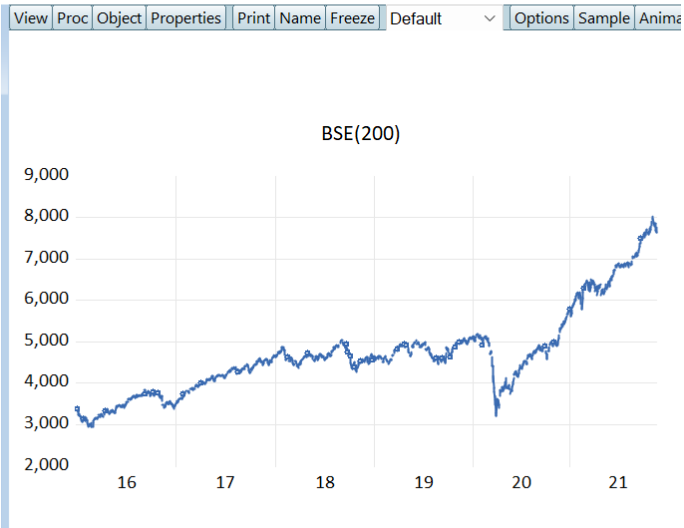
Name : Bhanu Dixit

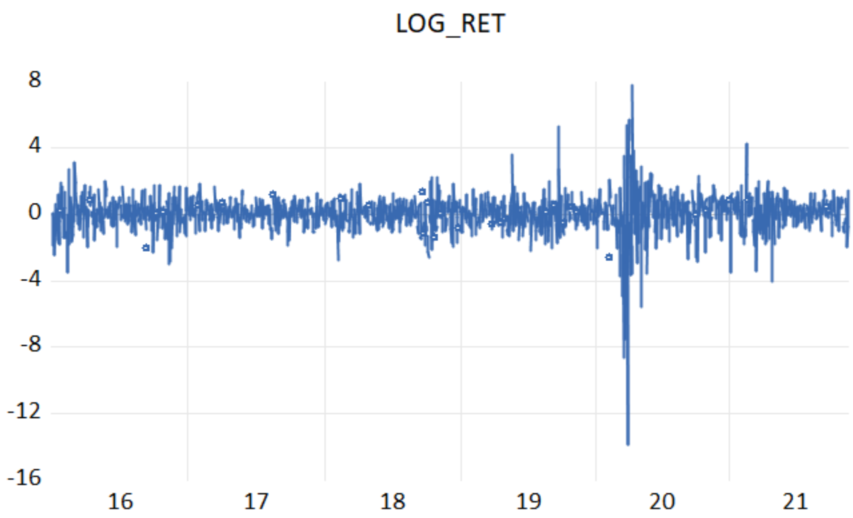
Roll Number:- 2023DSS1009(MSDSM/01/08)

BSE200(2016-20)

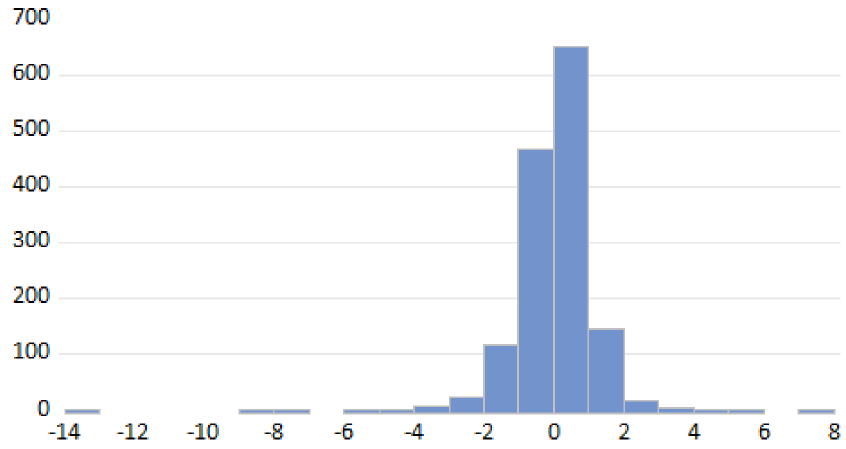
1.Plat for BSE 200 data. As we can see that this plot has a trend initially as follows:



But after one logarithmic difference it plots is as follows:-



2. Descriptive statsistc as as follows:



Series: LOG_RET
Sample 1/01/2016 11/19/2021
Observations 1443

Mean	0.057284
Median	0.126456
Maximum	7.804629
Minimum	-13.82776
Std. Dev.	1.110742
Skewness	-1.795195
Kurtosis	27.62887
Jarque-Bera	37245.78
Probability	0.000000

Max :

- This indicates that the maximum return in one day is 7.8 %

Min: -13.82776

- This indicates that the maximum negative return is one day is 13.8 %

Mean :

- The average value of log returns is 0.057 slightly positive, suggesting a small upward trend in returns over the observed period.

Median:

- The median is slightly higher than the mean, indicating a slight positive skew with value of 0.12.

Standard Deviation:

- Indicates a moderate level of volatility in the log returns with a value of 1.110742.

3. AR, MA , and ARIMA model are applicable to stationary series and we need to check the log returns are stationary or not. Therefore we will run a unit root test.

Null Hypothesis: LOG_RET has a unit root
Exogenous: Constant
Lag Length: 6 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.04945	0.0000
Test critical values:	1% level	-3.434695	
	5% level	-2.863346	
	10% level	-2.567780	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG_RET)
Method: Least Squares
Date: 11/21/24 Time: 15:27
Sample (adjusted): 1/13/2016 11/19/2021
Included observations: 1436 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_RET(-1)	-0.867834	0.066504	-13.04945	0.0000
D(LOG_RET(-1))	-0.122615	0.061410	-1.996664	0.0461
D(LOG_RET(-2))	-0.106198	0.057292	-1.853615	0.0640
D(LOG_RET(-3))	-0.084517	0.051498	-1.641172	0.1010
D(LOG_RET(-4))	-0.087124	0.045068	-1.933191	0.0534
D(LOG_RET(-5))	0.019993	0.037112	0.538740	0.5902
D(LOG_RET(-6))	-0.111402	0.026272	-4.240395	0.0000
C	0.053538	0.028978	1.847535	0.0649

R-squared	0.531611	Mean dependent var	0.001524
Adjusted R-squared	0.529315	S.D. dependent var	1.585461
S.E. of regression	1.087730	Akaike info criterion	3.011618
Sum squared resid	1689.546	Schwarz criterion	3.040975
Log likelihood	-2154.341	Hannan-Quinn criter.	3.022578

Fig :- One root Unit Test

As we can see, the p-value is less than the significance level, which means it is stationary.

Date: 11/21/24 Time: 15:32
Sample (adjusted): 1/04/2016 11/19/2021
Included observations: 1443 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.021	-0.021	0.6176	0.432
		2	0.033	0.032	2.1663	0.339
		3	0.025	0.027	3.0978	0.377
		4	-0.007	-0.007	3.1610	0.531
		5	0.115	0.113	22.300	0.000
		6	-0.133	-0.130	47.931	0.000
		7	0.118	0.111	68.099	0.000
		8	0.003	0.006	68.116	0.000
		9	-0.015	-0.013	68.431	0.000
		10	0.089	0.073	80.077	0.000
		11	-0.077	-0.051	88.749	0.000
		12	0.086	0.045	99.534	0.000
		13	-0.033	-0.007	101.14	0.000
		14	-0.046	-0.058	104.23	0.000
		15	0.052	0.034	108.24	0.000
		16	-0.047	-0.012	111.50	0.000

To find the value of p , for AR and MA model we need to plot the Correlogram

As we can see I the above graph we p=5 or p=6 could be chosen.

a) p= 5

Dependent Variable: LOG_RET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/21/24 Time: 15:52
Sample: 1/04/2016 11/19/2021
Included observations: 1443
Convergence achieved after 135 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056742	0.039379	1.440926	0.1498
AR(1)	-0.020012	0.015568	-1.285435	0.1988
AR(2)	0.030341	0.011486	2.641648	0.0083
AR(3)	0.022689	0.012302	1.844227	0.0654
AR(4)	-0.004289	0.014741	-0.290971	0.7711
AR(5)	0.113928	0.014299	7.967644	0.0000
SIGMASQ	1.214217	0.018300	66.35086	0.0000

R-squared	0.015147	Mean dependent var	0.057284
Adjusted R-squared	0.011032	S.D. dependent var	1.110742
S.E. of regression	1.104598	Akaike info criterion	3.041727
Sum squared resid	1752.115	Schwarz criterion	3.067314
Log likelihood	-2187.606	Hannan-Quinn criter.	3.051277
F-statistic	3.681052	Durbin-Watson stat	1.966834
Prob(F-statistic)	0.001237		

Inverted AR Roots	.66	.19+.60i	.19-.60i	-.53-.39i
	-.53+.39i			

1. **Durbin-Watson Statistic:**
   - Value = 1.966834
   - This statistic is close to 2, indicating little evidence of autocorrelation in the residuals.
2. **F-statistic:**
   - Value = 3.681052, with a p-value = 0.001237.
   - Indicates the model as a whole is statistically significant.

Now we will check residuals

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	9.67E+26	Prob. F(35,1407)	0.0000
Obs*R-squared	1443.000	Prob. Chi-Square(35)	0.0000
Scaled explained SS	15609.60	Prob. Chi-Square(35)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/21/24 Time: 16:02
Sample: 1/04/2016 11/19/2021
Included observations: 1443

This means this is heteroskedastic and we need to find some other model and we are getting similar results for p=6.

B) For MA model we take q= 6 based on correlogram

Dependent Variable: LOG_RET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/21/24 Time: 17:00
Sample: 1/04/2016 11/19/2021
Included observations: 1443
Convergence achieved after 111 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.057298	0.035102	1.632339	0.1028
MA(1)	0.014005	0.016883	0.829576	0.4069
MA(2)	0.012870	0.011351	1.133855	0.2570
MA(3)	0.024163	0.013621	1.773922	0.0763
MA(4)	0.009148	0.015786	0.579523	0.5623
MA(5)	0.087848	0.014737	5.961026	0.0000
MA(6)	-0.113537	0.014062	-8.074127	0.0000
SIGMASQ	1.201136	0.017642	68.08334	0.0000

R-squared	0.025758	Mean dependent var	0.057284
Adjusted R-squared	0.021005	S.D. dependent var	1.110742
S.E. of regression	1.099014	Akaike info criterion	3.032317
Sum squared resid	1733.239	Schwarz criterion	3.061559
Log likelihood	-2179.817	Hannan-Quinn criter.	3.043232
F-statistic	5.419924	Durbin-Watson stat	2.019515
Prob(F-statistic)	0.000004		

Inverted MA Roots	.60	.39+.56i	.39-.56i	-.32+.65i
	-.32-.65i	-.75		

The model seems to be significant, but we need to check Residuals as well.

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	2.09E+26	Prob. F(44,1398)	0.0000
Obs*R-squared	1443.000	Prob. Chi-Square(44)	0.0000
Scaled explained SS	16159.25	Prob. Chi-Square(44)	0.0000

This means residuals are heteroskedastic.

C) For Arima we run auto arima and found results as follows:

Automatic ARIMA Forecasting
Selected dependent variable: LOG_RET
Date: 11/21/24 Time: 17:12
Sample: 1/01/2016 11/19/2021
Included observations: 1443
Forecast length: 0
Model maximums: (4,4)2(0,0)
Regressors: C

Number of estimated ARMA models: 25
Number of non-converged estimations: 0
Selected ARMA model: (2,4)(0,0)
AIC value: 3.00765091016

ARMA Criteria Table

Model Selection Criteria Table
Dependent Variable: LOG_RET
Date: 11/21/24 Time: 17:12
Sample: 1/01/2016 11/19/2021
Included observations: 1443

Model	LogL	AIC*	BIC	HQ
(2,4)(0,0)	-2162.020132	3.007651	3.036893	3.018566
(4,4)(0,0)	-2160.235558	3.007949	3.044502	3.021593
(4,2)(0,0)	-2163.960728	3.010341	3.039582	3.021255
(2,3)(0,0)	-2166.201269	3.012060	3.037646	3.021610
(3,2)(0,0)	-2167.442511	3.013780	3.039367	3.023331
(3,4)(0,0)	-2165.846685	3.014341	3.047237	3.026620
(4,3)(0,0)	-2166.928519	3.015840	3.048737	3.028119

So we have some models with as(2,4),(4,4.)and (4,2) as minimum value of AIC.

So will make an equation dor each of them then check if they are valis or not.

For (2,4)

Dependent Variable: LOG_RET
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 11/21/24 Time: 17:20
Sample: 1/04/2016 11/19/2021
Included observations: 1443
Convergence achieved after 30 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056480	0.034515	1.636398	0.1020
AR(1)	0.019759	0.182516	0.108260	0.9138
AR(2)	0.763570	0.145224	5.257882	0.0000
MA(1)	-0.029096	0.184648	-0.157575	0.8748
MA(2)	-0.749300	0.145501	-5.149776	0.0000
MA(3)	0.076928	0.029444	2.612689	0.0091
MA(4)	-0.041223	0.037596	-1.096480	0.2731

R-squared	0.014168	Mean dependent var	0.057284
Adjusted R-squared	0.010049	S.D. dependent var	1.110742
S.E. of regression	1.105146	Akaike info criterion	3.042739
Sum squared resid	1753.857	Schwarz criterion	3.068325
Log likelihood	-2188.336	Hannan-Quinn criter.	3.052289
F-statistic	3.439694	Durbin-Watson stat	2.000430
Prob(F-statistic)	0.002237		

Inverted AR Roots	.88	-.86		
Inverted MA Roots	.86	.05+.22i	.05-.22i	-.92

And Residuals

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	73.29551	Prob. F(35,1407)	0.0000
Obs*R-squared	931.8908	Prob. Chi-Square(35)	0.0000
Scaled explained SS	9831.917	Prob. Chi-Square(35)	0.0000

For (4,4)

Dependent Variable: LOG_RET
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 11/21/24 Time: 17:22
Sample: 1/04/2016 11/19/2021
Included observations: 1443
Convergence achieved after 21 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056587	0.034518	1.639318	0.1014
AR(1)	-1.202635	0.225425	-5.334967	0.0000
AR(2)	0.127374	0.330827	0.385017	0.7003
AR(3)	1.099704	0.241921	4.545723	0.0000
AR(4)	0.481655	0.131447	3.664262	0.0003
MA(1)	1.237727	0.222081	5.573320	0.0000
MA(2)	-0.040657	0.347250	-0.117082	0.9068
MA(3)	-1.052663	0.275170	-3.825502	0.0001
MA(4)	-0.546540	0.125758	-4.345956	0.0000

R-squared	0.052036	Mean dependent var	0.057284
Adjusted R-squared	0.046747	S.D. dependent var	1.110742
S.E. of regression	1.084469	Akaike info criterion	3.006566
Sum squared resid	1686.488	Schwarz criterion	3.039463
Log likelihood	-2160.237	Hannan-Quinn criter.	3.018845
F-statistic	9.839408	Durbin-Watson stat	2.026966
Prob(F-statistic)	0.000000		

Inverted AR Roots	.91	-.60	-.76+.56i	-.76-.56i
Inverted MA Roots	.89	-.70+.59i	-.70-.59i	-.73

And Residuals

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	54.87486	Prob. F(54,1388)	0.0000
Obs*R-squared	982.6984	Prob. Chi-Square(54)	0.0000
Scaled explained SS	9535.457	Prob. Chi-Square(54)	0.0000

For (4,2)

Dependent Variable: LOG_RET
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 11/21/24 Time: 17:25
Sample: 1/04/2016 11/19/2021
Included observations: 1443
Convergence achieved after 32 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.057461	0.028397	2.023497	0.0432
AR(1)	-1.409358	0.053859	-26.16745	0.0000
AR(2)	-0.762008	0.063750	-11.95316	0.0000
AR(3)	-0.004384	0.047449	-0.092401	0.9264
AR(4)	-0.079291	0.030573	-2.593446	0.0096
MA(1)	1.425240	0.048033	29.67220	0.0000
MA(2)	0.806191	0.045600	17.67950	0.0000

R-squared	0.047099	Mean dependent var	0.057284
Adjusted R-squared	0.043118	S.D. dependent var	1.110742
S.E. of regression	1.086531	Akaike info criterion	3.008957
Sum squared resid	1695.271	Schwarz criterion	3.034544
Log likelihood	-2163.963	Hannan-Quinn criter.	3.018507
F-statistic	11.82956	Durbin-Watson stat	1.984237
Prob(F-statistic)	0.000000		

Inverted AR Roots	.07+.29i	.07-.29i	-.78-.55i	-.78+.55i
Inverted MA Roots	-.71-.55i	-.71+.55i		

For residuals

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	20.56558	Prob. F(35,1407)	0.0000
Obs*R-squared	488.3707	Prob. Chi-Square(35)	0.0000
Scaled explained SS	5265.806	Prob. Chi-Square(35)	0.0000

In all the above combinations of p and q, we got a provability value less than significant level but they are heteroskedastic

4.

We have check ar for 3 ,odels as (1,1) (1.,0) and (0,1) and results are as follows.

Dependent Variable: LOG_RET
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 11/21/24 Time: 17:39
Sample (adjusted): 1/05/2016 11/19/2021
Included observations: 1442 after adjustments
Convergence achieved after 32 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/04/2016
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.085942	0.022726	3.781627	0.0002
AR(1)	-0.029006	0.380403	-0.076250	0.9392
MA(1)	0.102203	0.379444	0.269350	0.7877
		Variance Equation		
C	0.028813	0.006719	4.288214	0.0000
RESID(-1)^2	0.106509	0.011324	9.405422	0.0000
GARCH(-1)	0.864213	0.015759	54.83967	0.0000

R-squared	-0.009428	Mean dependent var	0.058601
Adjusted R-squared	-0.010831	S.D. dependent var	1.110000
S.E. of regression	1.115995	Akaike info criterion	2.593417
Sum squared resid	1792.195	Schwarz criterion	2.615361
Log likelihood	-1863.854	Hannan-Quinn criter.	2.601608
Durbin-Watson stat	2.189595		

Inverted AR Roots	-.03
Inverted MA Roots	-.10

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	18386.85	Prob. F(27,1414)	0.0000
Obs*R-squared	1437.904	Prob. Chi-Square(27)	0.0000
Scaled explained SS	1993.238	Prob. Chi-Square(27)	0.0000

Dependent Variable: LOG_RET
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 11/21/24 Time: 17:48
Sample (adjusted): 1/05/2016 11/19/2021
Included observations: 1442 after adjustments
Convergence achieved after 23 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.085890	0.022786	3.769468	0.0002
AR(1)	0.072126	0.028968	2.489846	0.0128
		Variance Equation		
C	0.028884	0.006728	4.293033	0.0000
RESID(-1)^2	0.106433	0.011106	9.583142	0.0000
GARCH(-1)	0.864174	0.015690	55.07751	0.0000

R-squared	-0.008708	Mean dependent var	0.058601
Adjusted R-squared	-0.009409	S.D. dependent var	1.110000
S.E. of regression	1.115210	Akaike info criterion	2.592130
Sum squared resid	1790.918	Schwarz criterion	2.610416
Log likelihood	-1863.926	Hannan-Quinn criter.	2.598956
Durbin-Watson stat	2.186521		
Inverted AR Roots	.07		

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	25318.04	Prob. F(20,1421)	0.0000
Obs*R-squared	1437.965	Prob. Chi-Square(20)	0.0000
Scaled explained SS	1997.299	Prob. Chi-Square(20)	0.0000

Dependent Variable: LOG_RET
Method: ML ARCH - Normal distribution (Newton-Raphson / Marquardt steps)
Date: 11/21/24 Time: 17:50
Sample (adjusted): 1/04/2016 11/19/2021
Included observations: 1443 after adjustments
Convergence achieved after 11 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/01/2016
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.085637	0.022600	3.789277	0.0002
MA(1)	0.073219	0.029176	2.509572	0.0121
		Variance Equation		
C	0.028992	0.006703	4.325248	0.0000
RESID(-1)^2	0.106301	0.011097	9.578828	0.0000
GARCH(-1)	0.863915	0.015633	55.26241	0.0000

R-squared	-0.009303	Mean dependent var	0.057284
Adjusted R-squared	-0.010003	S.D. dependent var	1.110742
S.E. of regression	1.116283	Akaike info criterion	2.592798
Sum squared resid	1795.614	Schwarz criterion	2.611074
Log likelihood	-1865.704	Hannan-Quinn criter.	2.599620
Durbin-Watson stat	2.187355		
Inverted MA Roots	-.07		

Heteroskedasticity Test: ARCH

F-statistic	0.032775	Prob. F(1,1440)	0.8564
Obs*R-squared	0.032820	Prob. Chi-Square(1)	0.8562

Only the (0 1) model is able to justify homoscedastic . Else. All others are there heteroskedastic even after having a p-value less than the significance level.

R-squared: −0.009303-0.009303−0.009303

- Indicates poor explanatory power for the mean equation. This is common in financial return models where the focus is on modeling volatility rather than returns.

Log-likelihood: −1865.704-1865.704−1865.704

- Used to compare model fit across specifications.

Akaike (AIC) and Schwarz Criterion (SIC):

- AIC = 2.592798, SIC = 2.611074
- Lower values indicate better model fit.

Durbin-Watson Stat (2.18):

- Close to 2, indicating little to no autocorrelation in residuals.

This output summarizes the results of a GARCH (1,1) model for the dependent variable LOG_RET\text{LOG_RET}LOG_RET, which appears to be a time series of log returns. Here's the interpretation of the key results:

Mean Equation

The mean equation is modeled as:

LOG_RET=C+MA(1)+ϵt\text{LOG_RET} = C + \text{MA(1)} + \epsilon_tLOG_RET=C+MA(1)+ϵt

- **C (Constant):**
    - Coefficient = **0.085637**
    - ppp-value = **0.0002** (significant at the 1% level).
    - Interpretation: The constant term is positive and statistically significant, indicating a positive average return over the period.
- **MA(1) (Moving Average):**
    - Coefficient = **0.073219**
    - ppp-value = **0.0121** (significant at the 5% level).
    - Interpretation: The MA(1) term is significant, suggesting that the model accounts for short-term dependencies in the time series data.

Variance Equation (GARCH Model)

The variance equation is modeled as:

σt2=C+α·ϵt−12+β·σt−12\sigma_t^2 = C + \alpha \cdot \epsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2σt2=C+α·ϵt−12+β·σt−12

Where:

- ϵt−12\epsilon_{t-1}^2ϵt−12: Lagged squared residual (ARCH term)
- σt−12\sigma_{t-1}^2σt−12: Lagged conditional variance (GARCH term)
- **C (Constant in variance):**
    - Coefficient = **0.028992**
    - ppp-value = **0.0000** (significant).
    - Interpretation: This represents the long-term variance. It is positive and statistically significant, indicating the baseline level of volatility.
- **RESID(-1)^2 (ARCH term):**
    - Coefficient = **0.106301**
    - ppp-value = **0.0000** (significant).
    - Interpretation: The ARCH effect is significant, meaning that past shocks to returns influence current volatility.
- **GARCH(-1) (GARCH term):**
    - Coefficient = **0.863915**
    - ppp-value = **0.0000** (significant).
    - Interpretation: The GARCH effect is dominant, suggesting a strong persistence of volatility over time.

Sum of ARCH and GARCH coefficients (α+β\alpha + \betaα+β):

- α+β=0.106301+0.863915=0.970216\alpha + \beta = 0.106301 + 0.863915 = 0.970216α+β=0.106301+0.863915=0.970216
- This is less than 1, satisfying the stationarity condition of the GARCH model. However, the high value indicates persistent volatility.

Model Diagnostics

- **R-squared**: −0.009303-0.009303−0.009303

- Indicates poor explanatory power for the mean equation. This is common in financial return models where the focus is on modeling volatility rather than returns.

- **Log-likelihood**: −1865.704-1865.704−1865.704
  - Used to compare model fit across specifications.
- **Akaike (AIC) and Schwarz Criterion (SIC):**
  - AIC = 2.592798, SIC = 2.611074
  - Lower values indicate better model fit.
- **Durbin-Watson Stat (2.18):**
  - Close to 2, indicating little to no autocorrelation in residuals.

Conclusion

- The mean equation shows a significant constant and MA(1) term, indicating modest predictability in returns.
- GARRCH is better than ARIMA as it id homoscsdic.
- The variance equation effectively captures volatility clustering, with significant ARCH and GARCH terms and persistent volatility.
- Diagnostic statistics suggest that the model is well-specified for modeling volatility but has limited explanatory power for returns, as expected in financial time series analysis.