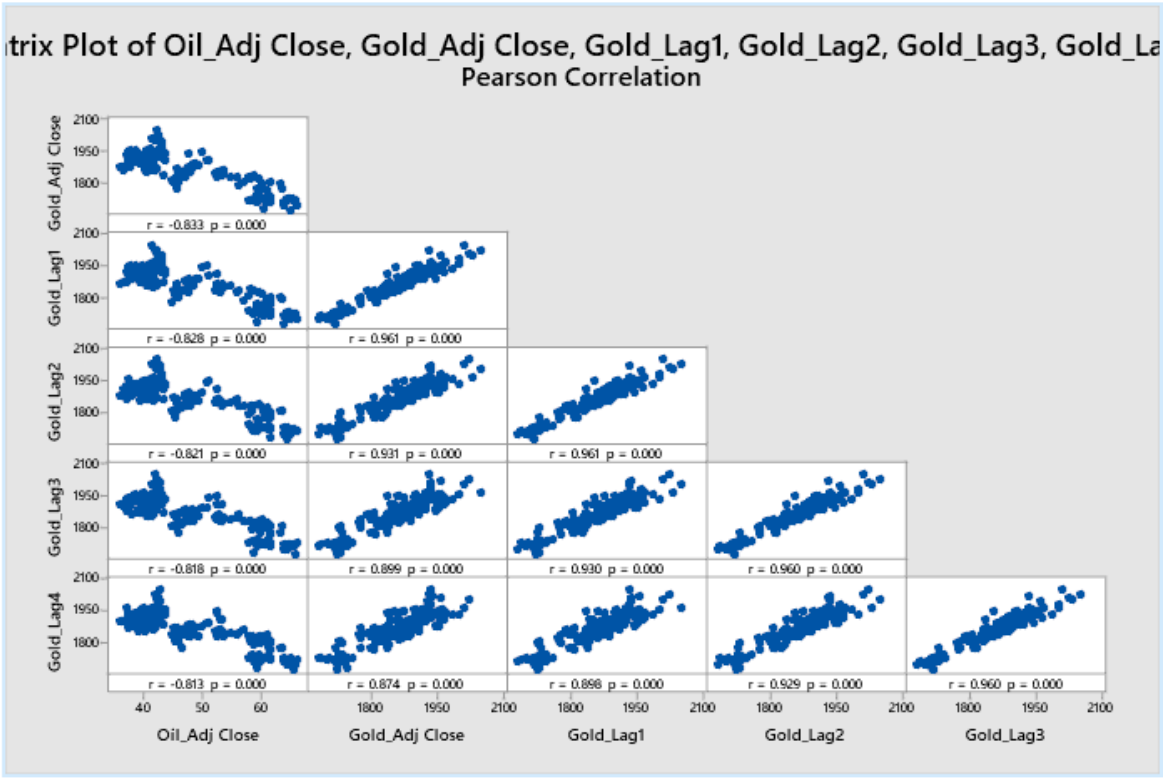
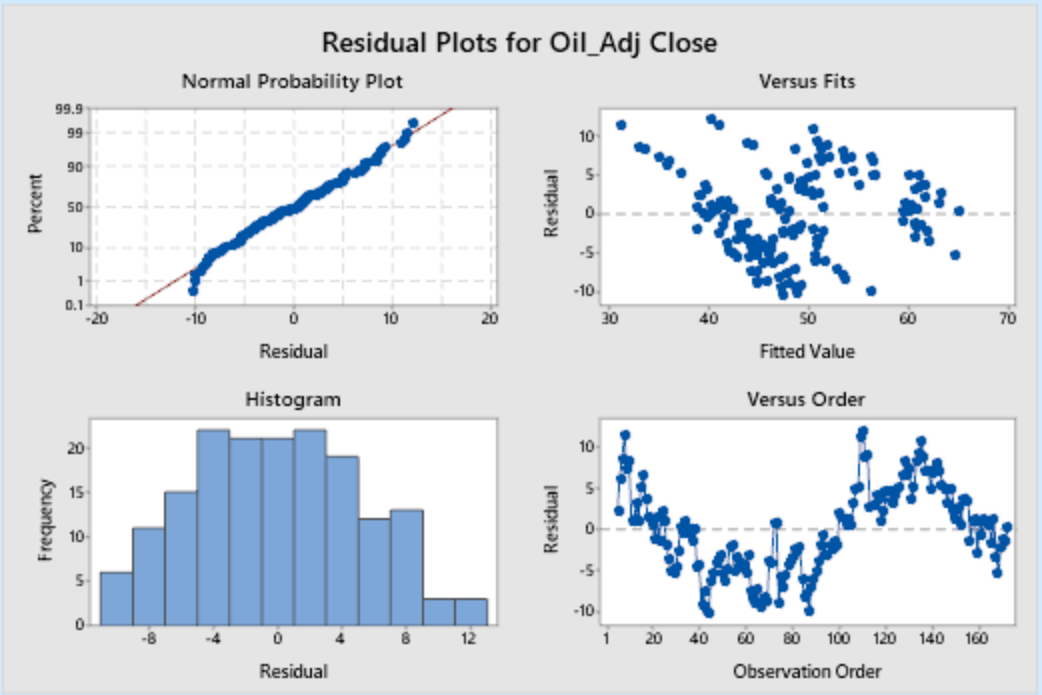
ARMA and ARIMA Modeling

A) Consider the gold price and its lag(1), lag(2), lag(3) and lag(4) variables, create a correlation matrix consisting of the oil price, the gold price and its lag variables. Comment and explain the relation between gold price and oil price.



It seems apparent that the correlation between gold price and oil price is a negative relationship (column 1). As oil price increases, gold price decreases - vice versa. Based on the correlation statistics (r), it seems like the non-lagged gold price correlates better negatively with oil price, because the magnitude of 'r' is larger.

B) To predict the oil price, we consider the gold price and its lag(4) variable in part a) as the predictors, suggested by the stepwise model process. Run the regression model. From the residual plots, is there any problem revealed in the model?



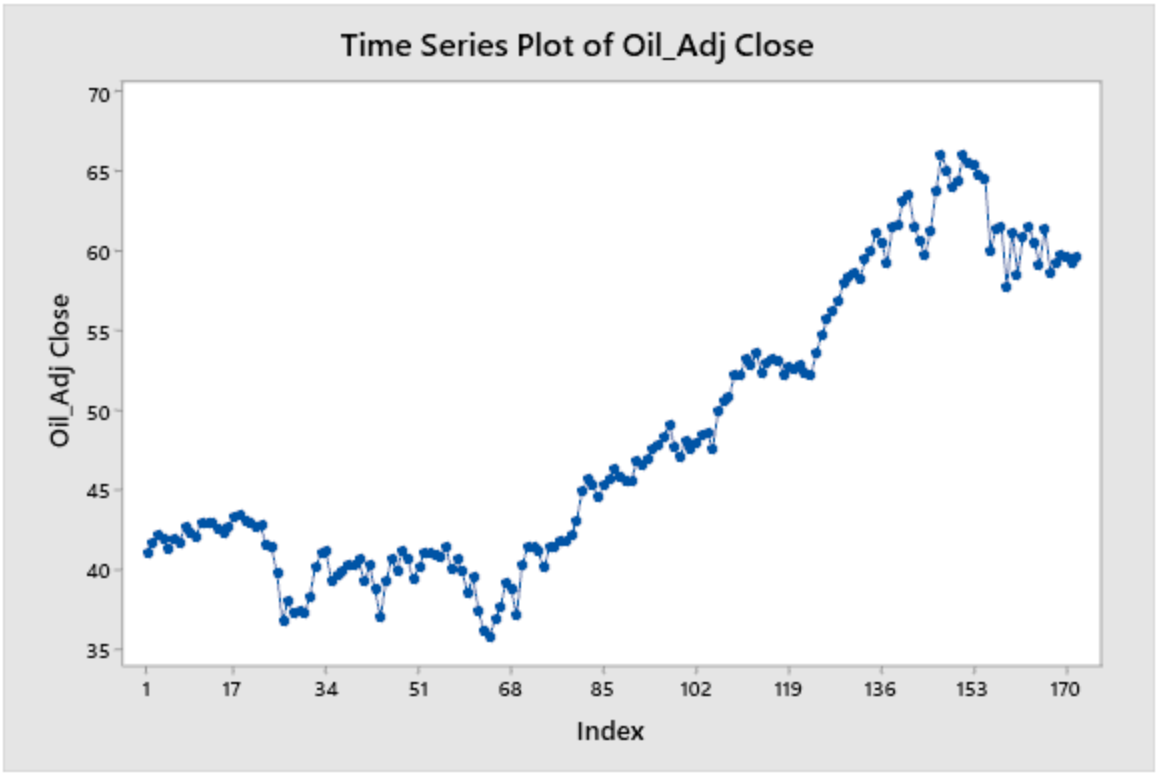
According to the Residual Plots:

Normal Probability Plot - shows that the model fits the data points almost perfectly.
Versus Fits - shows that the distribution is not random and that there is a clear pattern.
Histogram - shows abnormal behaving normal distribution.
Versus Order - shows that the data points are non-stationary and don't revert back to a mean.

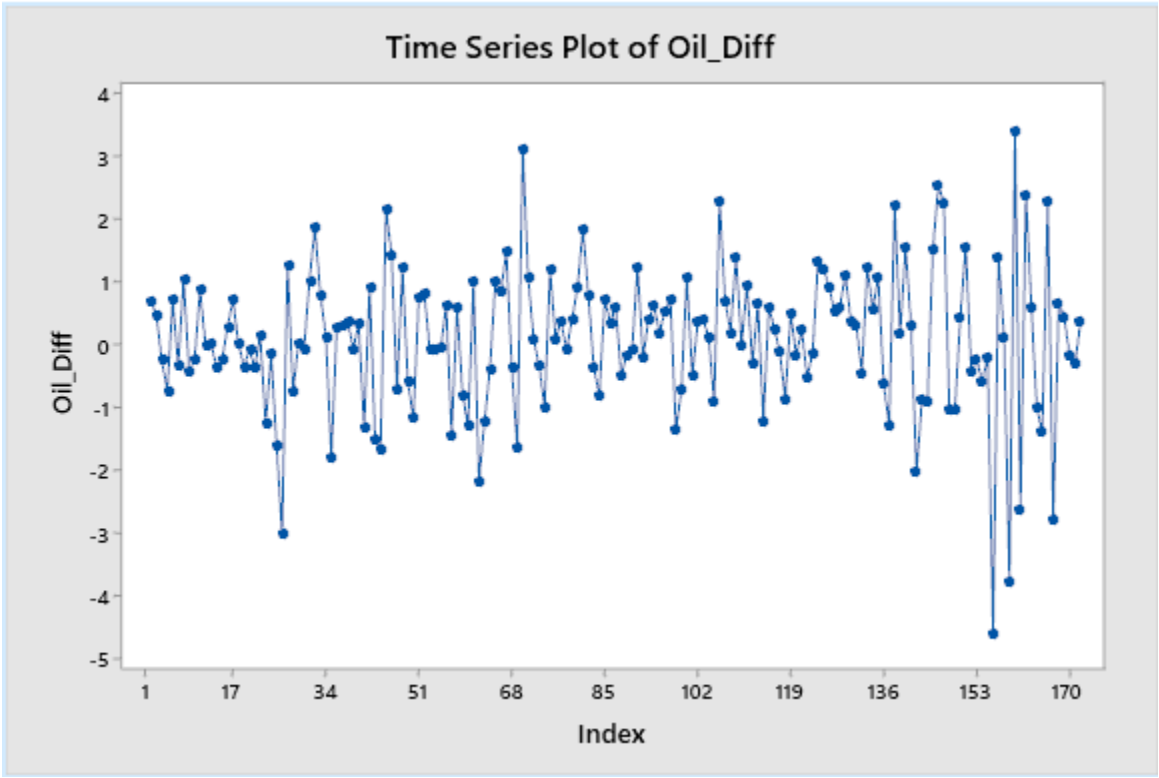
The model is not a good fit.

C) Next use the ARMA model to predict the oil price time series. Please decide to model either the original series or the 1st differences as a stationary time series. Explain briefly.

According to the two oil time series plots below, the original series is not a stationary time series, while the 1st difference series is a stationary time series. To use the ARMA model, it has to be done on a stationary time series with constant mean reversion. The 1st difference oil time series will be used.



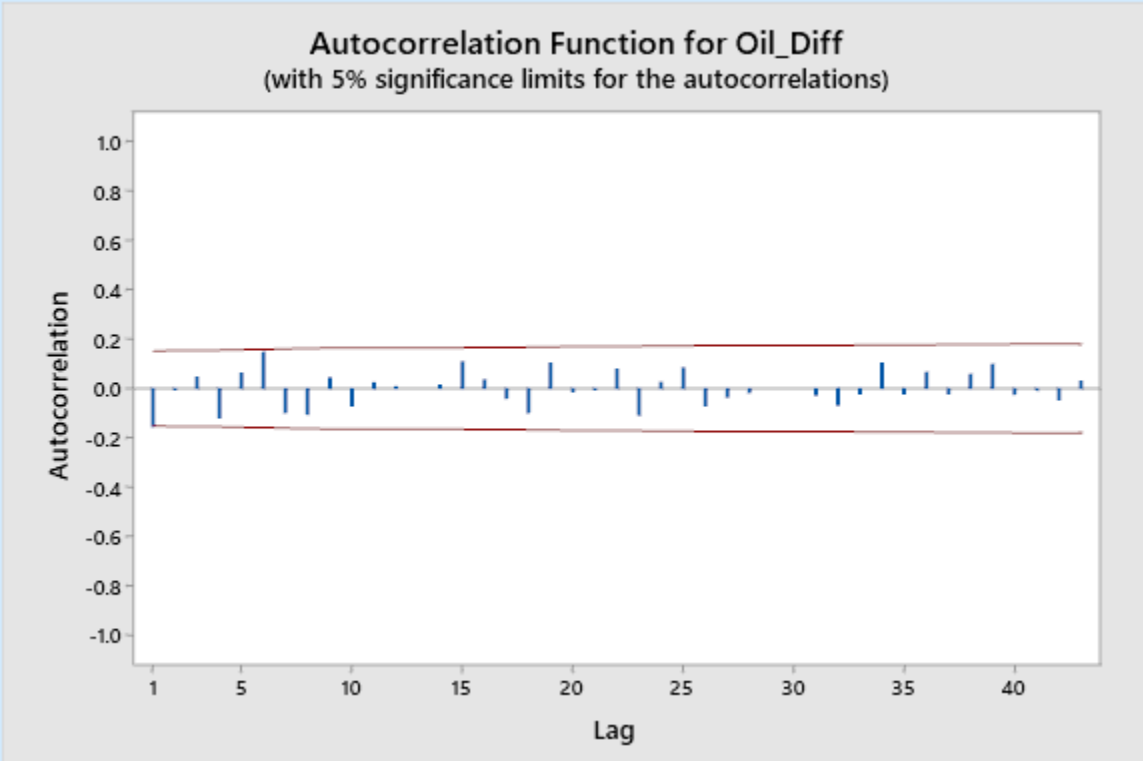
original time series of oil



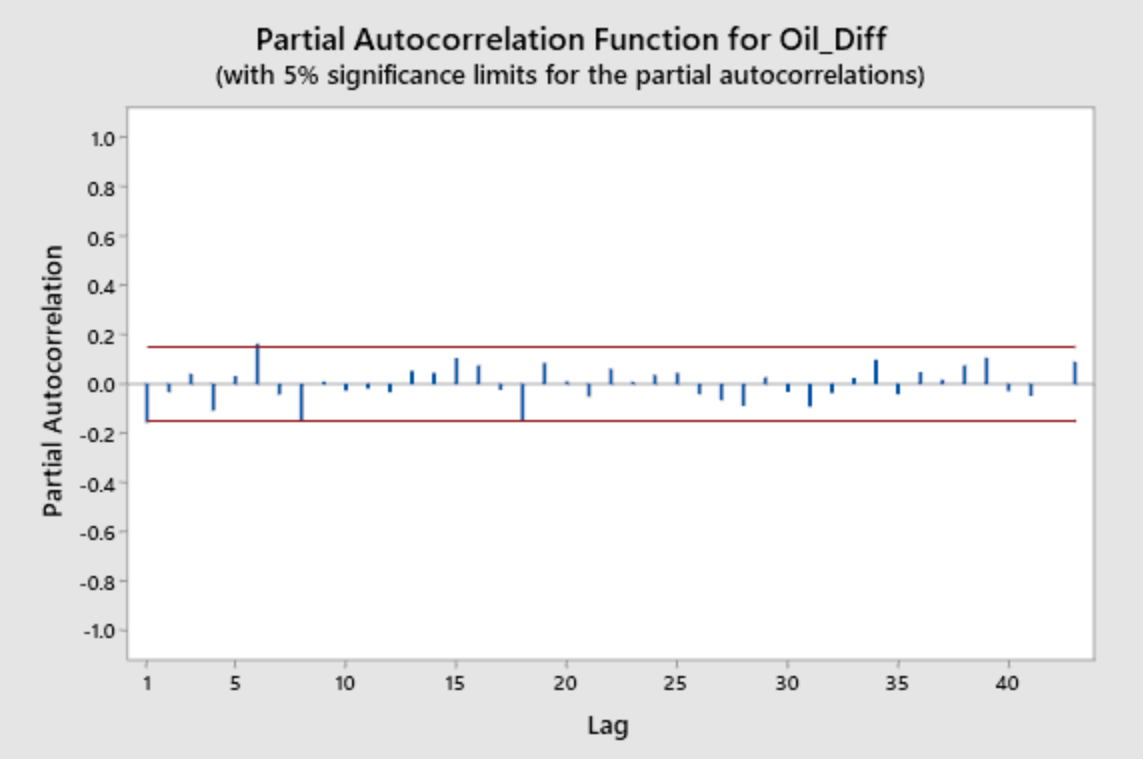
1st difference series of oil

D) Please use the SAC and SPAC of the series chosen in part c) to identify a few tentative ARMA models. Explain your choices briefly.

According to the Sample Autocorrelation and Sample Partial Autocorrelation graphs and their T-Statistics, there seems to be a significant spike at lag 4 (absolute T-Statistics > 2). Therefore, the tentative model used will be ARIMA (1,1,0) according to the SPAC, ARIMA(0,1,1) according to the SAC, and ARIMA (1,1,1) according to SPAC and SAC.



sample autocorrelation function



sample partial autocorrelation

Autocorrelation T-Statistics	*Partial Autocorrelation T-Statistics*

Autocorrelations

Lag	ACF	T	LBQ
1	-0.158197	-2.07	4.36
2	-0.009012	-0.12	4.37
3	0.048410	0.62	4.78
4	-0.121161	-1.54	7.38
5	0.065400	0.82	8.14

Partial Autocorrelations

Lag	PACF	T
1	-0.158197	-2.07
2	-0.034912	-0.46
3	0.042526	0.56
4	-0.109906	-1.44
5	0.031915	0.42

E) For models chosen in D), estimate the parameters and do the diagnostic checking. Also generate forecasts for the next 4 days after 12 April 2021.

In terms of the diagnostics, the ACF and PACF graphs lagged terms will be evaluated according to a threshold. In general, to determine the threshold the equation is ln(N). There are a total of 172 data points so the threshold would be ln(172) = 5.14749447681. To be sure, an arbitrary threshold value of 12 will also be checked.

ARIMA(1,1,0):

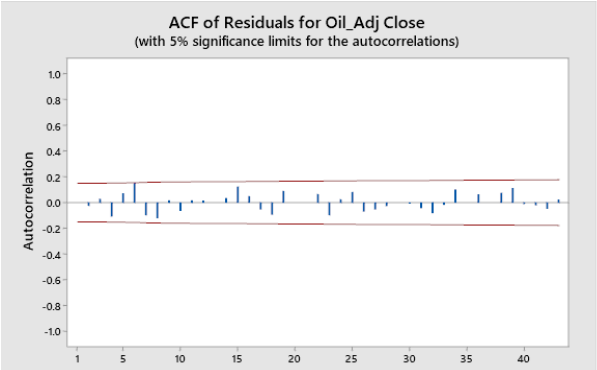
Forecasts from period 172

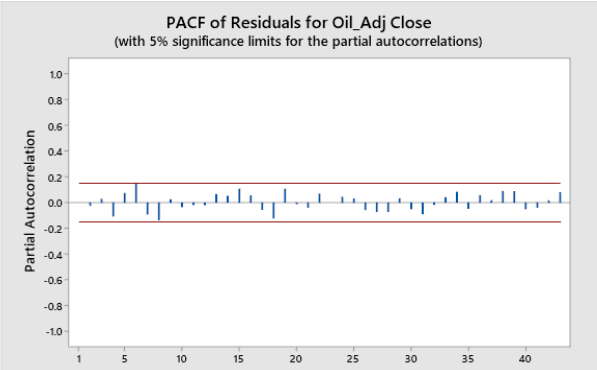
		95% Limits		
Period	Forecast	Lower	Upper	Actual
173	59.7656	57.5008	62.0304	
174	59.8810	56.9210	62.8411	
175	59.9886	56.4369	63.5402	
176	60.0974	56.0438	64.1509	

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.1585	0.0760	-2.09	0.038
Constant	0.1258	0.0883	1.42	0.156

In the ACF and PACF graphs the autocorrelation is not significant up to the thresholds value of ~5 and 12.





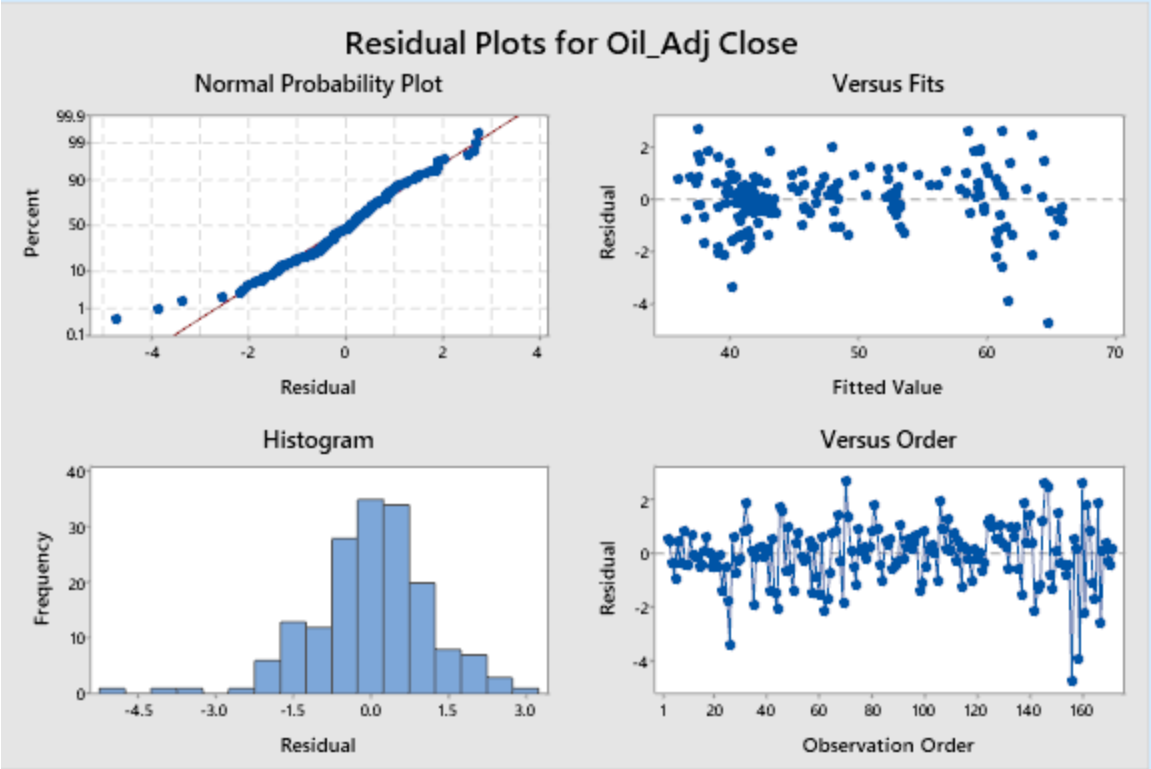
According to the Residual Plots:

Normal Probability Plot - shows that the model fits the data points pretty well except the first few points, which may be due to outliers.

Versus Fits - shows that the distribution is random and that there is no clear pattern.

Histogram - shows normally behaving normal distribution.

Versus Order - shows that the data points are stationary and reverts back to a mean.



The model is a good fit.

ARIMA(0,1,1):

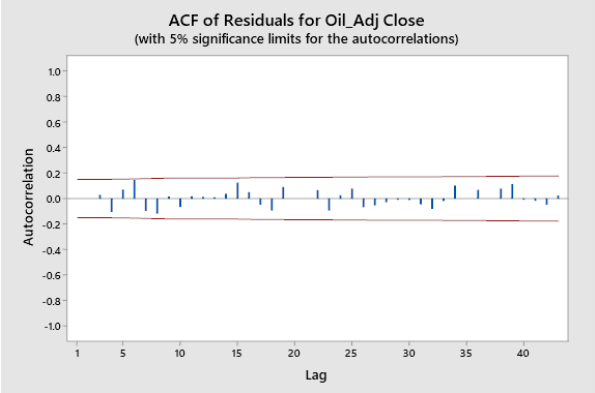
Forecasts from period 172

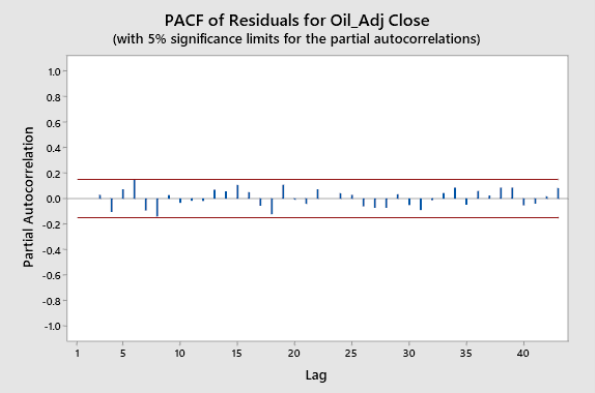
		95% Limits		
Period	Forecast	Lower	Upper	Actual
173	59.7754	57.5117	62.0392	
174	59.8839	56.9327	62.8351	
175	59.9924	56.4861	63.4988	
176	60.1009	56.1161	64.0858	

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.1636	0.0759	2.16	0.032
Constant	0.1085	0.0739	1.47	0.144

In the ACF and PACF graphs the autocorrelation is not significant up to the thresholds value of ~5 and 12.





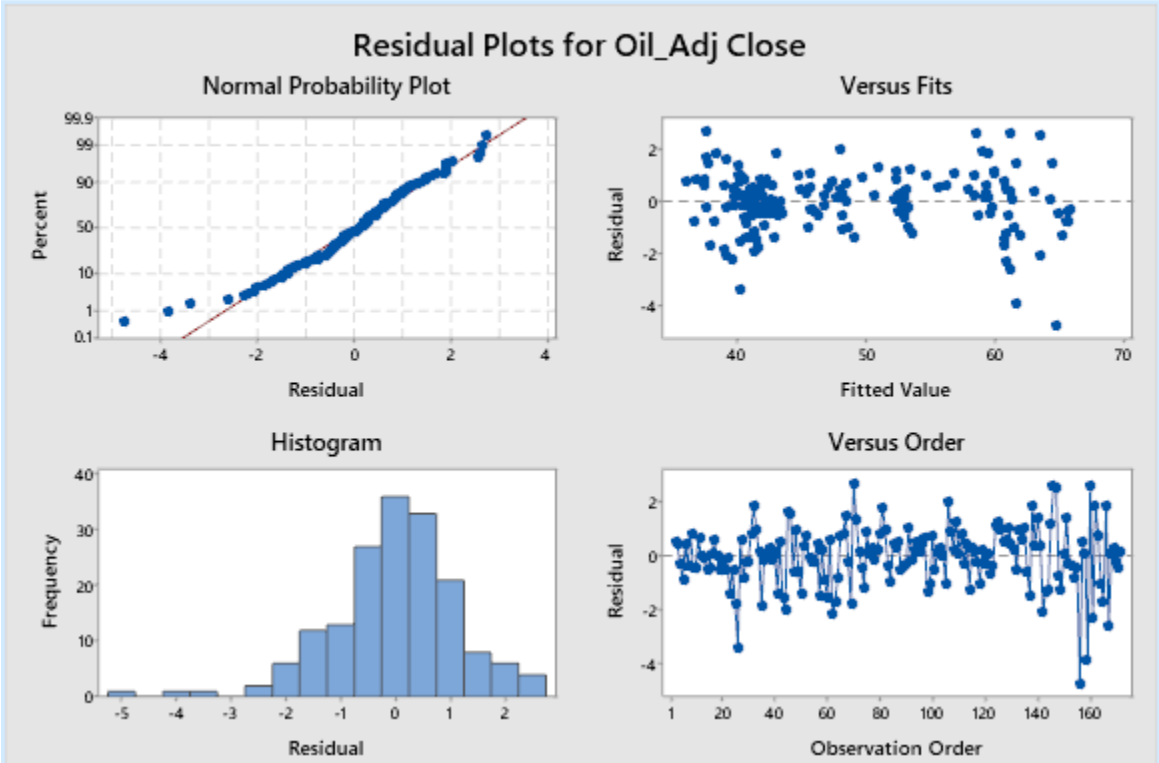
According to the Residual Plots:

Normal Probability Plot - shows that the model fits the data points pretty well except the first few points, which may be due to outliers.

Versus Fits - shows that the distribution is random and that there is no clear pattern.

Histogram - shows normally behaving normal distribution.

Versus Order - shows that the data points are stationary and reverts back to a mean.



The model is a good fit.

<u>ARIMA(1,1,1):</u>

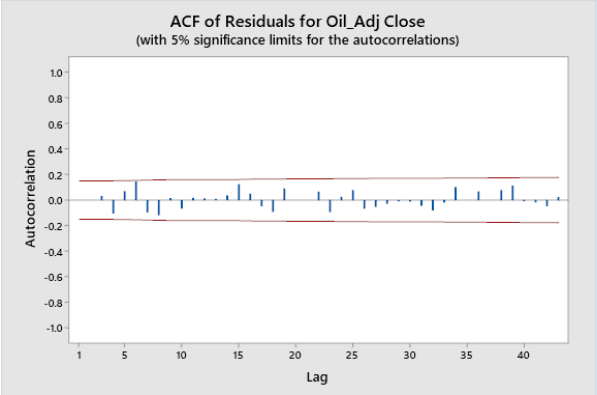
Forecasts from period 172

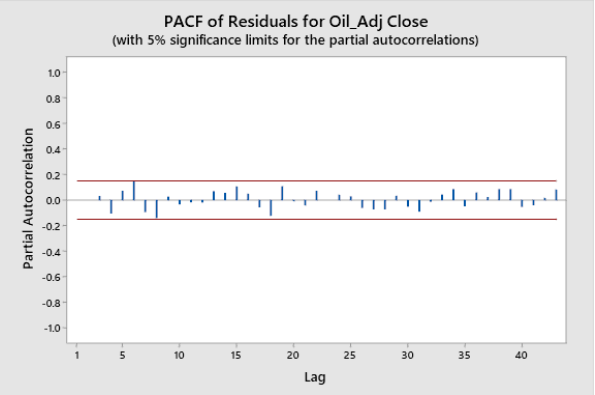
		95% Limits		
Period	Forecast	Lower	Upper	Actual
173	59.7761	57.5056	62.0466	
174	59.8843	56.9240	62.8446	
175	59.9928	56.4771	63.5085	
176	60.1013	56.1067	64.0959	

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.008	0.472	0.02	0.987
MA 1	0.171	0.465	0.37	0.713
Constant	0.1076	0.0734	1.47	0.144

In the ACF and PACF graphs the autocorrelation is not significant up to the thresholds value of ~5 and 12.





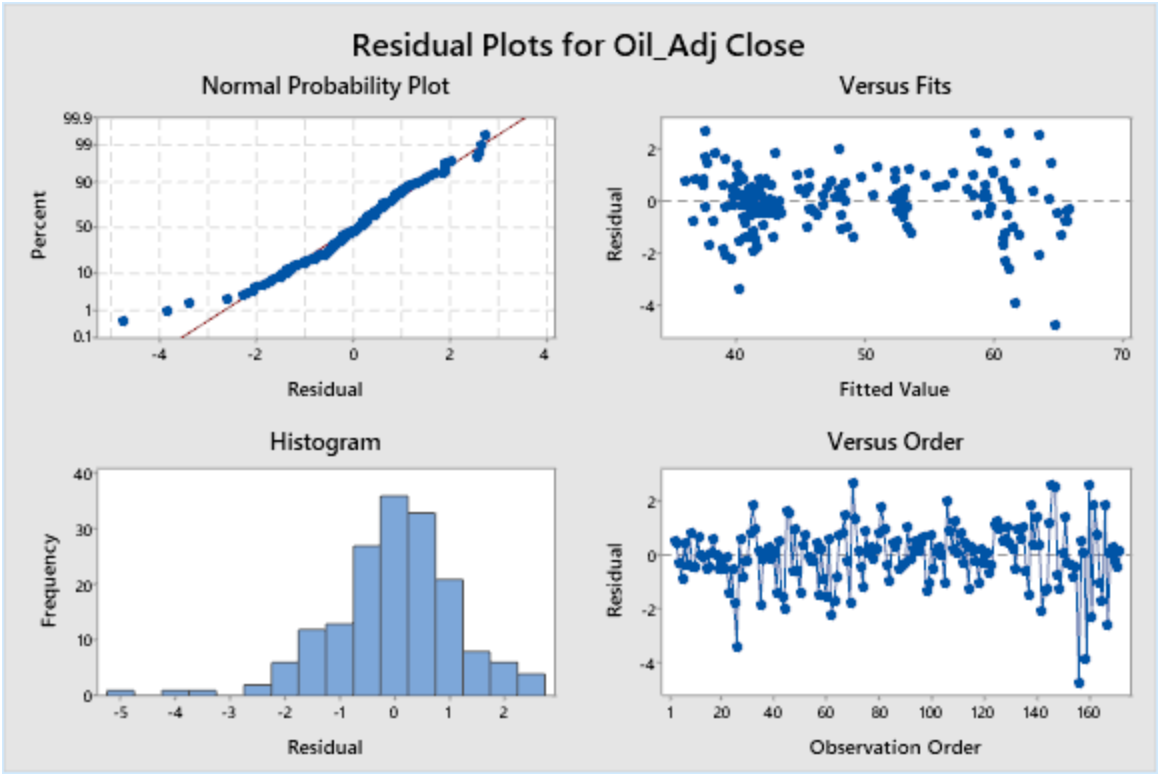
According to the Residual Plots:

Normal Probability Plot - shows that the model fits the data points pretty well except the first few points, which may be due to outliers.

Versus Fits - shows that the distribution is random and that there is no clear pattern.

Histogram - shows normally behaving normal distribution.

Versus Order - shows that the data points are stationary and reverts back to a mean.



The model is a good fit.

F) Based on the results in part E), choose the best ARMA model using MSE. Given the daily oil price from 13th April to 16th April 2021, calculate the forecast errors and comment on the prediction performance of the chosen model.

Model	*MSE*
ARIMA(1,1,0)	Residual Sums of Squares DF SS MS 169 225.563 1.33469 *Back forecasts excluded*
ARIMA(0,1,1)	Residual Sums of Squares DF SS MS 169 225.356 1.33347 *Back forecasts excluded*

ARIMA(1,1,1)	Residual Sums of Squares DF SS MS 168 225.356 1.34140 *Back forecasts excluded*

The slightly better performing model is ARIMA(0,1,1) with slightly less MSE.

Date	*Oil_Adj_Close (Actual)*	*Oil_Adj_Close (Predicted)*	*Forecast Error*
13/4/2021	60.18	59.7754	0.4046
14/4/2021	63.15	59.8839	3.2661
15/4/2021	63.46	59.9924	3.4676
16/4/2021	63.13	60.1009	3.0291

Mean Absolute Deviation (MAD): 2.54185
Mean Square Error (MSE): 8.007701735

Overall the prediction power of the model performed pretty well with very low MAD and MSE. The average distance between the data points and the mean is quite low based on MAD and the regression line is quite close to the data point based on MSE.