Financial Econometrics in R/Python

Assignment Three

Business School, Imperial College London

Author:

02512292 MohmedMaaeez Malek

04-12-2023

Contents

List of Tables

List of Figures

Preparation

Required Packages

```
#install.packages("...")
#library("...")
#install.packages("rugarch")
#install.packages("tseries")
#install.packages("knitr")
library(knitr)
library(tseries)
library(rugarch)
library(readxl)
library(dplyr)
library(ggplot2)
library(kableExtra)
library(margins)
library(randomForest)
library(e1071)
library(MASS)
library(class)
library(tree)
library(ggplot2)
library(gridExtra)
library(binom)
library(sandwich)
library(lmtest)

opts_chunk$set(tidy.opts=list(width.cutoff=70),tidy=TRUE)
```

Question (1)

Consider the daily simple returns of the S&P 500 composite index from January 1980 to December 2008. The index returns include dividend distributions. The data file is S&P500WeekDays which has 9 columns. The columns are (year, month, day, SP, M, T, W, H, F), where M, T, W, H, F denotes indicator variables for Monday to Friday, respectively. Use a regression model to study the effects of trading days on the index returns. What is the fitted model? Are the weekday effects significant in the returns at the 5% level? Use the Newey West estimator of the covariance matrix to obtain the t-ratio of regression estimates. Does the Newey West estimator change the conclusion of weekday effect? Fit an ARCH(1) and a GARCH(1,1) for the log returns. Assess the statistical significance of the coefficients and compute the unconditional variance given by the two models. What can you say?

Load Data

```r
daily_ret_data = read_excel("SP500WeekDays.xlsx")
```

```r
sp_data <- daily_ret_data[, c("sp", "M", "T", "W", "R", "F")]
model <- lm(sp ~ M + T + W + R + F, data = sp_data)
summary(model)
```

```
##
## Call:
## lm(formula = sp ~ M + T + W + R + F, data = sp_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.204627 -0.005214  0.000142  0.005379  0.115842
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0003290  0.0002923   1.125    0.260
## M           -0.0003711  0.0004184  -0.887    0.375
## T            0.0004114  0.0004107   1.002    0.317
## W            0.0003108  0.0004106   0.757    0.449
## R           -0.0002646  0.0004126  -0.641    0.521
## F                   NA         NA      NA       NA
##
## Residual standard error: 0.01117 on 7314 degrees of freedom
## Multiple R-squared:  0.0007543,  Adjusted R-squared:  0.0002078
## F-statistic:  1.38 on 4 and 7314 DF,  p-value: 0.238
```

Q. What is the fitted model? - Based on the data, the fitted linear regression model can be represented by the following equation: - The regression equation is:

$$\text{sp} = 0.0003290 - 0.0003711 \times \text{M} + 0.0004114 \times \text{T} + 0.0003108 \times \text{W} - 0.0002646 \times \text{R} + \text{NA} \times \text{F} + \text{error}$$

- Here, the coefficients (0.0003290, -0.0003711, 0.0004114, 0.0003108, -0.0002646) represent the estimated impact of coefficient and variables M, T, W, R (each trading day of week) relative to F (Friday), which serves as the baseline or reference category on the S&P 500 absolute daily returns in the dataset.
- The coefficient for Friday (F) is marked as "NA," indicating that it was automatically dropped from the model as, Friday is the reference category against which the effects of the other days of the week are compared in the regression model and it suggests that there is likely issue such as perfect multicollinearity or singularity.

Q. Are the weekday effects significant in the returns at the 5% level?

- The summary of the p-values for each trading day of the week in the simple linear regression model indicates that none of the p-values for the indicator variables (M, T, W, R) are below 0.05. Thus, the weekday effects are not statistically significant at the 5% level.

Q: Use the Newey West estimator of the covariance matrix to obtain the t-ratio of regression estimates.

```r
newey_west_results <- coeftest(model, vcov = NeweyWest(model))
print(newey_west_results)
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept)  0.00032896  0.00026007  1.2649   0.2060
## M           -0.00037109  0.00042757 -0.8679   0.3855
## T            0.00041140  0.00038543  1.0674   0.2858
## W            0.00031078  0.00038643  0.8042   0.4213
## R           -0.00026463  0.00038645 -0.6848   0.4935
```

Q. Does the Newey West estimator change the conclusion of weekday effect?

- None of the p-values for the intercept and variables M, T, W, R are below 0.05. So, the Newey-West results seem consistent with the simple linear regression model results, as there are no notable changes in the significance levels of the coefficients. The weekday effects remain statistically insignificant at the 5% level.

  Q. Fit an ARCH(1) and a GARCH(1,1) for the log returns. Assess the statistical significance of the coefficients and compute the unconditional variance given by the two models. What can you say?

```r
returns <- daily_ret_data$sp
log_returns <- log(1 + returns)
arch_model <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 0)),
                         mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
arch_fit <- ugarchfit(spec = arch_model, data = log_returns)
print(arch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,0)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## omega   0.000088    0.000002   47.043        0
## alpha1  0.287712    0.019936   14.432        0
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## omega   0.000088    0.000006  14.9644    0e+00
```

```
## alpha1  0.287712    0.062409   4.6101    4e-06
##
## LogLikelihood : 22974.43
##
## Information Criteria
## ------------------------------------
##
## Akaike       -6.2775
## Bayes        -6.2756
## Shibata      -6.2775
## Hannan-Quinn -6.2768
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.8156 0.36647
## Lag[2*(p+q)+(p+q)-1][2]    4.8585 0.04438
## Lag[4*(p+q)+(p+q)-1][5]    9.8568 0.01002
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                      1.661  0.1975
## Lag[2*(p+q)+(p+q)-1][2]   206.964  0.0000
## Lag[4*(p+q)+(p+q)-1][5]   492.973  0.0000
## d.o.f=1
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[2]     410.4 0.500 2.000       0
## ARCH Lag[4]     540.4 1.397 1.611       0
## ARCH Lag[6]     721.8 2.222 1.500       0
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  4.2261
## Individual Statistics:
## omega  3.625
## alpha1 2.091
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          0.61 0.749 1.07
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value     prob sig
## Sign Bias           0.5949 0.551910
## Negative Sign Bias  1.7382 0.082226   *
## Positive Sign Bias  1.2711 0.203748
## Joint Effect       11.9453 0.007573 ***
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
```

```
## 1    20     533.9   4.236e-101
## 2    30     550.4    1.215e-97
## 3    40     566.5    8.714e-95
## 4    50     596.7    3.361e-95
##
##
## Elapsed time : 0.08612609
```

```r
garch_model <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
                          mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
garch_fit <- ugarchfit(spec = garch_model, data = log_returns)
print(garch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## omega   0.000001    0.000001   1.3702  0.17063
## alpha1  0.070593    0.011500   6.1385  0.00000
## beta1   0.921161    0.012129  75.9455  0.00000
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## omega   0.000001    0.000014  0.08214 0.934536
## alpha1  0.070593    0.201276  0.35072 0.725795
## beta1   0.921161    0.209375  4.39959 0.000011
##
## LogLikelihood : 23835.37
##
## Information Criteria
## ------------------------------------
##
## Akaike       -6.5125
## Bayes        -6.5096
## Shibata      -6.5125
## Hannan-Quinn -6.5115
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic  p-value
## Lag[1]                      6.889 0.008674
## Lag[2*(p+q)+(p+q)-1][2]     7.085 0.011313
## Lag[4*(p+q)+(p+q)-1][5]    10.478 0.006935
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
```

```
##                          statistic p-value
## Lag[1]                       2.985 0.08405
## Lag[2*(p+q)+(p+q)-1][5]      3.313 0.35285
## Lag[4*(p+q)+(p+q)-1][9]      3.789 0.62480
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.1673 0.500 2.000  0.6825
## ARCH Lag[5]    0.1692 1.440 1.667  0.9726
## ARCH Lag[7]    0.5874 2.315 1.543  0.9698
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  439.1709
## Individual Statistics:
## omega  49.36373
## alpha1  0.06707
## beta1   0.08230
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          0.846 1.01 1.35
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value      prob sig
## Sign Bias           0.3173 7.510e-01
## Negative Sign Bias  3.9826 6.882e-05 ***
## Positive Sign Bias  2.4889 1.284e-02  **
## Joint Effect       35.3952 1.005e-07 ***
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     184.3    4.370e-29
## 2    30     206.1    1.312e-28
## 3    40     224.1    7.687e-28
## 4    50     238.5    1.019e-26
##
##
## Elapsed time : 0.07193398
```

- The results of ARCH(1) and GARCH(1,1) suggests that:
- **Parameter Significance:**
  - ARCH(1) model show statistically significant estimates for the intercept and coefficients (α_1). Whereas, GARCH(1,1) model show statistically significant estimates for the coefficients (α_1, β_1) but statistical insignificance for the intercept.
- **Unconditional Variance:**
  - The unconditional variance can be obtained from both the models as:
  - In GARCH(1,1) model the unconditional variance is obtained by:

$$\sigma^2 = \frac{\omega}{1 - \alpha_1}$$

```r
arch_omega <- coef(arch_fit)["omega"]
arch_alpha1 <- coef(arch_fit)["alpha1"]

# Unconditional Variance for ARCH(1)
arch_unconditional_variance <- arch_omega / (1 - arch_alpha1)
print("The unconditional variance of ARCH(1) is:")
```

```
## [1] "The unconditional variance of ARCH(1) is:"
```

```r
print(arch_unconditional_variance)
```

```
##        omega
## 0.0001228597
```

- In GARCH(1,1) model the unconditional variance is obtained by:

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}$$

  }

```r
garch_omega <- coef(garch_fit)["omega"]
garch_alpha1 <- coef(garch_fit)["alpha1"]
garch_beta1 <- coef(garch_fit)["beta1"]

# Unconditional Variance for GARCH(1,1)
garch_unconditional_variance <- garch_omega / (1 - garch_alpha1 - garch_beta1)
print("The unconditional variance of GARCH(1,1) is:")
```

```
## [1] "The unconditional variance of GARCH(1,1) is:"
```

```r
print(garch_unconditional_variance)
```

```
##        omega
## 0.0001377912
```

- **Comparison:**
  - Both models suggest that, on average, the volatility in the financial returns is persistent and exhibits clustering over time.
  - The GARCH(1,1) model provides a more nuanced prediction by considering not only the immediate past (ARCH) but also the persistence of volatility.

Question (2)

The file USMacro_Quarterly contains quarterly data on several macroeconomic series for the United States: the data are described in the file USMacro_Description. Compute t = ln(t), the logarithm of real GDP, and change in t, the quarterly growth rate of GDP. In the problems below, use the sample period 1955:1-2004:4 (where data before 1955 may be used, as necessary, as initial values in regressions).

Load Data

```
quarterly_data <- read_excel("USMacro_Quarterly.xls")
```

```
quarterly_data$ln_RealGDP <- log(quarterly_data$RealGDP)
quarterly_data$GDP_growth_rate <- c(NA, diff(quarterly_data$ln_RealGDP))
subset_data <- quarterly_data[quarterly_data$Date >= "1955:01" & quarterly_data$Date <= "2004:04", ]
```

Question (2) (i):

(a)

Question: Estimate the mean of ΔR_t

```
mean_change_in_Yt <- mean(subset_data$GDP_growth_rate, na.rm = TRUE)
print(mean_change_in_Yt)
```

```
## [1] 0.008258661
```

- As seen above, the mean of ΔR_t is 0.008258661

(b)

Question: Express the mean growth rate in percentage points at an annual rate

```
mean_change_in_Yt_annual <- mean_change_in_Yt * 400
print(mean_change_in_Yt_annual)
```

```
## [1] 3.303464
```

- As seen above, the mean growth rate in percentage points at an annual rate is 3.303464

(c)

Question: Estimate the standard deviation of ΔR_t. Express your answer in percentage points at an annual rate.

```
std_dev_change_in_Yt <- sd(subset_data$GDP_growth_rate, na.rm = TRUE)
std_dev_change_in_Yt_annual <- std_dev_change_in_Yt * sqrt(4)*100
print(std_dev_change_in_Yt_annual)
```

```
## [1] 1.84116
```

- As seen above, the standard deviation of ΔR_t in percentage points at an annual rate is 1.84116%

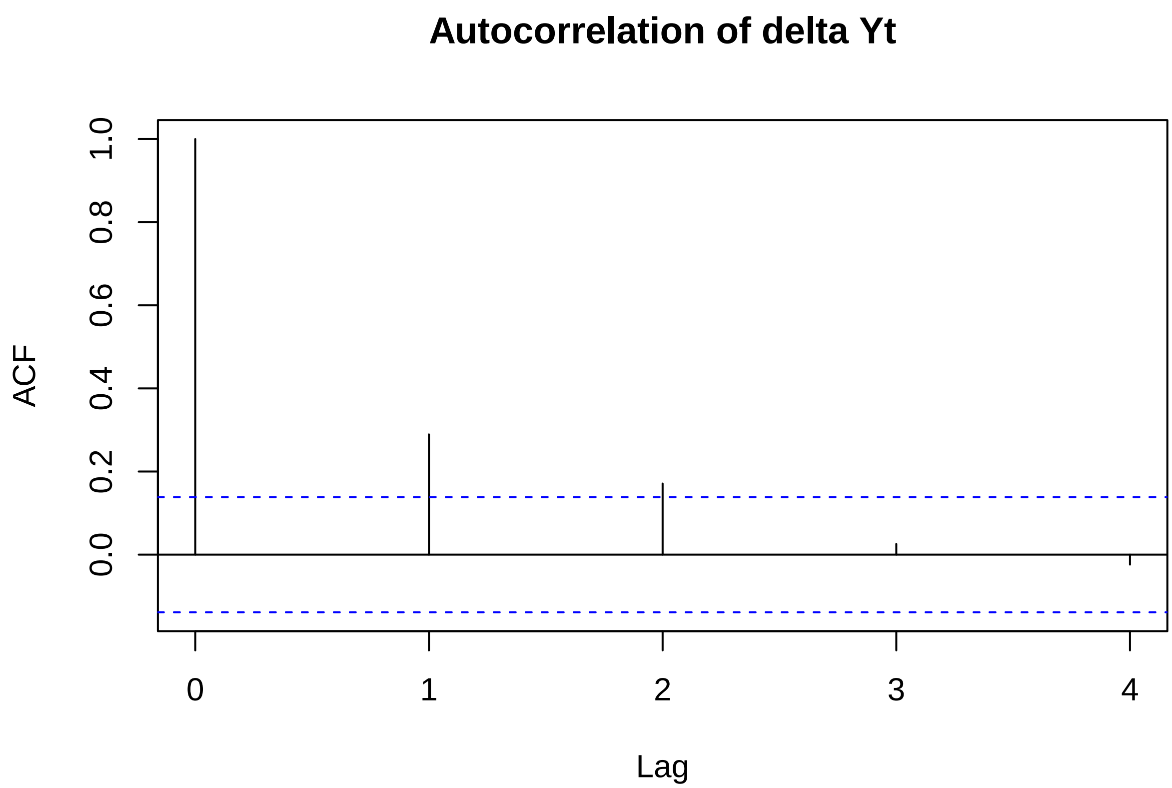

Figure 1: Autocorrelation of delta Yt

Table 1: Autocorrelation of delta Yt

Lag	Autocorrelation
0	1.0000000
1	0.2894087
2	0.1710898
3	0.0259453
4	-0.0239606

(d)

Question: Estimate the first four autocorrelations of ΔR_t. What are the units of autocorrelations?

- The Autocorrelations are unitless.

Question (2) (ii):

```
ar1_model <- arima(subset_data$GDP_growth_rate, order = c(1, 0, 0))
ar1_coefficient <- ar1_model$coef[1]
ar1_standard_error <- sqrt(ar1_model$var.coef[1])
df_1 <- length(subset_data$GDP_growth_rate) - length(coef(ar1_model))
t_statistic_AR1 <- ar1_coefficient / ar1_standard_error
critical_t_value_AR1 <- qt(0.975, df_1)
confidence_interval_AR1 <- c(ar1_coefficient - critical_t_value_AR1 * ar1_standard_error,
                             ar1_coefficient + critical_t_value_AR1 * ar1_standard_error)
is_significant_AR1 <- abs(t_statistic_AR1) > critical_t_value_AR1

cat("Estimated AR(1) coefficient:", ar1_coefficient, "\n")
```

```
## Estimated AR(1) coefficient: 0.2950893
```

```
cat("Standard error of AR(1) coefficient:", ar1_standard_error, "\n")
```

```
## Standard error of AR(1) coefficient: 0.06821476
```

```
cat("t-statistic:", t_statistic_AR1, "\n")
```

```
## t-statistic: 4.325886
```

```
cat("Degrees of freedom:", df_1, "\n")
```

```
## Degrees of freedom: 198
```

```
cat("95% Confidence interval lower bound for AR1 is", confidence_interval_AR1[1],
    "and the upper bound is ",confidence_interval_AR1[2], "\n")
```

```
## 95% Confidence interval lower bound for AR1 is 0.1605686 and the upper bound is  0.42961
```

```
cat(is_significant_AR1, ": The coefficient is statistically significantly different from zero")
```

```
## TRUE : The coefficient is statistically significantly different from zero
```

(a)

```
ar2_model <- arima(subset_data$GDP_growth_rate, order = c(2, 0, 0))
ar2_coefficients <- ar2_model$coef[2]
ar2_standard_errors <- sqrt(diag(ar2_model$var.coef)[2])
df_2 <- length(subset_data$GDP_growth_rate) - length(ar2_model$coef)
t_statistics_AR2 <- ar2_coefficients / ar2_standard_errors
```

```r
critical_t_value_AR2 <- qt(0.975, df_2)
confidence_interval_AR2 <- cbind(ar2_coefficients - critical_t_value_AR2 * ar2_standard_errors,
                                 ar2_coefficients + critical_t_value_AR2 * ar2_standard_errors)
is_significant_AR2 <- abs(t_statistics_AR2) > critical_t_value_AR2

cat("Estimated AR(2) coefficients:", ar2_coefficients, "\n")
```

```
## Estimated AR(2) coefficients: 0.09786695
```

```r
cat("Standard errors of AR(2) coefficients:", ar2_standard_errors, "\n")
```

```
## Standard errors of AR(2) coefficients: 0.07088085
```

```r
cat("t-statistics:", t_statistics_AR2, "\n")
```

```
## t-statistics: 1.380725
```

```r
cat("Degrees of freedom:", df_2, "\n")
```

```
## Degrees of freedom: 197
```

```r
cat("95% Confidence intervals for AR(2) coefficients:\n", confidence_interval_AR2, "\n")
```

```
## 95% Confidence intervals for AR(2) coefficients:
##  -0.0419157 0.2376496
```

```r
cat(is_significant_AR2, ": The coefficient is not statistically significantly different from zero")
```

```
## FALSE : The coefficient is not statistically significantly different from zero
```

```r
ar1_model <-arima(subset_data$GDP_growth_rate, order =c(1,0,0))
coeftest(ar1_model)
```

```
##
## z test of coefficients:
##
##             Estimate Std. Error z value  Pr(>|z|)
## ar1       0.29508928 0.06821476  4.3259 1.519e-05 ***
## intercept 0.00830326 0.00088184  9.4158 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ar2_model <-arima(subset_data$GDP_growth_rate, order =c(2,0,0))
coeftest(ar2_model)
```

```
##
## z test of coefficients:
##
##             Estimate Std. Error z value  Pr(>|z|)
## ar1       0.26688879 0.07085709  3.7666 0.0001655 ***
## ar2       0.09786695 0.07088085  1.3807 0.1673636
## intercept 0.00832678 0.00097205  8.5662 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- As seen from the results above, the AR(1) model has statistically significant coefficients for both ar1 and intercept. While, the AR(2) model has statistically significant coefficients for ar1 and intercept, but the coefficient ar2 is not statistically significant. Thus, there is no strong evidence to prefer AR(2) model over AR(1) model. So I would say that AR(1) is a reasonable choice. But looking further into it and trying another test to prove it again.

```
# Likelihood ratio test comparing AR(2) to AR(1)
lr_test_statistic <- 2 * (logLik(ar2_model) - logLik(ar1_model))
lr_p_value <- 1 - pchisq(lr_test_statistic, df = 2 - 1)  # df difference between models

cat("Likelihood ratio test p-value:", lr_p_value, "\n")
```

```
## Likelihood ratio test p-value: 0.1684685
```

```
# Information criteria
aic_ar1 <- AIC(ar1_model)
aic_ar2 <- AIC(ar2_model)

cat("AIC for AR(1) model:", aic_ar1, "\n")
```

```
## AIC for AR(1) model: -1320.454
```

```
cat("AIC for AR(2) model:", aic_ar2, "\n")
```

```
## AIC for AR(2) model: -1320.35
```

- As we can see, the likelihood ratio test p-value is 0.1684685, and the AIC values for the AR(1) and AR(2) models are -1320.454 and -1320.35, respectively.
- As the likelihood ratio test p-value is greater than the typical significance level of 0.05, there isn't strong evidence to reject the null hypothesis that the simpler model (AR(1)) is sufficient. The AIC values for both models are quite close, and in this case, the difference might not be substantial enough to strongly favor one model over the other. So, based on the likelihood ratio test and AIC values, there isn't a strong indication that the AR(2) model is significantly better than the AR(1) model.

(b)

```
ar3_model <- arima(subset_data$GDP_growth_rate, order = c(3, 0, 0))
ar3_coefficients <- ar3_model$coef[3]
ar3_standard_errors <- sqrt(diag(ar3_model$var.coef)[3])
df_3 <- length(subset_data$GDP_growth_rate) - length(ar3_model$coef)
t_statistics_AR3 <- ar3_coefficients / ar3_standard_errors
critical_t_value_AR3 <- qt(0.975, df_3)
confidence_interval_AR3 <- cbind(ar3_coefficients - critical_t_value_AR3 * ar3_standard_errors,
                                 ar3_coefficients + critical_t_value_AR3 * ar3_standard_errors)
is_significant_AR3 <- abs(t_statistics_AR3) > critical_t_value_AR3

cat("Estimated AR(3) coefficients:", ar3_coefficients, "\n")
```

```
## Estimated AR(3) coefficients: -0.05098659
```

```
cat("Standard errors of AR(3) coefficients:", ar3_standard_errors, "\n")
```

```
## Standard errors of AR(3) coefficients: 0.07096675
```

```r
cat("t-statistics:", t_statistics_AR3, "\n")
```

```
## t-statistics: -0.7184574
```

```r
cat("Degrees of freedom:", df_3, "\n")
```

```
## Degrees of freedom: 196
```

```r
cat("95% Confidence intervals for AR(3) coefficients:\n", confidence_interval_AR3, "\n")
```

```
## 95% Confidence intervals for AR(3) coefficients:
##  -0.190943 0.08896987
```

```r
cat(is_significant_AR3, ": The coefficient is not statistically significantly different from zero")
```

```
## FALSE : The coefficient is not statistically significantly different from zero
```

```r
ar4_model <- arima(subset_data$GDP_growth_rate, order = c(4, 0, 0))
ar4_coefficients <- ar4_model$coef[4]
ar4_standard_errors <- sqrt(diag(ar4_model$var.coef)[4])
df_4 <- length(subset_data$GDP_growth_rate) - length(ar4_model$coef)
t_statistics_AR4 <- ar4_coefficients / ar4_standard_errors
critical_t_value_AR4 <- qt(0.975, df_4)
confidence_interval_AR4 <- cbind(ar4_coefficients - critical_t_value_AR4 * ar4_standard_errors,
                              ar4_coefficients + critical_t_value_AR4 * ar4_standard_errors)
is_significant_AR4 <- abs(t_statistics_AR4) > critical_t_value_AR4

cat("Estimated AR(4) coefficients:", ar4_coefficients, "\n")
```

```
## Estimated AR(4) coefficients: -0.04026275
```

```r
cat("Standard errors of AR(4) coefficients:", ar4_standard_errors, "\n")
```

```
## Standard errors of AR(4) coefficients: 0.07087815
```

```r
cat("t-statistics:", t_statistics_AR4, "\n")
```

```
## t-statistics: -0.5680559
```

```r
cat("Degrees of freedom:", df_4, "\n")
```

```
## Degrees of freedom: 195
```

```r
cat("95% Confidence intervals for AR(4) coefficients:\n", confidence_interval_AR4, "\n")
```

```
## 95% Confidence intervals for AR(4) coefficients:
##  -0.1800489 0.09952342
```

```r
cat(is_significant_AR4, ": The coefficient is not statistically significantly different from zero")
```

```
## FALSE : The coefficient is not statistically significantly different from zero
```

```r
ar3_model <-arima(subset_data$GDP_growth_rate, order =c(3,0,0))
coeftest(ar3_model)
```

```
##
## z test of coefficients:
##
##              Estimate  Std. Error z value  Pr(>|z|)
## ar1        0.27179349  0.07108664  3.8234 0.0001316 ***
## ar2        0.11108622  0.07313364  1.5189 0.1287755
## ar3       -0.05098659  0.07096675 -0.7185 0.4724753
## intercept  0.00831382  0.00092441  8.9937 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ar4_model <-arima(subset_data$GDP_growth_rate, order =c(4,0,0))
coeftest(ar4_model)
```

```
##
## z test of coefficients:
##
##              Estimate  Std. Error z value  Pr(>|z|)
## ar1        0.26991282  0.07110282  3.7961  0.000147 ***
## ar2        0.11551946  0.07351707  1.5713  0.116106
## ar3       -0.04032775  0.07334730 -0.5498  0.582443
## ar4       -0.04026275  0.07087815 -0.5681  0.569997
## intercept  0.00830493  0.00088888  9.3432 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

BIC and AIC for each models:

```r
bic_values <- c(BIC(ar1_model), BIC(ar2_model), BIC(ar3_model), BIC(ar4_model))
aic_values <- c(AIC(ar1_model), AIC(ar2_model), AIC(ar3_model), AIC(ar4_model))

min_bic_order <- which.min(bic_values)
min_aic_order <- which.min(aic_values)

cat("BIC values:", bic_values, "\n")
```

```
## BIC values: -1310.559 -1307.157 -1302.374 -1297.398
```

```r
cat("AIC values:", aic_values, "\n")
```

```
## AIC values: -1320.454 -1320.35 -1318.866 -1317.188
```

```r
cat("Minimum BIC order:", min_bic_order, "\n")
```

```
## Minimum BIC order: 1
```

```r
cat("Minimum AIC order:", min_aic_order, "\n")
```

```
## Minimum AIC order: 1
```

- Based on the BIC values, the lag order with the minimum BIC is 1. Similarly, based on the AIC values, the lag order with the minimum AIC is also 1. Therefore, both BIC and AIC suggest choosing an AR(1) model as it has the lowest information criterion values.

Question (2) (iii):

```r
adf_test_ar1 <- adf.test(subset_data$GDP_growth_rate, alternative = "stationary", k = 1)
```

```
## Warning in adf.test(subset_data$GDP_growth_rate, alternative = "stationary", :
## p-value smaller than printed p-value
```

```r
print(adf_test_ar1)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  subset_data$GDP_growth_rate
## Dickey-Fuller = -7.6451, Lag order = 1, p-value = 0.01
## alternative hypothesis: stationary
```

- We got the test statistic value as -7.6451, and as we know, for ADF test, a more negative test statistic provides stronger evidence against the presence of a unit root. So there is evidence to suggest that the time series is stationary and does not have a unit root. This is a positive result for the purposes of time series analysis.

As an alternative, suppose that t is stationary around a deterministic trend. What are your conclusions?

Test for Stationarity Around a Deterministic Trend

Performing the Augmented Dickey-Fuller (ADF) test with the alternative hypothesis that Y_t is stationary around a deterministic trend.

```r
adf_test_det_trend <- adf.test(subset_data$GDP_growth_rate, alternative = "explosive", k = 1)
```

```
## Warning in adf.test(subset_data$GDP_growth_rate, alternative = "explosive", :
## p-value smaller than printed p-value
```

```r
print(adf_test_det_trend)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  subset_data$GDP_growth_rate
## Dickey-Fuller = -7.6451, Lag order = 1, p-value = 0.99
## alternative hypothesis: explosive
```

- The p-value of 0.99 suggests that we do not have enough evidence to reject the null hypothesis of a unit root with an explosive trend. Therefore, Y_t may not be stationary around a deterministic trend.

Conclusions of both ADF test results

`Stationarity Hypothesis`

The p-value of 0.01 is less than the common significance level of 0.05. Reject the null hypothesis of a unit root with stationarity. Evidence suggests that the series may be stationary.

`Explosive (Deterministic) Trend Hypothesis`

The p-value of 0.99 is greater than the common significance level of 0.05. Do not have enough evidence to reject the null hypothesis of a unit root with an explosive trend. The series may not be stationary around a deterministic trend.

`Overall`

The results are somewhat conflicting, indicating potential challenges in determining the stationarity of the series. Further analysis or consideration of alternative models may be warranted.

Question (2) (iv):

```r
ts_data <- ts(subset_data$GDP_growth_rate, frequency = 4)
arch_model <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 0)),
                          mean.model = list(armaOrder = c(0, 0)))

arch_fit <- ugarchfit(spec = arch_model, data = ts_data)
print(arch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,0)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.008621    0.000660  13.0611 0.000000
## omega   0.000063    0.000009   7.0282 0.000000
## alpha1  0.253276    0.124644   2.0320 0.042155
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.008621    0.000925   9.3189 0.000000
## omega   0.000063    0.000015   4.2424 0.000022
## alpha1  0.253276    0.125447   2.0190 0.043488
##
## LogLikelihood : 657.9129
##
## Information Criteria
## ------------------------------------
##
## Akaike       -6.5491
## Bayes        -6.4997
## Shibata      -6.5496
## Hannan-Quinn -6.5291
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic   p-value
## Lag[1]                      12.38 4.334e-04
## Lag[2*(p+q)+(p+q)-1][2]     16.27 4.244e-05
## Lag[4*(p+q)+(p+q)-1][5]     18.95 3.758e-05
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                   0.001335  0.9709
## Lag[2*(p+q)+(p+q)-1][2]  1.239589  0.4266
## Lag[4*(p+q)+(p+q)-1][5]  4.250589  0.2244
```

```
## d.o.f=1
##
## Weighted ARCH LM Tests
## ------------------------------------
##              Statistic Shape Scale P-Value
## ARCH Lag[2]      2.427 0.500 2.000 0.11923
## ARCH Lag[4]      5.169 1.397 1.611 0.08006
## ARCH Lag[6]      7.946 2.222 1.500 0.04267
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  1.7816
## Individual Statistics:
## mu     0.1889
## omega  1.4352
## alpha1 0.1279
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          0.846 1.01 1.35
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.5031 0.6154
## Negative Sign Bias  0.6708 0.5031
## Positive Sign Bias  0.8452 0.3990
## Joint Effect        1.3826 0.7096
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20      23.4       0.2202
## 2    30      27.7       0.5340
## 3    40      42.8       0.3114
## 4    50      49.5       0.4532
##
##
## Elapsed time : 0.03195286
```

- The results of ARCH(1) Model shows that
  - μ: Highly significant (p-value < 0.05)
  - ω: Highly statistically significant (p-value < 0.05)
  - α_1: Highly significant (p-value < 0.05)
- The unconditional variance is given by $\frac{\omega}{1-\alpha_1}$. In this case, the unconditional variance comes out to be 8.436852e-05

```r
garch_model <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
                           mean.model = list(armaOrder = c(0, 0)))

garch_fit <- ugarchfit(spec = garch_model, data = ts_data)
print(garch_fit)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
```

```
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : norm
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.008962    0.000748 11.97522 0.000000
## omega   0.000002    0.000005  0.44366 0.657288
## alpha1  0.213402    0.121834  1.75159 0.079845
## beta1   0.776384    0.111485  6.96405 0.000000
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.008962    0.001743  5.14213 0.000000
## omega   0.000002    0.000018  0.12776 0.898336
## alpha1  0.213402    0.337375  0.63254 0.527036
## beta1   0.776384    0.336232  2.30907 0.020939
##
## LogLikelihood : 670.0407
##
## Information Criteria
## ------------------------------------
##
## Akaike       -6.6604
## Bayes        -6.5944
## Shibata      -6.6612
## Hannan-Quinn -6.6337
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic   p-value
## Lag[1]                      9.217 0.0023976
## Lag[2*(p+q)+(p+q)-1][2]    13.843 0.0001851
## Lag[4*(p+q)+(p+q)-1][5]    17.357 0.0001025
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                    0.04685  0.8286
## Lag[2*(p+q)+(p+q)-1][5]   1.25146  0.8006
## Lag[4*(p+q)+(p+q)-1][9]   4.02725  0.5843
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.4659 0.500 2.000  0.4949
## ARCH Lag[5]    2.2328 1.440 1.667  0.4220
## ARCH Lag[7]    2.9930 2.315 1.543  0.5149
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  1.6069
```

```
## Individual Statistics:
## mu     0.2454
## omega  0.2098
## alpha1 0.1817
## beta1  0.2844
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.07 1.24 1.6
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias          0.01621 0.9871
## Negative Sign Bias 0.34715 0.7289
## Positive Sign Bias 1.14467 0.2537
## Joint Effect       3.10565 0.3756
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20      20.2       0.3826
## 2    30      29.8       0.4240
## 3    40      30.4       0.8363
## 4    50      38.5       0.8597
##
##
## Elapsed time : 0.02405095
```

- From the results of GARCH(1,1) model, we can observe that:
  - μ: Highly significant (p-value < 0.05)
  - ω: Not statistically significant (p-value > 0.05)
  - α_1: Not significant (p-value $= 0.079845$)
  - β_1: Highly significant (p-value < 0.05)
- The unconditional variance is given by $\frac{\omega}{1-\alpha_1-\beta_1}$. The unconditional variance might not be well-defined due to the values of α_1 and β_1.

Conclusions:

- The ARCH(1) and GARCH(1,1) models suggest that past squared observations help predict the current conditional variance.
- The models indicate statistically significant mean and volatility components, with high significance in the ARCH term.
- Unconditional variance may not be well-defined due to high values of α_1 and β_1 in case of GARCH(1,1).