VIETNAM NATIONAL UNIVERSITY – HO CHI MINH CITY
INTERNATIONAL UNIVERSITY
DEPARTMENT OF MATHEMATICS

FINANCIAL RISK MANAGEMENT 2, PROJECT REPORT

STATISTICAL ANALYSIS AND ARCH/GARCH ESTIMATION AND FORECASTING OF VALUE–AT–RISK ON HPG STOCK

Lecturer: Dr. Ta Quoc Bao

Group Members: Tran Viet Hang - MAMAIU18079
Nguyen Ha Uyen - MAMAIU19017
Nguyen Thi Bao Thy - MAMAIU18039
Nguyen Thi Cam Thanh - MAMAIU18073
Van Thi Hong Nhung - MAMAIU18069

Ho Chi Minh City, Vietnam
May, 2022

Contents

List of Figures

Chapter 1

Introduction

Value at risk (VaR) is a financial metric that quantifies the extent of possible financial losses within a firm, portfolio, or position over a specific time frame. This metric is most commonly used by investment and commercial banks to determine the extent and probabilities of potential losses in their institutional portfolios. For example, a financial firm may compute that an asset has a 5% one-month VaR of 3%, indicating that there is a 5% risk that the asset would lose value by 3% during the next month.

In this project, our group has made some statistical analyses on HPG stock. Our time series data consists of stock returns for 1248 days. We estimated Value-at-risk of the returns by using Autoregressive Conditional Heteroskedasticity (ARCH) Model and Generalized ARCH (GARCH) method, then compared each method's result and drew some conclusions. Finally, we performed the forecasting of VaR (5%) for the next 5 days.

By submitting this project report, we undersigned solemnly declare that the project report 'STATISTICAL ANALYSIS AND ARCH/GARCH ESTIMATION AND FORECASTING OF VALUE−AT−RISK ON HPG STOCK' is based on our own work carried out during the course 'Financial Risk Management 2' under the supervision of Dr. Ta Quoc Bao.

Chapter 2

Theoretical Background

2.1 Prices & Returns

Simple returns are computed as follows

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

To find the total return over a period **T** with **n** sub-periods, we need to compound the growth for each sub-period, as shown here:

$$P_f = P_0(1 + R_1)(1 + R_2)...(1 + R_n)$$

in which

- P_f: Final price
- P_0: Initial price
- R_x: Return for each sub-period

If each return R_x were identical, we could write this in a simplified exponential form:

$$P_f = P_0(1 + R_n)^n$$

The return in each sub-period is rarely identical, so compounding in this manner is impossible. For this reason, we turn our attention to logarithmic returns. Logarithmic returns measure the rate of exponential growth. Instead of measuring the percent of price change for each sub-period, we measure the exponent of its natural growth during that time. Later, we can add each sub-period's exponential growth to get the total growth for the period T.

The log returns (or continuously compounded returns) are given by the natural logarithm of simple returns,

$$y_t = \ln(1 + R_t)$$

and the $n-$period returns are consequently given by

$$R_t(n) = \left[\prod_{i=0}^{n}(1 + R_{t-i})\right] - 1 \quad \text{and} \quad r_t(n) = \ln(1 + R_t(n)).$$

While simple returns are commonly used in accounting, log returns are preferred in modelling and derivative pricing because of the two following differences:

- Log returns are symmetric while simple returns are not.
- Simple returns are not additive over time. Adding returns for multiple periods does not yield the total return over the total length of time. In other words, simple returns don't satisfy the property of time consistency. However, logarithmic returns are additive over time.

2.2 Statistical Inferences

- Skewness & kurtosis for a sample of size N from a population X :

$$\hat{s_3} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^3 \quad \text{and} \quad \hat{s_4} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^4;$$

  - **Skewness** can be quantified as a representation of the extent to which a given distribution varies from a normal distribution.
  - Skewness is used along with **Kurtosis** to better judge the likelihood of events falling in the tails of a probability distribution. Investors note right-skewness when judging a return distribution because it, like **excess kurtosis**, better represents the extremes of the data set rather than focusing solely on the average.

$$\kappa = Excess\ Kurtosis = Kurtosis - 3$$

- The Jarque-Bera Test compares the sample skewness and kurtosis of the data to 0 and 3, for normality testing;

$$JB = \frac{n}{6} \cdot \left(\hat{s_3}^2 + \frac{(\hat{s_4} - 3)^2}{4}\right) \sim \chi_2^2$$

- The Q-Q plot compares the data distribution against a reference theoretical one, and can also be used for normality testing.

2.3 Volatility Modeling

Measures of volatility are used in many important financial and economic models. It is a critical metric in the Value at risk calculation and is a one of the key variables in options pricing and asset allocation models. There are many different types of volatility models, including the autoregressive conditional heteroscedastic (ARCH) model of Engle (1982), the generalised autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986).

2.3.1 Autoregressive Conditional Heteroskedasticity (ARCH) Model

The basic idea of ARCH models is that (*i*) the shock to an asset return is serially uncorrelated, but dependent, and (*ii*) the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH(q) model assumes that,

$$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha_1 a_{t-1}^2 + ... + \alpha_q a_{t-q}^2 \end{aligned} \tag{2.1}$$

where ϵ_t is the residual, $\omega > 0, \alpha_i \geq 0$ for $i > 0$. The most common specification is ARCH(1):

$$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha_1 a_{t-1}^2 \end{aligned} \tag{2.2}$$

2.3.2 Generalised ARCH (GARCH) Model

Although the ARCH model has a simple functional form, it often requires many parameters to adequately describe the volatility process of a financial asset. To simplify the model, Bollerslev (1986) proposed a useful extension that lead to the development of the generalised ARCH (GARCH) model. Once again, where the mean equation in the time series y_t, may be described by $a_t = y_t - \mu_t$. Then we can show that, a_t may be described by a GARCH(p, q) model if

$$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{2.3}$$

where ϵ_t is the residual, $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$ and $\sum_{i,j=1}^{q,p} (\alpha_i + \beta_j) < 1$. The most common specification is GARCH(1,1):

$$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{2.4}$$

2.3.3 Volatility Forecasting

Forecasts from the ARCH model in equation (2.1) can be obtained recursively. Consider an ARCH(q) model at point t, where the 1-step ahead forecast of σ_t^2 is

$$\sigma_{t,1}^2 = \hat{\sigma}_{t+1}^2 = \omega + \alpha_1 a_t^2 + ... + \alpha_m a_{t-q+1}^2.$$

The 2-step ahead forecast is then

$$\sigma_{t,2}^2 = \mathbb{E}(\sigma_{t+2}^2) = \mathbb{E}(\omega + \alpha_1 a_{t+1} + \alpha_2 a_t + ... + \alpha_q a_{t-q+2}^2) = \omega + \alpha_1 \sigma_{t,1}^2 + \alpha_2 a_t^2 ... + \alpha_q a_{t+2-q}^2$$

and inductively, the $k-$step ahead forecast is

$$\sigma_{t,k}^2 = \omega + \sum_{i=1}^{k} \alpha_i \sigma_{t,k-i}^2 + \sum_{m=1}^{q-k} \alpha_{k+m} a_{t-m}^2.$$

Similarly for the GARCH(p,q) model, the $k-$step ahead forecast is

$$\sigma_{t,k}^2 = \omega + \sum_{i=1}^{k} \alpha_i \sigma_{t,k-i}^2 + \sum_{m=1}^{q-k} \alpha_{k+m} a_{t-m}^2 + \sum_{j=1}^{k} \beta_i \sigma_{t,k-i}^2 + \sum_{n=1}^{p-k} \beta_{k+n} \sigma_{t-n}^2.$$

2.4 Value-at-Risk (VaR)

Given some confidence $\alpha \in (0,1)$, the value-at-risk (VaR) of a portfolio with loss L at this confidence level is

$$VaR_\alpha(L) = \inf\{x : F_L(x) \geq \alpha\}$$

i.e. the smallest number x such that the probability that the loss exceeds x is no larger than $1-\alpha$. In words, VaR is simply a quantile of the loss distribution. If L follows the distribution $\mathcal{N}(\mu, \sigma^2)$, then its VaR is given by

$$VaR_\alpha(L) = \mu + \sigma \cdot \Phi^{-1}(\alpha).$$

For a process $\{y_t\}$, the estimated and $k-$day forecasted VaR at time t is given by

$$V\hat{a}R_\alpha(y_t) = \overline{y} + \hat{\sigma}_t \cdot \Phi^{-1}(\alpha)$$

and

$$VaR_\alpha(y_{t,k}) = \overline{y} + \sigma_{t,k} \cdot \Phi^{-1}(\alpha).$$

Chapter 3

Empirical Results

3.1 Statistical Analysis

Data of stock HPG price from 03/01/2017 to 31/12/2021 is taken into consideration.

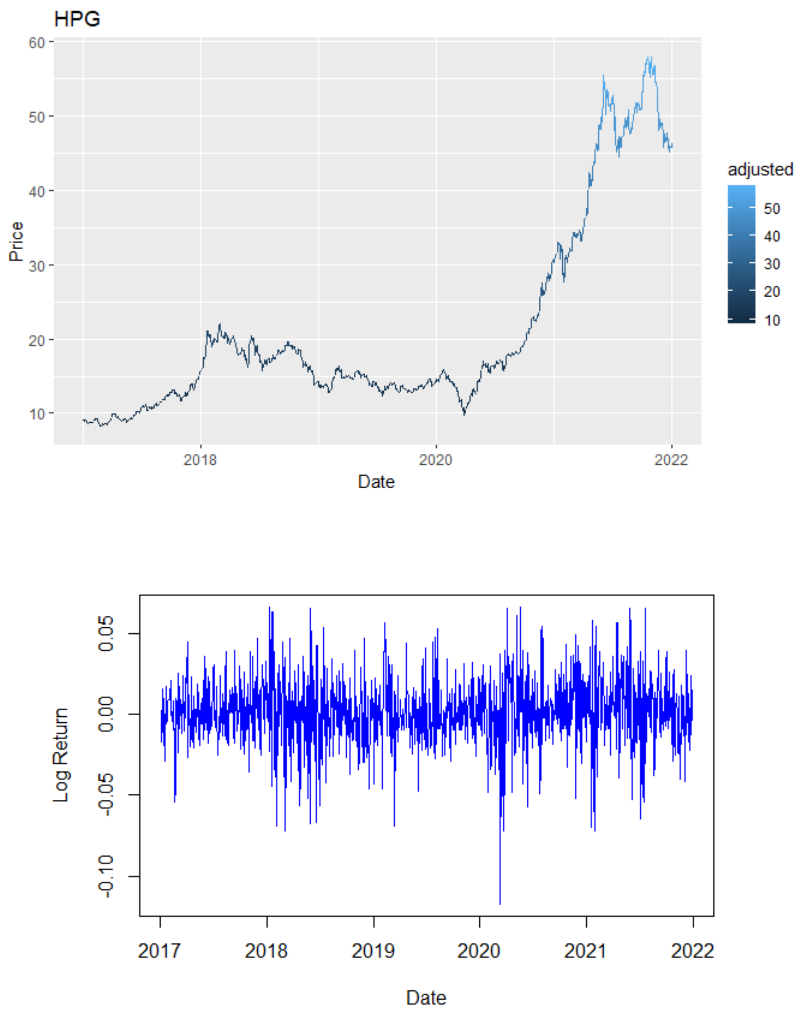


Figure 3.1: HPG Daily Adjusted Price & Log Return 03/01/2017 − 31/12/2021

The price chart shows that the price of HPG stock witnessed a slight decrease from 2019 to 2020, and then significantly increased toward 2022; in the short term, the stock price seems to fluctuate. Obviously, volatility clustering for the returns appears to resemble a non-stationary behavior, which implies that there is no applicable linear time series model for this data.

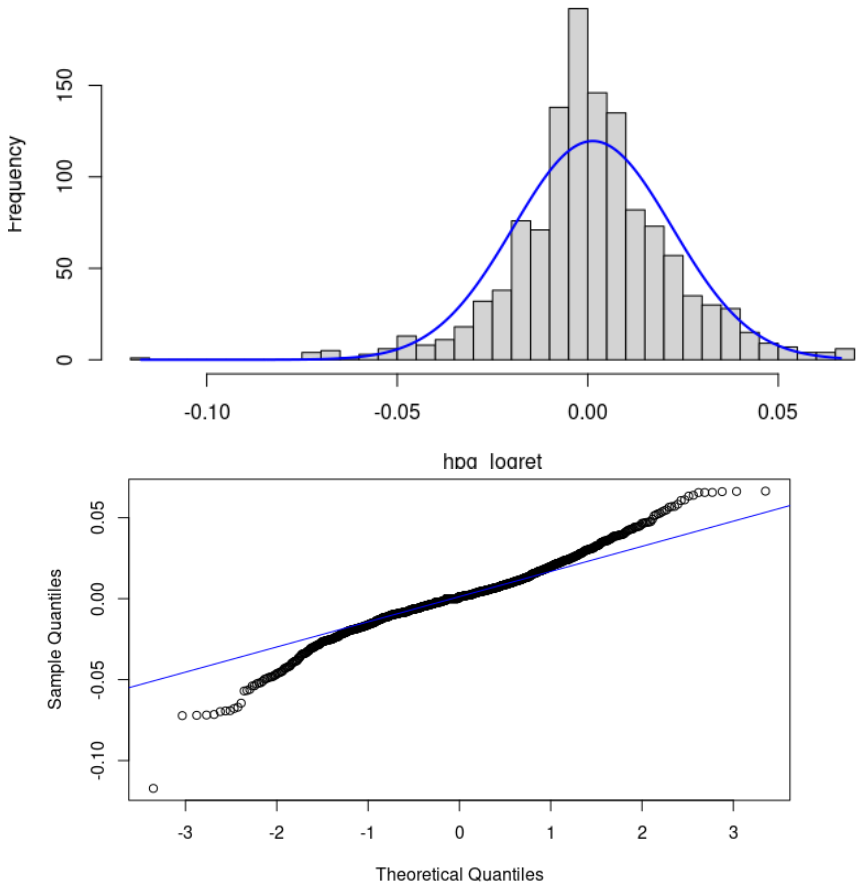


Figure 3.2: Histogram & Q-Q Plot of Log Returns

Observing the histogram and Q-Q plot shows that the log return distribution is too peaked in the middle and have a heavy left tail. The skewness is -0.2109911, which is slightly negative (left-skewed), this indicates that small gains are the norm, but large losses can occur, carrying the risk of bankruptcy. The excess Kurtosis κ is 2.026686, which is positive and hence a leptokurtic, it shows that this investment returns may be prone to extreme values. Therefore, an investment whose returns follow a leptokurtic distribution is considered to be risky. It means that big losses (as well as big gains) can occur. The Jarque-Bera test statistic is 222.85, which indicates that the log return is clearly not normally distributed.

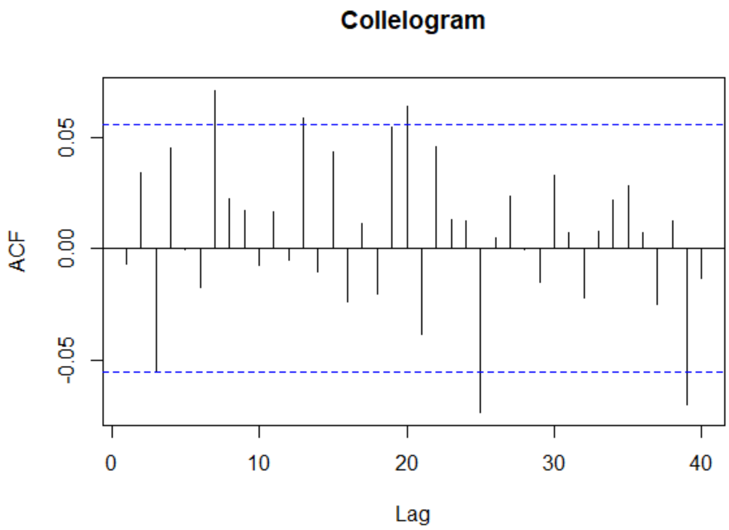


Figure 3.3: Autocorrelation Function (ACF) of Log Returns

The ACF at lags 3, 25 and 39 are significant, implying that the log returns indeed do not follow a stationary process.

3.2 ARCH Estimations for VaR

We use the ARCH(1) model in equation (2.2) for data-fitting.

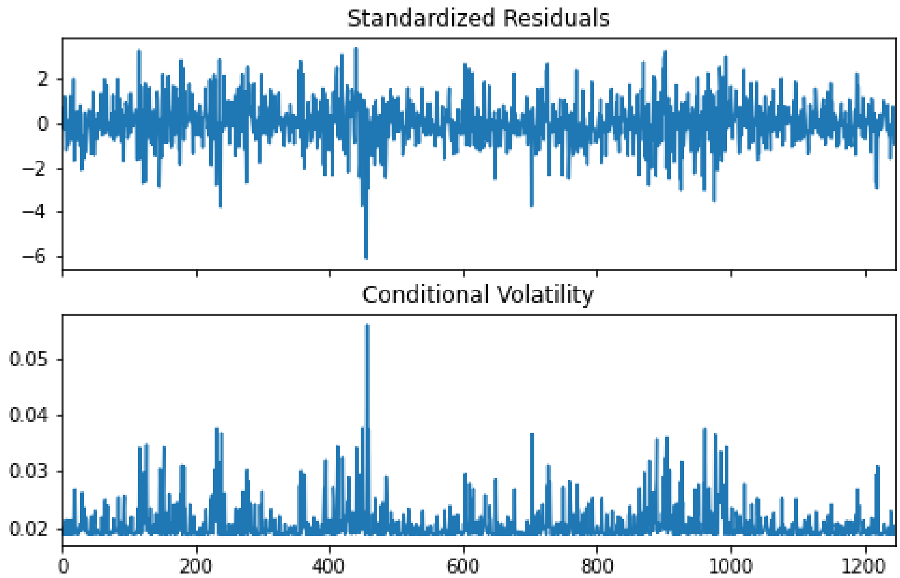
```
                      Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                       ARCH   Log-Likelihood:                3084.27
Distribution:                  Normal   AIC:                          -6162.53
Method:            Maximum Likelihood   BIC:                          -6147.14
                                        No. Observations:                 1248
Date:                Thu, May 19 2022   Df Residuals:                     1247
Time:                        15:48:32   Df Model:                            1
                                 Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu         1.3395e-03  5.653e-04      2.369  1.782e-02 [2.315e-04,2.448e-03]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega      3.5074e-04  2.643e-05     13.269  3.486e-40 [2.989e-04,4.025e-04]
alpha[1]       0.1975  4.836e-02      4.083  4.441e-05     [  0.103,  0.292]
============================================================================

Covariance estimator: robust
```

The estimated ARCH(1) model is

$$a_t = \sigma_t \epsilon_t,$$
$$\sigma_t^2 = 3.5074 \cdot 10^{-4} + 0.1975 a_{t-1}^2$$

Because p-value is smaller than 0.05, we reject the null hypothesis and confirm the existence of ARCH effects.



Plotting the standardized residuals and the conditional volatility shows some large (in magnitude) errors, even when standardized. It is clearly seen from the chart that volatility spike fairly steadily from time to time.

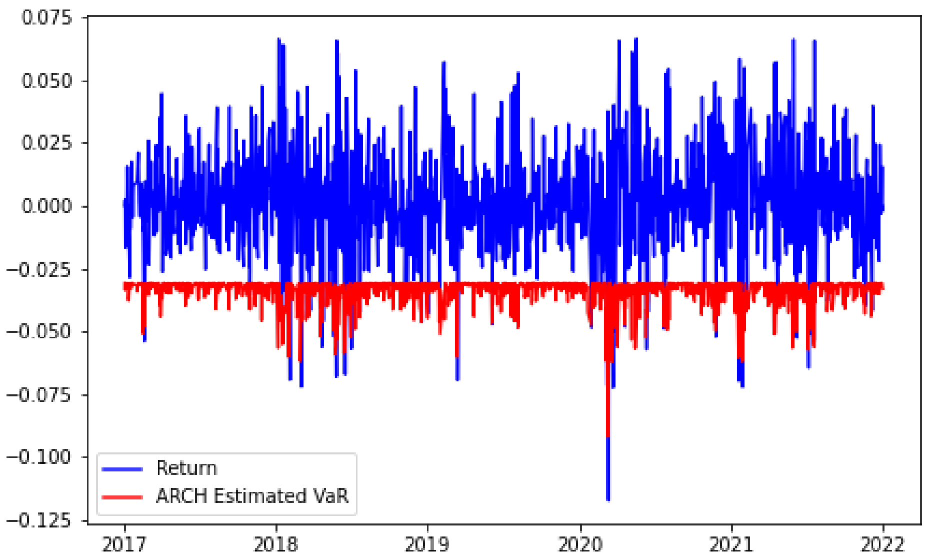


Figure 3.4: ARCH Estimated 5% VaR

The above graph demonstrates that ARCH(1) performed poorly on HPG log returns since the estimated VaR seemed to oscillate around the level of -0.03 and did not fit the volatility of the returns well.

3.3 GARCH Estimations for VaR

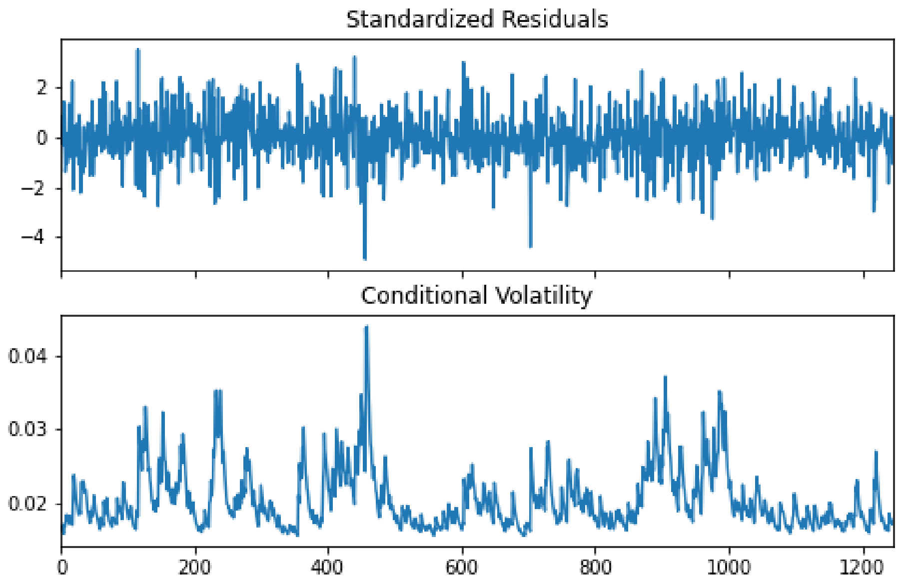
We use the GARCH(1,1) model in equation (2.4) for data-fitting.

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      GARCH   Log-Likelihood:                3125.02
Distribution:                  Normal   AIC:                          -6242.05
Method:            Maximum Likelihood   BIC:                          -6221.53
                                        No. Observations:                 1248
Date:                Thu, May 19 2022   Df Residuals:                     1247
Time:                        15:48:33   Df Model:                            1
                                Mean Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
mu         1.3746e-03  5.219e-04      2.634  8.448e-03 [3.516e-04,2.398e-03]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega      4.3299e-05  5.962e-06      7.262  3.801e-13 [3.161e-05,5.498e-05]
alpha[1]       0.1000  1.567e-02      6.380  1.775e-10   [6.928e-02,  0.131]
beta[1]        0.8000  2.153e-02     37.161 2.933e-302     [  0.758,  0.842]
============================================================================

Covariance estimator: robust
```

From the above table, both α_1 and β_1 are significantly different from zero. Therefore, it is reasonable to assume time-varying volatility of the residuals. The estimated GARCH(1,1) model is

$$a_t = \sigma_t \epsilon_t,$$
$$\sigma_t^2 = 4.3299 \cdot 10^{-5} + 0.1 a_{t-1}^2 + 0.8 \sigma_{t-1}^2$$



We can easily see that the standard residuals of GARCH model followed the same pattern as ARCH model above. However, GARCH model's residuals were lower than that of ARCH model, which means GARCH model fitted the data better than the other one. Moreover, the conditional volatility of GARCH model appeared to be more volatile.

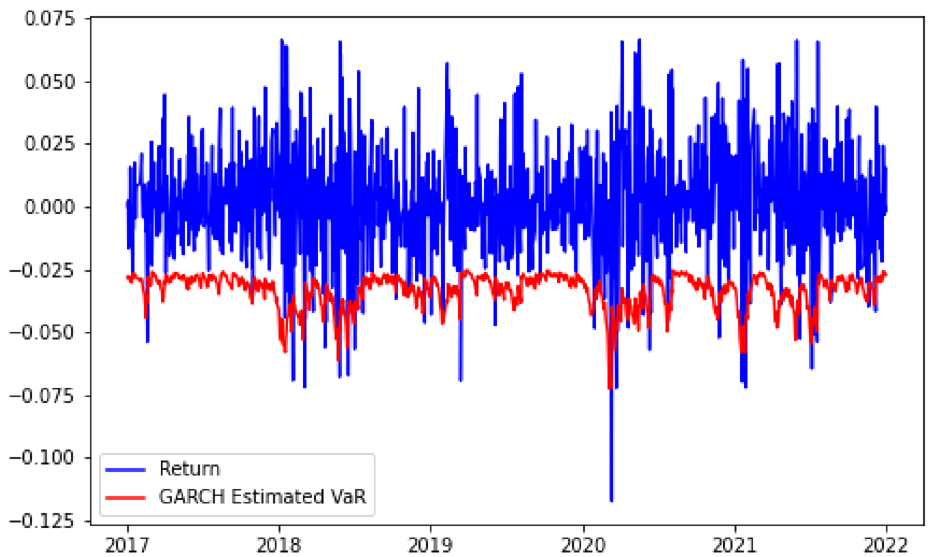


Figure 3.5: GARCH Estimated 5% VaR

It is noticeably that the estimated VaR using GARCH(1,1) model were more sensitive to the movement in stock returns. We can conclude that using GARCH(1,1) is much more effective when it comes to estimating VaR (5%) for HPG stock returns.

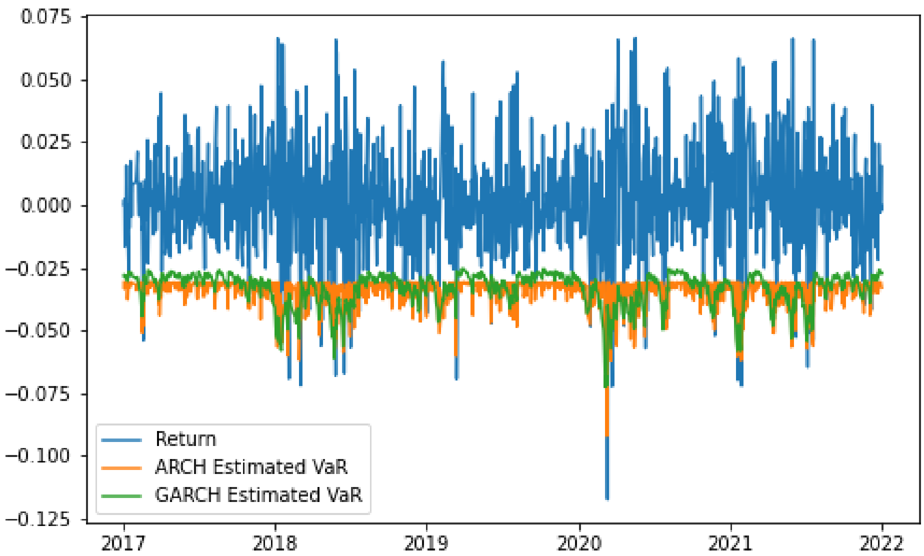


Figure 3.6: ARCH & GARCH Estimated 5% VaR

The above graph clearly displays the presence of VaR violations. The results show that all the techniques adopted for VaR forecasting are more or less risky. The number of VaR violations is counted and compared to the expected number of losses at the chosen confidence level. Therefore, GARCH approach seems to be an effective predictive tool in this particular case.

3.4 VaR Forecasting

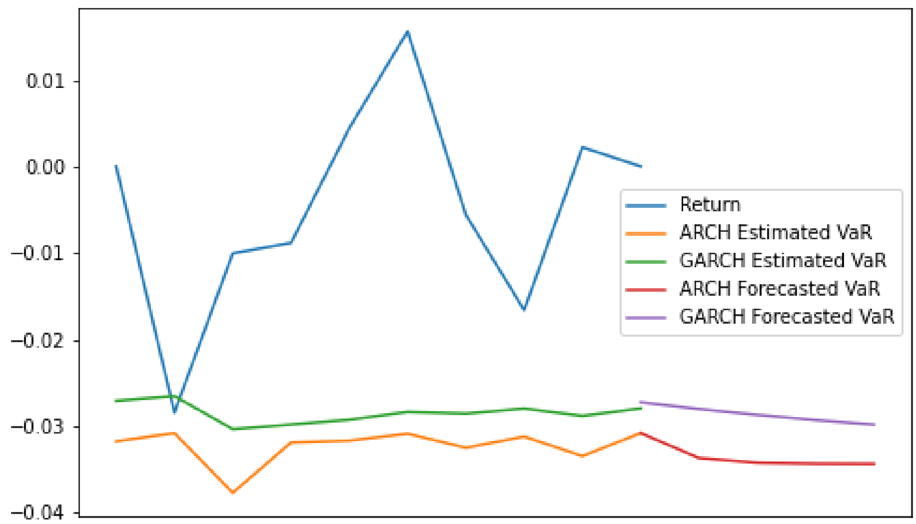


Figure 3.7: ARCH & GARCH 5−Day 5% VaR Forecast

VaR of the returns at 5% for the next 5 days have a tendency to gradually decline. While applying ARCH model yields the results to be -0.0308, -0.0337, -0.0343, -0.0344, -0.03438 respectively, GARCH model predicts the VaR to be -0.0273, -0.0280, -0.0287, -0.0293, -0.0298. The forecasted

VaR by GARCH model is slightly higher than ARCH model, indicating that GARCH model predicts the lowest amount of potential loss will be higher than ARCH model.

Bibliography

[1] Google Colaboratory session, with code & execution result available.
https://colab.research.google.com/drive/1hfUxzYMPXNd2ZaIXO7xMDCuR-RgOMoNt?usp=sharing