FINANCIAL ECONOMETRICS ASSIGNMENT (II)

NAME – MAMINA PANDA

MA ECONOMICS 2ND YEAR

DEPARTMENT – CITD

SUBMITTED TO – DR. SUMAN DAS

Assignment - Generalised ARCH (GARCH) model

INTRODUCTION

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) is a statistical model used in analysing time-series data when the variance error is assumed to be serially autocorrelated. These models assume that the variance of the error term is heteroskedastic (means error term is not constant).

Variance of error term is assumed to vary systematically and this is conditional on the average size of the error terms in previous periods, this means it has conditional heteroskedasticity, that's why it's said that error term follows an autoregressive moving average pattern.

GARCH models are mainly used in several types of financial data and is typically used to estimate volatility of returns of bonds, stocks and other such asset.

Why GARCH is better than ARCH model??

The GARCH model is more parsimonious than ARCH model as it uses very few parameters (most of the time just couple of parameters) to achieve what the ARCH model would need.

Steps to estimate GARCH (p, q) model

- We will estimate mean for the model using ARIMA model
- Then we will test for the existence of ARCH-GARCH effects from the estimated equation
- If the ARCH-GARCH effects is found to be there then we will re-estimate our model
- Large ARCH models are better identified by GARCH model.
- Finally, we conduct diagnostic test.

NOTE: If there is still heteroskedasticity in the model, it means that the specification we used our poor and we have to incorporate more ARCH/GARCH terms.

Source of the data

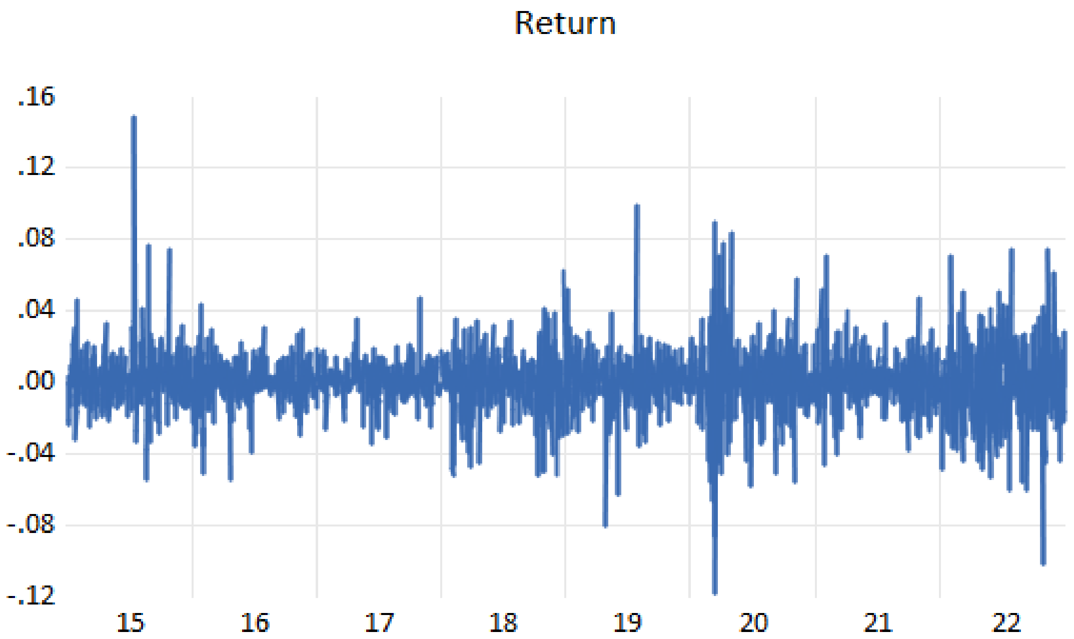
To conduct the analysis, I have taken the data of closing price from yahoo finance. The stock data which I have chosen is Google stock data called GOOG. The data ranges from 01-01-2015 to 31-12-2022. The total number of observations are 2014.

To calculate the return following formula has been used

In (closing price$_{(t)}$/closing price$_{(t-1)}$)

STEP 1: Mean Equation

First, we will check for the stationarity of the return of GOOG stock



From the graph it looks like our data on returns are stationary but we will solidify our result by running Dickey – Fuller test.

Augmented Dickey-Fuller test

Augmented Dickey-Fuller Unit Root Test on RETURN

Null Hypothesis: RETURN has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=25)

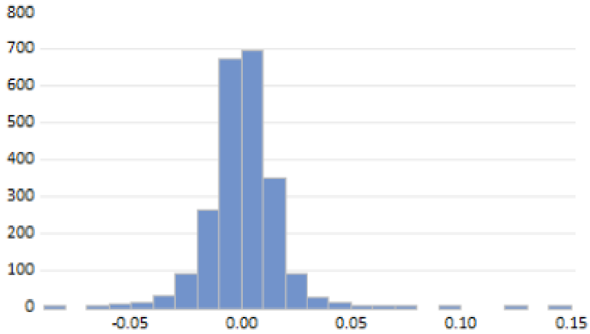
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-47.85071	0.0000
Test critical values:	1% level	-3.962607	
	5% level	-3.412042	
	10% level	-3.127931	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RETURN)
Method: Least Squares
Date: 10/31/23 Time: 14:18
Sample (adjusted): 1/05/2015 12/30/2022
Included observations: 2013 after adjustments

The table on the side shows us the result of Augmented Dickey Fuller Unit root test on return, and by looking at the p value we can say whether return series are stationary of not. Here we can see that p-value is coming out to be less than 0.05, hence we will reject the null hypothesis which implies that our return

Jarque – Berra Test – Test for Normality



Series: LOG_RETURNS
Sample 1/03/2011 1/10/2020
Observations 2263

Mean	0.000659
Median	0.000480
Maximum	0.148872
Minimum	-0.087493
Std. Dev.	0.015047
Skewness	0.621053
Kurtosis	14.91219
Jarque-Bera	13525.50
Probability	0.000000

Here we can see that the p value is less than 0.05, which implies the Ho (Null hypothesis) is rejected and the returns are not normally distributed.

Plot of ACF and PACF

Correlogram of RETURN

Date: 10/31/23 Time: 14:15
Sample: 1/02/2015 12/30/2022
Included observations: 2014

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.065	-0.065	8.4058	0.004
		2	-0.001	-0.005	8.4090	0.015
		3	-0.022	-0.022	9.3817	0.025
		4	-0.019	-0.022	10.110	0.039
		5	-0.003	-0.006	10.127	0.072
		6	-0.061	-0.062	17.609	0.007
		7	0.055	0.046	23.658	0.001
		8	-0.092	-0.088	40.701	0.000
		9	0.066	0.054	49.653	0.000
		10	-0.013	-0.008	49.998	0.000

The spikes which are crossing the confidence interval could be the potential candidates for our model (AR and MA).

Therefore, potential ARIMA structure could be ARIMA (1,0,1), ARIMA (6,0,6), ARIMA (6,0,1) ARIMA (1,0,6).

We will select the best model out of these four by comparing the values of Akaike information Criterion, Schwartz Criterion

Equation: AR1MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/31/23 Time: 14:27
Sample: 1/02/2015 12/30/2022
Included observations: 2014
Convergence achieved after 53 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000610	0.000341	1.790161	0.0736
AR(1)	0.733865	0.099434	7.380417	0.0000
MA(1)	-0.782446	0.090886	-8.609099	0.0000
SIGMASQ	0.000315	4.85E-06	64.91880	0.0000

R-squared	0.005090	Mean dependent var	0.000606
Adjusted R-squared	0.003605	S.D. dependent var	0.017801
S.E. of regression	0.017769	Akaike info criterion	-5.220701
Sum squared resid	0.634652	Schwarz criterion	-5.209563
Log likelihood	5261.246	Hannan-Quinn criter.	-5.216613
F-statistic	3.427831	Durbin-Watson stat	2.037612
Prob(F-statistic)	0.016482		

Inverted AR Roots	.73
Inverted MA Roots	.78

Equation: AR6MA6 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/31/23 Time: 14:29
Sample: 1/02/2015 12/30/2022
Included observations: 2014
Convergence achieved after 23 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000608	0.000379	1.605443	0.1086
AR(6)	-0.084663	0.303974	-0.278522	0.7806
MA(6)	0.023728	0.305997	0.077544	0.9382
SIGMASQ	0.000316	4.83E-06	65.35551	0.0000

R-squared	0.003727	Mean dependent var	0.000606
Adjusted R-squared	0.002240	S.D. dependent var	0.017801
S.E. of regression	0.017781	Akaike info criterion	-5.219327
Sum squared resid	0.635522	Schwarz criterion	-5.208189
Log likelihood	5259.862	Hannan-Quinn criter.	-5.215239
F-statistic	2.506431	Durbin-Watson stat	2.123605
Prob(F-statistic)	0.057376		

Inverted AR Roots	.57+.33i	.57-.33i	.00-.66i	-.00+.66i
	-.57+.33i	-.57-.33i		
Inverted MA Roots	.46-.27i	.46+.27i	.00+.54i	-.00-.54i
	-.46-.27i	-.46+.27i		

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/31/23 Time: 14:31
Sample: 1/02/2015 12/30/2022
Included observations: 2014
Convergence achieved after 20 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000609	0.000353	1.726267	0.0845
AR(6)	-0.058618	0.016324	-3.591020	0.0003
MA(1)	-0.063436	0.015182	-4.178288	0.0000
SIGMASQ	0.000314	4.81E-06	65.31225	0.0000

R-squared	0.007628	Mean dependent var	0.000606
Adjusted R-squared	0.006147	S.D. dependent var	0.017801
S.E. of regression	0.017747	Akaike info criterion	-5.223249
Sum squared resid	0.633033	Schwarz criterion	-5.212111
Log likelihood	5263.812	Hannan-Quinn criter.	-5.219161
F-statistic	5.150052	Durbin-Watson stat	1.998717
Prob(F-statistic)	0.001508		

Inverted AR Roots	.54-.31i	.54+.31i	.00-.62i	-.00+.62i
	-.54+.31i	-.54-.31i		
Inverted MA Roots	.06			

Equation: AR1MA6 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/31/23 Time: 14:31
Sample: 1/02/2015 12/30/2022
Included observations: 2014
Convergence achieved after 21 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000609	0.000354	1.719864	0.0856
AR(1)	-0.062145	0.014831	-4.190266	0.0000
MA(6)	-0.057428	0.016404	-3.500921	0.0005
SIGMASQ	0.000314	4.81E-06	65.30019	0.0000

R-squared	0.007503	Mean dependent var	0.000606
Adjusted R-squared	0.006021	S.D. dependent var	0.017801
S.E. of regression	0.017748	Akaike info criterion	-5.223123
Sum squared resid	0.633113	Schwarz criterion	-5.211986
Log likelihood	5263.685	Hannan-Quinn criter.	-5.219035
F-statistic	5.064761	Durbin-Watson stat	2.001485
Prob(F-statistic)	0.001700		

Inverted AR Roots	-.06			
Inverted MA Roots	.62	.31-.54i	.31+.54i	-.31-.54i
	-.31+.54i	-.62		

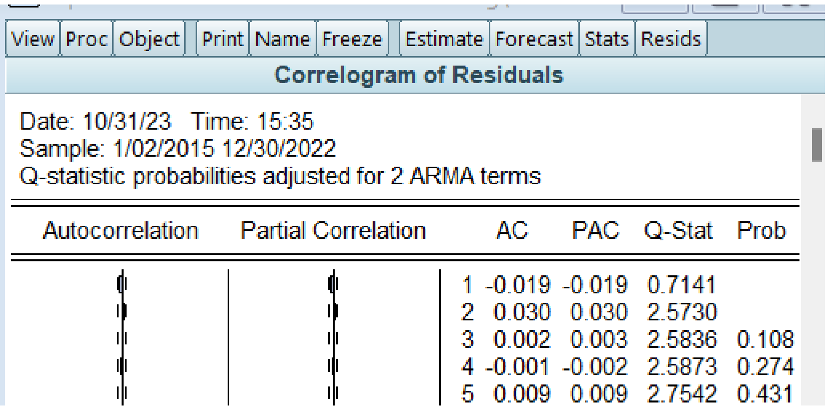
To be able to chose which ARMA model will fit the best we will look at the significance level of AR and MA, adjusted R square (higher the better) and H-Q criteria (smaller the better). Hence by comparing all three the best model will be ARIMA (6,0,1)

Hence our mean equation will be

$$y_t = 0.0006 - 0.0586y_{t-2} - 0.0634\varepsilon_{t-1} + \varepsilon_t$$

Diagnostics test

In diagnostic checking I have to check whether the residuals are white noise or not and whether the process is covariance stationary and invertible



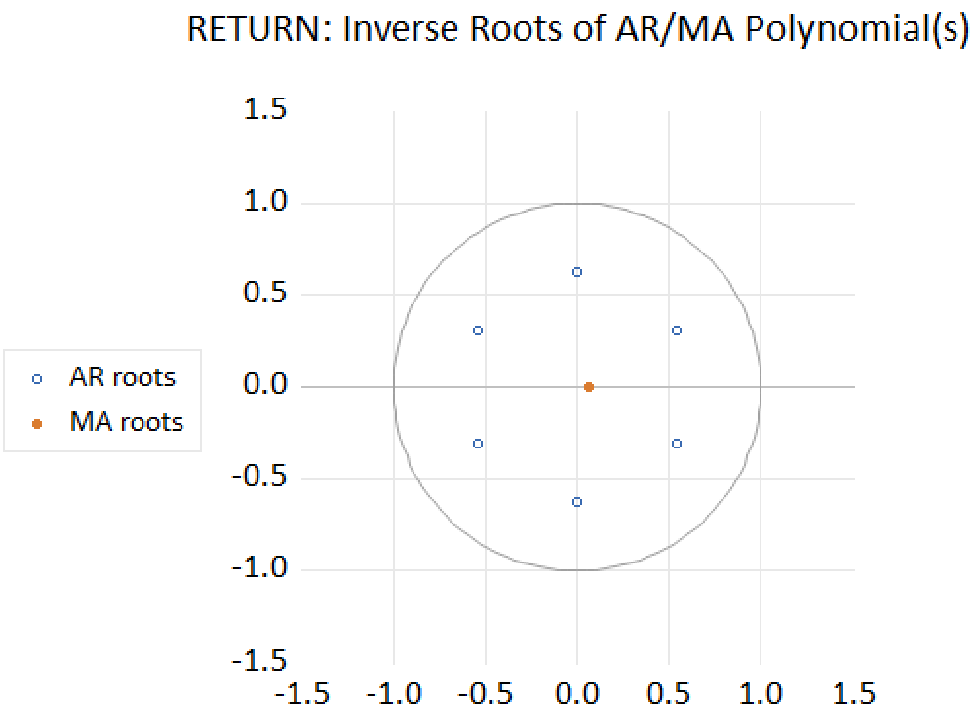
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals

Date: 10/31/23 Time: 15:35
Sample: 1/02/2015 12/30/2022
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.7141	
		2	0.030	0.030	2.5730	
		3	0.002	0.003	2.5836	0.108
		4	-0.001	-0.002	2.5873	0.274
		5	0.009	0.009	2.7542	0.431

Since none of the spikes in ACF and PACF are crossing the confidence interval hence all the residuals are white noise.



It is evident from the above diagram that both AR and MA roots are lying within unit root circle, it means ARMA process is stationary and invertible.

Checking the existence of ARCH effect

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	41.41284	Prob. F(1,2011)	0.0000
Obs*R-squared	40.61758	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/01/23 Time: 11:52
Sample (adjusted): 1/05/2015 12/30/2022
Included observations: 2013 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000270	2.15E-05	12.57059	0.0000
RESID^2(-1)	0.142048	0.022073	6.435281	0.0000

R-squared	0.020178	Mean dependent var	0.000314
Adjusted R-squared	0.019690	S.D. dependent var	0.000920
S.E. of regression	0.000911	Akaike info criterion	-11.16259
Sum squared resid	0.001670	Schwarz criterion	-11.15702
Log likelihood	11237.15	Hannan-Quinn criter.	-11.16055
F-statistic	41.41284	Durbin-Watson stat	2.031453
Prob(F-statistic)	0.000000		

$H0$: There is no ARCH effect on the error term

$H0$: There is ARCH effect on the error term

We can see that the p value is less than 0.05, so we will reject the null hypothesis and accept the alternative hypothesis that there is ARCH effect on the error term. Now we can estimate the GARCH model.

ARCH order

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals Squared

Date: 11/01/23 Time: 12:00
Sample: 1/02/2015 12/30/2022
Included observations: 2014

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.142	0.142	40.694	0.000
		2	0.129	0.111	74.423	0.000
		3	0.115	0.085	101.01	0.000
		4	0.106	0.071	123.90	0.000
		5	0.118	0.079	151.85	0.000
		6	0.104	0.058	173.56	0.000
		7	0.067	0.017	182.65	0.000
		8	0.059	0.014	189.78	0.000
		9	0.055	0.014	195.99	0.000
		10	0.060	0.022	203.24	0.000

To determine the GARCH order we will look at the correlogram of Partial correlation and we can clearly see that 1st 5 lags are crossing the confidence interval, hence we will check by adding these lags one after another.

Adding 3 lags

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	27.43798	Prob. F(3,2007)	0.0000
Obs*R-squared	79.22856	Prob. Chi-Square(3)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/01/23 Time: 12:11
Sample (adjusted): 1/07/2015 12/30/2022
Included observations: 2011 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000219	2.27E-05	9.635290	0.0000
RESID^2(-1)	0.116708	0.022240	5.247612	0.0000
RESID^2(-2)	0.100646	0.022281	4.517198	0.0000
RESID^2(-3)	0.085375	0.022240	3.838739	0.0001

R-squared	0.039398	Mean dependent var	0.000314
Adjusted R-squared	0.037962	S.D. dependent var	0.000921
S.E. of regression	0.000903	Akaike info criterion	-11.17950
Sum squared resid	0.001637	Schwarz criterion	-11.16835
Log likelihood	11244.98	Hannan-Quinn criter.	-11.17540
F-statistic	27.43798	Durbin-Watson stat	2.011784
Prob(F-statistic)	0.000000		

Adding 4 lags

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	23.18707	Prob. F(4,2005)	0.0000
Obs*R-squared	88.86865	Prob. Chi-Square(4)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/01/23 Time: 12:04
Sample (adjusted): 1/08/2015 12/30/2022
Included observations: 2010 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000204	2.32E-05	8.777074	0.0000
RESID^2(-1)	0.110738	0.022277	4.970924	0.0000
RESID^2(-2)	0.093565	0.022348	4.186770	0.0000
RESID^2(-3)	0.077044	0.022348	3.447495	0.0006
RESID^2(-4)	0.070714	0.022277	3.174369	0.0015

R-squared	0.044213	Mean dependent var	0.000314
Adjusted R-squared	0.042306	S.D. dependent var	0.000921
S.E. of regression	0.000901	Akaike info criterion	-11.18308
Sum squared resid	0.001629	Schwarz criterion	-11.16914
Log likelihood	11244.00	Hannan-Quinn criter.	-11.17796
F-statistic	23.18707	Durbin-Watson stat	2.011091
Prob(F-statistic)	0.000000		

Adding 5 lags

Equation: AR6MA1 Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	21.13217	Prob. F(5,2003)	0.0000
Obs*R-squared	100.6670	Prob. Chi-Square(5)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/01/23 Time: 12:03
Sample (adjusted): 1/09/2015 12/30/2022
Included observations: 2009 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000188	2.36E-05	7.963130	0.0000
RESID^2(-1)	0.105128	0.022275	4.719498	0.0000
RESID^2(-2)	0.087572	0.022357	3.917066	0.0001
RESID^2(-3)	0.069739	0.022387	3.115194	0.0019
RESID^2(-4)	0.061953	0.022356	2.771278	0.0056
RESID^2(-5)	0.078543	0.022274	3.526204	0.0004

R-squared	0.050108	Mean dependent var	0.000315
Adjusted R-squared	0.047737	S.D. dependent var	0.000921
S.E. of regression	0.000899	Akaike info criterion	-11.18783
Sum squared resid	0.001618	Schwarz criterion	-11.17109
Log likelihood	11244.17	Hannan-Quinn criter.	-11.18168
F-statistic	21.13217	Durbin-Watson stat	2.008995

In the very first estimation I added three lags, in the next estimation I added four lags and at the last I added five lags. By comparing all the three tables we can conclude that adding four lags is the most efficient for our estimation. In the last table (5 lags) we can see that adding the 5th lag is making the 4th lag insignificant, hence we will not go with adding 5 lags.

We already have the mean equation. Also fitting in the variance equation, we will get:

ht = 0.000204 + 0.110 (ht-1)^2 + 0.093 (ht-2)^2 +0.077(ht-3)^2 + 0.070(ht-4)^2

We can check that the coefficients of the variance term are all positive and adding the coefficients is not exceeding 1.

The coefficient of variance adds up to 0.35. The persistence of volatility is higher the closer it is to 1.

Estimation of GARCH (1,1)

Very often GARCH (1,1) models can be good alternative to high order ARCH model and for volatility these GARCH models are simple and robust.

Equation: UNTITLED Workfile: GOOG::Goog\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 11/01/23 Time: 12:56
Sample (adjusted): 1/12/2015 12/30/2022
Included observations: 2008 after adjustments
Convergence achieved after 28 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/09/2015
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	0.000829	0.000274	3.023483	0.0025
AR(6)	-0.012786	0.024002	-0.532689	0.5942
MA(1)	-0.015085	0.026151	-0.576853	0.5640
Variance Equation				
C	1.40E-05	1.38E-06	10.16838	0.0000
RESID(-1)^2	0.109480	0.006805	16.08771	0.0000
GARCH(-1)	0.851086	0.009833	86.55729	0.0000

R-squared	0.003008	Mean dependent var	0.000636
Adjusted R-squared	0.002014	S.D. dependent var	0.017811
S.E. of regression	0.017793	Akaike info criterion	-5.391578
Sum squared resid	0.634735	Schwarz criterion	-5.374830
Log likelihood	5419.144	Hannan-Quinn criter.	-5.385430

We can clearly see that the p value of ARCH and GARCH terms are less than 0.05 (it is actually zero), and there we will reject the null hypothesis and conclude that they are statistically significant.

Diagnostic testing

In diagnostic checking we will check that, is there any arch effect on the residuals that were estimated from the GARCH model. To test ARCH effect, ARCH-LM test will be used.

$H0$: There is no ARCH effect on the error term

$H0$: There is ARCH effect in the error term

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	0.010670	Prob. F(1,2005)	0.9177
Obs*R-squared	0.010681	Prob. Chi-Square(1)	0.9177

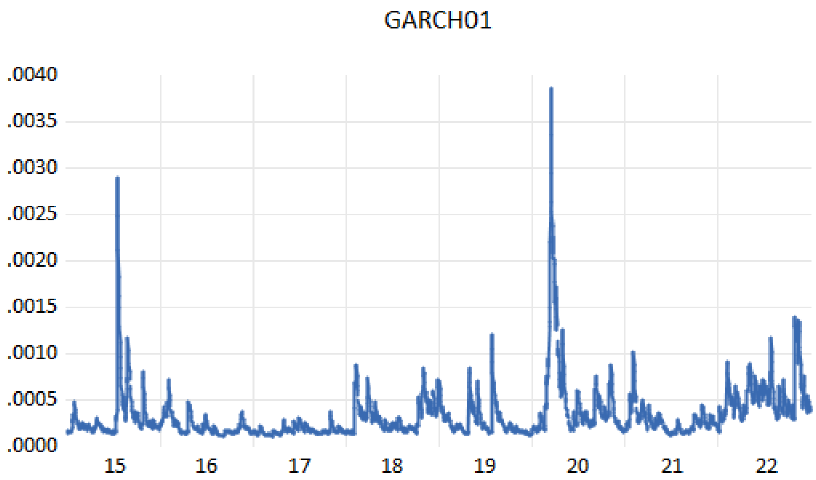
Test Equation:
Dependent Variable: WGT_RESID^2
Method: Least Squares
Date: 11/01/23 Time: 13:11
Sample (adjusted): 1/13/2015 12/30/2022
Included observations: 2007 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.002511	0.070152	14.29064	0.0000
WGT_RESID^2(-1)	-0.002307	0.022333	-0.103298	0.9177

R-squared	0.000005	Mean dependent var	1.000203
Adjusted R-squared	-0.000493	S.D. dependent var	2.978350
S.E. of regression	2.979085	Akaike info criterion	5.022105
Sum squared resid	17794.27	Schwarz criterion	5.027690
Log likelihood	-5037.683	Hannan-Quinn criter.	5.024156
F-statistic	0.010670	Durbin-Watson stat	2.000052
Prob(F-statistic)	0.917737		

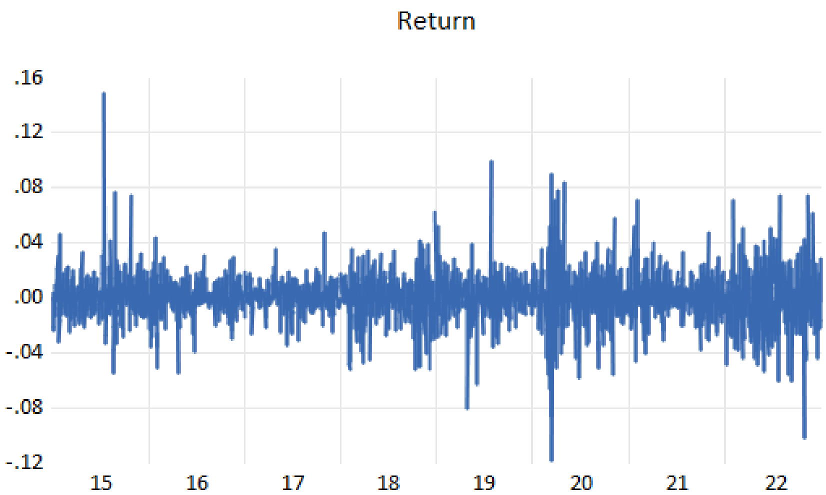
We can check from the table above that p value is greater than 0.05, we will accept the null hypothesis that there is no ARCH effect on the error term. So, our model is correct we need not to further specify the model. Therefore, GARCH (1,1) model mitigates Heteroskedasticity.

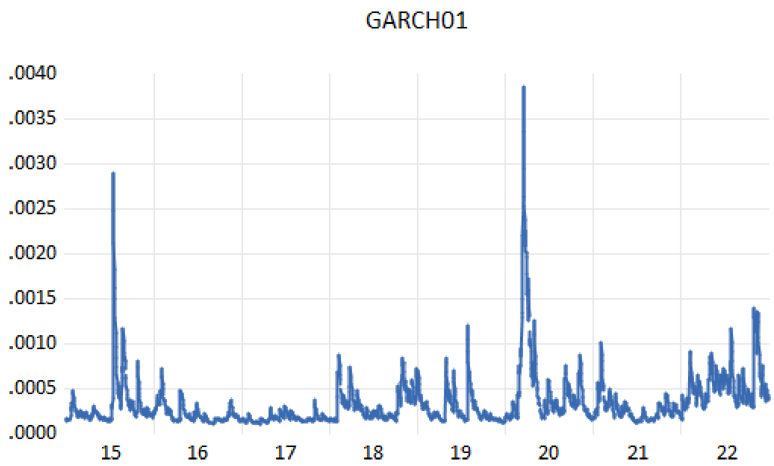
Conditional variance graph



From the above graph we can see that high variance is followed by high variance and low variance is followed by low variance, hence there is volatility clustering

Comparing conditional variance graph and return series





We can clearly see that return series and conditional variance graph are very similar and have spikes in the same time interval

Comment on spikes in the return series and conditional variance graph

In 2015 Chinese government was encouraging investments which made the investment bubble which later burst and effected a lot of economies. Hence investors started selling shares globally as a result China growth slowed down and there was a decrease in price of petroleum. The US corporations are also watching sales and profits decline as overseas business is hurt by the strong dollar, falling oil prices and uncertainty over China's economy. Hence in 2015 U.S. economy growth was very slow and industrial production was negative over 12 months. As compare U.S. the global economy was even weaker.

There are large spikes during 2019-2020, we all know global pandemic hit during this period. The economic impact of the covid -19 in the US has been very disruptive and has severely affected all sectors of the economy. In 2020, the U.S. GDP contracted at a 3.5% annualized rate. It was the biggest contraction since 1946