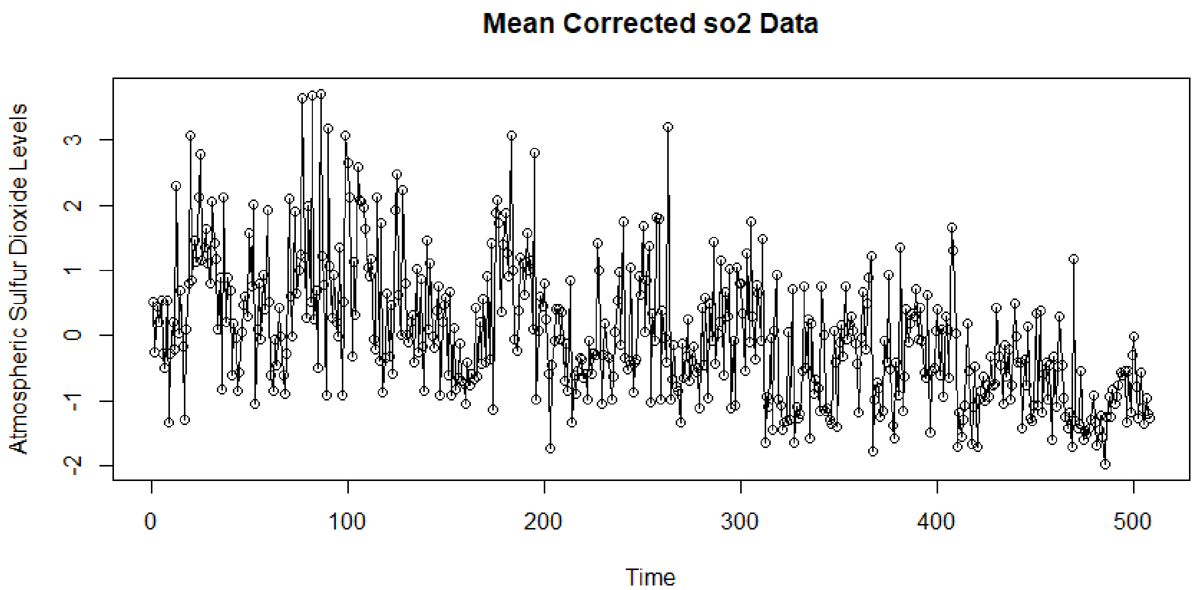
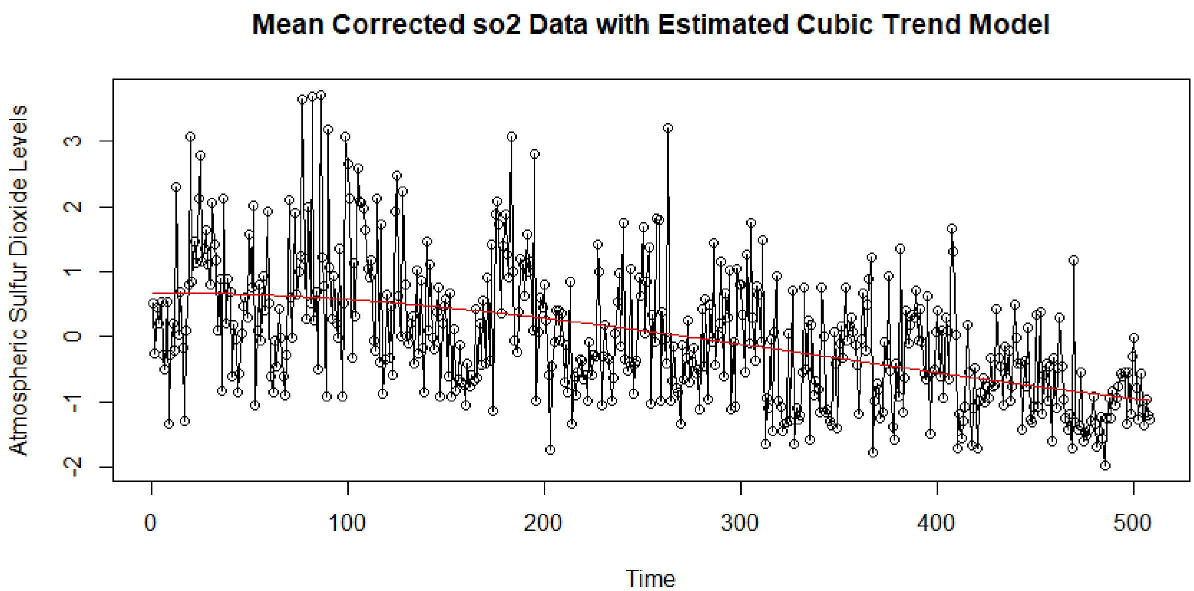
Determination of d:

The data I am using is so2 data which is data from monitoring atmospheric sulfur dioxide levels.



Doesn't look stationary because variance seems to be decreasing over time and looks like there is a downward linear trend.

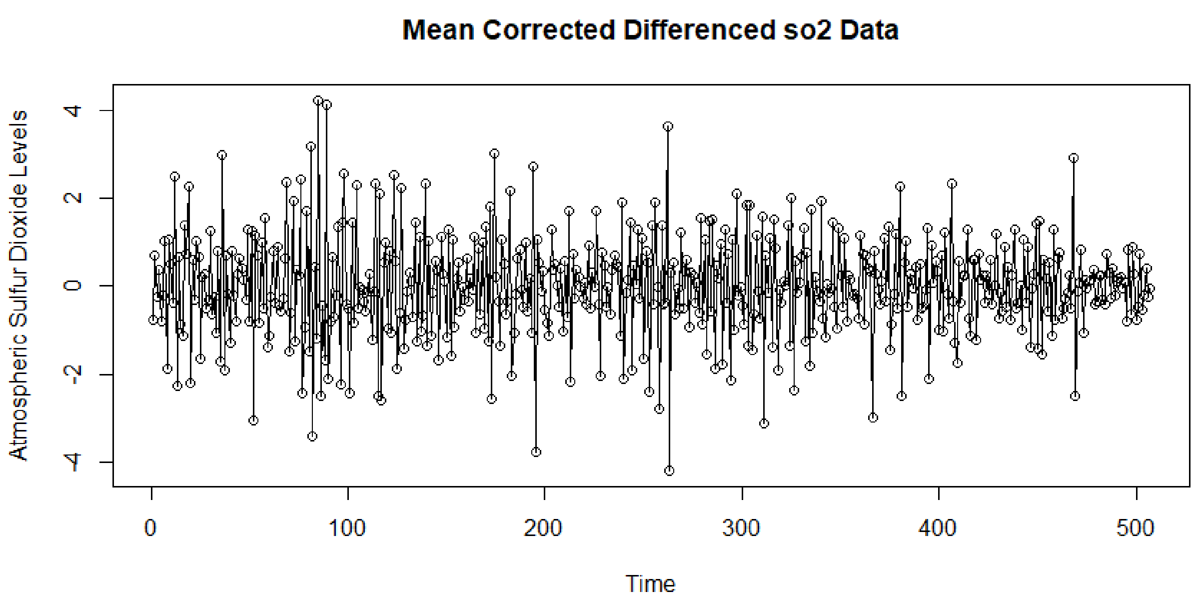


Coefficients:

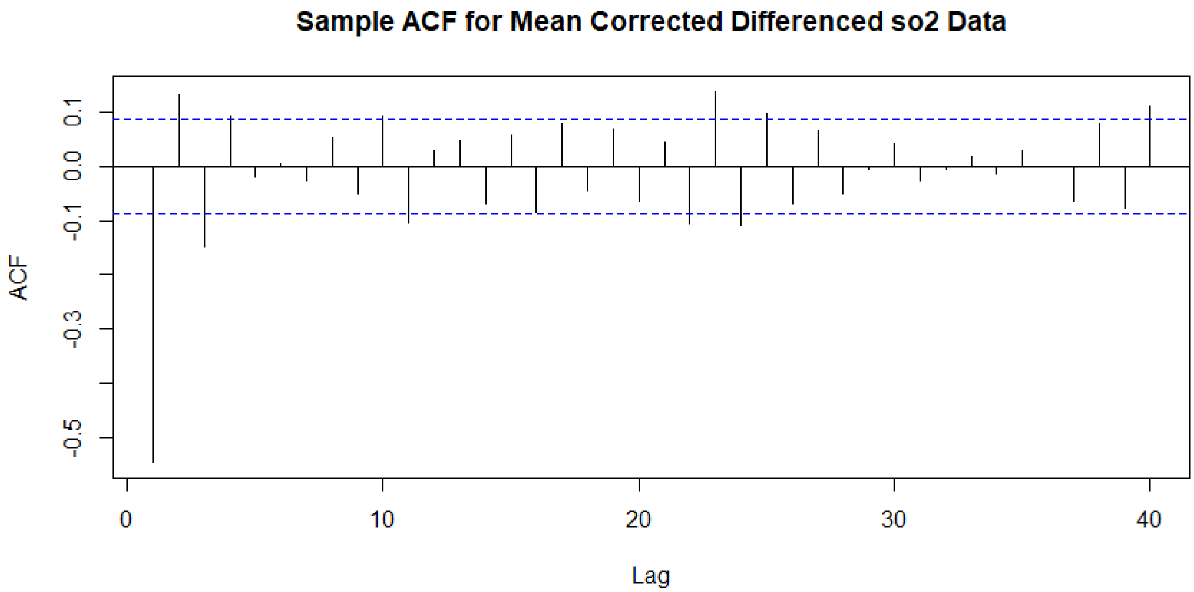
(Intercept)	time	time2	time3
6.717e-01	2.206e-04	-1.329e-05	1.274e-08

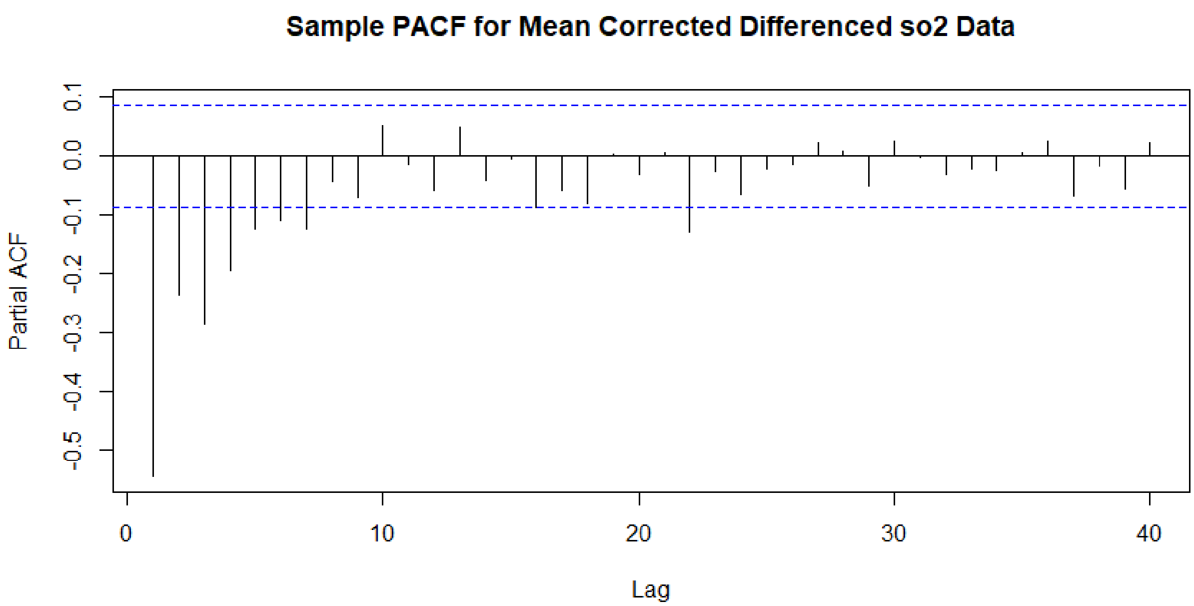
d = 1

The reason I chose this d is because of the relative sizes of the coefficients. The coefficients in front of the quadratic and cubic terms are very small in comparison to the coefficient in front of the linear term. Therefore, this is a good indicator that the trend is linear which means we only need to difference the data once to remove a linear trend.



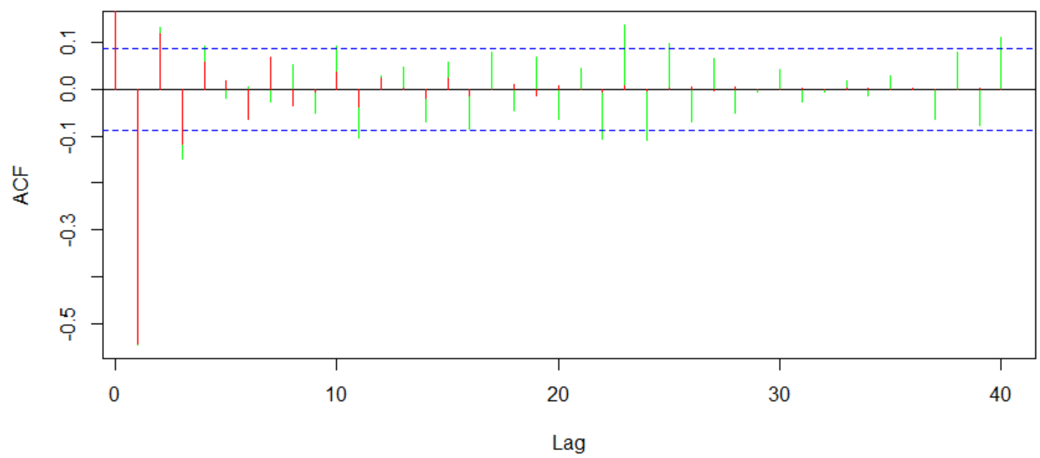
Determination of p and q for the mean-corrected differenced data:



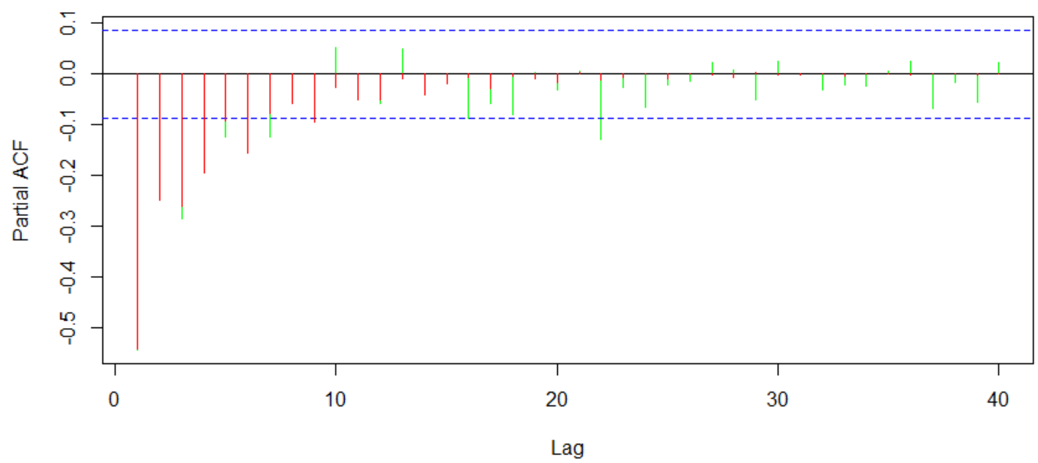


The sample ACF suggests cut off at lag 4 although could even say at lag 3 since at lag 4 it's barely over the bound. The sample PACF sort of looks like exponential decay. These suggest using an MA(4) or MA(3) model so we could start with an ARMA(p,4) or ARMA(p,3).

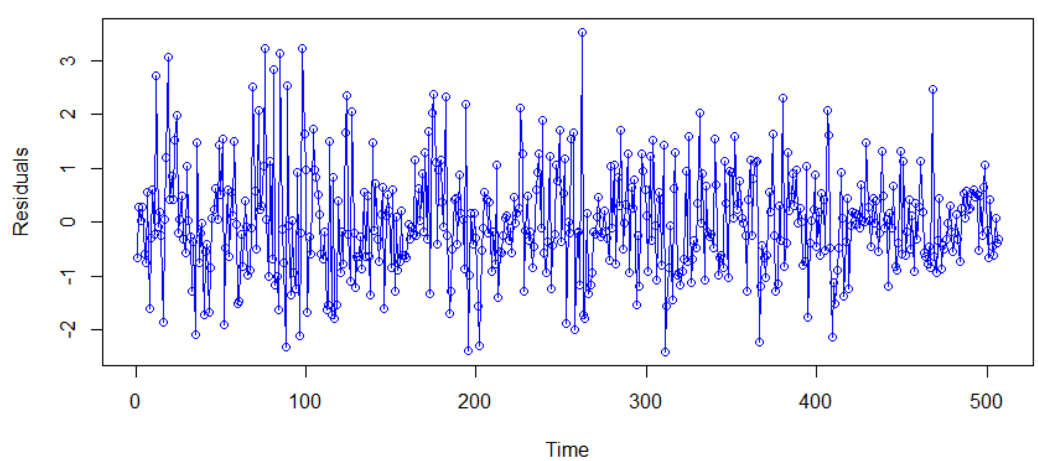
Sample ACF and ARMA(2,3) Model ACF



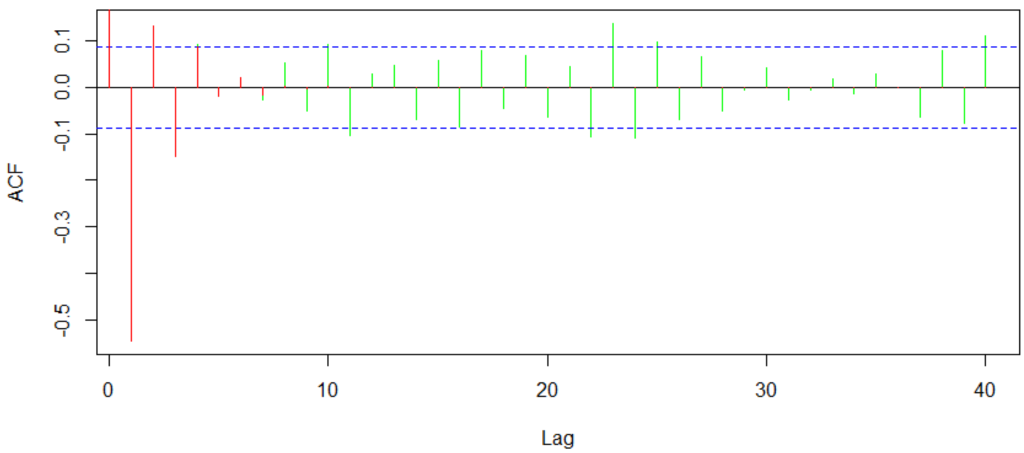
Sample PACF and ARMA(2,3) Model PACF



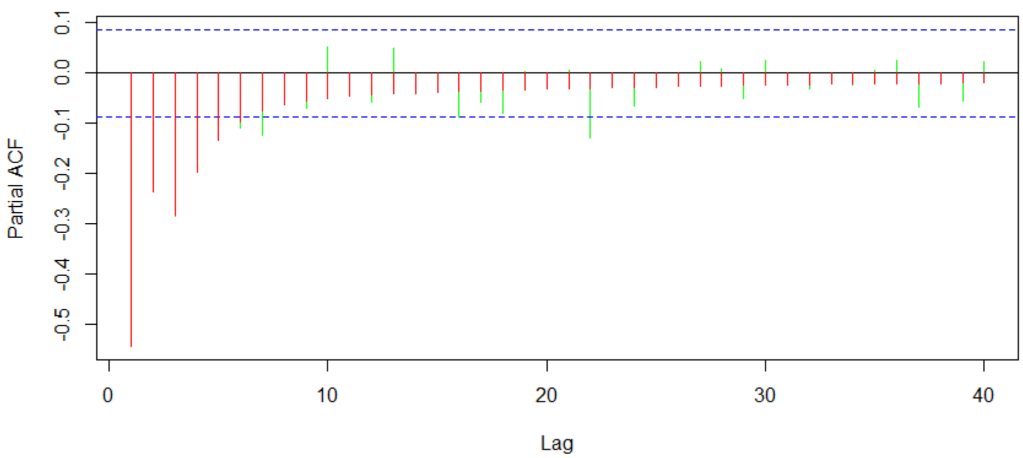
Standardized Residuals for ARMA(2,3)



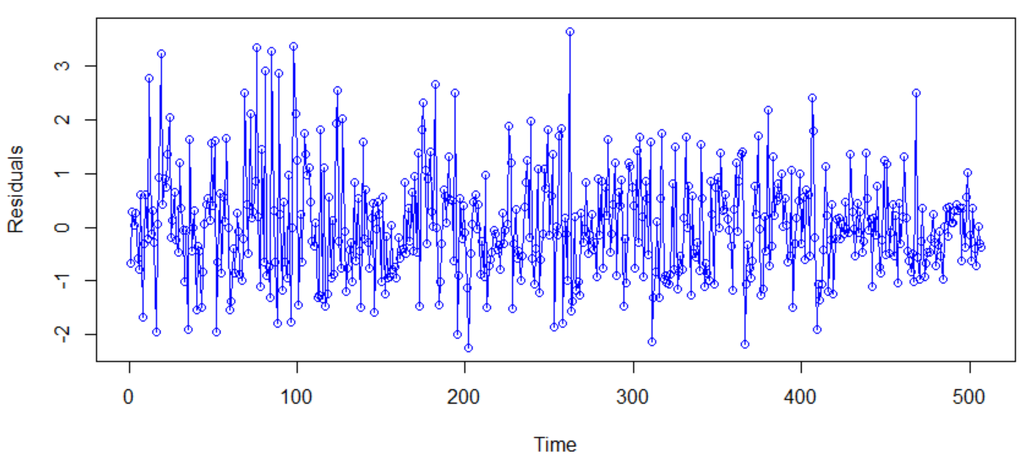
Sample ACF and ARMA(4,3) Model ACF

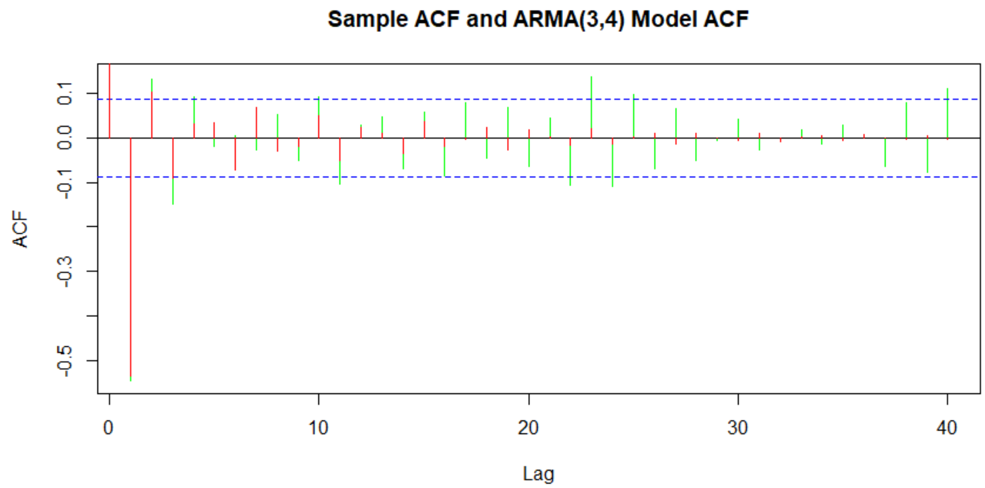


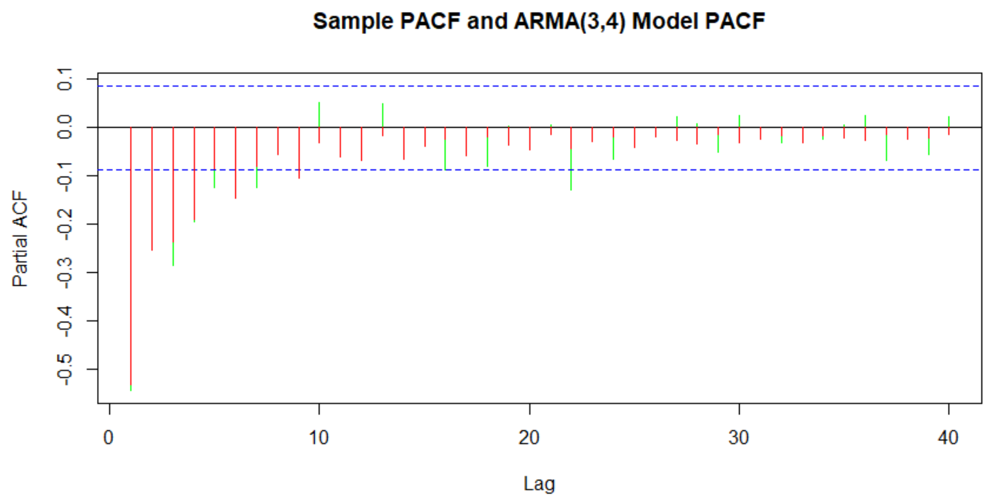
Sample PACF and ARMA(4,3) Model PACF

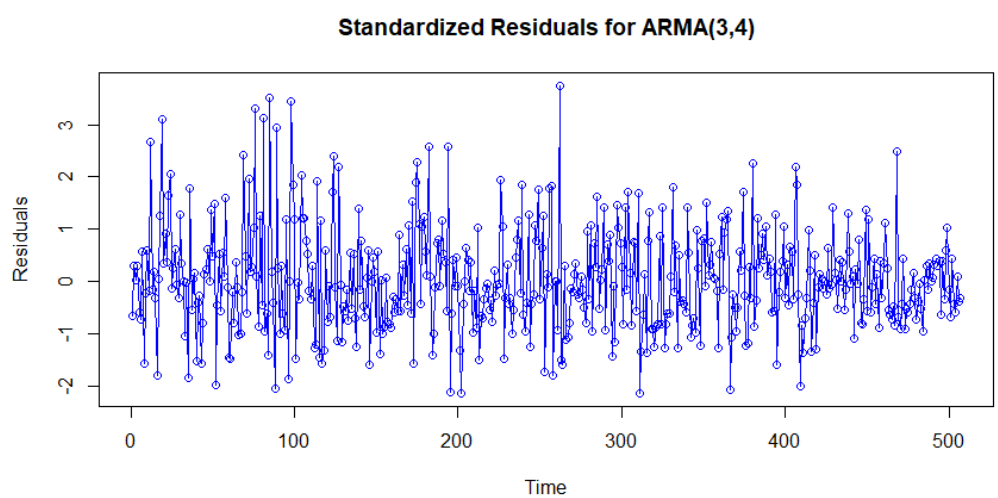


Standardized Residuals for ARMA(4,3)

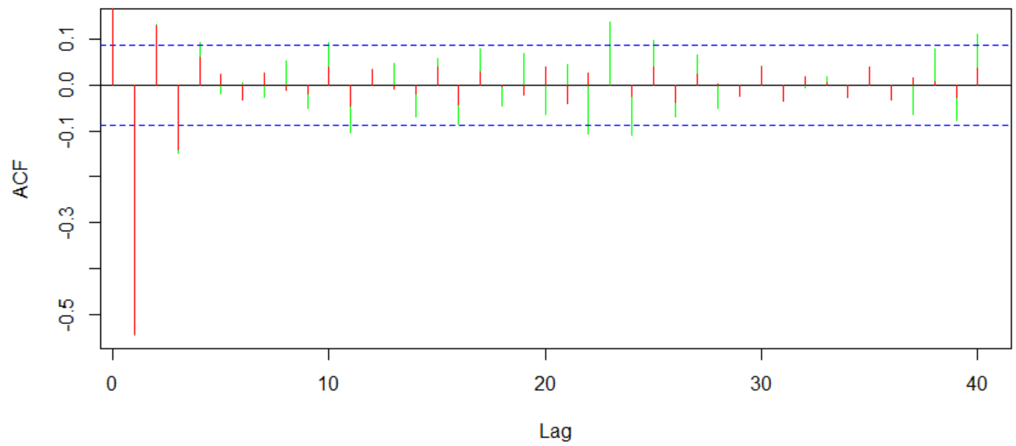



Sample ACF and ARMA(3,4) Model ACF
ACF
0.1
0.0
-0.1
-0.3
-0.5
0
10
20
30
40
Lag

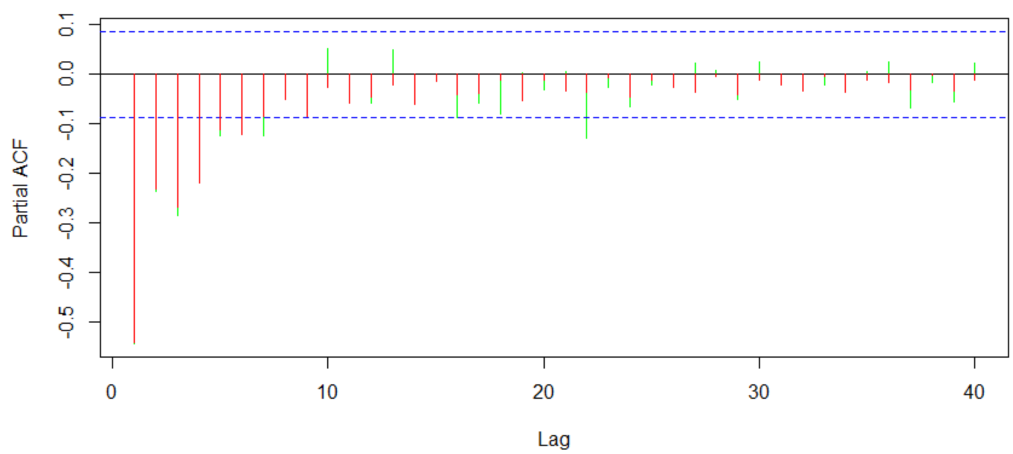

Sample PACF and ARMA(3,4) Model PACF
Partial ACF
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
0
10
20
30
40
Lag


Standardized Residuals for ARMA(3,4)
Residuals
3
2
1
0
-1
-2
0
100
200
300
400
500
Time

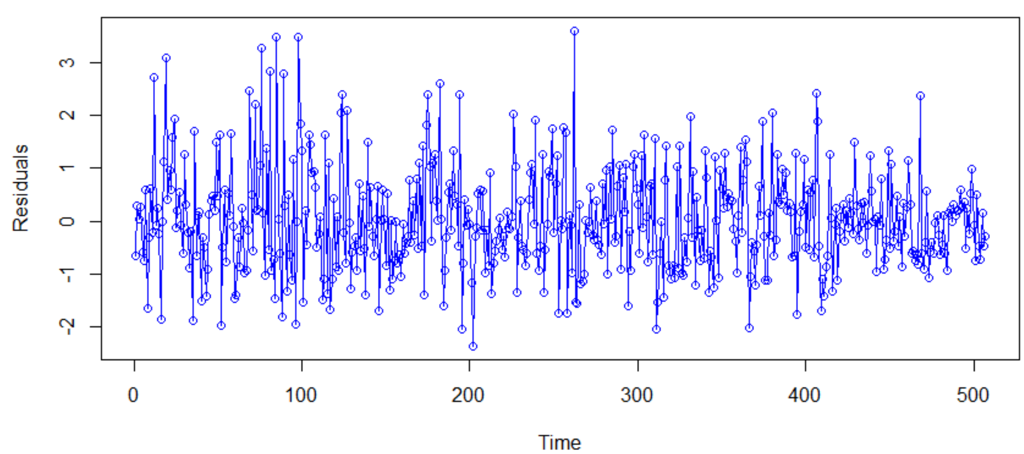
Sample ACF and ARMA(6,4) Model ACF



Sample PACF and ARMA(6,4) Model PACF



Standardized Residuals for ARMA(6,4)



ARMA(2,3): aic = 1319.47
ARMA(4,3): aic = 1307.03
ARMA(3,4): aic = 1308.13
ARMA(6,4): aic = 1304.13

p = 4 and q = 3

By analyzing the plots of the sample ACF/PACF I was able to get a rough idea for a starting value for q. Since the sample ACF suggested cut off at lag 4 or lag 3 and the sample PACF looked like exponential decay. These indicated using an MA(4) or MA(3) model so I could start with an ARMA(p,4) or ARMA(p,3). I first checked the ARMA(p,4) for varying p keeping note of the small aic values and tried an ARMA(3,4) and ARMA(6,4). The ARMA(3,4) was not a good fit as the model ACF/PACF did not have good agreement with the leading significant values of the sample ACF/PACF. The ARMA(6,4) agreements were a lot better however at the cost of a more complex model. I then began checking ARMA(p,3) which looked better than when I was checking ARMA(p,4) and again keeping note of the small aic values. I tried an ARMA(2,3) and the agreements of the model ACF/PACF were not great. Finally, I fit an ARMA(4,3) model which had the smallest aic when I was checking ARMA(p,3) models and its model ACF/PACF had the best agreements with the leading significant values of the sample ACF/PACF than the others. This led me to find my value for p, hence coming to an ARMA(4,3) as my model of choice.

Using MLE to fit the ARMA(p,q) model for the chosen p and q and analyzing the model:

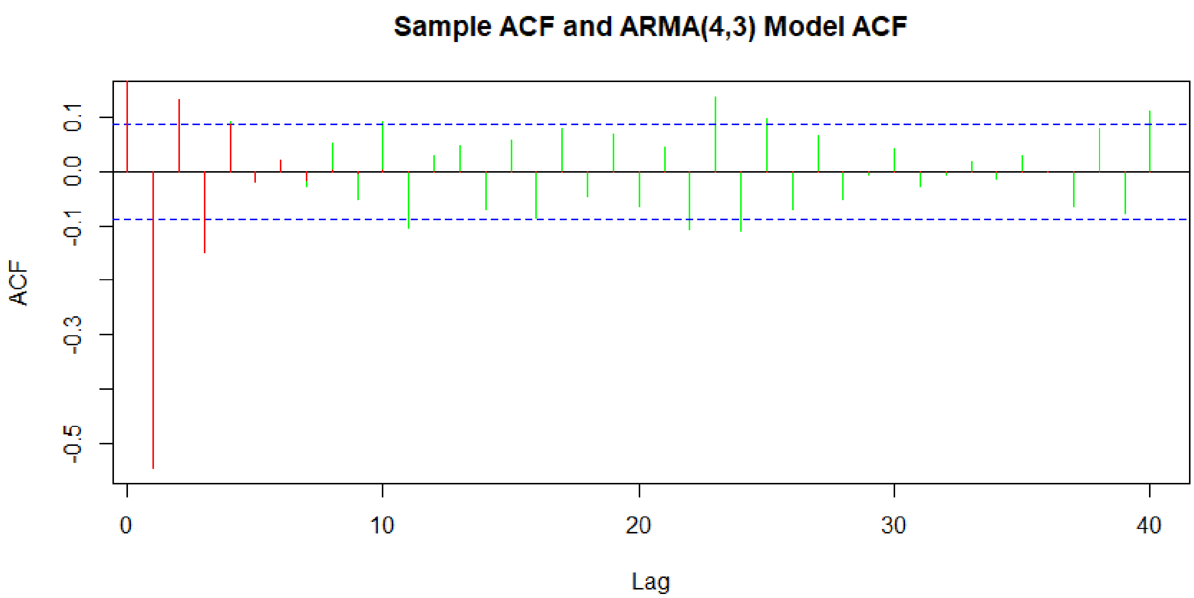
p = 4
d = 1
q = 3

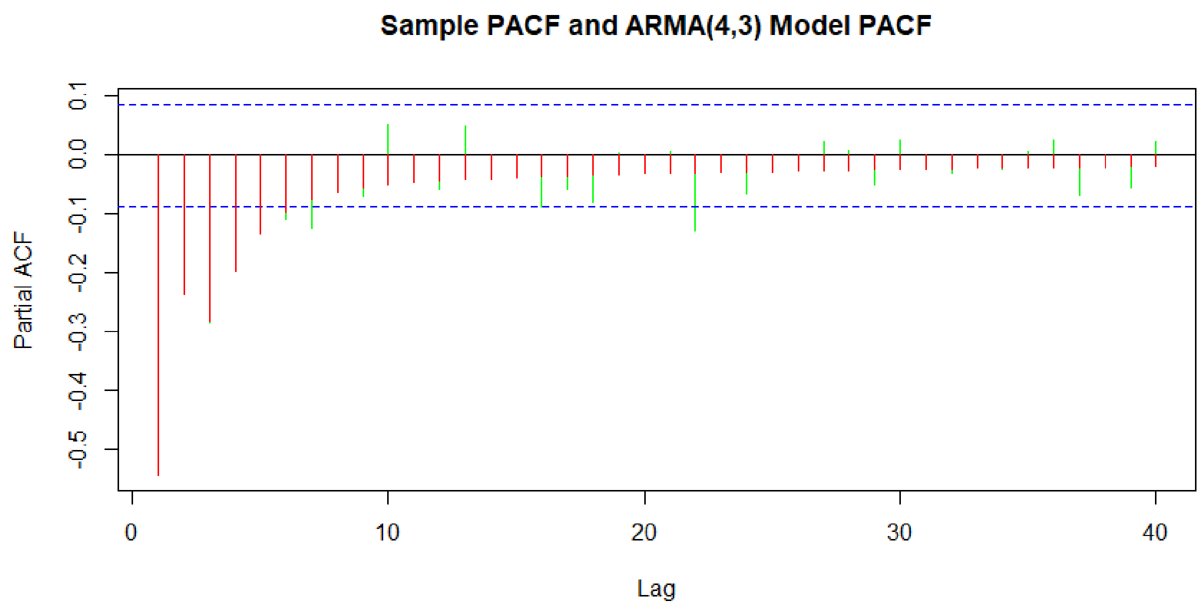
Coefficients:

	ar1	ar2	ar3	ar4	ma1	ma2	ma3	intercept
	0.8421	-0.0164	-0.1255	0.1434	-1.7430	0.8604	-0.1175	-0.0035
s.e.	0.5326	0.4471	0.0792	0.0564	0.5398	0.9346	0.3997	0.0006

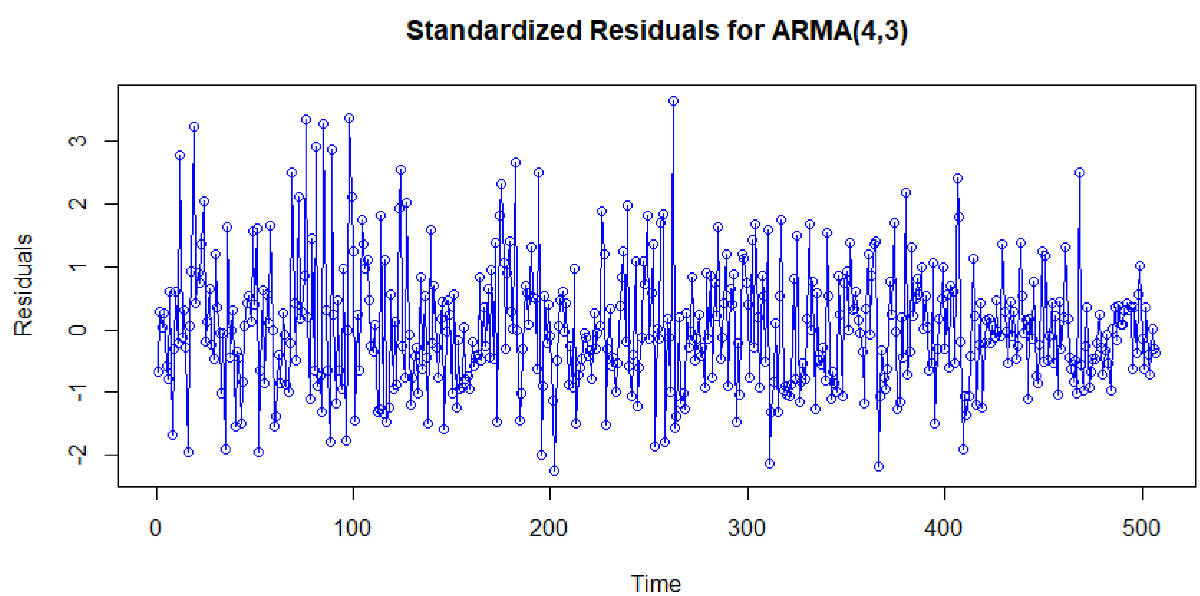
sigma^2 estimated as 0.7401

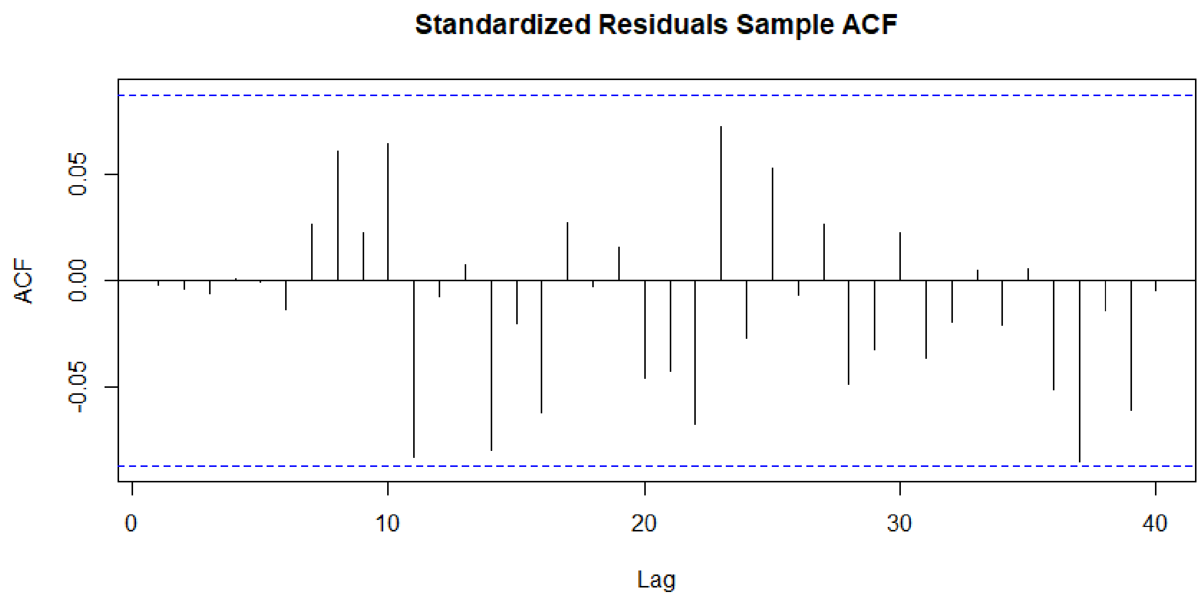
aic = 1307.03

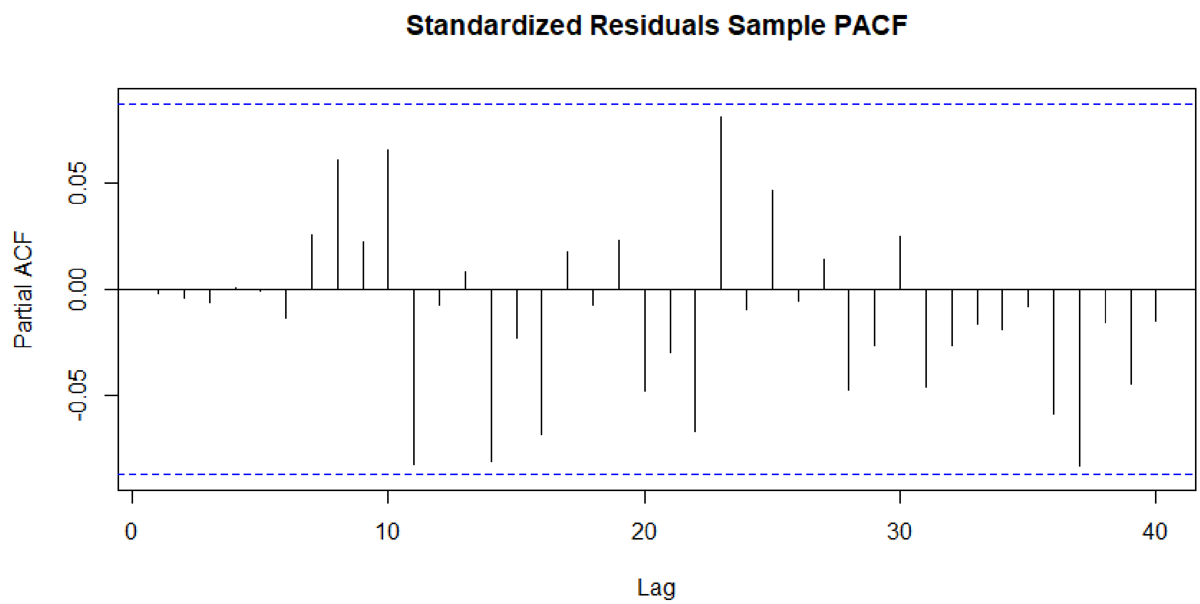




The model fit is pretty good, as we see good agreement in the leading significant values of the model ACF and model PACF. The model ACF is a little off at lag = 4 and the model PACF at lag = 3 and lag = 6 but quite a bit off at lag = 7, however I still think this is okay since it did a good job capturing the significant values before. Overall, this is a solid fit.







The standardized residuals appear stationary and look like noise which is good. It means our model captured most of the correlation structure in our data and indicates it's a good fit. The plot of the standardized residuals sample ACF/PACF have no significant values, therefore they both support the iid hypothesis.

The Ljung-Box statistic QLB = 15.52 with p-value 0.746 > 0.05. This does not provide sufficient evidence to reject the iid hypothesis.
The Mcleod-Li statistic QML = 54.42 with p-value 1e-04 < 0.05. This provides sufficient evidence to reject the iid hypothesis.

My conclusion is that the residuals are samples from an iid times series. The reason I conclude this is because the plots of the standardized residuals sample ACF/PACF both support the iid hypothesis and the Ljung-Box statistic does not support rejection of the iid hypothesis. Even though the Mcleod-Li statistic does support rejection of the iid hypothesis I don't believe this alone is enough to completely reject it considering the other test didn't and both the plots of the standardized residuals ACF/PACF had no significant values which support the iid hypothesis.

Overall, I believe our ARMA(4,3) model is a great fit. The model ACF/PACF had good agreements with majority of the leading significant values of the sample ACF/PACF. The residuals appear stationary and look like noise, as well as I concluded the residuals support the iid hypothesis. I came to this conclusion since there were no significant values in our residuals sample ACF/PACF and the Ljung-Box statistic did not support rejection of the hypothesis. After analyzing all these aspects, I believe our model fit is great.

Using the estimated model to make a forecast: