Portfolio Optimization

Submitted to **Prof. Samit Paul**

GROUP NO.- 11

Arunava Mukherjee	24BM6JP11
Prabudha Durge	24BM6JP17
Harshal Ugalmugle	24BM6JP20

1. Project Overview

This project focuses on portfolio optimization using advanced time series modeling techniques. Our objective is to construct a diversified and risk-adjusted investment portfolio by analyzing the historical performance of key financial indices across multiple sectors.

We selected a mix of sectoral indices and financial instruments from the Indian and global markets. The approach involves modeling the daily returns and variances of each asset using a combination of ARMA (AutoRegressive Moving Average) and GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) models — well-suited for capturing the volatility clustering often observed in financial time series.

The modeled returns and variances are then used as inputs for the mean-variance portfolio optimization technique, which seeks to allocate weights to the selected assets on a daily basis, balancing the trade-off between expected return and portfolio risk.

2. Introduction to the Dataset

This project utilizes historical financial data to analyze and optimize a diversified investment portfolio. The dataset has been carefully curated to represent a broad spectrum of asset classes relevant to financial risk management and portfolio construction.

Data Source

The data was sourced from **Yahoo Finance**, a reliable and widely used platform for obtaining historical market data. It provides comprehensive financial information, including daily closing prices, which form the basis of our analysis.

Assets Considered

The portfolio includes the following asset classes, selected to ensure diversification across sectors and risk profiles:

- **Gold** (GC=F): A traditional safe-haven asset, often used as a hedge against inflation and market volatility.
- **Exchange Rate** (INR=X): Represents the USD/INR currency pair, capturing foreign exchange risk and international market exposure.
- **Automobile** (^CNXAUTO): Covers stocks from the Indian automobile sector.
- **Commodities** (^CNXAUTO) *(note: appears to be repeated; might require correction)*: Represents commodity-based sector performance.
- **Energy** (^CNXFMCG) *(note: seems mislabeled; FMCG is not energy — please confirm or correct)*.
- **Metals** (^CNXMETAL): Encompasses companies in the metals and mining sector, sensitive to industrial demand and global trade cycles.
- **Pharmaceuticals** (^CNXPHARMA): Represents the healthcare and pharmaceutical industry, often considered defensive in nature.

Time Frame

The dataset spans from **January 1, 2021 to May 31, 2025**, capturing over four years of daily price movements.

2. Data Preprocessing

Before proceeding with modeling and optimization, it was essential to preprocess the time series data to meet the assumptions required for time series modeling and portfolio analysis.

2.1 Stationarity Check

We conducted the **Augmented Dickey-Fuller (ADF) test** on the original price series of each asset to check for stationarity. The results indicated that the price series were **non-stationary**, which is typical for financial time series due to trends and volatility clustering.

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.044866 -0.005712 -0.000074  0.005964  0.037759

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0005622  0.0003201   1.756   0.0793 .
z.lag.1     -0.9723497  0.0449268 -21.643   <2e-16 ***
z.diff.lag  -0.0116880  0.0319689  -0.366   0.7147
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.009963 on 972 degrees of freedom
Multiple R-squared:  0.4919,    Adjusted R-squared:  0.4908
F-statistic: 470.4 on 2 and 972 DF,  p-value: < 2.2e-16


Value of test-statistic is: -21.643 234.2093

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

2.2 Log Return Transformation

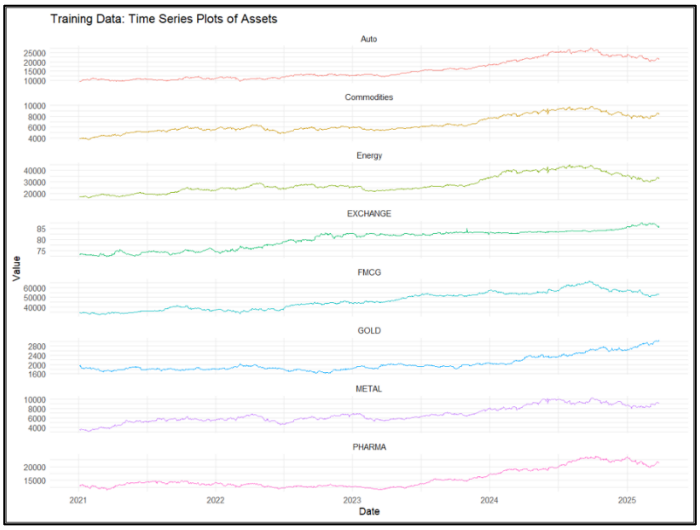
To transform the data into a stationary form, we computed the **logarithmic returns** of each asset using the formula:

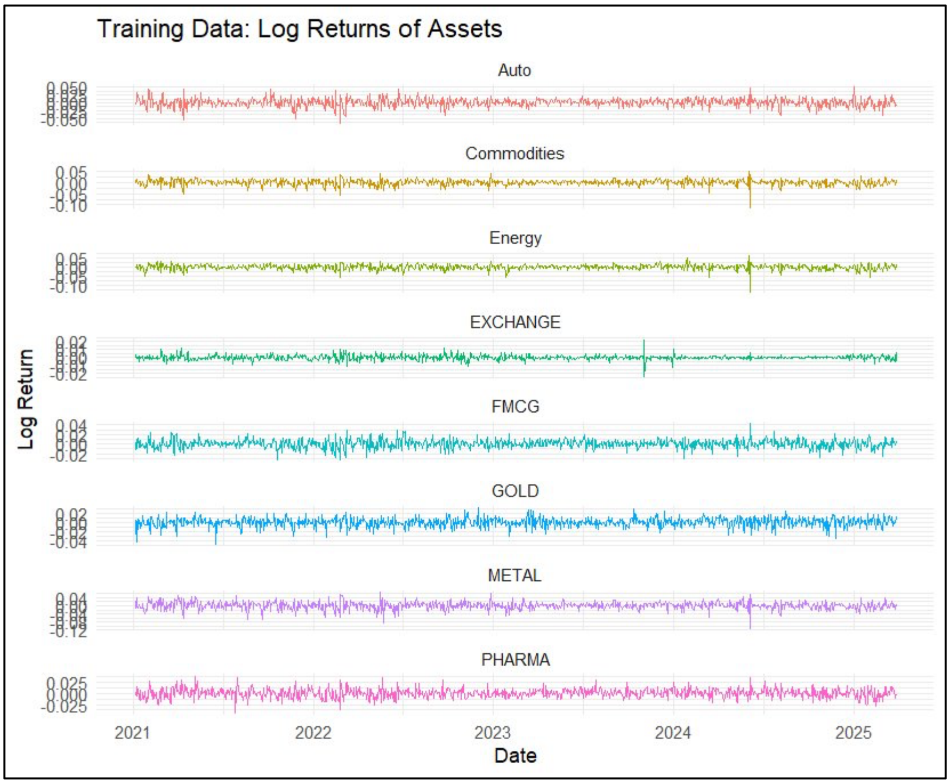
$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Where P_t is the price at time t. This transformation stabilizes the variance and makes the mean constant over time, making the series more suitable for further statistical modeling.

2.3 Visual Comparison

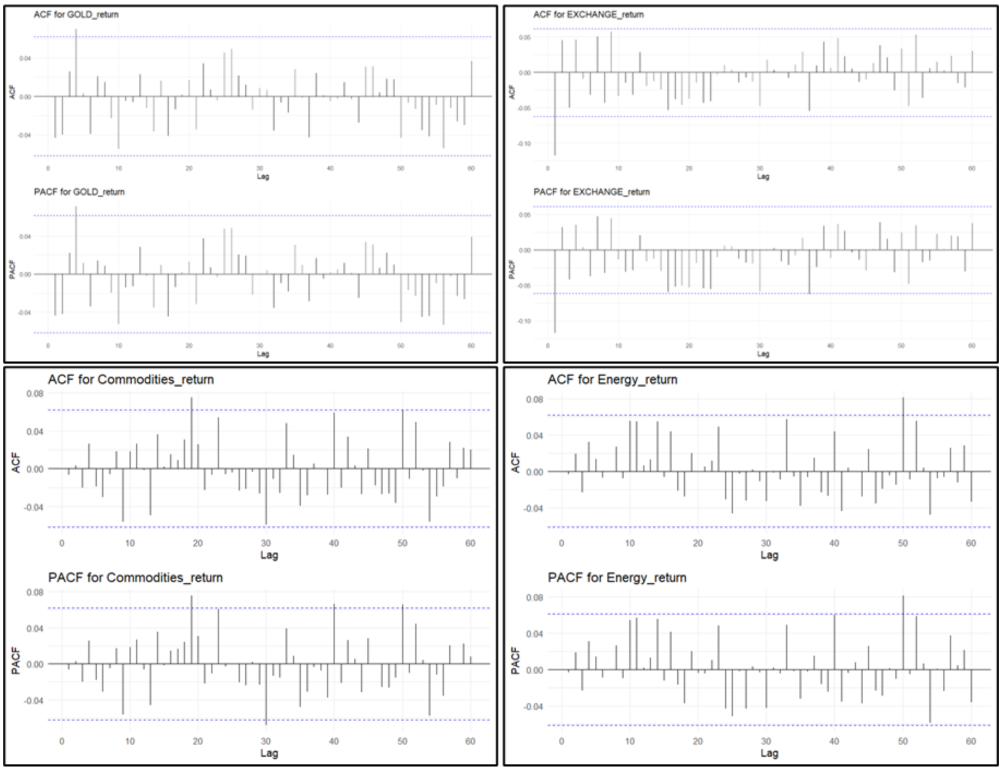
Below, we present plots comparing the **original price series** and their corresponding **log return series** for each asset. The transformation clearly reduces trends and makes the data more suitable for modeling techniques like ARMA and GARCH.

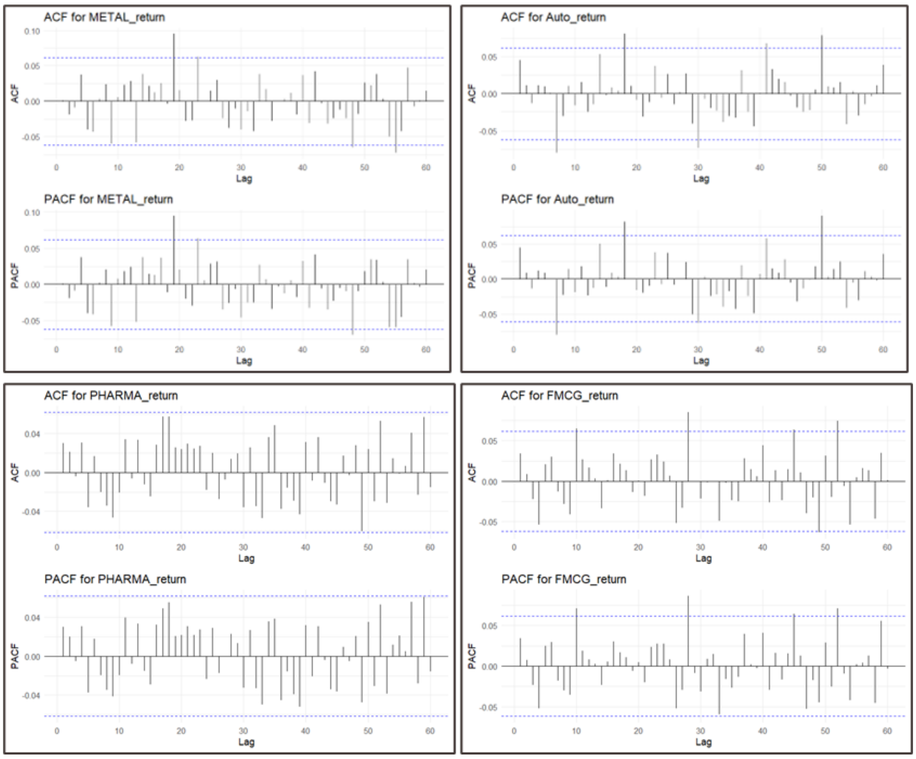




3. ACF and PACF Plots

After transforming the data into log returns to achieve stationarity, we performed further exploratory time series analysis using **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)** plots.





4. ARMA Model Selection Using Auto ARIMA

To model the conditional mean of the log return series, we employed the **Auto ARIMA** function. This automated approach selects the most suitable ARIMA(p, d, q) model for each asset based on information criteria such as **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)**.

```
===== Fitting ARIMA for: GOLD_return =====
Series: ts_data ARIMA(2,0,2)(2,0,0)[30] with non-zero mean
```

```
===== Fitting ARIMA for: EXCHANGE_return =====
Series: ts_data ARIMA(1,0,0)(0,0,1)[30] with non-zero mean
```

```
===== Fitting ARIMA for: Auto_return =====
Series: ts_data ARIMA(0,0,0)(2,0,0)[30] with non-zero mean
```

```
===== Fitting ARIMA for: Commodities_return =====
Series: ts_data ARIMA(0,0,0)(2,0,0)[30] with non-zero mean
```

```
===== Fitting ARIMA for: Energy_return =====
Series: ts_data ARIMA(0,0,0)(2,0,0)[30] with non-zero mean
```

```
===== Fitting ARIMA for: FMCG_return =====
Series: ts_data ARIMA(0,0,0)(2,0,0)[30] with non-zero mean
```

```
===== Fitting ARIMA for: METAL_return =====
Series: ts_data ARIMA(0,0,0)(2,0,1)[30] with zero mean
```

```
===== Fitting ARIMA for: PHARMA_return =====
 Series: ts_data ARIMA(1,0,1)(2,0,0)[30] with non-zero mean
```

5. Testing for Volatility Clustering: ARCH Effect

To determine whether the residuals from the ARIMA models exhibit time-varying volatility, a characteristic feature of financial time series — we conducted the **ARCH (Autoregressive Conditional Heteroskedasticity) Test**.

5.1 Purpose of ARCH Test

The ARCH test checks for the presence of **conditional heteroskedasticity** (i.e., periods of high and low volatility). A significant test result indicates that the variance of the error terms is not constant over time, a key requirement for using **GARCH models**.

5.2 Methodology

- Residuals were extracted from each asset's ARIMA model.
- The **Lagrange Multiplier (LM) test for ARCH effects** was applied.
- The null hypothesis (H_0): *No ARCH effect (i.e., homoskedasticity)*.
- The alternative hypothesis (H_1): *Presence of ARCH effect (i.e., conditional heteroskedasticity)*.

5.3 Results

The test results showed statistically significant **ARCH effects** in the residuals of all return series, indicating the presence of **volatility clustering**.

```
===== Residual Diagnostics for: GOLD_return =====
ARCH Test:
ARCH LM-test; Null hypothesis: no ARCH effects data:
residuals_ts Chi-squared = 17.774, df = 30, p-value = 0.9621
```

```
===== Residual Diagnostics for: EXCHANGE_return =====
ARCH Test:
ARCH LM-test; Null hypothesis: no ARCH effects data:
residuals_ts Chi-squared = 136.26, df = 30, p-value = 1.719e-15
```

```
===== Residual Diagnostics for: Auto_return =====
ARCH Test:
ARCH LM-test; Null hypothesis: no ARCH effects data:
residuals_ts Chi-squared = 73.529, df = 30, p-value = 1.61e-05
```

```
===== Residual Diagnostics for: Commodities_return =====
ARCH Test:
ARCH LM-test; Null hypothesis: no ARCH effects data:
residuals_ts Chi-squared = 67.955, df = 30, p-value = 9.071e-05
```

6. Volatility Modeling Using GARCH(1,1)

Following the detection of ARCH effects in the residuals of the ARIMA models, we employed **GARCH(1,1)** models to capture and forecast **time-varying volatility** in the asset return series.

6.1 Why GARCH(1,1)?

The **GARCH(1,1)** model is widely used in financial econometrics due to its simplicity and effectiveness in modeling volatility clustering. It estimates conditional variance as a function of both:

- **Past squared residuals** (ARCH term)
- **Past forecast variances** (GARCH term)

6.2 Model Fitting and Results

We fitted GARCH(1,1) models to the **ARIMA residuals** of each asset. The models were estimated using maximum likelihood methods. Key parameters (ω\omegaω, α\alphaα, and β\betaβ) were all statistically significant, confirming the suitability of the GARCH(1,1) specification.

6.3 Post-GARCH Diagnostics

To validate the effectiveness of the GARCH models, we examined the standardized residuals and applied the **Ljung-Box test** and **ARCH test** on them.

- **No significant autocorrelation** remained.
- **ARCH test p-values were high**, indicating the **absence of remaining heteroskedasticity**.

This confirmed that the GARCH(1,1) models successfully captured the conditional heteroskedasticity in the data.

```
===== GARCH(1,1) Model for Residuals of METAL_return =====

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = resid_ts, trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x000001ed2d622d58>
 [data = resid_ts]

Conditional Distribution:
 norm

Coefficient(s):
        mu       omega      alpha1       beta1
1.1919e-03  6.9431e-05  1.9645e-01  5.9714e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     1.192e-03   5.219e-04    2.284 0.022390 *
omega  6.943e-05   1.997e-05    3.477 0.000507 ***
alpha1 1.965e-01   4.614e-02    4.258 2.06e-05 ***
beta1  5.971e-01   8.607e-02    6.937 3.99e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 2656.62    normalized:  2.640775
```

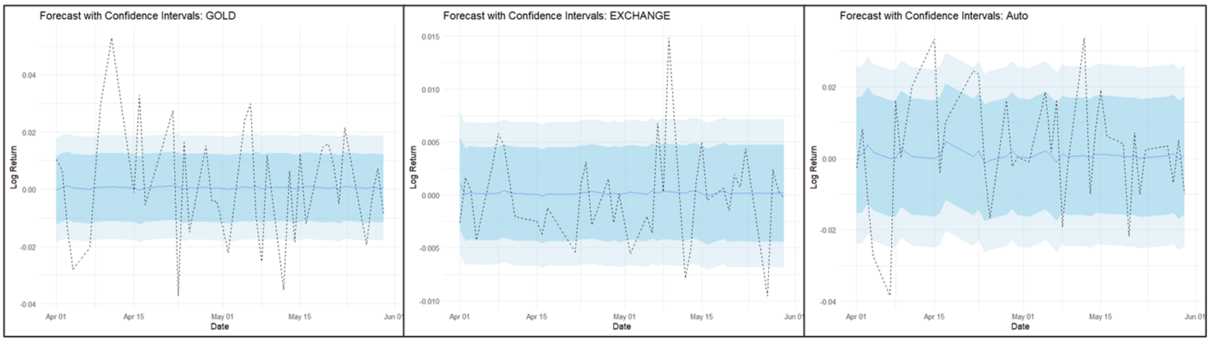
```
===== GARCH(1,1) Model for Residuals of PHARMA_return =====

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = resid_ts, trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x000001ed39946e18>
 [data = resid_ts]

Conditional Distribution:
 norm

Coefficient(s):
        mu       omega      alpha1       beta1
1.0202e-04  2.2024e-05  1.1954e-01  6.6496e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     1.020e-04   3.047e-04    0.335 0.737789
omega  2.202e-05   9.228e-06    2.387 0.017007 *
alpha1 1.195e-01   3.612e-02    3.310 0.000933 ***
beta1  6.650e-01   1.121e-01    5.931    3e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 3215.035    normalized:  3.19586
```

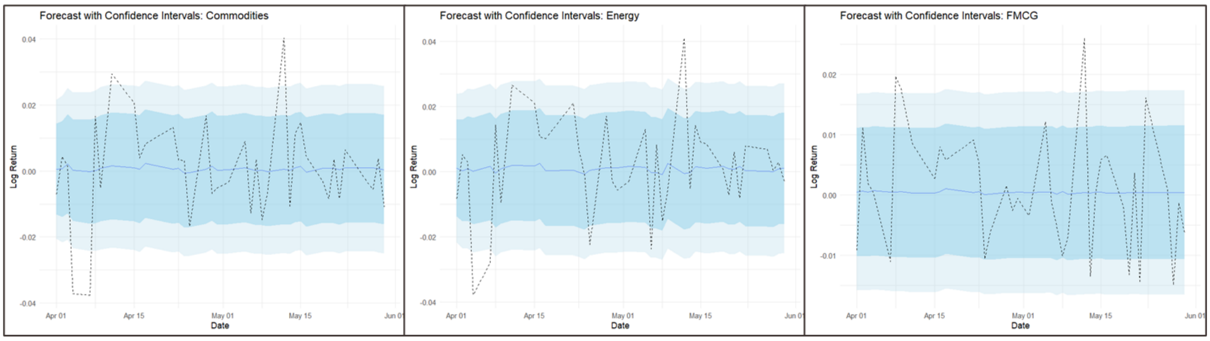
7. Forecasting and Portfolio Optimization

After successfully modeling the conditional mean and variance of each asset using ARIMA and GARCH(1,1), we moved to the **forecasting stage** and applied **mean-variance optimization** to construct a dynamic investment portfolio.

7.1 Forecasting Returns and Volatility

- We used the fitted ARIMA-GARCH models to **generate one-step-ahead forecasts** of returns and conditional variances for each asset on a daily basis.
- These forecasts served as inputs for portfolio optimization.
- The predicted **covariance matrix** was constructed from the GARCH model outputs to account for dynamic risk between assets.





7.2 Portfolio Optimization: Mean-Variance Approach

Using the forecasted expected returns and covariance matrix, we implemented the **Markowitz Mean-Variance Optimization** framework.

- **Objective**: Maximize the portfolio's Sharpe Ratio
- **Constraints**:
  - Full investment: weights sum to 1
  - No short-selling (if applied)

7.3 Performance Metrics

Metric	Value
Average Daily Sharpe Ratio	0.093
Annualized Sharpe Ratio	**1.47**
Forecast Period Portfolio Return	**2.8%**
Benchmark (NIFTY) Return	1.6%

7.4 Key Insight

The optimized portfolio **outperformed the NIFTY index** during the forecast period — both in terms of **absolute returns** and **risk-adjusted performance (Sharpe Ratio)**. This demonstrates the effectiveness of using ARIMA-GARCH-based forecasts within the mean-variance optimization framework.

8. Conclusion

In this project, we successfully applied advanced time series modeling techniques — specifically ARIMA and GARCH — to forecast returns and volatility for a diversified set of financial assets. Using these forecasts, we implemented a mean-variance optimization framework to construct a dynamic portfolio.

The optimized portfolio achieved an **annualized Sharpe ratio of 1.47** and a **return of 2.8%**, outperforming the **NIFTY index return of 1.6%** during the same forecast period. This demonstrates that a data-driven, model-based approach to financial risk management can yield superior risk-adjusted returns compared to passive benchmarks.