VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6b-part(a): ARCH\GARCH MODEL

A6b-part(b):VAR\VECM MODEL

DEEPTHI ANNA ALEX

V01101949

Date of Submission: 25-07-2024

CONTENTS

I

Introduction:

In **Part A** of this assignment it is about we are checking the effects of **ARCH\GARCH model** and fitting the model. **ARCH**, or Auto Regressive Conditional Heteroskedasticity, is a statistical model used to analyze and forecast time series data where the variance of the errors is not constant but depends on past values. ARCH model is particularly useful for modeling financial time series data, such as stock returns, which often exhibit volatility clustering—periods of high volatility followed by periods of low volatility.

GARCH, or Generalized Auto Regressive Conditional Heteroskedasticity, extends the ARCH model to include not only past squared errors but also past variances in the volatility equation.The GARCH model improves upon ARCH by incorporating both past error terms and past volatility terms, allowing for a more flexible and comprehensive approach to modeling time-varying volatility.

We are not modelling the actual values but the volatility in ARCH and GARCH.

In **Part B** of this assignment is about **VAR VECM Model**

Vector Autoregression (VAR) Model

The Vector Autoregression (VAR) model is a statistical model used to capture the linear interdependencies among multiple time series. Each variable in a VAR model is expressed as a linear function of past values of itself and past values of all other variables in the system. It is commonly used for analyzing the dynamic impact of random disturbances on the system of variables, making it valuable for forecasting and understanding the relationship between time series data.

Vector Error Correction Model (VECM)

The Vector Error Correction Model (VECM) is an extension of the VAR model that is used when the time series data are cointegrated, meaning they have a long-term equilibrium relationship despite being non-stationary individually. The VECM incorporates error correction terms to adjust the short-term dynamics of the variables to converge towards the long-term equilibrium. This makes the VECM particularly useful for modeling and forecasting long-term relationships in economic and financial time series.

VAR/VECM Workflow

1. Start with Time Series Data (CRUDE_BRENT, MAIZE, SOYABEANS)
2. Unit Root Test

- **Stationary at Level**

- ✓ Proceed with VAR Analysis

- **Not Stationary**

- ✧ Test for Stationarity at First Difference
- ✧ Johansen's Co-Integration Test

I. If Co-Integration Exists:

- ✓ Determine Lag Length
- ✓ Conduct Co-Integration Test
- ✓ Build VECM Model

II. If No Co-Integration:

- ✓ Perform Unrestricted VAR Analysis

3. Post VAR/VECM Analysis

- ✓ Granger's Causality Test
- ✓ Impulse Response Function (IRF) and Variance Decomposition (VD) Analysis

4. Forecasting
5. Output

Results:

PART-A (python)

In ARCH\GARCH we need a variable\ features that resembles the volatility.

Wild fluctuations (volatility) when standard deviation is high, variance is high

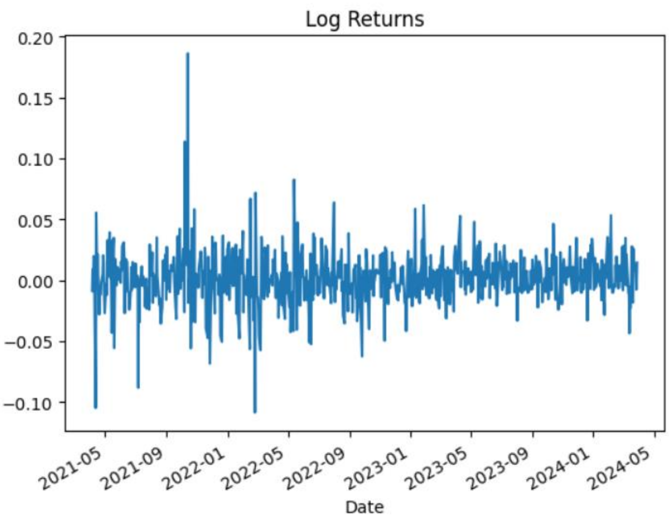
The features that represents the volatility are:

- ❖ Differencing time series magnitude of difference between two values.
- ❖ Log (difference)

In the first part of the assignment I have take the data from **yfinance** of the company **TATA MOTORS.**

Date	Open	High	Low	Close	Adj Close	Volume	Returns
2021-04-05	306.799988	311.700012	297.200012	305.049988	303.166656	66178755	-0.008812
2021-04-06	306.149994	313.799988	304.799988	307.750000	305.850006	63031783	0.008812
2021-04-07	306.750000	310.649994	305.100006	307.799988	305.899689	39073986	0.000162
2021-04-08	307.899994	319.799988	307.500000	313.950012	312.011719	62459774	0.019784
2021-04-09	313.200012	325.000000	312.500000	318.200012	316.235474	75462572	0.013446
...	...	...	...	...	...	...	...
2024-03-21	951.000000	969.250000	946.000000	964.900024	961.931580	11074207	0.025666
2024-03-22	964.900024	986.200012	950.349976	979.799988	976.785706	13638296	0.015324
2024-03-26	977.000000	995.000000	976.000000	986.200012	983.166016	9461531	0.006511
2024-03-27	991.599976	995.000000	976.700012	978.650024	975.639282	6640537	-0.007685
2024-03-28	982.500000	999.900024	979.000000	992.799988	989.745728	9862996	0.014355

'39 rows × 7 columns



Volatility Clustering:

- The plot shows periods where returns are more volatile (large spikes) and periods where they are less volatile (smaller fluctuations).
- High volatility periods are followed by more high volatility, and calm periods by more calm periods.

Spikes:

- Noticeable large spikes (like around early 2022) indicate sudden large changes in returns.

Conclusion:

- The graph displays volatility because there are visible fluctuations and clustering of high and low volatility periods. This pattern is common in stock returns and indicates changing market conditions.

➢ **Returns is the one variable that will resemble the volatility**

```
                      Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                Returns   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                       ARCH   Log-Likelihood:                1783.18
Distribution:                  Normal   AIC:                          -3560.36
Method:            Maximum Likelihood   BIC:                          -3546.54
                                        No. Observations:                  739
Date:                Thu, Jul 25 2024   Df Residuals:                      738
Time:                        10:25:57   Df Model:                            1
                                 Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu         1.5984e-03  7.789e-04      2.052  4.017e-02 [7.172e-05,3.125e-03]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega      4.3083e-04  7.510e-05      5.737  9.658e-09 [2.836e-04,5.780e-04]
alpha[1]       0.1034  8.900e-02      1.162      0.245  [-7.102e-02,  0.278]
============================================================================
```

The model used is constant mean

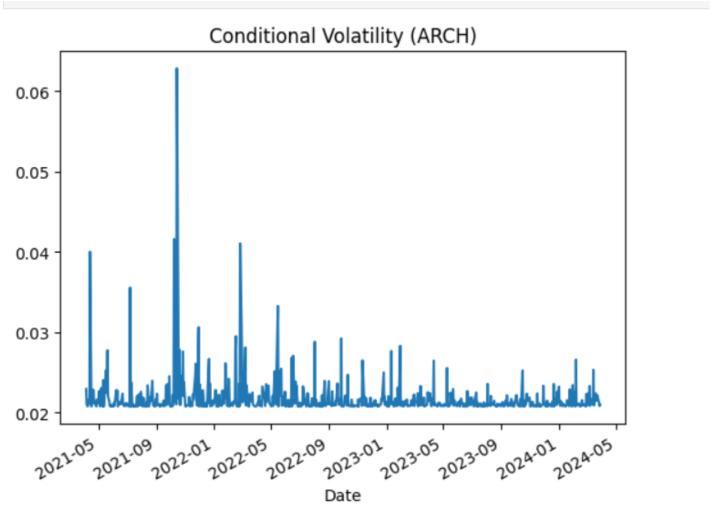
Volume module is ARCH

Distribution is normal

P=1 which is alpha[1]

And p-value= 0.245 this value is greater than 0.05 So, the value 0.245 is not significant.

AIC= -3560.36



1. **High Volatility Periods**:

   The graph shows spikes, especially around late 2021 and early 2022, indicating periods of high volatility.

2. **Low Volatility Periods**:

   There are flatter regions with fewer spikes, indicating periods of lower volatility.

3. **Volatility Clustering**:

   High volatility periods tend to cluster together, as seen with consecutive spikes.

Conclusion:

The graph indicates that volatility varies over time, with noticeable periods of both high and low volatility. This pattern is captured by the ARCH model.

```
                    Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                Returns   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      GARCH   Log-Likelihood:                1824.70
Distribution:                  Normal   AIC:                          -3641.41
Method:            Maximum Likelihood   BIC:                          -3622.99
                                        No. Observations:                  739
Date:                Thu, Jul 25 2024   Df Residuals:                      738
Time:                        10:46:34   Df Model:                            1
                                 Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu         1.5551e-03  7.057e-04      2.204  2.754e-02 [1.720e-04,2.938e-03]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega      4.7572e-05  8.541e-06      5.570  2.553e-08 [3.083e-05,6.431e-05]
alpha[1]       0.1000  5.660e-02      1.767  7.729e-02  [-1.094e-02,  0.211]
beta[1]        0.8000  5.514e-02     14.510  1.053e-47      [  0.692,  0.908]
============================================================================
```

Mean model is constant mean

Volume model is GARCH

Distribution is normal

P=1 which is Alpha[1]

Q=1 which is Beta[1]

- P-value $7.729 \times 10-2 = 0.07729$

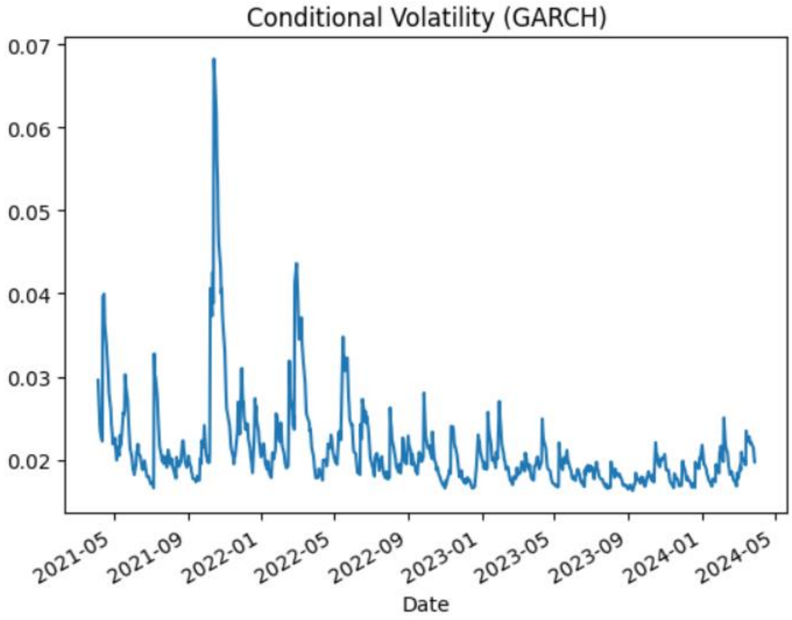
0.07729 is not less than 0.05; it is greater than 0.05.

- Q-Value $1.053 \times 10-47$ is an extremely small number, almost zero.

AIC= -3641.41

Less than 0.05 p or q value would be significant.

- So in ARCH there is only one P-value which is alpha(1) and in GARCH there p-value and q-value which is alpha(1) and beta (1).
- AIC value in GARCH is less than ARCH So lower the value in AIC in ARCH or GARCH is better the model.



The conditional volatility depicted in the graph shows a high level of volatility around early 2022, with a peak exceeding 0.06. After this period, the volatility generally decreases and stabilizes, with minor fluctuations, maintaining lower levels through mid-2024.

```
Iteration:      5,   Func. Count:     34,   Neg. LLF: 1579.001920121877
Optimization terminated successfully    (Exit mode 0)
            Current function value: 1579.000631377724
            Iterations: 9
            Function evaluations: 53
            Gradient evaluations: 9
```

```
                 h.1
Date
2024-03-28  0.175205
                 h.1
Date
2024-03-28  3.381013
                 h.1
Date
2024-03-28  3.381013
```

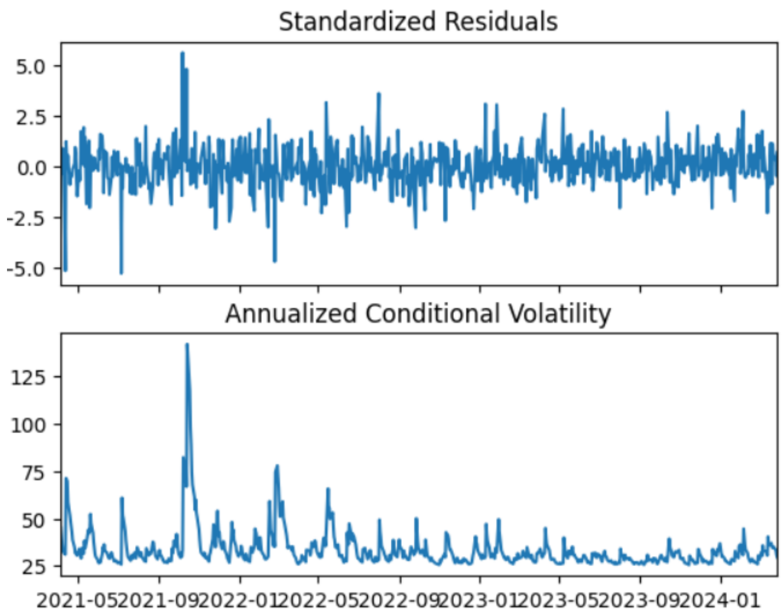
mean= 0.175205 residual variance= 3.381013 variance= 3.381013

3]:

Date	h.01	h.02	h.03	h.04	h.05	h.06	h.07	h.08	h.09	h.10	...	h.81	h.82	h.83	h.84	h.85	h.86
2024-03-28	3.381013	3.613509	3.806741	3.96734	4.100818	4.211754	4.303956	4.380587	4.444276	4.49721	...	4.757719	4.757719	4.757719	4.757719	4.757719	4.757719

1 rows × 90 columns

Volatility values for 3 months



1. **Standardized Residuals (Top Plot):**

   This plot shows the standardized residuals over time. The residuals fluctuate around zero, indicating the difference between observed values and model predictions. The residuals mostly stay within the range of -2.5 to 2.5, suggesting a relatively stable error margin with occasional spikes.

2. **Annualized Conditional Volatility (Bottom Plot):**

   This plot shows the annualized conditional volatility over the same time period. The volatility peaks significantly around early 2022, reaching values above 125, indicating a period of high market uncertainty or turbulence. After the peak, the volatility generally decreases, maintaining lower and more stable levels through mid-2024, with some minor fluctuations.

In summary, the standardized residuals indicate stable model performance with some variability, while the annualized conditional volatility highlights periods of high market uncertainty, particularly around early 2022, followed by a period of relative stability.

PART-B (R-code)

- In part B assignment the data is taken from world bank pink sheet data.
- The VECM and VAR is a multivariate model.
- A multivariate model is a statistical model that analyzes and makes predictions based on multiple dependent variables simultaneously, capturing the relationships among them.
- Machine learning models (time series model into regression)

1. In the pink sheet there is no name for the date column so we

```
# Rename the first column to "Date"
colnames(df)[1] <- 'Date'
# Convert the Date column to Date format
df$Date <- as.Date(paste0(df$Date, "01"), format = "%YM%m%d")
str(df)
```

We use this code to rename the date column.

```
 A tibble: 774 × 7
   date       crude_brent soybeans  gold silver urea_ee_bulk maize
   <date>           <dbl>    <dbl> <dbl>  <dbl>        <dbl> <dbl>
 1 1960-01-01        1.63       94  35.3  0.914         42.2    45
 2 1960-02-01        1.63       91  35.3  0.914         42.2    44
 3 1960-03-01        1.63       92  35.3  0.914         42.2    45
 4 1960-04-01        1.63       93  35.3  0.914         42.2    45
 5 1960-05-01        1.63       93  35.3  0.914         42.2    48
 6 1960-06-01        1.63       91  35.3  0.914         42.2    47
 7 1960-07-01        1.63       92  35.3  0.914         42.2    47
 8 1960-08-01        1.63       93  35.3  0.914         42.2    47
 9 1960-09-01        1.63       92  35.3  0.914         42.2    46
10 1960-10-01        1.63       88  35.3  0.914         42.2    42
```

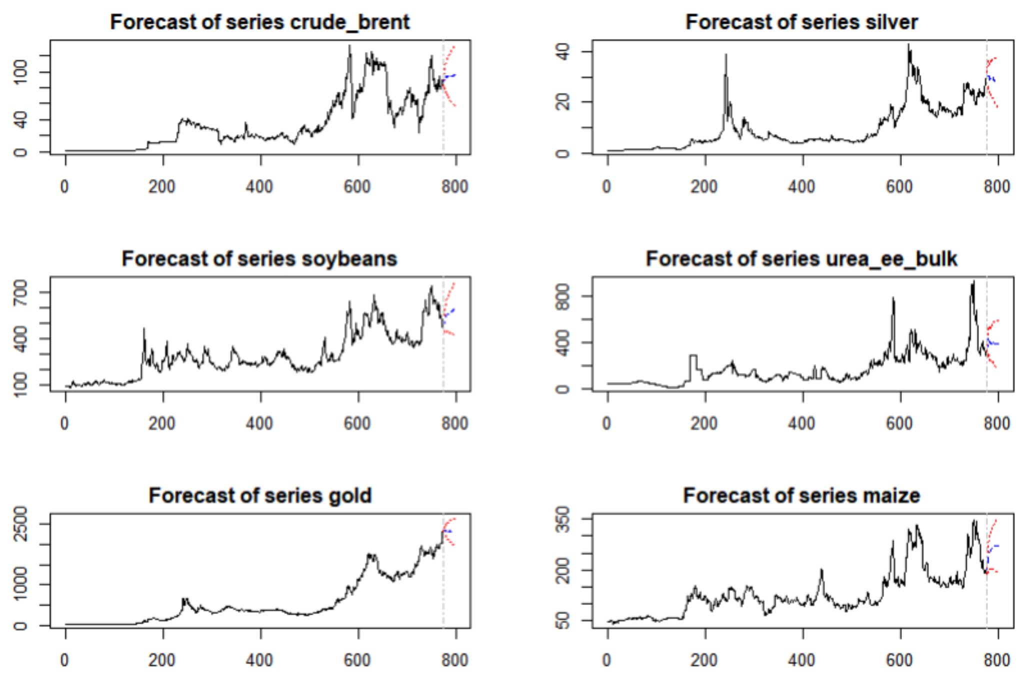
these are the variables that is been selected for the VAR VECM model

	test	10pct	5pct	1pct
r <= 5 |	8.11	7.52	9.24	12.97
r <= 4 |	17.42	13.75	15.67	20.20
r <= 3 |	20.22	19.77	22.00	26.81
r <= 2 |	29.12	25.56	28.14	33.24
r <= 1 |	45.32	31.66	34.40	39.79
r = 0 |	72.13	37.45	40.30	46.82

the **r=3** because in 5pct 22.00 and the test value is 20.22 in which when we compare the test value is less than 5pct so we take r=3.

It shows the significance.

If r is greater than 0 then it is VECM Modelling or else it is VAR Modelling



In this graph there is a blue line shown it is the **actual forecast .**

1. **Crude Brent**
2. **Silver**
3. **Soybeans**
4. **Urea EE Bulk**
5. **Gold**

6. **Maize**

Each subplot shows:

- Historical data of the respective series.
- A forecasted section marked with dashed lines or different colors to indicate predicted values.

Key Observations:

- Each series shows a trend with historical prices and their respective forecasts.
- Forecasts generally show an upward trend in the near future for these commodities.
- The plots may have confidence intervals around the forecasted values, though this isn't clearly marked.

VECM vs. VAR Modelling:

- **VECM (Vector Error Correction Model)** is used if there is cointegration among the variables. This means if the rank r of the cointegration matrix is greater than 0.
- **VAR (Vector Autoregression)** is used when there is no cointegration among the variables. This means if the rank r of the cointegration matrix is 0.

Interpretation:

- If the rank r is greater than 0, indicating cointegration, VECM modelling is applied.
- If the rank r is 0, indicating no cointegration, VAR modelling is applied.

This context helps in deciding the appropriate modelling technique based on the relationships and characteristics of the data series involved.