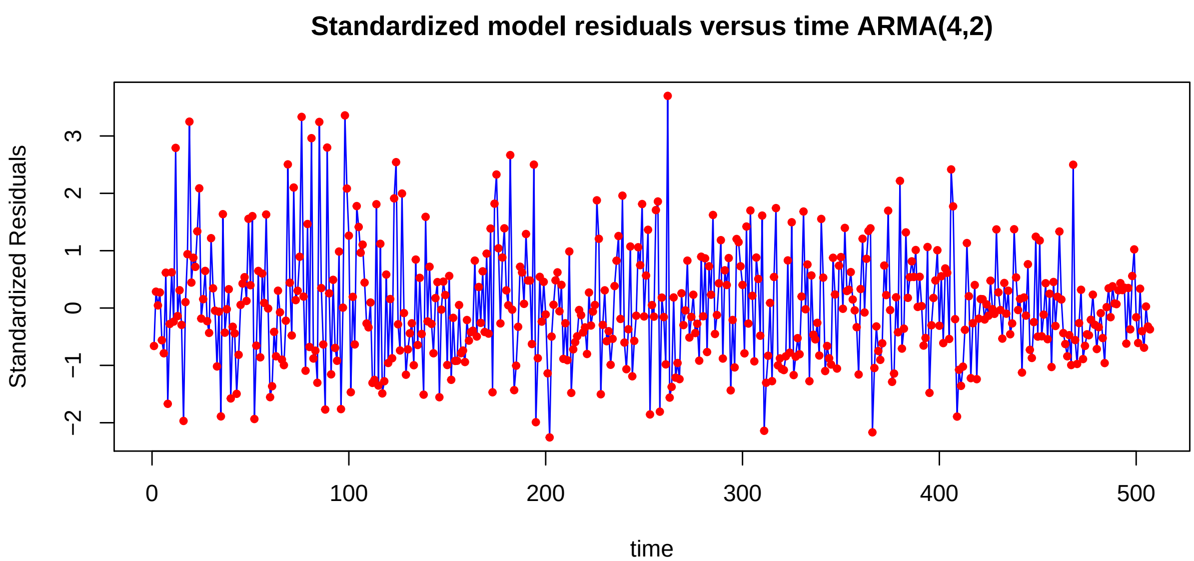
```r
# AIC
arma42fit$aic
```

```
## [1] 1307.177
```

```r
# ARMA(4,2) sample acf/pacf and model acf/pacf
par(mfrow = c(1,2))
plot(sampleacf, main = "", col = 'green', lwd = 2)
lines(x = 0:40, ARMAacf(ar = c(6.953413e-01,1.130086e-01,-1.096243e-01,1.319973e-01),
                        ma = c(-1.593950e+00,5.939507e-01), lag.max = 40),
      type = 'h', col = 'red')
title("Sample ACF Model ACF ARMA(4,2)",line=0.5)

plot(samplepacf, main = "", col = 'green', lwd = 2)
lines(ARMAacf(ar = c(6.953413e-01,1.130086e-01,-1.096243e-01,1.319973e-01),
              ma = c(-1.593950e+00,5.939507e-01), lag.max = 40, pacf = TRUE),
      type = 'h', col = 'red')
title("Sample PACF Model PACF ARMA(4,2)", line=0.5)
```

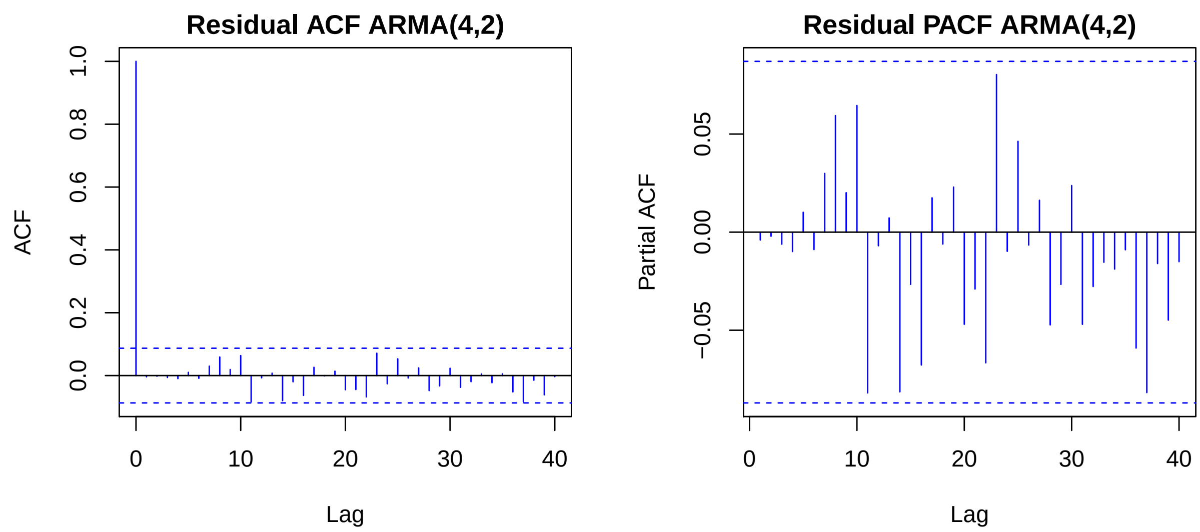



```
# ARMA(4,2) standardized residuals
par(mfrow = c(1,1))
arma42resid = rstandard(arma42fit)
plot(arma42resid, type = 'l', col = 'blue',
    xlab = 'time', ylab = 'Standardized Residuals',
    main = '')
title("Standardized model residuals versus time ARMA(4,2)")
points(arma42resid, pch = 20, col = 'red')
```



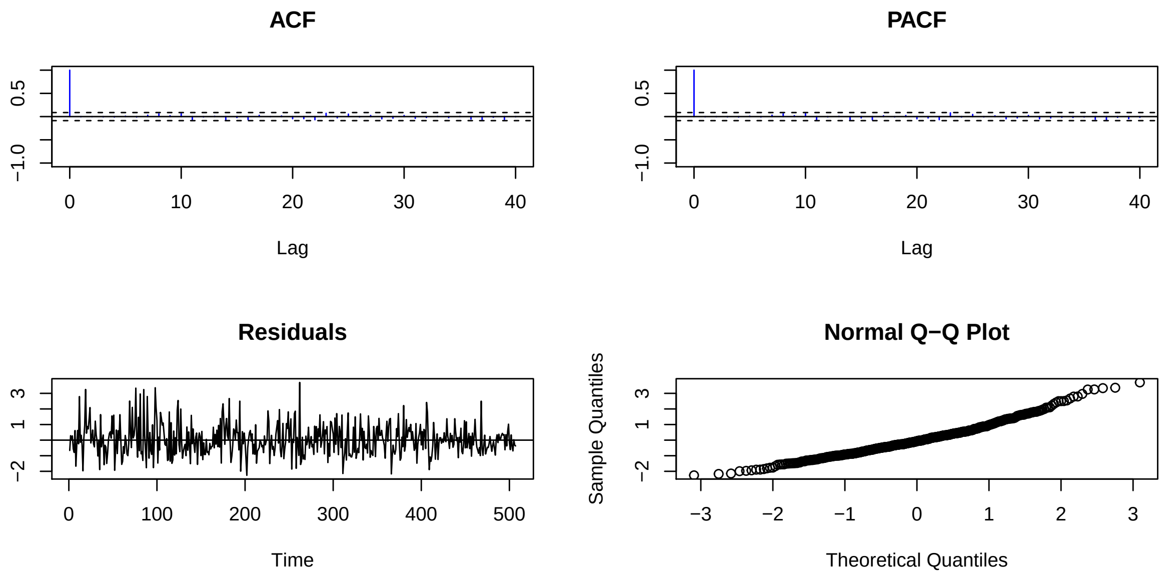

```
# ARMA(4,2) Residual acf/pacf
par(mfrow = c(1,2))
arma42residacf = stats::acf(arma42resid, lag.max = 40, plot = F)
plot(arma42residacf, main = "", col = 'blue')
title("Residual ACF ARMA(4,2)",line=0.5)
```

```r
arma42residpacf = stats::pacf(arma42resid, lag.max = 40, plot = F)
plot(arma42residpacf, main = "", col = 'blue')
title("Residual PACF ARMA(4,2)", line=0.5)
```

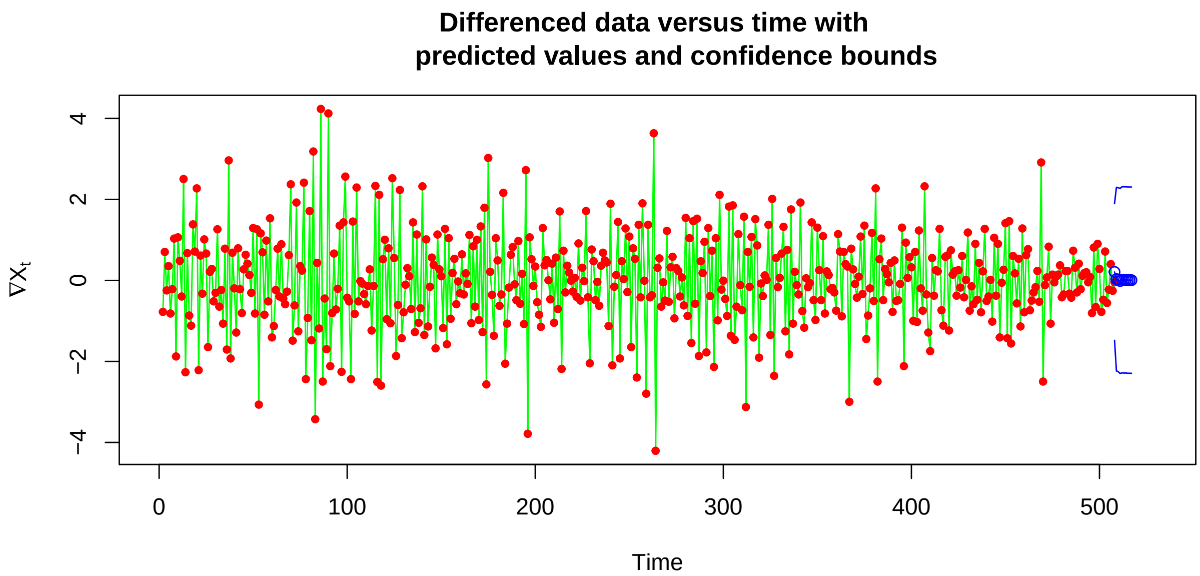



```r
# ARMA(4,2) Statistic tests
test(arma42resid)
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     15.65    0.7384
## McLeod-Li Q                Q ~ chisq(20)     53.26     1e-04 *
## Turning points T  (T-336.7)/9.5 ~ N(0,1)       345    0.3792
## Diff signs S        (S-253)/6.5 ~ N(0,1)       260     0.282
## Rank P       (P-64135.5)/1905.5 ~ N(0,1)     63045    0.5671
```




```
# plots with predicted value and Confidence bounds
par(mfrow = c(1,1))
pred = forecast::forecast(arma42fit)
plot(x = 2:508, y = diffso2,
     type = 'l', col = 'green',
     xlim = c(0,530),
     xlab = 'Time', ylab = expression(paste(nabla,X[t])),
     main = 'Differenced data versus time with
     predicted values and confidence bounds')
points(x = 2:508, y = diffso2,
       pch = 20, col = 'red')
lines(pred$mean, col = 'blue')
points(pred$mean, col = 'blue')
lines(pred$upper[,2], col = 'blue')
lines(pred$lower[,2], col = 'blue')
```



Differenced data versus time with
predicted values and confidence bounds
∇X_t
4
2
0
−2
−4
0
100
200
300
400
500
Time