Empirical Finance: Assignment 3

By Group 15

Lanlan Hou	2801069
Rick van der Ploeg	2782774
Sebastiaan van Dijen	2769505
Wa Chak Sou	2796840

School of Business and Economics

17th November 2025

QUESTION 1: GARCH modeling

a. AR(1)-GARCH(1,1) model estimation

Parameter	**Estimate**	**Std. Error**	***t*-value**	**Pr(> |t|)**
μ	8.103131e−06	3.261587e−04	0.02484413	9.801793e−01
ϕ_1	−9.302675e−03	3.004835e−02	−0.30959016	7.568726e−01
ω	4.653160e−06	4.555086e−06	1.02153071	3.070031e−01
α_1	7.917190e−02	1.147816e−02	6.89761330	5.288436e−12
β_1	8.958792e−01	3.705583e−03	241.76472950	0.000000e+00

Table 1.1.1: The estimation of parameters using AR(1)-GARCH(1,1) under Normal distribution

Table 1.1.1 shows the estimation of parameters for an AR(1)-GARCH(1,1) model under the normal distribution assumption. For the coefficients of AR(1) model for the returns, the p-values for both μ and ϕ_1, which are 0.9802 and 0.7569 respectively, are significantly larger than the significance level ($\alpha = 0.05$). It reveals that there was no average gain or loss and the returns process does not have autocorrelation. For the coefficients of the GARCH(1,1) model for the variance, both $\alpha_1 = 0.07917$ and $\beta_1 = 0.8959$ are highly significant since their p-values are much smaller than the significance level. With the sum of $\alpha_1 + \beta_1 \approx 0.9751$, this that shocks to volatility decay slowly, which means the volatility is highly persistent. In contrast, the constant-term ω is not significant since its p-value is larger than the significance level.

The unconditional variance is given by:

$$\mathbb{E}[\sigma^2] = \frac{w}{1-\alpha-\beta} = \frac{0.00000465316}{1-0.0791719-0.8958792} \approx 0.0001865076216$$

Therefore, for the unconditional volatility (the standard deviation of returns) can be computed by:

$$\mathbb{E}[\sigma] = \sqrt{\mathbb{E}[\sigma^2]} = \sqrt{0.0001865076216} \approx 0.01365677933$$

b. Forecast and dynamics behavior

To forecast the returns and volatilities for h periods ahead, they can be obtained by deviation recursively. The forecast of Returns with AR(1)-GARCH(1,1) Model can be computed as:

$$\mathbb{E}_T[y_{T+h}] = \phi_1^h y_T + \sum_{i=0}^{h-1} \phi_1^i \mu$$

Using the estimated parameters, the 1-step ahead volatility is deterministic as there is no uncertainty left in the model. It can be computed as:

$$\mathbb{E}_T[\sigma_{T+1}^2] = \omega + \alpha_1 \epsilon_T^2 + \beta_1 \sigma_T^2$$

where $\epsilon_t = \sigma_t e_t$ with $e_t \sim N(0,1)$. Consequently, the N-step ahead variance forecast can be computed by:

$$\mathbb{E}_T[\sigma_{T+h}^2] = (\alpha_1+\beta_1)^{h-1}\sigma_{T+1}^2 + \sum_{i=0}^{h-2}(\alpha_1+\beta_1)^i \omega$$

The corresponding standardized returns for the next five years can be defined as:

$$y_{t,standardized} = \frac{y_t - \mathbb{E}_T[y_t]}{\sqrt{\mathbb{E}_T[\sigma_t^2]}}$$

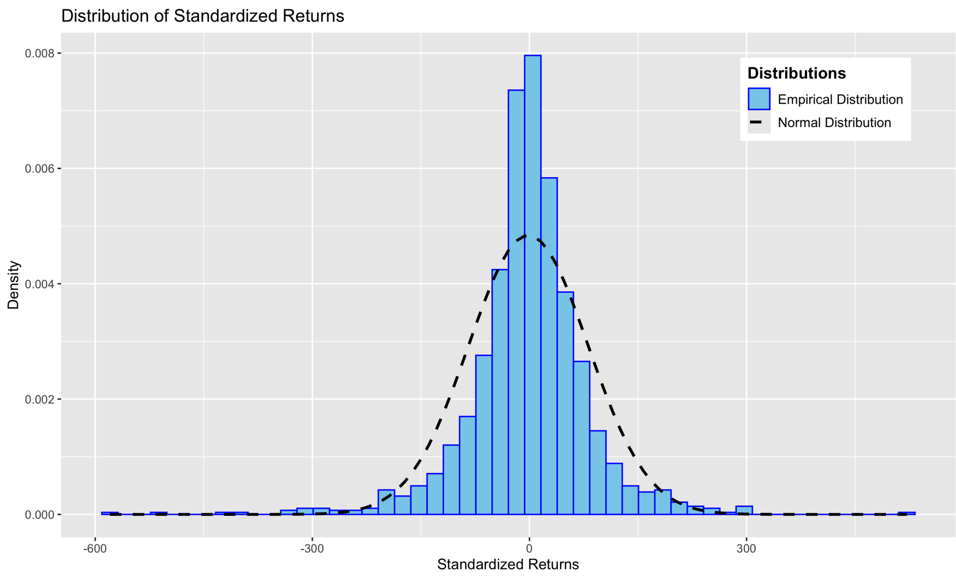

Figure 1.2.1 Distribution of Standardized Daily Log Returns of HSBC (November 2015 to November 2020)

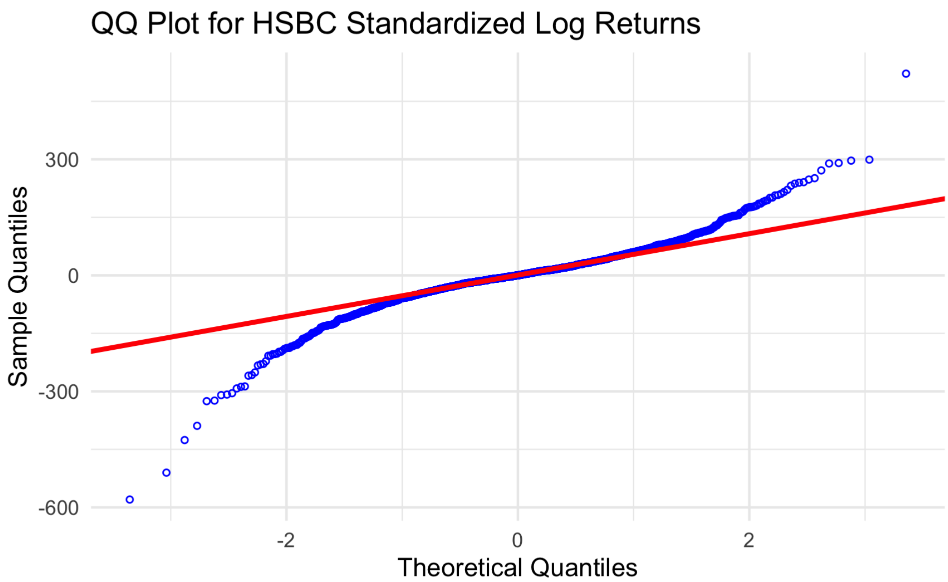

Figure 1.2.2 QQ plot of Standardized Daily Log Returns of HSBC (November 2015 to November 2020)

Figure 1.2.1 shows the distribution of Standardized daily log returns of HSBC and Figure 1.2.2 reveals the corresponding QQ plot. The QQ plot implies the rejection of normality by showing the clear deviations from the theoretical normal line, particularly in the tails. It confirms that the fat-tailed dynamics are still present in the model even under the assumption of normally distributed errors.

Question 2: Improving the model

a. AR(1)-GARCH(1,1) model with Student's t distributed returns

Parameter	Estimate	Std. Error	t-value	Pr($>\|t\|$)
μ	2.227953e−04	3.027526e−04	0.7358989	4.617922e−01
ϕ_1	−3.217405e−02	2.839195e−02	−1.1332104	2.571259e−01
ω	4.085769e−06	3.232341e−06	1.2640279	2.062200e−01
α_1	6.610598e−02	1.789162e−02	3.6948005	2.200595e−04
β_1	9.117124e−01	2.352615e−02	38.7531553	0.000000e+00
ν	6.291618e+00	1.075171e+00	5.8517390	4.864595e−09

Table 2.1.1: The estimation of parameters using AR(1)-GARCH(1,1) under Student-t distribution

Table 2.1.1 reports the estimation of parameters for the AR(1)–GARCH(1,1) model under the Student's t distribution. Similar to the assumption of Normal distribution, the coefficients in the AR(1) model are not statistically significant. The p-values for both μ and ϕ_1 ($p_\mu = 0.4618$ and $p_{\phi_1} = 0.2571$) are greater than 5% significance level, even they are slightly smaller than the p-values in *Normal* distribution. This again suggests that the daily returns process do not exhibit a significant average drift and show no autocorrelation. For the GARCH(1,1) model for the variance, the coefficients $\alpha_1 = 0.06611$ and $\beta_1 = 0.9117$ have extremely small p-values, which indicates highly significant. At the same time, the sum $\alpha_1 + \beta_1 \approx 0.9778$ again confirms that volatility remains highly persistent and decays slowly over time. In contrast, the constant term ω with its p-value $p_\omega = 0.2062$ is not statistically significant as it is greater than the significance level.

Finally, the estimated degrees of freedom is $\nu = 6.2916$, with the extreme small p-value. It means that the Student's t distribution, which has the pattern of heavy tails distribution, is highly significant and can capture heavy tail.

b. Likelihood-Ratio test

To check how well one model performs compared to another model, a Likelihood Ratio Test can be performed. There are two models, one is restricted and the other unrestricted. The restricted model is nested inside the unrestricted model. Denote U = Unrestricted and R = Restricted, by definition:

$$log\mathcal{L}_R \leq log\mathcal{L}_U$$

The hypotheses for the test are formulated as follows:

H_0 : Both the restricted and unrestricted model fit the data equally well,

H_1 : The unrestricted model fits the data better.

The test statistic can be computed as two times the difference between the unrestricted and restricted likelihood:

$$\text{LR} = 2(log\mathcal{L}_U - log\mathcal{L}_R) \sim \chi^2_\nu$$

The degrees of freedom ν is equal to the number of restrictions.

Unrestricted	Restricted	Test
AR(1)–GARCH(1,1) - Student's t	**AR(1)–GARCH(1,1) - Normal**	$H_0 : \nu \to \infty$

Table 2.2.1: The setting of Likelihood Ratio Test for Normal vs Student-t GARCH models

LogL_Norm	LogL_t	Test Stat.	DF	Critical Value (5%)	p-value
3701.575	3731.245	59.34143	1	3.841459	1.321165e−14

Table 2.2.2: The Likelihood Ratio Test for Normal vs Student-t GARCH models

Table 2.2.1 illustrates the setting of the Likelihood Ratio test. The unrestricted model is the AR(1)–GARCH(1,1) model with Student's t distribution and the restricted model the model is with Normal distribution. It is because Normal distribution is nested by Student's t distribution as the degree of freedom of the corresponding t distribution ν approaching to infinity. Consequently, the number of restriction for the Likelihood Ratio test is 1 for the χ^2 distribution.

Table 2.2.2 illustrates the result of the Likelihood-Ratio test between the AR(1)–GARCH(1,1) model under the Normal and Student's t distribution. Since the test statistic (LR = 59.341) is much larger than the critical value ($\chi^2_{1,0.95} = 3.841$), and the p-value is highly significant, as a result, reject the null hypothesis H_0 that both models fit equally well, and accept the alternative H_1 that the AR(1)–GARCH(1,1) model under the assumption of Student's t distribution fits the data better.

Appendix

```r
# Load required packages
library(openxlsx)   # Excel file handling
library(lubridate)  # Date/time utilities
library(quantmod)   # Financial data & modeling
library(rugarch)    # GARCH/AR-GARCH models
library(ggplot2)    # Data visualization

raw <- read.xlsx('student_groups_stocks.xlsx', sheet = 1)

groupNumber <- 15
nameOfStock <- raw$Stock.Name[groupNumber]
startDate <- raw$Start.Date[groupNumber]
endDate <- raw$`End.Date.(+10y)`[groupNumber]

# compute the start date of the final five years
originalDate <- as.Date(startDate)
endDate_fiveYears <- originalDate %m+% years(5)

# extract the daily stock price from source
getSymbols(nameOfStock, src = 'yahoo', from = startDate, to = endDate)

# extract the data of first five years
firstFiveYear <- window(HSBC, start = startDate, end = endDate_fiveYears)

# extract the Adjusted Close for the first five years
adjustedPrice_firstFiveYears<- as.numeric(firstFiveYear[,'HSBC.Adjusted'])

# compute the Daily Log Returns for the first five years
compoundedReturns_firstFiveYears <- numeric(length(adjustedPrice_firstFiveYears) - 1)
for(i in (2:length(adjustedPrice_firstFiveYears))){
  compoundedReturns_firstFiveYears[i-1] <- log(adjustedPrice_firstFiveYears[i] / adjustedPrice_firstFiveYears[i-1])
}

# specify AR(1)-GARCH(1,1) model under normal distribution
ar1garch11_spec_norm <- ugarchspec(variance.model = list(model = 'sGARCH',
                                                         garchOrder = c(1, 1)),
                        mean.model = list(armaOrder = c(1, 0),
                        include.mean = TRUE),
                        distribution.model = 'norm')

ar1garch11_fit_norm <- ugarchfit(spec = ar1garch11_spec_norm,
                                 data = compoundedReturns_firstFiveYears)

ar1garch11_fit_norm@fit$matcoef
```

Figure 1: Question 1(a)

```r
# extract the data of last five years
lastFiveYear <- window(HSBC, start = endDate_fiveYears, end = endDate)

```

```
# extract the Adjusted Close for the next five years
adjustedPrice_lastFiveYears<- as.numeric(lastFiveYear[,'HSBC.Adjusted'])

# compute the Daily Log Returns for the next five years
compoundedReturns_lastFiveYears <- numeric(length(adjustedPrice_lastFiveYears) - 1)
for(i in (2:length(adjustedPrice_lastFiveYears))){
  compoundedReturns_lastFiveYears[i-1] <- log(adjustedPrice_lastFiveYears[i] / adjustedPrice_lastFiveYears[i-1])
}

# extract the estimated parameters
mu_norm <- coef(ar1garch11_fit_norm)['mu']
ar1_norm <- coef(ar1garch11_fit_norm)['ar1']
omege_norm <- coef(ar1garch11_fit_norm)['omega']
alpha1_norm <- coef(ar1garch11_fit_norm)['alpha1']
beta1_norm <- coef(ar1garch11_fit_norm)['beta1']
sum_alphaBeta <- alpha1_norm + beta1_norm

t <- length(compoundedReturns_firstFiveYears)
returns_t <- compoundedReturns_firstFiveYears[t]
eps_t <- ar1garch11_fit_norm@fit$residuals[t]
var_t <- ar1garch11_fit_norm@fit$var[t]

# compute the 1-step ahead variance
oneAhead_var <- omege_norm + alpha1_norm * (eps_t^2) + beta1_norm * (var_t)

forecast_returns <- numeric(length(compoundedReturns_lastFiveYears))
forecast_variance <- numeric(length(compoundedReturns_lastFiveYears))

# compute the returns and variance forecast for the next 5 years
for (i in 1:length(compoundedReturns_lastFiveYears)){
  # Forecast for Log Returns
  forecast_returns[i] <- (sum(sapply(0:(i-1), function(j) ar1_norm^j)) * mu_norm) + ar1_norm^i * returns_t

  # Forecast for Variance
  if (i == 1) {
    forecast_variance[i] <- oneAhead_var
  } else {
    forecast_variance[i] <- (sum(sapply(0:(i-2),
                                        function(j) sum_alphaBeta^j))) * omege_norm + sum_alphaBeta^(i-1) * oneAhead_var
  }
}

# standardize the Daily Log Returns
standardized_returns <- (compoundedReturns_lastFiveYears - forecast_returns) / sqrt(forecast_variance)

# Distribution of Standardized Returns
returns_df <- data.frame(standardizedReturns = standardized_returns)

ggplot(returns_df, aes(x = standardizedReturns)) +
  geom_histogram(aes(y = after_stat(density), color = 'Empirical
```

```
    Distribution'),
                 bins = 50, fill = 'skyblue') +
  stat_function(fun = dnorm,
                args = list(mean = mean(standardized_returns, na.rm =
                    TRUE),
                            sd = sd(standardized_returns, na.rm = TRUE)
                                ),
                aes(color = 'Normal Distribution'), linewidth = 1,
                    linetype = 'dashed') +
  scale_color_manual(name = "Distributions",
                     values = c('Normal Distribution' = 'black',
                                'Empirical Distribution' = 'blue')) +
  labs(title = 'Distribution of Standardized Returns',
       x = 'Standardized Returns', y = 'Density') +
  theme(legend.position = c(0.95, 0.95),
        legend.justification = c("right", "top"),
        legend.title = element_text(size = 12, face = "bold"),
        legend.text = element_text(size = 10))
ggsave('figures/standardizedReturns_distribution.png', width = 10,
    height = 6, dpi = 300)

# QQ Plot of Standardized Returns
ggplot(returns_df, aes(sample = standardizedReturns)) +
  stat_qq(shape = 1, color = 'blue', size = 1) +
  stat_qq_line(color = 'red', linewidth = 1) +
  labs(title = 'QQ Plot for HSBC Standardized Log Returns',
       x = 'Theoretical Quantiles', y = 'Sample Quantiles') +
  theme_minimal()
ggsave('figures/standardizedReturns_qqplot.png', dpi = 300)
```

Figure 2: Question 1(b)

```
# specify AR(1)-GARCH(1,1) model under student t distribution
ar1garch11_spec_t <- ugarchspec(variance.model = list(model = 'sGARCH',
                                                      garchOrder = c(1,
                                                          1)),
                                mean.model = list(armaOrder = c(1, 0),
                                include.mean = TRUE),
                                distribution.model = 'std')

ar1garch11_fit_t <- ugarchfit(spec = ar1garch11_spec_t,
                              data = compoundedReturns_firstFiveYears)

ar1garch11_fit_t@fit$matcoef
```

Figure 3: Question 2(a)

```
# compute the likelihood
logLikelihood_norm <- (likelihood(ar1garch11_fit_norm)) # restricted
logLikelihood_t <- (likelihood(ar1garch11_fit_t)) # unrestricted

# compute the test statistic
LR  <- -2 * (logLikelihood_norm - logLikelihood_t)

degreeOfFreedom  <- 1
criticalValue <- qchisq(0.95, df = degreeOfFreedom)
pValue <- 1 - pchisq(LR, df = degreeOfFreedom)

```

```
likelihoodRatioTest_result_df <- data.frame(LogL_Norm = logLikelihood_
   norm,
                                              LogL_t = logLikelihood_t,
                                              TestStatistic = LR,
                                              DF = degreeOfFreedom,
                                              CriticalValue =
                                                 criticalValue,
                                              PValue = pValue)
likelihoodRatioTest_result_df
```

Figure 4: Question 2(b)