FINANCIAL ECONOMECTRICS

Assignment – 1

To – Dr. Mandira Sarma

By - Mamina Panda

CITD (MA – Economics)

3rd - Semester

INTRODUCTION

In this assignment we will use Box – Jenkins method to forecast daily return of Google Limited Liability Company. I will use E-views for this task.

The Box Jenkin model assumes that the given time series is stationary so if our time series is not stationary then we have to make it stationary by taking 1st difference. So if our model is stationary then we will try to estimate ARMA model otherwise we will estimate AR(I)MA model.

Box and Jenkins introduced three steps method to select appropriate models for estimating and forecasting univariate models. Those three steps are:

- Identification
- Estimation
- Diagnostic

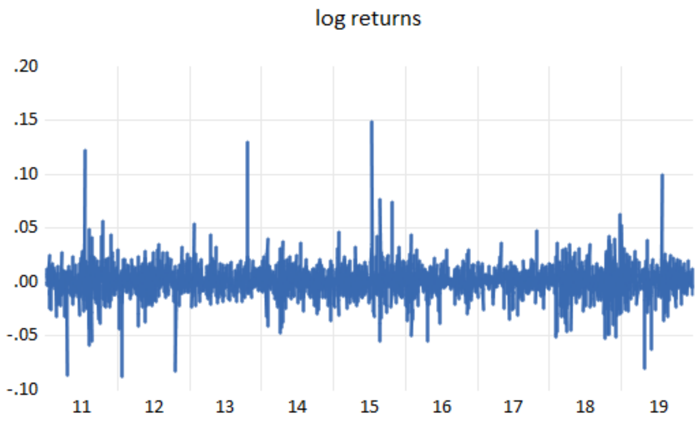
Data

I have taken the data of closing price of Google Limited Liability company from Yahoo Finance website from 3rd Jan 2011 to 30th Dec 2019, name of the stock exchange is GOOG. My total observations are 2264, and the variable which we are dealing with is return, to calculate the return following formula has been used

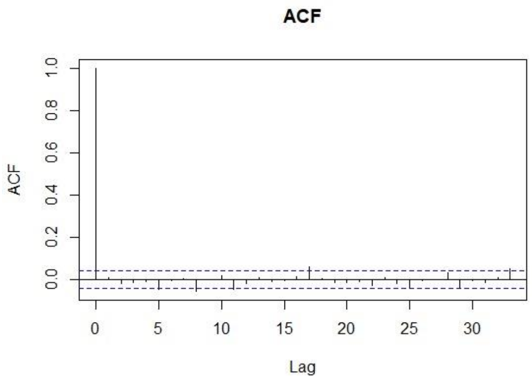
ln (closing $price_{(t)}$/closing $price_{(t-1)}$)

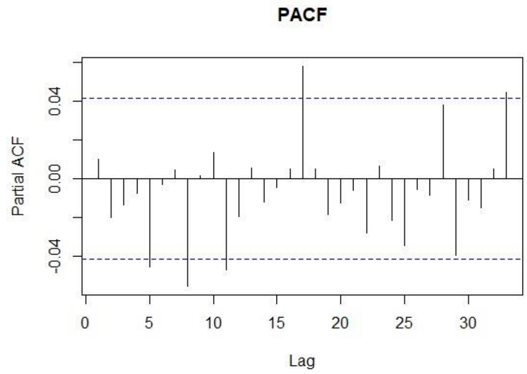
IDENTIFICATION

In this step we will check for the stationarity of our asset return



From the graph it looks like returns are stationary. But we will confirm our result from ACF and PACF model and Dickey fuller test.





PACF is showing significant spikes at various numbers, in ACF even though the spikes are not too high but still may of them are crossing the confidence interval. The PACF cuts off tell us about the AR term, while the ACF typically cuts off at the indicated number of MA. Since return series has stationary behaviour, hence d=0. To create the proper ARMA model, we must consider various combinations of all the lag terms.

Augmented Dickey fuller test

Null Hypothesis: LOG_RETURNS has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=26)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-47.03748	0.0000
Test critical values:	1% level	-3.962103	
	5% level	-3.411795	
	10% level	-3.127785	

*MacKinnon (1996) one-sided p-values.

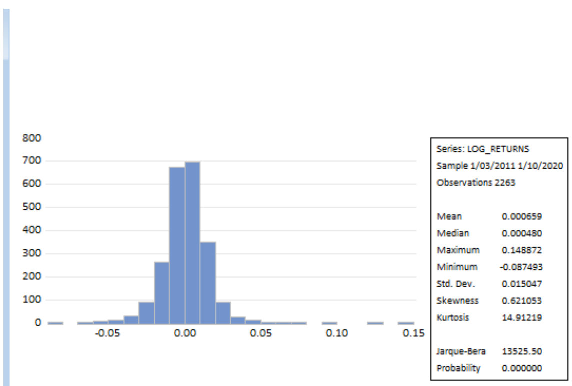
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG_RETURNS)
Method: Least Squares
Date: 10/04/23 Time: 11:50
Sample (adjusted): 1/04/2011 12/30/2019
Included observations: 2262 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_RETURNS(-1)	-0.989758	0.021042	-47.03748	0.0000
C	0.000553	0.000633	0.873016	0.3827
@TREND("1/03/2011")	8.75E-08	4.85E-07	0.180476	0.8568

R-squared	0.494803	Mean dependent var	-5.18E-06
Adjusted R-squared	0.494356	S.D. dependent var	0.021174
S.E. of regression	0.015056	Akaike info criterion	-5.552690
Sum squared resid	0.512111	Schwarz criterion	-5.545098
Log likelihood	6283.092	Hannan-Quinn criter.	-5.549920
F-statistic	1106.263	Durbin-Watson stat	1.999269
Prob(F-statistic)	0.000000		

To solidify this result, we will run Augmented Dickey Fuller test. The test result is given on left side and we can see that the P- value is coming out to be zero which smaller than 0.05 so we will reject the null hypothesis which means our return series is stationary.

JARQUE-BERRA TEST



Ho: Returns are normally distributed.

Ha: Returns are not normally distributed

Since p value is less than 0.05 hence, we will reject the null hypothesis which means returns are not normally distributed.

ESTIMATION

Correlogram of return

Correlogram of LOG_RETURNS

Date: 10/04/23 Time: 12:03
Sample: 1/03/2011 12/30/2019
Included observations: 2263

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.010	0.010	0.2383	0.625
		2	-0.020	-0.020	1.1129	0.573
		3	-0.014	-0.014	1.5546	0.670
		4	-0.007	-0.007	1.6707	0.796
		5	-0.045	-0.045	6.2177	0.286
		6	-0.004	-0.003	6.2469	0.396
		7	0.007	0.005	6.3451	0.500
		8	-0.053	-0.055	12.808	0.119
		9	0.001	0.001	12.809	0.171
		10	0.017	0.013	13.503	0.197
		11	-0.044	-0.047	17.959	0.083
		12	-0.021	-0.019	18.922	0.090
		13	0.011	0.006	19.215	0.117
		14	-0.009	-0.012	19.411	0.150
		15	-0.005	-0.004	19.478	0.193
		16	0.012	0.005	19.824	0.228
		17	0.060	0.058	28.106	0.044
		18	0.003	0.005	28.131	0.060
		19	-0.015	-0.018	28.643	0.072
		20	-0.012	-0.013	28.992	0.088
		21	-0.010	-0.006	29.206	0.109
		22	0.029	0.028	31.146	0.093

We will use correlogram to determine what are the possible models. With the autocorrelation part we are going to determine MA component and with Partial autocorrelation part we will determine AR component.

By looking at the correlogram some of the possible models are possible models are:

ARMA (5,8) ARMA (8,5) ARMA (8,8) ARMA (8,11)

ARMA (5,11) ARMA (11,5)

Note: I have tried other possibilities also but I have added only 10 in the assignment for the sake of comparison.

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 12:08
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.043007	0.018854	-2.281000	0.0226
MA(8)	-0.050384	0.018903	-2.665430	0.0077
SIGMASQ	0.000226	2.55E-06	88.69158	0.0000

R-squared	0.002517	Mean dependent var	0.000659
Adjusted R-squared	0.001634	S.D. dependent var	0.015047
S.E. of regression	0.015035	Akaike info criterion	-5.555520
Sum squared resid	0.510883	Schwarz criterion	-5.547931
Log likelihood	6289.070	Hannan-Quinn criter.	-5.552751
Durbin-Watson stat	1.974561		

Inverted AR Roots	.43-.31i	.43+.31i	-.16+.51i	-.16-.51i
	-.53			
Inverted MA Roots	.69	.49+.49i	.49-.49i	-.00+.69i
	-.00-.69i	-.49+.49i	-.49-.49i	-.69

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 09:19
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 13 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(8)	-0.051336	0.018632	-2.755241	0.0059
MA(5)	-0.041416	0.018895	-2.191868	0.0285
SIGMASQ	0.000226	2.55E-06	88.58761	0.0000

R-squared	0.002505	Mean dependent var	0.000659
Adjusted R-squared	0.001622	S.D. dependent var	0.015047
S.E. of regression	0.015035	Akaike info criterion	-5.555507
Sum squared resid	0.510889	Schwarz criterion	-5.547919
Log likelihood	6289.057	Hannan-Quinn criter.	-5.552738
Durbin-Watson stat	1.974933		

Inverted AR Roots	.64-.26i	.64+.26i	.26+.64i	.26-.64i
	-.26+.64i	-.26-.64i	-.64-.26i	-.64+.26i
Inverted MA Roots	.53	.16-.50i	.16+.50i	-.43+.31i
	-.43-.31i			

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 12:16
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.042217	0.018397	-2.294759	0.0218
MA(11)	-0.043779	0.020550	-2.130365	0.0332
SIGMASQ	0.000226	2.54E-06	89.02702	0.0000

R-squared	0.001767	Mean dependent var	0.000659
Adjusted R-squared	0.000884	S.D. dependent var	0.015047
S.E. of regression	0.015041	Akaike info criterion	-5.554768
Sum squared resid	0.511267	Schwarz criterion	-5.547179
Log likelihood	6288.220	Hannan-Quinn criter.	-5.551999
Durbin-Watson stat	1.975649		

Inverted AR Roots	.43-.31i	.43+.31i	-.16+.51i	-.16-.51i
	-.53			
Inverted MA Roots	.75	.63+.41i	.63-.41i	.31+.68i
	.31-.68i	-.11+.74i	-.11-.74i	-.49-.57i
	-.49+.57i	-.72+.21i	-.72-.21i	

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 12:17
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(11)	-0.041851	0.020543	-2.037258	0.0417
MA(5)	-0.040851	0.018437	-2.215676	0.0268
SIGMASQ	0.000226	2.54E-06	88.99155	0.0000

R-squared	0.001617	Mean dependent var	0.000659
Adjusted R-squared	0.000733	S.D. dependent var	0.015047
S.E. of regression	0.015042	Akaike info criterion	-5.554619
Sum squared resid	0.511344	Schwarz criterion	-5.547030
Log likelihood	6288.051	Hannan-Quinn criter.	-5.551850
Durbin-Watson stat	1.975666		

Inverted AR Roots	.72+.21i	.72-.21i	.49+.57i	.49-.57i
	.11+.74i	.11-.74i	-.31+.68i	-.31-.68i
	-.63+.41i	-.63-.41i	-.75	
Inverted MA Roots	.53	.16+.50i	.16-.50i	-.43-.31i
	-.43+.31i			

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 16:08
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 26 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.461396	0.370890	-1.244023	0.2136
MA(5)	0.420445	0.378356	1.111240	0.2666
SIGMASQ	0.000226	2.56E-06	88.24512	0.0000

R-squared	0.000224	Mean dependent var	0.000659
Adjusted R-squared	-0.000661	S.D. dependent var	0.015047
S.E. of regression	0.015052	Akaike info criterion	-5.553231
Sum squared resid	0.512058	Schwarz criterion	-5.545642
Log likelihood	6286.480	Hannan-Quinn criter.	-5.550462
Durbin-Watson stat	1.974313		

Inverted AR Roots	.69-.50i	.69+.50i	-.26+.81i	-.26-.81i
	-.86			
Inverted MA Roots	.68-.49i	.68+.49i	-.26+.80i	-.26-.80i
	-.84			

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 09:19
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 12 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(8)	-0.346181	0.312407	-1.108108	0.2679
MA(8)	0.296071	0.318648	0.929147	0.3529
SIGMASQ	0.000226	2.57E-06	88.13679	0.0000

R-squared	0.000953	Mean dependent var	0.000659
Adjusted R-squared	0.000069	S.D. dependent var	0.015047
S.E. of regression	0.015047	Akaike info criterion	-5.553955
Sum squared resid	0.511684	Schwarz criterion	-5.546366
Log likelihood	6287.300	Hannan-Quinn criter.	-5.551186
Durbin-Watson stat	1.976124		

Inverted AR Roots	.81+.34i	.81-.34i	.34+.81i	.34-.81i
	-.34-.81i	-.34+.81i	-.81-.34i	-.81+.34i
Inverted MA Roots	.79-.33i	.79+.33i	.33+.79i	.33-.79i
	-.33+.79i	-.33-.79i	-.79-.33i	-.79+.33i

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 12:26
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(8)	-0.052576	0.018579	-2.829864	0.0047
MA(11)	-0.046086	0.020598	-2.237452	0.0254
SIGMASQ	0.000226	2.55E-06	88.59815	0.0000

R-squared	0.002758	Mean dependent var	0.000659
Adjusted R-squared	0.001876	S.D. dependent var	0.015047
S.E. of regression	0.015033	Akaike info criterion	-5.555755
Sum squared resid	0.510759	Schwarz criterion	-5.548166
Log likelihood	6289.336	Hannan-Quinn criter.	-5.552986
Durbin-Watson stat	1.974066		

Inverted AR Roots	.64-.26i	.64+.26i	.26+.64i	.26-.64i
	-.26+.64i	-.26-.64i	-.64+.26i	-.64-.26i
Inverted MA Roots	.76	.64+.41i	.64-.41i	.31+.69i
	.31-.69i	-.11+.75i	-.11-.75i	-.50-.57i
	-.50+.57i	-.73+.21i	-.73-.21i	

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 12:28
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(11)	-0.043851	0.020576	-2.131214	0.0332
MA(8)	-0.051492	0.018802	-2.738675	0.0062
SIGMASQ	0.000226	2.55E-06	88.62991	0.0000

R-squared	0.002596	Mean dependent var	0.000659
Adjusted R-squared	0.001713	S.D. dependent var	0.015047
S.E. of regression	0.015035	Akaike info criterion	-5.555593
Sum squared resid	0.510843	Schwarz criterion	-5.548004
Log likelihood	6289.153	Hannan-Quinn criter.	-5.552824
Durbin-Watson stat	1.973780		

Inverted AR Roots	.72+.21i	.72-.21i	.49+.57i	.49-.57i
	.11+.74i	.11-.74i	-.31+.68i	-.31-.68i
	-.63+.41i	-.63-.41i	-.75	
Inverted MA Roots	.69	.49+.49i	.49-.49i	-.00+.69i
	-.00-.69i	-.49+.49i	-.49-.49i	-.69

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 15:56
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.042217	0.018397	-2.294759	0.0218
MA(11)	-0.043779	0.020550	-2.130365	0.0332
SIGMASQ	0.000226	2.54E-06	89.02702	0.0000

R-squared	0.001767	Mean dependent var	0.000659
Adjusted R-squared	0.000884	S.D. dependent var	0.015047
S.E. of regression	0.015041	Akaike info criterion	-5.554768
Sum squared resid	0.511267	Schwarz criterion	-5.547179
Log likelihood	6288.220	Hannan-Quinn criter.	-5.551999
Durbin-Watson stat	1.975649		

Inverted AR Roots	.43-.31i	.43+.31i	-.16+.51i	-.16-.51i
	-.53			
Inverted MA Roots	.75	.63+.41i	.63-.41i	.31+.68i
	.31-.68i	-.11+.74i	-.11-.74i	-.49-.57i
	-.49+.57i	-.72+.21i	-.72-.21i	

Dependent Variable: LOG_RETURNS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/04/23 Time: 16:13
Sample: 1/03/2011 12/30/2019
Included observations: 2263
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(11)	-0.041851	0.020543	-2.037258	0.0417
MA(5)	-0.040851	0.018437	-2.215676	0.0268
SIGMASQ	0.000226	2.54E-06	88.99155	0.0000

R-squared	0.001617	Mean dependent var	0.000659
Adjusted R-squared	0.000733	S.D. dependent var	0.015047
S.E. of regression	0.015042	Akaike info criterion	-5.554619
Sum squared resid	0.511344	Schwarz criterion	-5.547030
Log likelihood	6288.051	Hannan-Quinn criter.	-5.551850
Durbin-Watson stat	1.975666		

Inverted AR Roots	.72+.21i	.72-.21i	.49+.57i	.49-.57i
	.11+.74i	.11-.74i	-.31+.68i	-.31-.68i
	-.63+.41i	-.63-.41i	-.75	
Inverted MA Roots	.53	.16+.50i	.16-.50i	-.43-.31i
	-.43+.31i			

To be able chose for which ARMA model will work we will look at the P value of the AR and MA for the significance (for it to be significant it should be less than 0.05), we will look at the adjusted R square (Higher the better) and we will also look at the Akaike info criterion, Schwarz criterion and Hannan-Quinn criteria (smaller all of these three will be better will be our model). By comparing all these things in the above analysis best model could be ARMA (8,11), which has highest adjusted R square, AR and MA are significant and values of Akaike, Schwarz and Hannan are smaller.

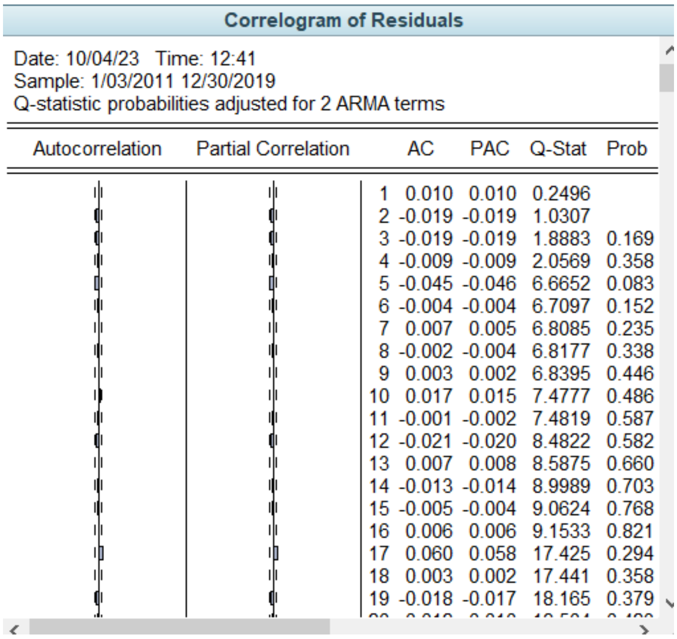
DIAGNOSTIC AND FORECASTING

We have our potential candidate ARMA (8,11)

Requirement for a stable univariate process:

1. Residual of the model are white noise (Ljung – Box Q test)
2. Check if the estimated ARMA process is (covariance) stationary: AR roots should lie inside the unit circle.
3. Check if the estimated ARMA process is invertible: all MA roots should lie inside the unit circle.

Let's check whether the residual is white noise



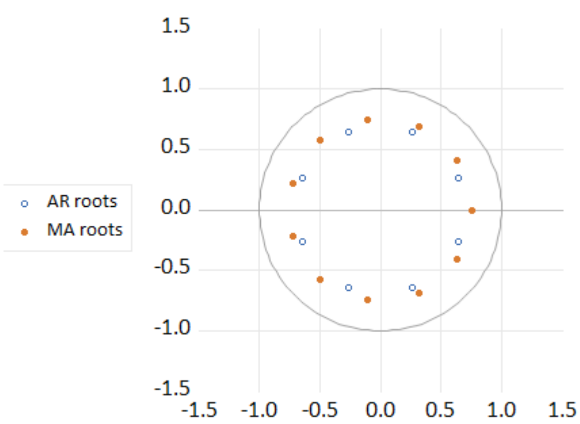
Correlogram of Residuals

Date: 10/04/23 Time: 12:41
Sample: 1/03/2011 12/30/2019
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.010	0.010	0.2496	
		2	-0.019	-0.019	1.0307	
		3	-0.019	-0.019	1.8883	0.169
		4	-0.009	-0.009	2.0569	0.358
		5	-0.045	-0.046	6.6652	0.083
		6	-0.004	-0.004	6.7097	0.152
		7	0.007	0.005	6.8085	0.235
		8	-0.002	-0.004	6.8177	0.338
		9	0.003	0.002	6.8395	0.446
		10	0.017	0.015	7.4777	0.486
		11	-0.001	-0.002	7.4819	0.587
		12	-0.021	-0.020	8.4822	0.582
		13	0.007	0.008	8.5875	0.660
		14	-0.013	-0.014	8.9989	0.703
		15	-0.005	-0.004	9.0624	0.768
		16	0.006	0.006	9.1533	0.821
		17	0.060	0.058	17.425	0.294
		18	0.003	0.002	17.441	0.358
		19	-0.018	-0.017	18.165	0.379

We can clearly see from the correlogram that none of spike in ACF and PACF is crossing the confidence interval hence residuals are indeed a white noise. The residuals' ACF and PACF lags, all fall within the 95% confidence interval, making them all statistically Insignificant.

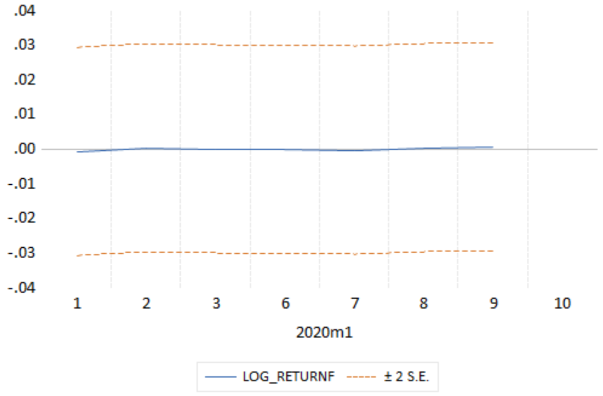
LOG_RETURNS: Inverse Roots of AR/MA Polynomial(s)

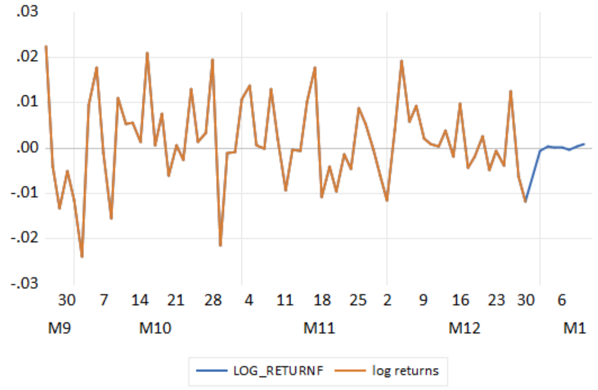


It is evident from the diagram that both AR roots and MA roots are lying within unit circle, it means ARMA process is stationary and invertible. And this implies that all the three conditions are satisfied for stable univariate process.

FORECAST

We forecasted the returns using ARMA (8,11) and compared them to the actual return. So, we have forecasted from 2nd Jan 2020 to 8th Jan 2020. The plot of the forecasted vs actual values is shown below





Comparison of actual and predicted value

Date	Actual Value	Predicted value	Mean Absolute percentage error
02-01-2020	0.023104	0.00046607	0.9798
03-01-2020	-0.00492	0.0000822	0.98329
06-01-2020	0.024358	0.0000446	0.9981
07-01-2020	0.00785	-0.000449	0.9428
08-01-2020	0.010984	0.000335	0.9695

We have calculated the mean absolute error percentage to calculate the see the accuracy of our forecasted value, closer the value is to zero, the better our forecast is and we can see that mean absolute percentage error is pretty high which means and is very close to one which means our forecasted values does not have much accuracy. Now there can be several other factors which are affecting the returns which are not taken into account. Also there can be external shock, which was not predictable. In my case that external shock could be Covid 19.