# The comparison of forecasting performance of historical volatility versus realized volatility

by

© **Linkai Huang**

A practicum submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Applied Statistics.

Department of Mathematics and Statistics

Memorial University

St. John's, Newfoundland and Labrador, Canada

# Abstract

When forecasting stock market volatility with a standard volatility method (GARCH), it is common that the forecast evaluation criteria often suggests that the realized volatility (the sum of squared high-frequency returns) has a better prediction performance compared to the historical volatility (extracted from the close-to-close return). Since many extensions of the GARCH model have been developed, we follow the previous works to compare the historical volatility with many new GARCH family models (i.e., EGARCH, TGARCH, and APARCH model) and realized volatility with the ARMA model. Our analysis is based on the S&P 500 index from August $1^{st}$, 2018 to February $1^{st}$, 2019 (127 trading days), and the data has been separated into an estimation period (90 trading days) and an evaluation period (37 trading days). In the evaluation period, by taking realized volatility as the proxy of the true volatility, our empirical result shows that the realized volatility with ARMA model provides more accurate predictions, compared to the historical volatility with the GARCH family models.

# Table of contents

**Title page** **i**

**Abstract** **ii**

**Table of contents** **iii**

**List of tables** **v**

**List of figures** **vi**

**1 Introduction** **1**

**2 Stock return and models** **4**

2.1 Volatility . . . . . 4

2.2 Daily return . . . . . 5

2.3 Realized volatility . . . . . 6

2.4 Autoregressive conditional heteroskedasticity model and its extension . 11

2.4.1 ARCH model . . . . . 11

2.4.2 GARCH model . . . . . 16

2.4.3 The extensions of GARCH model . . . . . 19

**3 Data and empirical results** **24**

3.1 Data selection . . . . . 24
3.2 Descriptive statistic . . . . . 25
3.3 Forecast methodology and criteria . . . . . 26
3.4 Volatility model of the daily returns . . . . . 28
3.5 ARMA model of the realized volatility . . . . . 41

**4 Forecast evaluation and conclusion** **46**

**Bibliography** **49**

# List of tables


3.1 Summary statistics of the daily return and realized volatility of S&P 500 index from August 1st, 2018 to December 7th, 2019. . . . . 25

3.2 Summary statistics of the sample trend model. . . . . 30

3.3 Summary statistics of the sample periodic model. . . . . 31

3.4 Summary statistics of the AR(8) model. . . . . 34

3.5 Summary statistics of the MA(9) model. . . . . 35

3.6 Summary statistics of the ARMA(2,2) model. . . . . 35

3.7 Summary statistics of the AR(8), MA(9) and ARMA(2,2) model once zero-coefficients are removed from the models. . . . . 36

3.8 Summary statistics of the GARCH(1,1) model of sample squared residuals $a_t^2$. . . . . 40

3.9 Summary statistics of the AR(10) model with sample $D_t$. . . . . 43

3.10 Summary statistics of the MA(9) model with sample $D_t$. . . . . 43

3.11 Summary statistics of the AR(10) and MA(9) model with sample $D_t$ once zero-coefficients are removed from the models. . . . . 44

4.1 Evaluation criteria: HMAE and HRMSE with prediction results. . . . . 47

# List of figures


3.1 The daily return (first row) and daily realized volatility (second row) with linear plot (first column), histogram (second column) and auto-correlogram (third column). The graphs relate to the data of S&P 500 index from Aug 1st, 2018 to Dec 7th, 2019. . . . . 26

3.2 The logged daily close price of S&P 500 index from August 1st, 2018 to December 7th, 2019. . . . . 29

3.3 The time series and auto-correlogram of the innovation $z_t$. . . . . 30

3.4 The fitted series of our periodic model and innovation $z_t$ (first row); Residuals from the fitted periodic model (second row). . . . . 32

3.5 The time series of the sample daily return $r_t$. . . . . 33

3.6 The auto-correlogram and partial auto-correlogram of the sample daily return series: $r_t$. . . . . 34

3.7 The auto-correlogram and partial auto-correlogram of the residuals from AR(8) (first row), MA(9) (second row) and ARMA(2,2) (third row) models. . . . . 37

3.8 The auto-correlogram and partial auto-correlogram of the squared residual series from AR(8): $a_t^2$. . . . . 37

3.9 Model checking of the GARCH (1,1) model in equation (3.1). . . . . 39

3.10 The time series, auto-correlogram and partial auto-correlogram of the sample realized volatility $RV_t$ from August $1^{st}$, 2018 to December $7^{th}$, 2019. . . . . 41

3.11 The time series, auto-correlogram and partial auto-correlogram of the first difference of the sample realized volatility: $D_t$. . . . . 42

3.12 The auto-correlogram and partial auto-correlogram of the residuals from AR(10) model (first row) and MA(9) model (second row) about the sample series $D_t$. . . . . 45

3.13 The time series of the residuals from MA(9) model about sample $RV_t$. . 45

4.1 One-day ahead forecasting results of historical volatility (first row) and realized volatility (second row) from the evaluation period (displayed in red dash). The black solid line presents the realized volatility. . . . 47

# Chapter 1

# Introduction

Volatility is an index for measuring a stock's trading price changes over time. It is thought to be a good measurement of risk: higher volatility indicates higher risk and vice versa.

Volatility is not directly observable, but it can be estimated through the return, which is defined as a measure of stock price change. In finance, volatility is normally measured as the (conditional) standard deviation or variance of the return. Investors often measure and forecast the daily volatility through daily returns, which measure the profit of holding a stock over a day. However, many researchers argued that using the daily return with standard volatility models (generalized autoregressive conditional heteroskedasticity: GARCH- family models) provides a poor forecasting ( e.g., Koopman et al., 2005 ). Andersen and Bollerslev (1998) argued that in a 24-hours currency exchange market, the standard volatility model does provide a good estimate and forecast of the true volatility, but the associated data, the innovation of the daily return $a = \sigma\epsilon$, is very noisy due to the error term $\epsilon$. They also pointed out that the sum of the squared intraday returns provides a better measurement of the true volatility theoretically and empirically. Martens (2002) showed that the sum of squared intraday returns could also be used to measure the true daily volatility of a stock, but the data need to be adjusted first since the stock market do not trade on a 24-hours basis. With GARCH(1,1) model, it is shown that using the rescaled sum of squared intraday returns improves not only the measurement, but also the prediction of the daily volatility of a stock.

In the last two decades, because of the availability of high-frequency data, realized

volatility has gained popularity not only among researchers but also investors. However, in contrast to the simple daily return, which can be obtained online free of cost, access to accurate high-frequency data is often expensive. Consequently, many small investors still prefer to use the daily return to estimate and forecast daily volatility, and it is still the most popular approach among practitioners.

To overcome the shortcomings of the standard GARCH model, many new models, which are the extensions of the standard GARCH model, have been developed in the past two decades. Nelson (1991) proposed the exponential-GARCH model, which allows for asymmetric effects between positive and negative shock (leverage effect). Ding et al. (1993) proposed the asymmetric power autoregressive conditional heteroscedastic (APARCH) model which also allows the leverage effect. And it has been reported as significantly improve the goodness of fit of the model.

The previous studies showed that the sum of squared intraday returns has a good performance on measuring and predicting the true volatility, but they only used the standard GARCH model for the comparison, and other models from the GARCH family were not considered in their studies.

For this practicum, we obtained the Standard & Poor's 500 (S&P 500) index from August 01, 2018 to February 01, 2019 with 127 trading days for empirical data. Opening, closing and tick-by-tick prices were available for all trading days in our sample. The one-step-ahead predictions were generated by sub-samples which contain 90 observations each through the last 37 trading days. We used the GARCH family models with the daily return data to conduct forecasting. This process involved model selection, which means we tried different models and select the one which could better explain the data. Using the autoregressive moving average (ARMA) model, we conducted forecasting based on the realized volatility, the results from which are considered as the benchmark to compare with. Forecast evaluation was based on two loss functions: heteroscedasticity adjusted mean square error (HRMSE) and heteroscedasticity mean absolute error (HMAE), which are also used in the studies of Andersen et al. (1999) and Martens (2002). Since the true volatility is not directly observable, the realized volatility was taken as the estimation of the true volatility, and used in the loss function.

The result shows that the volatility forecasting conducted based on the realized volatility does provide a better forecast performance compared to the forecasting

using the daily return. The daily return is not able to estimate and forecast the daily volatility accurately.

The remainder of this practicum is organized as follows. In the next chapter, we present some considerations about the daily return, realized volatility, and the GARCH family models. In Chapter 3, we introduce our sample set and discuss the forecasting methodology. Meanwhile, we offer examples for explaining our forecast methodology in detail. The forecast evaluation and conclusions are presented in Chapter 4.

# Chapter 2

# Stock return and models

## 2.1 Volatility

Stocks, are common financial assets. Investors purchase stocks which they think will increase value. Meanwhile, they sell stocks that to be believed at risk. The behavior of investors can influence the demand and supply of stocks, that ultimately may affect the prices. Price fluctuation is a common phenomenon in stock markets as a result of investors adjusting their opinions constantly due to newly released information such as updated economic data, companies leadership, policy moves, and political change.

Volatility can be defined as an index of variation of a stock's trading price over time. It is a good measurement of risk: higher volatility indicates higher risk and vice versa. In finance, volatility is normally measured as the conditional standard deviation or the variance of the return (Tsay, 2014).

Historical volatility, realized volatility and implied volatility are the three most popular indices of volatility. While historical volatility measures daily return by using close to close price, realized volatility measures the price variability of intraday returns. In that way, realized volatility is an index of intraday price risk.

## 2.2 Daily return

Historical volatility measures the underlying securities by tracing the price changes during a certain period. To calculate it, we normally use "return".

Assume the closing price of a stock on time $t$ and $t-1$ are $P_t$ and $P_{t-1}$, respectively. Suppose a stock does not experience any dividends taking during the period from time $t-1$ to time $t$. The simple gross return $R_t + 1$ is defined as:

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

where $R_t$ is the simple return. Notice that, when $R_t$ is small, we have that $\log(1+R_t) \approx R_t$ (Tsay, 2014). Because of this, we refer to logarithm of the ratio $\frac{P_t}{P_{t-1}}$ as the return. In percentage, the return is expressed as

$$r_t = 100 \log(R_t + 1) = 100(\log(P_t) - \log(P_{t-1}))$$

If we let the time $t$ be day $t$, the $r_t$ presents the daily return.

One property of this transformation for the return $R_t$ is its additivity through multiperiod returns (Tsay, 2014). For example, The sum of $k$ single-period returns has the same form of the one for a single period return,

$$\begin{aligned} r_t[k] &= 100 \log(R_t[k] + 1) \\ &= 100 \log(\frac{P_t}{P_{t-k}}) \\ &= 100 \log(\frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \\ &= \sum_{i=0}^{k-1} r_{t-i} \end{aligned} \tag{2.1}$$

In general, a $k$-period simple gross return is defined as the product of $k$ one-period simple gross return:

$$\begin{aligned} 1 + R_t[k] &= \frac{P_t}{P_{t-k}} \\ &= \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1}) \ldots (1 + R_{t-k+1}) \\ &= \prod_{i=0}^{k-1} (1 + R_{t-i}) \end{aligned} \tag{2.2}$$

Simple gross return exhibits an equivalent property as the additivity of the return, that is valid when the value of simple return is small. For example, $k = 2$ by (2.2):

$$\begin{aligned} R_t[2] &= (1 + R_t)(1 + R_{t-1}) - 1 \\ &= R_t + R_{t-1} + R_t R_{t-1} \end{aligned} \tag{2.3}$$

When $R_t$ and $R_{t-1}$ are small, $R_t R_{t-1} \approx 0$. Thus, $R_t[2] \approx R_t + R_{t-1}$. In general, for any $k$ period:

$$R_t[k] \approx \sum_{i=0}^{k-1} R_{t-i}$$

## 2.3 Realized volatility

French et al. (1987) proposed a method of estimating the volatility of low-frequency return through high-frequency data. In recent years, this approach has attracted the interest of many people due to the availability of high-frequency data. Some studies point out that the daily realized volatility estimates, which is constructed from intraday returns, perform better on forecasting than the volatility estimates based on simple daily return.

In realized volatility, it is assumed that there are $n$ equally spaced intraday returns through day $t$, and the length of each space is determined as $\Delta_t$. Thus the trading period in day t is $T = \Delta_t d$. Let $r_{t,i}$ be the return on the market over the $i^{th}$ period on day $t$:

$$r_{t,i} = 100(\log(P_{t,i}) - \log(P_{t,i-1})); \quad i = 1, \ldots, d$$

$P_{t,i}$ is the stock price at the end of $i^{th}$ time interval on day t. Realized volatility assume $r_{t,i}$ follows the model:

$$r_{t,i} = \mu_i \Delta_t + \delta_i \epsilon_i \sqrt{\Delta_t} \quad i = 1, \ldots, d$$

Here, $\mu_i$ and $\delta_i$ are the drift rate and diffusion rate of $r_{t,i}$; $\epsilon_i$ are i.i.d random variables independent of the information available on day $t-1$, $\mathcal{F}_{t-1}$ and it follows a standard normal distribution. From equation (2.1):

$$\begin{aligned} r_t[n] &= \sum_{i=1}^{d} r_{t,i} \\ &= \sum_{i=1}^{d} (\mu_i \Delta_t + \delta_i \epsilon_i \sqrt{\Delta_t}) \end{aligned} \tag{2.4}$$

Let us assume that all information on day $t-1$ is available $\mathcal{F}_{t-1}$, and intraday returns are uncorrelated $Cov(r_{t,i}, r_{t,j}|\mathcal{F}_{t-1}) = 0, i \neq j$, we can get the realized volatility ($RV_t$), which is the conditional variance of the sum of $d$ returns on day $t$:

$$\begin{aligned} RV_t &= Var(r_t[d]|\mathcal{F}_{t-1}) \\ &= Var(\sum_{i=1}^{d} (\mu_i \Delta_t + \delta_i \epsilon_i \sqrt{\Delta_t})|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d} Var((\mu_i \Delta_t + \delta_i \epsilon_i \sqrt{\Delta_t})|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d} Var(\delta_i \epsilon_i \sqrt{\Delta_t}|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d} \Delta_t Var(\delta_i \epsilon_i|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d} \delta_i^2 \Delta_t Var(\epsilon_i) \\ &= \sum_{i=1}^{d} \delta_i^2 \Delta_t \end{aligned} \tag{2.5}$$

In stock markets, the drift $\mu_i$ is assumed to be close to zero when time interval $\Delta_t$

is small. Thus,

$$r_{t,i} \approx \delta_i \epsilon_i \sqrt{\Delta_t}$$

and,

$$\begin{aligned} E[\sum_{i=1}^{d}(r_{t,i})^2|\mathcal{F}_{t-1}] &= E[\sum_{i=1}^{d}\delta_i^2\epsilon_i^2\Delta_t|\mathcal{F}_{t-1}] \\ &= \sum_{i=1}^{d}\Delta_t\delta_i^2 E(\epsilon_t^2|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d}\Delta_t\delta_i^2 Var(\epsilon_t|\mathcal{F}_{t-1}) \\ &= \sum_{i=1}^{d}\delta_i^2\Delta_t \end{aligned} \tag{2.6}$$

Hence, the realized volatility $\sum_{i=1}^{d}(r_{t,i})^2$ is the conditional unbiased estimator of $\Delta_t\sigma_t^2 = \sum_{i=1}^{d}\delta_i^2\Delta_t$.

We can approximate the above equation to obtain the realized variance as:

$$RV_t \approx \sum_{i=1}^{d} r_{t,i}^2 \tag{2.7}$$

Although tick-by-tick returns is the finest interval returns which we can use in practice, the observed returns are distorted by microstructure noise from the market. Two common examples of microstructure noise are nonsynchronous trading and bid-ask bounce.

Nonsynchronous trading is caused by timing effects and trading effects. Timing effects occur due to different time zones of stock markets or different schedules of stock trading, while trading effects relate to infrequency trading, which means that stocks are not traded every consecutive interval (Miller et al., 1994). Bid-ask bounce refers to the situation that a stock price jumps up and down between ask-price and bid-price in a limited time.

Such noises may cause bias when we are using the empirical quadratic variation to estimate the underlying volatility, and the problem becomes more serious when the intervals become finer. Accordingly, an appropriate time interval must be sought in

order to acquire the bias-corrected data. Five-minute returns are generally acknowledged as the highest frequency returns which avoid most of the distortions from the effect of the microstructure (Areal and Taylor, 2002).

Andersen and Bollerslev (1998) constructed realized volatility estimates via cumulative squared intraday returns. They conducted simulations based on five-minute currency exchange rate of Deutsche Mark-U.S. Dollar and Japanese Yen-U.S. Dollar from October 1st, 1987 to September 30th, 1992, which excluded weekends due to the closure of the market. In their study, 288 squared five-minute intraday returns were summed as the daily realized volatility estimate, and total 260 daily realized volatility estimates were included:

$$RV_t = \sum_{i=1}^{288} r_{t,i}^2 \quad t = 1, \ldots, 260$$

However, the stock market is not like the currency exchange market which operates 24 hours a day. In fact, different stock markets also have different lengths of daily operation time. For example, New York Stock Exchange (NYSE) has regular trading hours from 9:30 am to 4 pm, Eastern Standard Time (EST); while Nasdaq also operates pre-market trading hours from 4 am to 9:30 am, and after-hours trading extends from 4 pm to 8 pm, EST. Thus, the number of 5-minute returns depends on stock market operations time with the absence of overnight 5-minutes returns and weekend returns. Under this circumstance, it is reasonable to consider the realized volatility as the combination of 5-minutes daytime returns and an overnight return (Martens, 2002).

$$RV_t = r_{t,0}^2 + \sum_{i=1}^{d} r_{t,i}^2$$

Here, and $r_{t,i}$ represents the $i^{th}$ intraday return on day $t$; $r_{t,0}$ represents the overnight return on day $t$: $r_{t,0} = \log(P_{t,0}) - \log(P_{t-1,d})$, $P_{t,0}$ is the opening price on day $t$ and $P_{t-1,d}$ is the closing price on day $t-1$.

The overnight return is more "noisy" than daytime 5-minutes return. Martens (2002) pointed out that market-related news, which is mostly international news, commonly released during night time, causing the changes of returns to be relatively large during the evening, comparing to daytime trading hours. However, an overnight return is not capable of showing all changes occur during a night. For example, after

an intense, volatile night, the open price of day $t$ is the same as the close price of the day $t-1$. Accordingly, the overnight return is 0. In order to account for the situation that nontrading hours stock returns are more volatile compares to trading hours stock returns, Martens (2002) suggested to remove the close-to-open return $r_{t,0}$ and construct a rescaled sum of intraday returns for $RV_t$:

$$RV_t = (1+c)\sum_{i=1}^{d} r_{t,i}^2$$

$c$ is a constant parameter so that $RV_t$ measures daily volatility.

In the absence of overnight return, it is reasonable to consider $c\sum_{i=1}^{d} r_{t,i}^2$ as an estimate of the sum of overnight intraday returns. Martens (2002) recommended to measure $c$ as:

$$c = \frac{Var_{co}}{Var_{oc}}$$

$$Var_{oc} = \frac{1}{N}\sum_{t=1}^{N}(\log(P_{t,d}) - \log(P_{t,0}))^2$$

$$Var_{co} = \frac{1}{N}\sum_{t=1}^{N}(\log(P_{t,0}) - \log(P_{t-1,d}))^2$$

We can correspondingly obtain realized volatility as:

$$RV_t = \frac{Var_{oc} + Var_{co}}{Var_{oc}}\sum_{i=1}^{d} r_{t,i}^2 \tag{2.8}$$

Notice that $Var_{oc}$ and $Var_{co}$ are the average squared open-to-close return and the average squared close-to-open return, respectively from day 1 to day $N$.

Another method, which is proposed by Areal and Taylor (2002), assigned different weights to intraday squared returns, according to the distribution of each period squared return in the total daily intraday squared returns. Hansen and Lunde (2002) proposed an alternative approach by removing the close-to-open return from (2.8) and define $1+c$ as the proportion of average variance of daily intraday squared returns in the average daily intraday squared returns $\frac{1}{N}\sum_{i=1}^{N}(\sum_{i=1}^{d} r_{t,i}^2)$.

## 2.4 Autoregressive conditional heteroskedasticity model and its extension

### 2.4.1 ARCH model

From section (2.2), we defined the stock return at time $t$ as $r_t$, $r_t = 100(\log(P_t) - \log(P_{t-1}))$. In order to understand $r_t$ in a proper perspective, we can first consider its conditional mean $\mu_t$ and variance $\sigma_t^2$:

$$\mu_t = E(r_t|\mathcal{F}_{t-1}), \;\; \sigma_t^2 = Var(r_t|\mathcal{F}_{t-1}) = E((r_t - \mu_t)^2|\mathcal{F}_{t-1})$$

Note that $\mathcal{F}_{t-1}$ is the information available at time $t-1$ (Tsay, 2014).

In finance, the mean and variance of an asset return play important roles in risk measurement. If an asset return evolves in a continuous manner (Tsay, 2014) and jump is rare, the mean and variance of the return is predictable. Understanding the return evolution helps us predict the price changes and control the associated risks. In this project, our goal is to understand the evolution pattern of the conditional variance and build an appropriate model for $\sigma_t^2$.

In regular time series, we normally assume the error term comes from a white noise process, which means errors are uncorrelated with mean and variance constant. However, in financial time series, it is very likely that volatility is at least autoregressive and conditionally heteroskedastic. Autoregression refers to a time series model that uses observations from previous time steps as input to a regression equation to predict the value at the next time step. Heteroskedasticity refers to the situation that the collection of variables have different variances (overtime, in our case). Therefore, we need a proper model to describe this phenomenon.

Autoregressive conditional heteroskedasticity(ARCH) model is one of the most popular models for analyzing heteroskedastic data, which is proposed by Engle (1982). The ARCH model with $n$ lags, ARCH(n), is expressed as:

$$r_t = \mu_t + a_t, \;\; a_t = \sigma_t \epsilon_t, \;\; \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{n} \alpha_i a_{t-i}^2$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\sum_{i=1}^{n} a_i < 1$.

We can obtain some basic properties of $a_t$ from above expression.

$$E(a_t) = E(\sigma_t \epsilon_t) = E(\sigma_t)E(\epsilon_t) = 0$$

$$Var(a_t) = E(a_t^2) = E(\sigma_t^2 \epsilon_t^2) = E(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{n} \alpha_i E(a_{t-i}^2) \tag{2.9}$$

In order to estimate $E(\sigma_t^2)$, we need to first introduce an important concept in time series, which is called weak stationarity. A time series $X_t$ is said to be weakly stationary if it has invariant first and second moments, i.e, $E(X_t) = \mu$, $Var(X_t) = \vartheta^2$, $\forall t \in \mathbb{Z}$, and the covariance between $X_t$ and $X_{t-k}$ only depends on time lag $k$: $Cov(X_t, X_{t-k}) = \gamma(k)$ (Tsay, 2014).

In an ARCH model, we assume $a_t$ varies in a fixed range. Statistically, it indicates that $\{a_t\}$ is a weak stationary time series (Tsay, 2014).

Let $E(a_t^2) = E(\sigma_t^2) = E(a_{t-i}^2) = \varsigma^2$, from (2.9):

$$E(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{n} \alpha_i E(a_{t-i}^2)$$

$$\Rightarrow \varsigma^2 = \alpha_0 + \sum_{i=1}^{n} \alpha_i \varsigma^2$$

$$\Rightarrow (1 - \sum_{i=1}^{n} \alpha_i)\varsigma^2 = \alpha_0$$

$$\Rightarrow \varsigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{n} \alpha_i}$$

Due to the fact that $\varsigma^2 \geq 0$ and $\alpha_0 > 0$, we require $1 - \sum_{i=1}^{n} \alpha_i > 0$, thus $0 \leq \sum_{i=1}^{n} \alpha_i < 1$.

Because of this linear inequality constraint, we normally consider the lag n to be smaller than 3.

If we consider $n = 3$, give an ARCH(3) model:

$$a_t = \sigma_t \epsilon_t, \;\; \epsilon_t \sim N(0,1), \;\; \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \alpha_3 a_{t-3}^2$$

with parameter constraints: $\alpha_1, \alpha_2, \alpha_3 > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 < 1$. From the constraint, we know that it is possible that at least one of the three parameters $\alpha_p$ is smaller than 0.33. Under this circumstance, the dependency of $\sigma_t^2$ on $a_{t-p}^2$ is too weak to be considered in this ARCH model(3). Therefore, in practice, the ARCH model of order two is more parsimonious than a model of higher order, then it is preferred.

ARCH(1) is the most popular model among practitioners for the use of ARCH model. It can be expressed as:

$$a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1), \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \tag{2.10}$$

where $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$.

Combine the (2.9) and (2.10), we can obtain the unconditional mean, variance and covariance of $a_t$ of ARCH(1) model:

$$E(a_t) = 0, \quad Var(a_t) = \frac{a_0}{1 - a_1},$$

$$\begin{aligned} Cov(a_t, a_{t-s}) &= Cov(\sigma_t \epsilon_t, \sigma_{t-s} \epsilon_{t-s}) \\ &= E[(\sigma_t \epsilon_t - 0)(\sigma_{t-s} \epsilon_{t-s} - 0)] \\ &= E(\sigma_t \epsilon_t \sigma_{t-s} \epsilon_{t-s}) \\ &= E(\epsilon_t) E(\sigma_t \sigma_{t-s} \epsilon_{t-s}) \\ &= 0. \end{aligned} \tag{2.11}$$

In applications, it is sometimes required the existence of higher order moments of $a_t$ and additional constraints (Tsay, 2014). For example, for study the tail behavior of $a_t$, the fourth moments of it is required. With the assumption of $\epsilon_t$ following a standard normal distribution:

$$\begin{aligned} E(a_t^4) = E(\sigma_t^4 \epsilon_t^4) &= E[\epsilon_t^4 (\alpha_0 + \alpha_1 a_{t-1}^2)^2] \\ &= E[\epsilon_t^4 (\alpha_0^2 + 2\alpha_0 \alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4)] \\ &= E(\epsilon_t^4) E(\alpha_0^2 + 2\alpha_0 \alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4)] \\ &= 3E(\alpha_0^2 + 2\alpha_0 \alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4) \\ &= 3[\alpha_0^2 + 2\alpha_0 \alpha_1 E(a_{t-1}^2) + \alpha_1^2 E(a_{t-1}^4)] \end{aligned} \tag{2.12}$$

Assume $a_t$ is also fourth moment invariant, $E(a_t^4) = M_4 > 0$, from (2.11):

$$M_4 = 3(\alpha_0^2 + 2\alpha_0\alpha_1\frac{\alpha_0}{1-\alpha_1} + \alpha_1^2 M_4) > 0$$

$$\Rightarrow M_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} > 0$$

$$\Rightarrow 1 - 3\alpha_1^2 > 0$$

$$\Rightarrow \alpha_1^2 < \frac{1}{3}$$

$$\Rightarrow 0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$$

Therefore, the unconditional kurtosis of $a_t$ is:

$$\frac{E(a_t^4)}{[Var(a_t)]^2} = \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)}\frac{(1-\alpha_1)^2}{\alpha_0^2} = \frac{3(1-\alpha_1)(1+\alpha_1)}{(1-3\alpha_1^2)} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}$$

$$\Rightarrow \frac{E(a_t^4)}{[Var(a_t)]^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} > 3$$

The kurtosis of a normal distribution is 3. Therefore, the distribution of $a_t$ has a heavier tail when compared to a standard normal distribution, which indicates that the $a_t$ from ARCH(1) is more likely to produce "outliers" in contrast to a variable following a standard normal distribution (Tsay, 2014).

In (2.10), the equation captures the phenomenon that a large shock $a_{t-1}$ is normally followed by another large shock $a_t$, although the influence of the past shocks will decrease as time goes by. This identity is called cluster volatility in financial time series, which can cause volatility to be overestimated (Tsay, 2014).

The estimation of ARCH(1) is often conducted through maximum likelihood estimation under the assumption of normality of $\epsilon_t$. Due to the fact that $\epsilon_t$ is a white noise process, $E(a_t|\mathcal{F}_{t-1}) = E(\sigma_t|\mathcal{F}_{t-1})E(\epsilon_t|\mathcal{F}_{t-1}) = 0$ and

$$Var(a_t|\mathcal{F}_{t-1}) = Var(\sigma_t\epsilon_t|\mathcal{F}_{t-1}) = E[\sigma_t^2\epsilon_t^2|\mathcal{F}_{t-1}] = E(\sigma_t^2|\mathcal{F}_{t-1})E(\epsilon_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2|\mathcal{F}_{t-1})$$

$$\Rightarrow Var(a_t|\mathcal{F}_{t-1}) = E(\sigma_t^2|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 a_{t-1}^2 = \sigma_t^2,$$

the conditional distribution of $a_t$: $a_t|\mathcal{F}_{t-1}$ follows a normal distribution with mean 0

and variance $\sigma_t^2$. The conditional density function of $a_t$ is expressed as:

$$f(a_t|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{a_t^2}{2\sigma_t^2})$$

The joint density function $f(a_T \dots a_t \dots a_1|a_0)$ is:

$$\begin{aligned} f(a_T \dots a_t \dots a_1|a_0) &= f(a_T|a_{T-1} \dots a_0) \dots f(a_t|a_{t-1} \dots a_0) \dots f(a_1|a_0) \\ &= f(a_T|\mathcal{F}_{T-1}) \dots f(a_t|\mathcal{F}_{t-1}) \dots f(a_1|\mathcal{F}_0) \\ &= \prod_{t=1}^{T} f(a_t|\mathcal{F}_{t-1}) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{a_t^2}{2\sigma_t^2}) \\ &= (2\pi)^{-\frac{T}{2}} \prod_{t=1}^{T} (\sigma_t^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}\sum_{t=1}^{T}\frac{a_t^2}{\sigma_t^2}) \end{aligned} \tag{2.13}$$

$$\begin{aligned} \ell(a_T \dots a_t \dots a_1|a_0) &= \log f(a_T \dots a_t \dots a_1|a_0) \\ &= -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{a_t^2}{\sigma_t^2} \end{aligned} \tag{2.14}$$

Plug (2.10) into the above function:

$$\ell(a_T \dots a_t \dots a_1|a_0) = -\frac{1}{2}\sum_{t=1}^{T}[\log(\alpha_0 + \alpha_1 a_{t-1}^2) + \frac{a_t^2}{\alpha_0 + \alpha_1 a_{t-1}^2}] - \frac{T}{2}\log(2\pi)$$

We can obtain the maximum likelihood estimates of $\alpha_0$ and $\alpha_1$: $\hat{\alpha}_0$ and $\hat{\alpha}_1$, by maximizing $\ell(a_T \dots a_t \dots a_1|a_0)$.

The forecast of $a_t$ is simple. Consider the a one-step ahead forecast of ARCH(1), all information is available at time $t$.

$$a_t(1) = \sigma_t(1)\epsilon_{t+1} \quad \sigma_t^2(1) = \hat{\alpha}_0 + \hat{\alpha}_1 a_t^2$$

$a_t(1)$ and $\sigma_t(1)$ denote the forecast of $a_{t+1}$ and $\sigma_{t+1}$, $t = 1, \dots, T$.

When the forecast step $l > 2$, the forecasting model can be simplified (Tsay, 2014). Consider $l = 2$:

$$\sigma_t^2(2) = \alpha_0 + \alpha_1 a_t^2(1) = \alpha_0 + \alpha_1 \sigma_t^2(1)\epsilon_{t+1}^2$$

As $E(\epsilon_{t+1}^2|\mathcal{F}_t) = 1$, the 2-step ahead forecast becomes:

$$\sigma_t^2(2) = \alpha_0 + \alpha_1 \sigma_t^2(1)$$

Let $l = 3$:

$$\sigma_t^2(3) = \alpha_0 + \alpha_1 a_t^2(2) = \alpha_0 + \alpha_1 \sigma_t^2(2)\epsilon_{t+2}^2 = \alpha_0 + \alpha_1[\alpha_0 + \alpha_1 \sigma_t^2(1)]\epsilon_{t+2}^2$$

with the assumption of $\epsilon_{t+2}^2 \sim N(0,1)$, $E(\epsilon_{t+2}^2|\mathcal{F}_{t+1}) = 1$. The 3-step ahead forecast satisfies:

$$\sigma_t^2(3) = \alpha_0 + \alpha_0\alpha_1 + \alpha_1^2 \sigma_t^2(1)$$

In general, when the forecast step $l > 1$:

$$\sigma_t^2(l) = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \cdots + \alpha_1^{l-2}) + \alpha_1^{t-1}\sigma_t^2(1) = \alpha_0 \frac{(1 - \alpha_1^{l-1})}{1 - \alpha_1} + \alpha_1^{l-1}\sigma_t^2(1)$$

Therefore,

$$\sigma_t^2(l) \to \frac{\alpha_0}{1 - \alpha_1}, \text{ as } l \to \infty$$

.

### 2.4.2 GARCH model

ARCH model is simple, but not always efficient. In practice, it is often either not able to the variation well enough or requires too many parameters. An extension of ARCH model is proposed by Bollerslev(1986), which is called generalized autoregressive conditional heteroskedasticity (GARCH) model. This model allows the past volatilities to affect the present volatility. GARCH model can be written as:

$$\begin{aligned} r_t &= \mu_t + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{2.15}$$

with parameters constraints $\sum_{i=1}^{m}\alpha_i + \sum_{j=1}^{s}\beta_j < 1$ and $\alpha_0 > 0,\ \alpha_i, \beta_j \geq 0$.

In order to understand the parameter constraints and the properties of (2.15), let us define a new parameter $\psi_t = a_t^2 - \sigma_t^2$, so that $\sigma_t^2 = a_t^2 - \psi_t$ and $\sigma_{t-j}^2 = a_{t-j}^2 - \psi_{t-j}$ (Tsay, 2014). Therefore, $E(\psi_t|\mathcal{F}_{t-1}) = E(a_t^2 - \sigma_t^2|\mathcal{F}_{t-1}) = E(a_t^2|\mathcal{F}_{t-1}) - E(\sigma_t^2|\mathcal{F}_{t-1}) = 0$. Let us plug the new parameter $\psi_t$ into (2.16):

$$\begin{aligned}
a_t^2 - \psi_t &= \alpha_0 + \sum_{i=1}^{m}\alpha_i a_{t-i}^2 + \sum_{j=1}^{s}\beta_j(a_{t-j}^2 - \psi_{t-j}) \\
\Rightarrow a_t^2 - \psi_t &= \alpha_0 + \sum_{i=1}^{m}\alpha_i a_{t-i}^2 + \sum_{j=1}^{s}\beta_j a_{t-j}^2 - \sum_{j=1}^{s}\beta_j\psi_{t-j} \\
\Rightarrow a_t^2 - \sum_{i=1}^{m}\alpha_i a_{t-i}^2 - \sum_{j=1}^{s}\beta_j a_{t-j}^2 &= \alpha_0 + \psi_t - \sum_{j=1}^{s}\beta_j\psi_{t-j} \\
\Rightarrow E(a_t^2) - \sum_{i=1}^{m}\alpha_i E(a_{t-i}^2) - \sum_{j=1}^{s}\beta_j E(a_{t-j}^2) &= \alpha_0 + E(\psi_t) - \sum_{j=1}^{s}\beta_j E(\psi_{t-j})
\end{aligned} \tag{2.16}$$

In GARCH model, we assume $\{a_t\}$ is a weak stationary series, which implies that $E(a_t^2) = E(a_{t-i}^2)$. By (2.17):

$$\begin{aligned}
E(a_t^2) - \sum_{i=1}^{m}\alpha_i E(a_{t-i}^2) - \sum_{j=1}^{s}\beta_j E(a_{t-j}^2) &= \alpha_0 + E(\psi_t) - \sum_{j=1}^{s}\beta_j E(\psi_{t-j}) \\
\Rightarrow (1 - \sum_{i=1}^{m}\alpha_i - \sum_{j=1}^{s}\beta_j)E(a_t^2) &= \alpha_0 \\
\Rightarrow E(a_t^2) &= \frac{\alpha_0}{(1 - \sum_{i=1}^{m}\alpha_i - \sum_{j=1}^{s}\beta_j)}
\end{aligned} \tag{2.17}$$

$E(a_t^2) \geq 0$, and $\alpha_0 > 0$, thus $1 - \sum_{i=1}^{m}\alpha_i - \sum_{j=1}^{s}\beta_j > 0$, which implies $0 \leq \sum_{i=1}^{m}\alpha_i + \sum_{j=1}^{s}\beta_j < 1$.

In practice, we often consider GARCH(1,1) as our target model:

$$\begin{aligned}
r_t = \mu_t + a_t,\ \ a_t = \sigma_t\epsilon_t,\ \ \epsilon_t \sim N(0,1) \\
\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1\sigma_{t-1}^2
\end{aligned} \tag{2.18}$$

where $\alpha_0 > 0; \alpha_1, \beta_1 \geq 0$ and $0 \leq \alpha_1 + \beta_1 < 1$.

The properties of GARCH(1,1) model can be easily obtained:

$$E(a_t) = 0,\ Cov(a_t, a_{t-s}) = 0$$

$$\begin{aligned} Var(a_t) = E(a_t^2) = E(\sigma_t^2) &= E(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2) \\ \Rightarrow E(a_t^2) &= \alpha_0 + \alpha_1 E(a_{t-1}^2) + \beta_1 E(\sigma_{t-1}^2) \\ \Rightarrow E(a_t^2) &= \alpha_0 + \alpha_1 E(a_{t-1}^2) + \beta_1 E(a_{t-1}^2) \\ \Rightarrow E(a_t^2) &= \frac{\alpha_0}{1-\alpha_1-\beta_1} \end{aligned}$$

In order to calculate the unconditional kurtosis of $a_t$: $\kappa_4$, we need to know the expression of $a_t^4$:

$$\begin{aligned} E(a_t^4) &= E[\epsilon^4(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2] \\ &= E(\epsilon_t^4) E(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2 \\ &= 3E[\alpha_0^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_0\alpha_1 a_{t-1}^2 + 2\alpha_0\beta_1\sigma_{t-1}^2 + 2\alpha_1\beta_1 a_{t-1}^2\sigma_{t-1}^2] \\ &= 3[\alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \beta_1^2 E(\sigma_{t-1}^4) + 2\alpha_0\alpha_1 E(a_{t-1}^2) + 2\alpha_0\beta_1 E(\sigma_{t-1}^2) + 2\alpha_1\beta_1 E(a_{t-1}^2\sigma_{t-1}^2)] \\ &= 3[\alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \beta_1^2 E(\sigma_{t-1}^4) + 2\alpha_0\alpha_1 E(a_{t-1}^2) + 2\alpha_0\beta_1 E(\sigma_{t-1}^2) + 2\alpha_1\beta_1 E(\epsilon_{t-1}^2\sigma_{t-1}^4)] \\ &= 3[\alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \beta_1^2 E(\sigma_{t-1}^4) + 2\alpha_0\alpha_1 E(a_{t-1}^2) + 2\alpha_0\beta_1 E(\sigma_{t-1}^2) + 2\alpha_1\beta_1 E(\sigma_{t-1}^4)] \\ &= 3[\alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \beta_1(\beta_1 + 2\alpha_1) E(\sigma_{t-1}^4) + 2\alpha_0\alpha_1 E(a_{t-1}^2) + 2\alpha_0\beta_1 E(\sigma_{t-1}^2)] \end{aligned} \tag{2.19}$$

Same as ARCH(1) model, we assume $a_t$ is fourth moment invariant, thus $E(a_t^4) = E(a_{t-1}^4)$. Meanwhile, $E(a_{t-1}^4) = E(\sigma_{t-1}^4\epsilon_{t-1}^4) = 3E(\sigma_{t-1}^4)$; $E(a_t^2) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$. From (2.19):

$$\begin{aligned} E(a_t^4) &= 3[\alpha_0^2 + \alpha_1^2 E(a_{t-1}^4) + \frac{1}{3}\beta_1(\beta_1 + 2\alpha_1) E(a_{t-1}^4) + 2\alpha_0(\alpha_1 + \beta_1) E(a_{t-1}^2)] \\ &\Rightarrow (1 - 3\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1) E(a_t^4) = 3\alpha_0^2 + 6\alpha_0(\alpha_1 + \beta_1)\frac{\alpha_0}{1-\alpha_1-\beta_1} \\ &\Rightarrow E(a_t^4) = \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1-\alpha_1-\beta_1)[1 - (\alpha_1+\beta_1)^2 - 2\alpha_1^2]} \end{aligned} \tag{2.20}$$

The kurtosis of the GARCH(1,1) is:

$$\kappa_4 = \frac{E(a_t^4)}{[Var(a_t)]^2} = \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1-\alpha_1-\beta_1)[1 - (\alpha_1+\beta_1)^2 - 2\alpha_1^2]} \frac{(1-\alpha_1-\beta_1)^2}{\alpha_0^2} = \frac{3[1 - (\alpha_1+\beta_1)^2]}{1 - (\alpha_1+\beta_1)^2 - 2\alpha_1^2}$$

if $1-(\alpha_1+\beta_1)^2-2\alpha_1^2>0$ and $\alpha_1\neq 0$, $1-(\alpha_1+\beta_1)^2-2\alpha_1^2<1-(\alpha_1+\beta_1)^2$. Therefore, $\kappa_4>3$, which means the distribution of GARCH(1,1) has a heavier tail than normal distribution. Notice that, as the extension of ARCH(1), GARCH(1,1) also experiences volatility cluster (Tsay, 2014).

The estimation of $\alpha_0$,$\alpha_1$ and $\beta_1$ is simple. We can use the same idea as estimate ARCH(1): use the conditional distribution of $a_t|\mathcal{F}_{t-1}$ to obtain the joint probability density function of $a_t$, under the normality assumption of $\epsilon_t$. Then maximize the joint likelihood to obtain $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$. Due to the space limitation of this project, we will not discuss the details here, see Tsay (2014).

As ARCH(1), the forecast of GARCH(1,1) can be simplified. Consider 2-steps ahead forecast $\sigma_t^2(2)$:

$$\begin{aligned}\sigma_t^2(2) &= \alpha_0+\alpha_1 a_t^2(1)+\beta_1\sigma_t^2(1)\\ &= \alpha_0+\alpha_1\sigma_t^2(1)\epsilon_{t+1}^2+\beta_1\sigma_t^2(1)\\ &= \alpha_0+(\alpha_1+\beta_1)\sigma_t^2(1)+\alpha_1\sigma_t^2(1)(\epsilon_{t+1}^2-1)\end{aligned} \tag{2.21}$$

Here, we notice that $\epsilon_{t+1}^2\sim N(0,1)$, which means $E(\epsilon_{t+1}^2-1)=E(\epsilon_{t+1}^2)-1=0$. According, we can obtain our 2-step ahead forecast $\sigma_t^2(2)=\alpha_0+(\alpha_1+\beta_1)\sigma_t^2(1)$.

We can extend the result to $l$-step ahead forecast :

$$\sigma_t^2(l)=\alpha_0\frac{1-(\alpha_1+\beta_1)^{l-1}}{1-\alpha_1-\beta_1}+(\alpha_1+\beta_1)^{l-1}\sigma_t^2(1)$$

Therefore, the $l$-step forecast of volatility $\sigma_t^2(l)$ only depends on the forecast origin $t$ and its volatility $\sigma_t^2$.

$$\text{And } \sigma_t^2(l)\to\frac{\alpha_0}{1-\alpha_1-\beta_1} \text{ as } l\to\infty$$

.

### 2.4.3 The extensions of GARCH model

Many extension of the GARCH model have been developed since it was first introduced.

### The GARCH-M model

In finance, we can obtain GARCH-M model by considering the direct impact of the volatility of the return. Here, the "M" stands for "in the mean" (Tsay, 2005). The GARCH-M(1,1) model is expressed as:

$$r_t = \mu + c\sigma_t^2 + a_t, \;\; a_t = \sigma_t \epsilon_t, \;\; \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Here, $\mu$ and $c$ are constant parameters and $\alpha_0 > 0$; $\alpha_1, \; \beta_1 \geq 0$.

We usually refer to "c" as a risk premium parameter. If "c" is positive, the return $r_t$ is positively related to its volatility and vice versa.

The formulation of the GARCH-M(1,1) model implies that there are serial correlations in $r_t$, which are introduced by $\sigma_t^2$ (Tsay, 2014). If the risk parameter $c = 0$, the serial correlation will not longer exist.

### The EGARCH model

The original GARCH model requires its parameters to be non-negative. This restriction has been criticized as being over restrictive. In order to address this problem, Nelson (1991) proposed Exponential GARCH (EGARCH) model, which has no restriction on $\alpha_i$ and $\beta_j$; it also allows for asymmetric effects between the positive and negative shocks. EGARCH(1,1) can be defined as:

$$r_t = \mu_t + a_t, \;\; a_t = \sigma_t \epsilon_t, \;\; \epsilon_t \sim N(0,1)$$

$$(1 - \alpha_1 B)(\log(\sigma_t^2) - \alpha_0) = Bg(\epsilon_t)$$

Here, $\alpha_0$ and $\alpha_1$ are parameters. $B$ is a lag operator, which means $B\log(\sigma_t^2) = \log(\sigma_{t-1}^2)$ and $Bg(\epsilon_t) = g(\epsilon_{t-1})$.

The asymmetric effect is determined by $g(\epsilon_t)$, which is defined as the following function:

$$g(\epsilon_t) = \delta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)]$$

where $\delta$ and $\gamma$ are parameters.

Compared to the GARCH model, the conditional variance $\sigma_t^2$ has become $\log(\sigma_t^2)$ in the EGARCH model, which allows negative value. Therefore, the constraints on the parameters are relaxed (Tsay, 2014). The asymmetric effect of the shocks is introduced by $g(\epsilon_{t-1})$. To better understand this character, we can express $\log(\sigma_t^2)$ from the EGARCH(1,1) in another way:

$$\begin{aligned} (1-\alpha_1 B)(\log(\sigma_t^2)-\alpha_0) &= Bg(\epsilon_t) \\ \Rightarrow \log(\sigma_t^2)-\alpha_0-\alpha_1(\log(\sigma_{t-1}^2)-\alpha_0) &= g(\epsilon_{t-1}) \\ \Rightarrow \log(\sigma_t^2)-\alpha_1\log(\sigma_{t-1}^2) &= \alpha_0(1-\alpha_1)+g(\epsilon_{t-1}) \end{aligned} \tag{2.22}$$

Since $\epsilon_t \sim N(0,1)$, $|\epsilon_t|$ is following a folded normal distribution, $E(|\epsilon_t|) = \sqrt{\frac{2}{\pi}}$. Thus, $g(\epsilon_{t-1})$ can be rewritten as:

$$g(\epsilon_{t-1}) = \begin{cases} \delta\epsilon_{t-1}+\gamma\epsilon_{t-1}-\gamma\sqrt{\frac{2}{\pi}}, & \text{if } \epsilon_{t-1}>0 \\ \delta\epsilon_{t-1}-\gamma\epsilon_{t-1}-\gamma\sqrt{\frac{2}{\pi}}, & \text{if } \epsilon_{t-1}<0 \\ -\gamma\sqrt{\frac{2}{\pi}}, & \text{if } \epsilon_{t-1}=0 \end{cases} \tag{2.23}$$

Note that the sign of $\epsilon_t$ depends on $a_t$.

Plug (2.23) into (2.22):

$$\log(\sigma_t^2) = \begin{cases} \alpha_1\log(\sigma_{t-1}^2)+\left[\alpha_0(1-\alpha_1)-\gamma\sqrt{\frac{2}{\pi}}\right]+(\delta+\gamma)\frac{a_{t-1}}{\sigma_{t-1}}, & \text{if } a_{t-1}>0 \\ \alpha_1\log(\sigma_{t-1}^2)+\left[\alpha_0(1-\alpha_1)-\gamma\sqrt{\frac{2}{\pi}}\right]+(\delta-\gamma)\frac{a_{t-1}}{\sigma_{t-1}}, & \text{if } a_{t-1}<0 \\ \alpha_1\log(\sigma_{t-1}^2)+\left[\alpha_0(1-\alpha_1)-\gamma\sqrt{\frac{2}{\pi}}\right], & \text{if } a_{t-1}=0 \end{cases} \tag{2.24}$$

In (2.24), we can transform $\frac{a_{t-1}}{\sigma_{t-1}}$ to $|\frac{a_{t-1}}{\sigma_{t-1}}|$ and take exponential on both sides:

$$\sigma_t^2 = \begin{cases} \sigma_{t-1}^{2\alpha_1}\eta\exp[(\gamma+\delta)|\frac{a_{t-1}}{\sigma_{t-1}}|], & \text{if } a_{t-1}>0 \\ \sigma_{t-1}^{2\alpha_1}\eta\exp[(\gamma-\delta)|\frac{a_{t-1}}{\sigma_{t-1}}|], & \text{if } a_{t-1}<0 \\ \sigma_{t-1}^{2\alpha_1}\eta, & \text{if } a_{t-1}=0 \end{cases}$$

Here, $\eta = \exp\left[\alpha_0(1-\alpha_1)-\gamma\sqrt{\frac{2}{\pi}}\right]$.

The coefficient $-\delta$ shows that positive and negative $a_{t-1}$ offer asymmetric effects to $\sigma_t^2$.

### The TGARCH model

Threshold GARCH (TGARCH) model, which was proposed by Zakoian (1994), is another useful model in financial time series for the volatility with asymmetric responses to positive and negative returns. In practice, the volatility of a stock return often rises higher in response to negative returns other than positive. In financial time series, we call this phenomenon leverage effect(Ding et al., 1993).

The TGARCH(1,1) model is defined as:

$$r_t = \mu_t + a_t,\ a_t = \sigma_t \epsilon_t,\ \ \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma_1 \mathbb{I}_{t-1}) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$\alpha_0 > 0;\ \alpha_1,\ \gamma_1,\ \beta_1 \geq 0$. $\mathbb{I}_{t-1}$ is an indicator function:

$$\mathbb{I}_{t-1} = \begin{cases} 1, & \text{if } a_{t-1} < 0 \\ 0, & \text{otherwise} \end{cases}$$

From the model, it is easy to see that a negative $a_{t-1}$ contributes $(\alpha_1 + \gamma_1) a_{t-1}^2$ to $\sigma_t^2$ while a positive $a_{t-1}$ only contributes $a_1 a_{t-1}^2$.

### The APARCH model

The TGARCH model belongs to the family of asymmetric power autoregressive conditional heteroscedastic (APARCH) model which was proposed by Ding et al. (1993). The APARCH(1,$\theta$,1) model is commonly used in practice. It is expressed as:

$$r_t = \mu_t + a_t,\ a_t = \sigma_t \epsilon_t,\ \ \epsilon_t \sim N(0,1)$$

$$\sigma_t^\theta = \alpha_0 + \alpha_1 (|a_{t-1}| + \gamma_1 a_{t-1})^\theta + \beta_1 \sigma_{t-1}^2$$

Note that $\alpha_0 > 0; \theta, \alpha_1, \beta_1 \geq 0$ and $-1 < \gamma_1 < 1$.

Notice that, when $\gamma_1 = 0$, the APARCH model does not have the asymmetric effect of $a_t$; positive and negative $a_{t-1}$ have the same effect on $\sigma_t^\theta$. Parameter $\theta$ plays an essential role in the APARCH model, but it does not have a specific interpretation.

Ding et al. (1993) mentioned that there is no any obvious reason for the conditional

variance $\sigma_t^2$ to be linearly correlated with $a_t^2$ in the GARCH model, and their proposed general model covers most cases. Tsay (2014) pointed out that although there is no good interpretation for $\theta$, it does improve the goodness of fit.

# Chapter 3

# Data and empirical results

## 3.1 Data selection

S&P 500 index is a stock market indicator whose objective is to measure market capitalization of the leading 500 companies that trade in the American stock markets. The S&P 500 index is based on the last transaction price of each stock. This index covers about 80% available market capitalization, provides a broad view of U.S. financial market health.

For this practicum, we use the data of S&P 500 index from August 01, 2018 to February 01, 2019 (127 days), excluding weekends and holidays, through the Bloomberg terminal. The opening and closing times of S&P 500 index are the same as NYSE, which is 9:00 and 4:15 pm, respectively. Although some stocks in S&P 500 index have pre-market and post-market trading sessions, the majority trades at any time during the trading period. Therefore, we do not consider price changes out of the normal trading period.

We divide our observation into two periods, one for estimation and the other for evaluation. This means that the parameters are estimated based on the data from the estimation period, and one-step-ahead forecasts are based on the models produced in the estimation section. This method is introduced in the work of Hasen et al. (2005). Because the availability of real data in the evaluation period, we can use this for examining the forecasting performance of our models.

For intraday data, we separate the normal trading time of a day into 81 five-minute

intervals. The last price in every five minutes interval is selected as the interval closing price. The estimation of the weight parameter in realized volatility model, $\hat{c} \approx 0.2609$, which is higher than what Martens (2002) obtained from 1990-1994 S&P 500 future index, $\hat{c} \approx 0.2099$. It indicates that the overnight returns are more volatile in late 2018, than in early 1990s.

## 3.2 Descriptive statistic

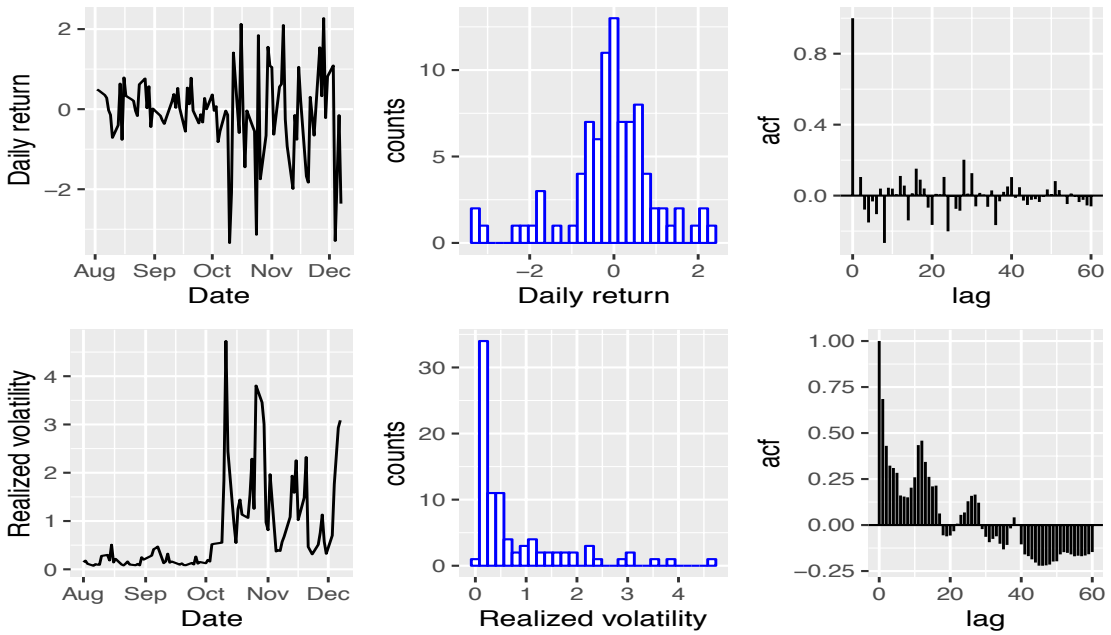
The graphs of Figure 3.1 are the time series, histogram, and auto-correlogram of the daily return and realized volatility for the data from the estimation period. The summary statistics are presented in Table 3.1.

The daily return exhibits signs of volatility after October 2018. Seven large shocks can be observed: four of them are positive, and the others are negative. The auto-correlogram shows there is no strong correlation in the return series except the seventh lag.

The realized volatility also experienced a significant increase in October 2018. Before this date, the volatility was stable and below 0.5. Afterward, it abruptly changed up and down between 0.4 and 5. The auto-correlogram structure presents a weak periodic pattern in the series.

| | Daily return | Realized volatility |
|---|---|---|
| Mean | -0.074 | 0.792 |
| Minimum | -3.342 | 0.074 |
| Maximum | 2.271 | 4.729 |
| Standard deviation | 1.060 | 0.964 |
| Skewness | -0.752 | 1.857 |
| Kurtosis | 1.586 | 3.211 |

Table 3.1: Summary statistics of the daily return and realized volatility of S&P 500 index from August 1st, 2018 to December 7th, 2019.




s

Figure 3.1: The daily return (first row) and daily realized volatility (second row) with linear plot (first column), histogram (second column) and auto-correlogram (third column). The graphs relate to the data of S&P 500 index from Aug 1st, 2018 to Dec 7th, 2019.

## 3.3 Forecast methodology and criteria

As mentioned in the previous section, we separate our data into two periods: one for estimation and the other for evaluation. Our forecasting was conducted as follows. We obtained 37 sub-samples which contained 90 days of observations each. An individual sub-sample was used for estimating one-day-ahead volatility. For example, the first sub-sample contained data from the $1^{st}$ of August to $7^{th}$ of December 2018; we use the model whose estimates are based on this sub-sample to forecast the volatility of $10^{th}$ of December 2018. The second sub-sample, which contained observations from $2^{nd}$ of August to $10^{th}$ of December 2018, was used for building a model to forecast the volatility of $11^{th}$ of December 2018. This forecasting method used the idea of "rolling windows": we conducted forecasting based on the sub-samples. A sub-sample moves one trading day ahead each time with a constant sub-sample size 90. The process repeats until we obtain 37 out-of-sample forecast result. Please note that the models, which are estimated based on the sub-samples, belong to the GARCH model family.

The next step was to evaluate the performance of our forecasting. Although the true volatility $\sigma_t^2$ is not observable, the realized volatility $RV_t$ is an observable unbiased estimator of $\Delta_t\sigma_t^2$, which has been proved in Chapter 2. Therefore, it is reasonable to use $RV_t$ to evaluate our forecast performance. Here, we consider two forecast criteria:

$$\begin{aligned} HRMSE &= \sqrt{\frac{1}{37}\sum_{t=91}^{127}\left(1-\frac{RV_t}{\Delta_t\hat{\sigma}_t^2}\right)^2} \\ HMAE &= \frac{1}{37}\sum_{t=91}^{127}\left|1-\frac{RV_t}{\Delta_t\hat{\sigma}_t^2}\right| \end{aligned} \tag{3.1}$$

Here, $\Delta_t\hat{\sigma}_t^2$ represents the forecast volatility of day $t$. The above two forecast criteria are also considered by Andersen et al. (1999) and Martens (2002).

Now, we need a benchmark to be used to compare those two forecast criteria for GARCH models. Therefore, with the realized volatility as the data of our benchmark, we process one-day ahead forecast through ARMA model, then use equation (3.1) to calculate the benchmark forecast criteria.

Here, we introduce an autoregressive (AR) model (Shumway and Stoffer, 2017):

$$x_t = \phi_0 + \sum_{i=1}^{k}\phi_i x_{t-i} + a_t$$

where $a_t \sim N(0, \sigma_a^2)$.

In this model, The value of $x_t$ only depends on the past lagged series $\{x_{t-i}\},\ i = 1, \ldots, k$.

Another possibility is to use the moving average (MA) model (Shumway and Stoffer, 2017):

$$x_t = \alpha_0 - \sum_{j=1}^{s}\alpha_j a_{t-j}^2 + a_t$$

where $\{a_t\} \sim N(0, \sigma_a^2)$. In MA model, the value of $x_t$ depends on the past error term $\{a_{t-j}\},\ j = 1, \ldots, s$.

Although those two models look different, they can be easily transformed into each other by expanding the series $\{x_{t-i}\}$.

In practice, we often require a high order of AR or MA model with many parameters to describe the data adequately. By combining the AR and MA process, we can often achieve an appropriate model with a relatively small amount of parameters (Tsay, 2014). This model is named as the autoregressive moving average (ARMA) model:

$$x_t = \phi_0 + \sum_{i=1}^{k} \phi_i x_{t-i} + a_t - \sum_{j=1}^{s} \alpha_j a_{t-j}$$

$a_t$ is the white noise process with mean 0 and variance $\sigma_a^2$ at time $t$. In this model, the value of $x_t$ does not only depend on the past lagged values $x_{t-1},\ x_{t-2}, \ldots, x_{t-i}$ but also the error terms from the current and past periods, which improves the efficiency of predicting (Tsay, 2014).

It is worth noticing that the volatility $\sigma_t^2$ from GARCH model can be written in a ARMA-like model form (Tsay, 2014). From equation (2.16), let $\psi_t = a_t^2 - \sigma_t^2$:

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \\ &= \alpha_0 + \sum_{i=1}^{m} \alpha_i (\sigma_{t-i}^2 + \psi_{t-i}) + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \\ &= \alpha_0 + \sum_{i=1}^{m} \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{m} \alpha_i \psi_{t-i} \end{aligned} \tag{3.2}$$

It is obvious that $\psi_t$ is not a white noise process, thus the above transformation is not exactly a ARMA model.

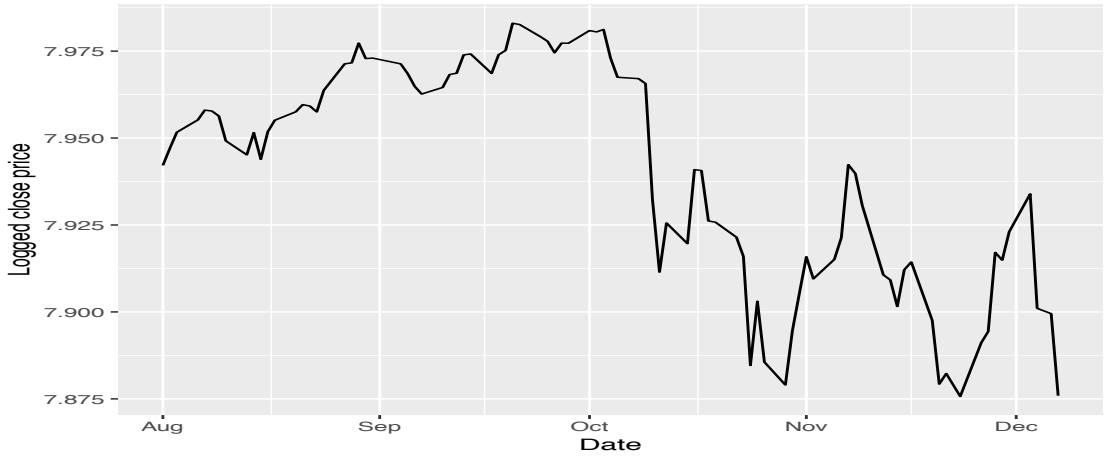
## 3.4 Volatility model of the daily returns

In this section, we will give an example of conducting one-step-ahead forecast of volatilities. In this example, we try to forecast the volatility of December $10^{th}$ 2018 (day 91) with the daily return data.

Figure 3.2 shows the time series of the logged daily close price of S&P 500 index during the estimation period. This graph has several features. First, the logged close price shows a downward trend; it experienced a rapid drop at the beginning of October, 2018. Second, the price exhibits a weak cyclical pattern with an variable

duration. A possible explanation for this weak periodic pattern is that it is an artifact due to the data size. However, since the project is about forecasting volatility in a short term; it is common that the time series shows a weak or no cyclic pattern.

Before building a model for volatility, we first consider removing the time trend and periodic pattern.




s

Figure 3.2: The logged daily close price of S&P 500 index from August 1st, 2018 to December 7th, 2019.

We begin with removing the time trend from our data. Let $C_t$ be the logged close price at time t, and use time index t as an explanatory:

$$C_t = \alpha_0 + \alpha_1 t + z_t$$

Here, $z_t$ is the innovation from the time series at time $t$.

Table 3.2 is the summary statistics of the fitted linear regression model:

$$C_t = 7.98 - 0.0009\, t + z_t$$

with standard error 0.004712 and $8.993 * 10^{-5}$, respectively. The intercept is positive and the time slope is negative, while they are both significant. The standard error of $z_t$ is 0.02216.

Figure 3.3 shows the time series and auto-correlogram of the innovation series $z_t$. From the time series plot in Figure 3.3, it is obvious that the time trend has

Call:
lm(formula = logclose ∼ time)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|
| -0.046274 | -0.017494 | -0.002738 | 0.016663 | 0.040219 |

Coefficients:

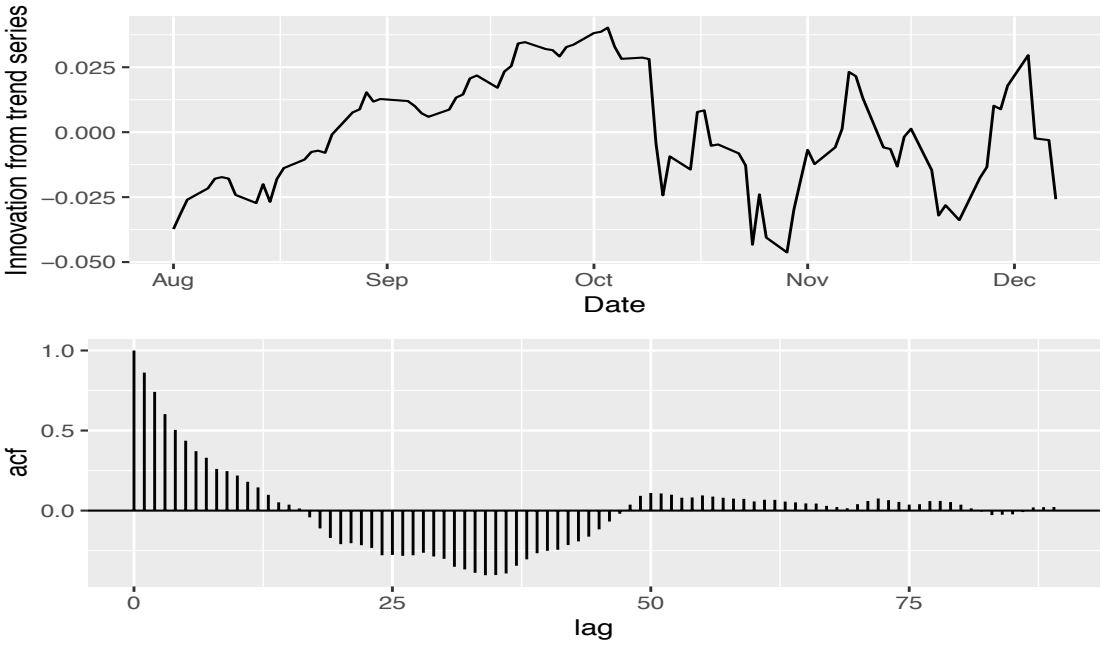
| | Estimate | Std. Error | t value | P(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 7.980e+00 | 4.712e-03 | 1693.649 | < 2e-16 *** |
| time | -8.736e-04 | 8.993e-05 | -9.714 | 1.39e-15 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02216 on 88 degrees of freedom
Multiple R-squared: 0.5174, Adjusted R-squared: 0.5119
F-statistic: 94.35 on 1 and 88 DF, p-value: 1.393e-15

Table 3.2: Summary statistics of the sample trend model.

been removed. In the auto-correlogram, the correlation first decay slowly; then starts to increase since the 16th lag. This phenomenon occur in circles, which suggests a periodical pattern in the series; now we will try to remove it.




s

Figure 3.3: The time series and auto-correlogram of the innovation $z_t$.

The auto-correlogram in Figure 3.3 suggests every circle contains 62 observations. Let $w$ as a frequency index: $w = \frac{1}{62}$ and $t$ as a time index: $t = 1, \ldots, 90$. The periodic

model is defined as (Shumway and Stoffer, 2017):

$$p_t = \beta_0 + \beta_1 \cos(2\pi wt) + \beta_2 \sin(2\pi wt)$$

Please note that $z_t = p_t + x_t$. $x_t$ is the logged close price after removing time trend and periodic pattern.

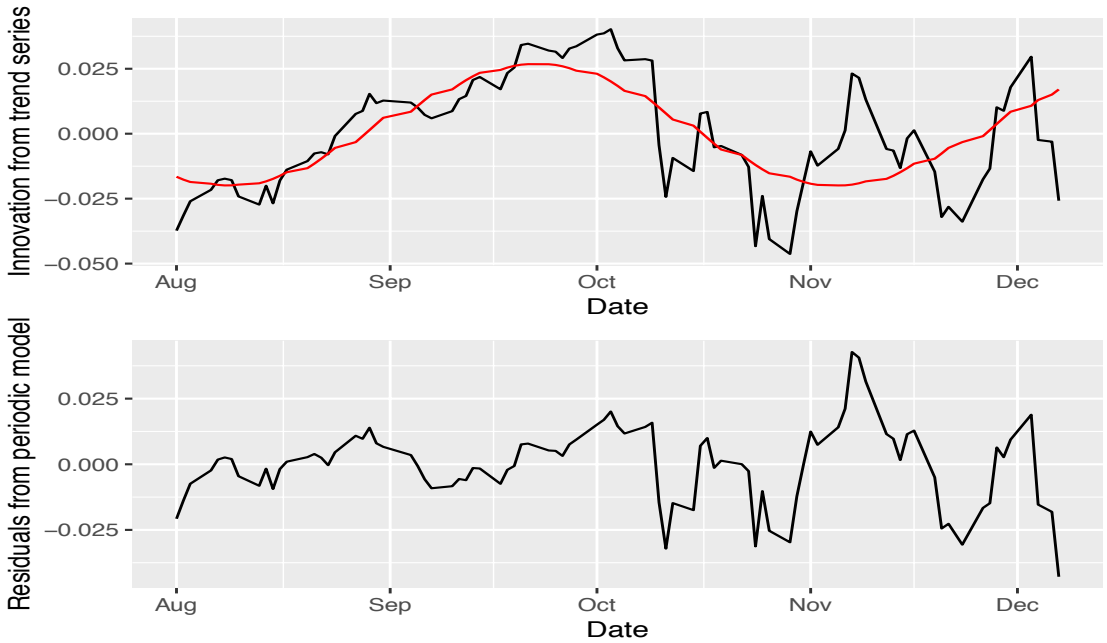
Table 3.3 shows the summary statistics of our periodic model. Under 5% significant level, $\beta_0, \beta_1$ and $\beta_2$ are all significant. The standard error of residuals is 0.01509. The fitted periodic model can be written as:

$$p_t = 0.0034 - 0.0187 \cos(2\pi wt) - 0.0140 \sin(2\pi wt)$$

Figure 3.4 maps the fitted series (the smooth line in the top graph) and residuals (bottom graph) of the periodic model.

| Call: | | | | |
|---|---|---|---|---|
| lm(formula = m1$residuals ~ cos(2*pi/62*time) + sin(2*pi/62*time)) | | | | |
| Residuals: | | | | |
| Min | 1Q | Median | 3Q | Max |
| -0.042836 | -0.008003 | 0.001489 | 0.009395 | 0.042753 |
| Coefficients: | | | | |
| | Estimate | Std. Error | t value | P(> \|t\|) |
| (Intercept) | 0.003436 | 0.001669 | 2.059 | 0.0425 * |
| cos(2*pi/62*time) | -0.018671 | 0.002288 | -8.160 | 2.33e-12 *** |
| sin(2*pi/62*time) | -0.014029 | 0.002321 | -6.043 | 3.66e-08 *** |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |
| Residual standard error: 0.01509 on 87 degrees of freedom | | | | |
| Multiple R-squared: 0.5417, Adjusted R-squared: 0.5311 | | | | |
| F-statistic: 51.41 on 2 and 87 DF, p-value: 1.826e-15 | | | | |

Table 3.3: Summary statistics of the sample periodic model.




s

Figure 3.4: The fitted series of our periodic model and innovation $z_t$ (first row); Residuals from the fitted periodic model (second row).

Now, we can express the daily return as:

$$r_t = 100(\log(P_{d,t}) - \log(P_{d,t-1}))$$

Here, $\log(P_{d,t})$ is the modified logged close price on day $t$, which has been detrended and deperiodiced.

In section 2.4.2, we consider the return series $r_t$ satisfies:

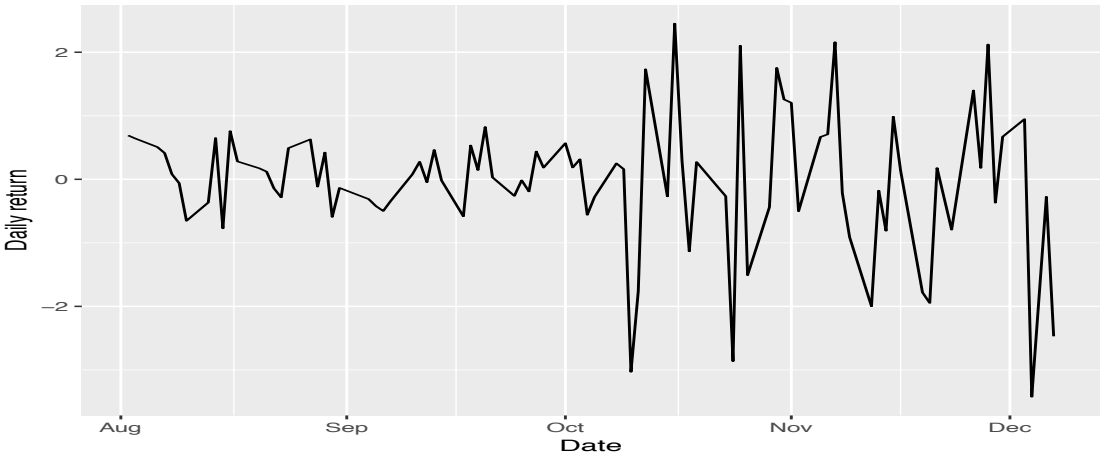
$$r_t = \mu_t + a_t,\ a_t = \sigma_t \epsilon_t,\ \epsilon_t \sim N(0,1)$$

In order to predict $a_t$, we first need to remove $\mu_t$, which is the conditional mean of $r_t$: $\mu_t = E(r_t|\mathcal{F}_{t-1})$, from the series.

From Figure 3.5, it is obvious that the mean of $r_t$ moves up and down over time. Since we already eliminated the time trend and periodical pattern from $r_t$, it is possible that other factors influence our series.

In the statistical jargon we often say that a time series $x_t$ is auto-correlated at

lag-k if $\rho_k = \frac{Cov(x_t, x_{t-k})}{\sqrt{Var(x_t)Var(x_{t-k})}} \neq 0$, which indicates that $x_t$ depends on $x_{t-k}$. The collection of $\rho_k$ is called: auto-correlation function (ACF). In contrast, by controlling the contribution from the values of shorter lags $x_{t-i},\ 0 < i < k$, partial correlation function (partial ACF) gives the partial correlation between $x_t$ and $x_{t-k}$ with their own value (Tsay 2005).

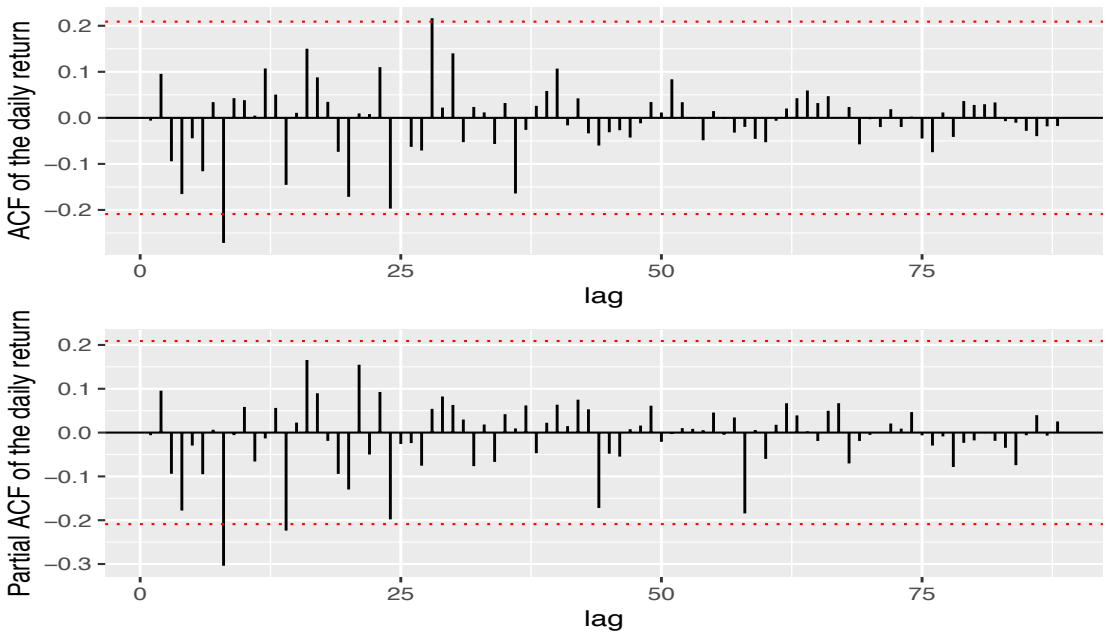



s

Figure 3.5: The time series of the sample daily return $r_t$.

In Figure 3.6, the first row and second row present the sample ACF and PACF of the daily return series, respectively. The ACFs and PACFs do not appear any high values. The eighth lag is significantly correlated, which should be considered in modeling the mean series $\mu_t$. Here, we consider three models for $\mu_t$: AR model, MA model and ARMA model.

Tsay (2014) mentioned that ACF is useful for identify MA process because ACF cuts off at lag-p for a MA(p) series. For AR process, PACF cuts off at lag-q for a AR(q) series. According to those rules, we initially consider AR(9), MA(9) as the candidate models for our sample. ARMA process combines AR and MA processes. Comparing to AR and MA models, it can often obtain a good fit of data with a smaller number of parameters. In our case, we initially considered ARMA(4,4) as a candidate model for covering a sufficient amount of correlations.




s

Figure 3.6: The auto-correlogram and partial auto-correlogram of the sample daily return series: $r_t$.

Through model fitting, we notice that the coefficients of some lags in the models are not significant. After removing their corresponding coefficients and refitting the models, we obtain an AR(8), a MA(9) and an ARMA(2,2) model.

| Model | Estimate | Std. Error | t value | P($>\|t\|$) |
|---|---|---|---|---|
| AR(8): | | | | |
| ar1 | -0.0301 | 0.0100 | -0.3011 | 0.7633 |
| ar2 | 0.0424 | 0.1008 | 0.4210 | 0.6738 |
| ar3 | -0.1370 | 0.1058 | -1.2942 | 0.1956 |
| ar4 | -0.2325 | 0.1077 | -2.1590 | 0.0308 * |
| ar5 | -0.0993 | 0.1063 | -0.9341 | 0.3502 |
| ar6 | -0.1210 | 0.1051 | -1.1518 | 0.2494 |
| ar7 | 0.0002 | 0.1063 | 0.0015 | 0.9988 |
| ar8 | -0.3606 | 0.1053 | -3.4256 | 0.0006 *** |
| intercept | -0.0122 | 0.0536 | -0.2272 | 0.8202 |
| $\sigma^2$ estimated as 0.8986: log-likelihood = -122.27, aic = 264.55 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.4: Summary statistics of the AR(8) model.

| Model | Estimate | Std. Error | t value | P($>\|t\|$) |
|---|---|---|---|---|
| MA(9): | | | | |
| ma1 | 0.0039 | 0.1316 | 0.0295 | 0.9764 |
| ma2 | -0.0602 | 0.1156 | -0.5208 | 0.6025 |
| ma3 | -0.3737 | 0.1306 | -2.8621 | 0.0042 ** |
| ma4 | -0.1940 | 0.1331 | -1.4568 | 0.1452 |
| ma5 | -0.1347 | 0.1106 | -1.2181 | 0.2232 |
| ma6 | -0.2601 | 0.1090 | -2.3866 | 0.0170 * |
| ma7 | 0.2141 | 0.1304 | 1.6415 | 0.1007 |
| ma8 | -0.4857 | 0.1178 | -4.1230 | 3.739e-05 *** |
| ma9 | 0.2904 | 0.1137 | 2.5546 | 0.0106 * |
| intercept | -0.0035 | 0.0131 | -0.2635 | 0.7922 |
| $\sigma^2$ estimated as 0.7098: log-likelihood = -116.5, aic = 255.01 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.5: Summary statistics of the MA(9) model.

| Model | Estimate | Std. Error | t value | P($>\|t\|$) |
|---|---|---|---|---|
| ARMA(2,2): | | | | |
| ar1 | -0.1745 | 0.1948 | -0.8960 | 0.3703 |
| ar2 | -0.7067 | 0.2135 | -3.3102 | 0.0009 *** |
| ma1 | 0.1784 | 0.1416 | 1.2598 | 0.2077 |
| ma2 | 0.8695 | 0.1640 | 5.3021 | 1.145e-07 *** |
| intercept | -0.0259 | 0.1180 | -0.2193 | 0.8264 |
| $\sigma^2$ estimated as 1.047: log-likelihood = -128.54, aic = 269.08 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.6: Summary statistics of the ARMA(2,2) model.

Table 3.4, Table 3.5 and Table 3.6 show the summary statistics of the AR(8), MA(9) and ARMA(2,2) models, respectively. After removing the insignificant coefficients, we refit those three models again. The summary statistics are presented in Table 3.7.

By comparing the log-likelihood of those three models with zero-coefficients removed, we notice that the performance of the AR(8) model is better than the ARMA(2,2) model, but worse than MA(9) model. This result is understandable considering the MA(9) model contains two more lags than the AR(8) model. Furthermore, according to the AIC, MA(9) is still the most parsimonious model and it should be preferred over AR(8) and ARMA(2,2).

| Model | Estimate | Std. Error | t value | P($>\|t\|$) |
|---|---|---|---|---|
| AR(8): $r_t = \phi_1 r_{t-4} + \phi_2 r_{t-8} + a_t$ | | | | |
| ar4 | -0.2338 | 0.1088 | -2.1480 | 0.0317 * |
| ar8 | -0.3503 | 0.1068 | -3.2798 | 0.0010 ** |
| ar1, ar2, ar3, ar5, ar6, ar7 and intercept are insignificant, whose coefficients are set to be 0 | | | | |
| $\sigma^2$ estimated as 0.9452: log-likelihood = -124.36, aic = 252.73 | | | | |
| MA(9): $r_t = \alpha_1 a_{t-3} + \alpha_2 a_{t-6} + \alpha_3 a_{t-8} + \alpha_4 a_{t-9} + a_t$ | | | | |
| ma3 | -0.3198 | 0.1077 | -2.9686 | 0.0030 ** |
| ma6 | -0.3082 | 0.1169 | -2.6370 | 0.0084 *** |
| ma8 | -0.5327 | 0.0928 | -5.7423 | 9.34e-09 *** |
| ma9 | 0.3205 | 0.1158 | 2.7682 | 0.0056 ** |
| ma1, ma2, ma4, ma5, ma7 and intercept are insignificant, whose coefficients are set to be 0 | | | | |
| $\sigma^2$ estimated as 0.79: log-likelihood = -119.43, aic = 246.85 | | | | |
| ARMA(2,2): $r_t = \phi_1 r_{t-2} + \alpha_1 a_{t-2} + a_t$ | | | | |
| ar2 | -0.5393 | 0.2688 | -2.0063 | 0.0448 * |
| ma2 | 0.7003 | 0.2224 | 3.1490 | 0.0016 ** |
| ar1, ma1 and intercept are insignificant, whose coefficients are set to be 0 | | | | |
| $\sigma^2$ estimated as 1.065: log-likelihood = -129.17, aic = 262.34 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.7: Summary statistics of the AR(8), MA(9) and ARMA(2,2) model once zero-coefficients are removed from the models.

Figure 3.7 shows the ACF and PACF of the residuals $a_t$ from the refitted AR(8) model (first row), MA(9) model (second row) and ARMA(2,2) model (third row). According to Figure 3.7, MA(9) performs better on removing auto-correlations while AR(8) is better on removing partial auto-correlations. To this end, it seems that either MA(9) or AR(8) is a good choice for our sample. Because AR(8) has lower number of parameters, I am going to select it as the model of our data.

Therefore, our fitted model is:

$$r_t = -0.2338 r_{t-1} - 0.3503 r_{t-8} + a_t$$

Here $\{a_t\}$ is a white noise process with mean 0 and variance $\sigma_t^2$.

Table 3.7 and Figure 3.7 (first row) shows that the AR(8) model describes the data efficiently, and the residual series $a_t$ presents no significant correlation.

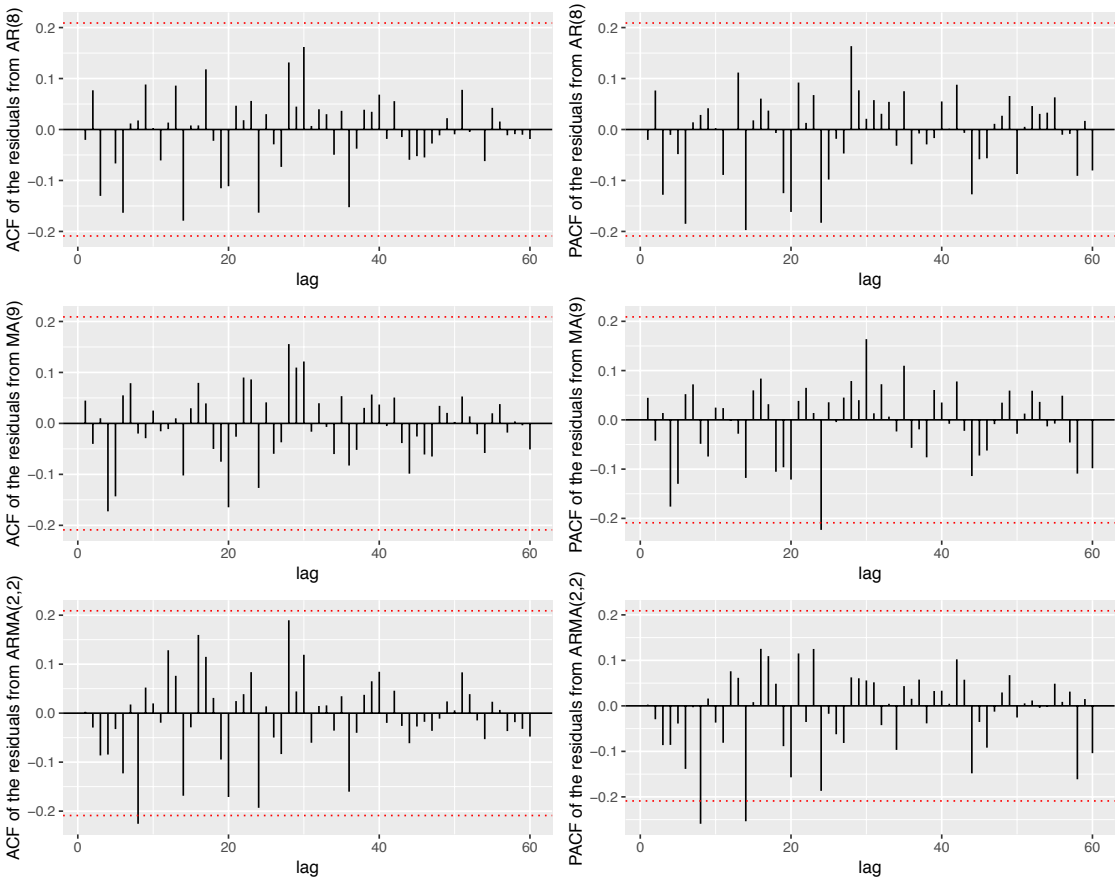
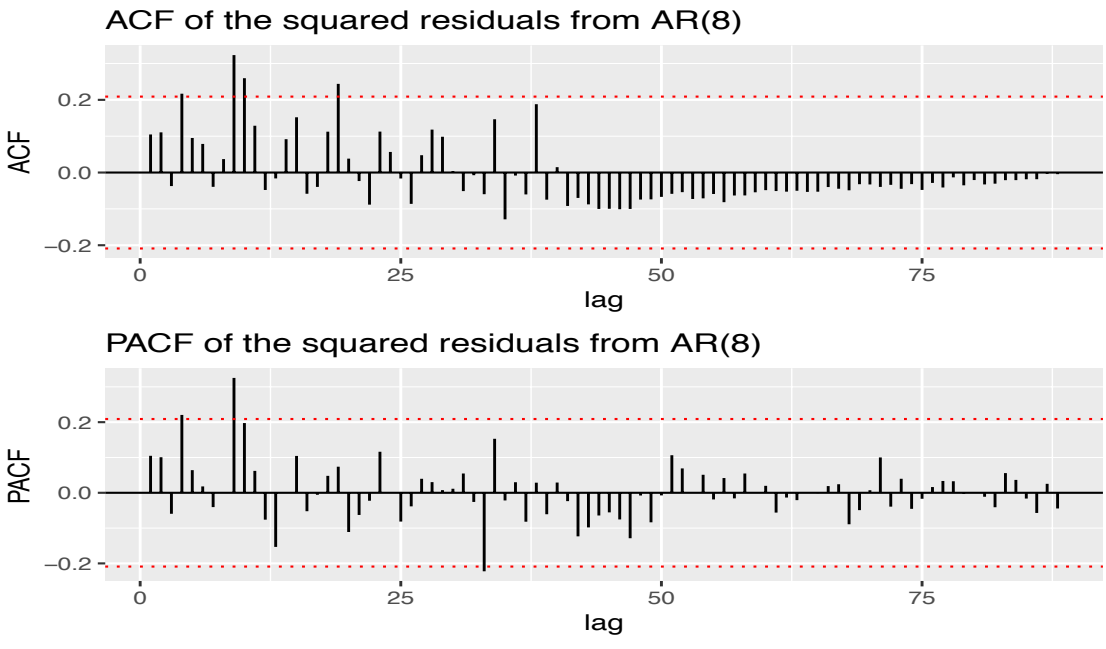



Figure 3.7: The auto-correlogram and partial auto-correlogram of the residuals from AR(8) (first row), MA(9) (second row) and ARMA(2,2) (third row) models.




s

Figure 3.8: The auto-correlogram and partial auto-correlogram of the squared residual series from AR(8): $a_t^2$.

Since we already modeled the mean series of $r_t$: $\mu_t$, $\{a_t\}$ can be easily substracted as the model residuals. Meanwhile, according to equation (2.18),

$$E(a_t^2) = E(\sigma_t^2)$$

where $\sigma_t^2$ is the volatility.

Figure 3.8 shows the sample ACF and PACF of the series $a_t^2$. Clearly, both plots indicate that the squared residual series $a_t^2$ may have some serial correlations. Consequently, the sample ACF and PACF of $a_t^2$ suggests that the volatility $\sigma_t^2$ could be modeled as ARCH effects (Tsay 2005).

We will start with a simple model: GARCH(1,1) for volatility. The model is specified as:

$$r_t = \mu_t + a_t,\ a_t = \sigma_t \epsilon_t,\ \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

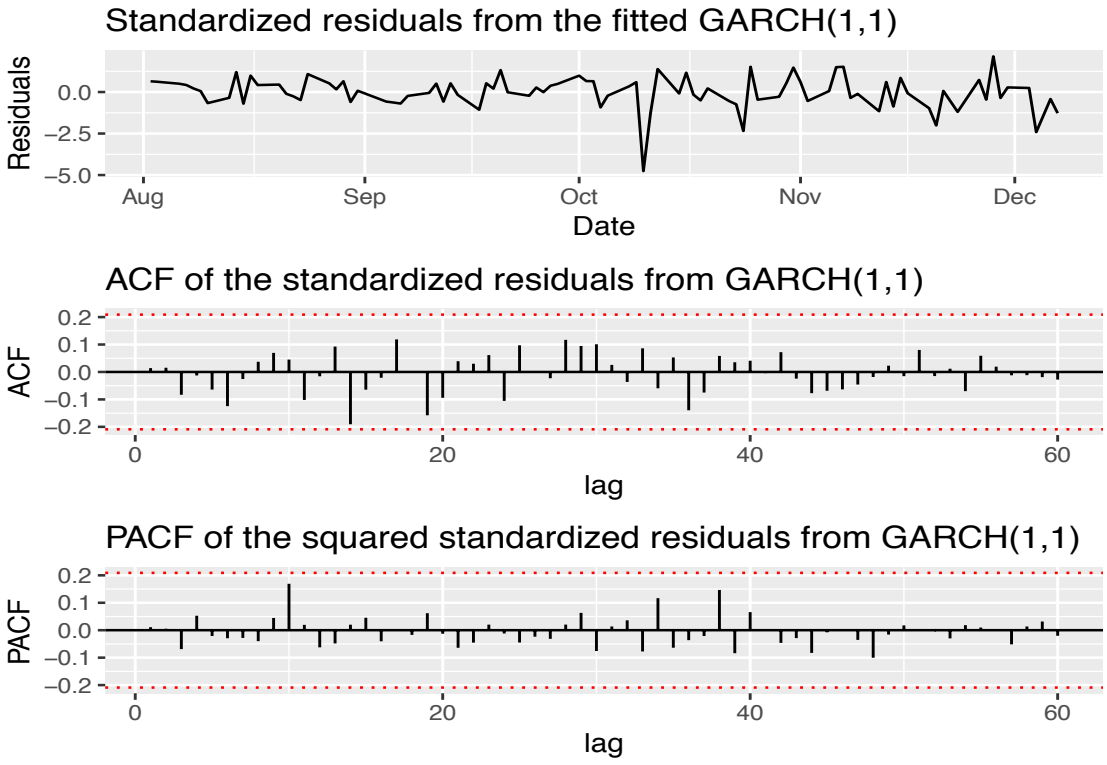
According to Table 3.7 and Table 3.8, our fitted model is:

$$\begin{aligned} r_t &= -0.2338 r_{t-1} - 0.3503 r_{t-8} + a_t,\ a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.1307 a_{t-1}^2 + 0.8489 \sigma_{t-1}^2 \end{aligned} \tag{3.3}$$

The estimates of $\alpha_1$ and $\beta_1$ are all significant.

Figure 3.9 shows the sample time series, ACF and PACF of the standardized residuals series: $\epsilon_t = \frac{a_t}{\sigma_t}$.

The ACF and PACF plots suggest that no serial correlation remains in $\epsilon_t$ or $\epsilon_t^2$. In Table 3.8, the Ljung-Box test of the standardized residuals gives Q(10)=3.515, Q(20)=15.4627. Ljung-Box statistics of $\epsilon_t^2$ series also shows Q(10)=4.2168, and Q(20)=6.72. Consequently, the AR(8)-GARCH(1,1) model in equation (3.3) is adequate for describing the auto-correlation and conditional heteroscedasticity of our sample under 95% confidence level.




s

Figure 3.9: Model checking of the GARCH (1,1) model in equation (3.1).

| Formula= $\sim$ garch(1,1) | | | | |
|---|---|---|---|---|
| Conditional Distribution: Normal | | | | |
| Error Analysis: | | | | |
| | Estimate | Std. Error | t value | P($>\|t\|$) |
| $\alpha_0$ | 0.0446 | 0.0266 | 1.678 | 0.0933 . |
| $\alpha_1$ | 0.1307 | 0.0663 | 1.970 | 0.0489 * |
| $\beta_1$ | 0.8489 | 0.0664 | 12.793 | <2e-16 *** |
| Log-likelihood: -116.704; normalized: -1.3113; AIC: 2.68; BIC: 2.7739 | | | | |
| Standardised Residuals Tests: | | | | |
| | | | Statistic | p-Value |
| Jarque-Bera Test | R | $\chi^2$ | 131.3341 | 0 |
| Shapiro-Wilk Test | R | W | 0.9134 | 1.8751e-05 |
| Ljung-Box Test | R | Q(10) | 3.5149 | 0.9666 |
| Ljung-Box Test | R | Q(15) | 9.9062 | 0.8256 |
| Ljung-Box Test | R | Q(20) | 15.4627 | 0.7493 |
| Ljung-Box Test | $R^2$ | Q(10) | 4.2168 | 0.937 |
| Ljung-Box Test | $R^2$ | Q(15) | 5.4486 | 0.9876 |
| Ljung-Box Test | $R^2$ | Q(20) | 6.72 | 0.9975 |
| LM Arch Test | R | $TR^2$ | 3.7885 | 0.987 |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.8: Summary statistics of the GARCH(1,1) model of sample squared residuals $a_t^2$.

From R output, $\hat{a}_{90} = -1.8248$ and $\hat{\sigma}_{90} = 2.0595$. With equation (3.1), we can obtain the volatility of $10^{th}$ of December, 2018 (day 91):

$$\hat{\sigma}_{91}^2 = 0.1307(-1.8248)^2 + 0.8489(2.0595)^2 = 4.0359$$

If we repeat the process presented above, we can obtain one-day-ahead forecast volatilities $\hat{\sigma}_t^2$ from $11^{th}$ of December 2018 to $1^{st}$ of February 2019.

## 3.5 ARMA model of the realized volatility

In this section, we forecast the realized volatility of December $10^{th}$, 2018 (day 91) $RV_{91}$; this case is considered as an example of processing one-step-ahead forecast of realized volatility.

First, we check the time series, ACF and PACF of the sample realized volatility $RV_t$, $t = 1, \ldots, 90$, which are shown in Figure 3.10.

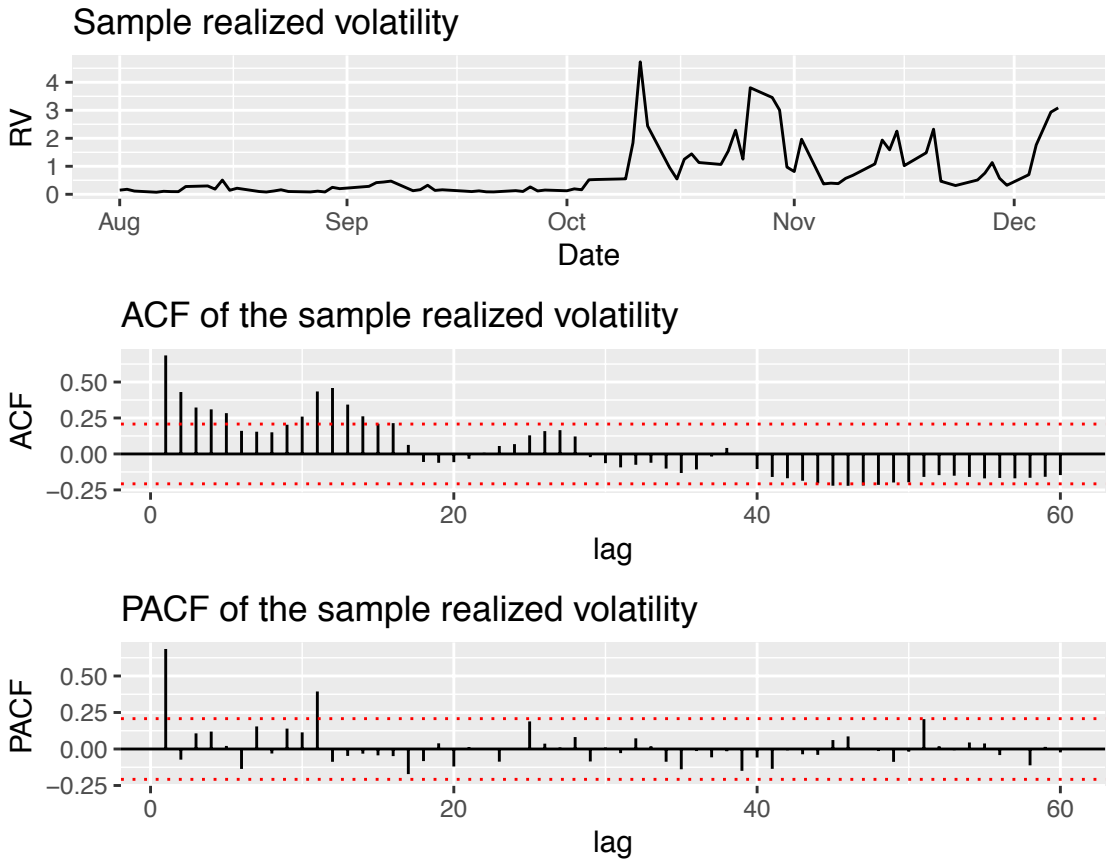



Figure 3.10: The time series, auto-correlogram and partial auto-correlogram of the sample realized volatility $RV_t$ from August $1^{st}$, 2018 to December $7^{th}$, 2019.

From the time series in Figure 3.10, the realized volatility does not appear any linear trend or periodical pattern, which is also supported by the sample ACF. However, in the sample ACF and PACF, the first lag presents a high value, which suggests us to take the first difference of the series: $D_t = RV_t - RV_{t-1}$ (Tsay, 2014). Figure 3.11

presents the time series, auto-correlogram and partial auto-correlogram of the first difference of the sample realized volatility.

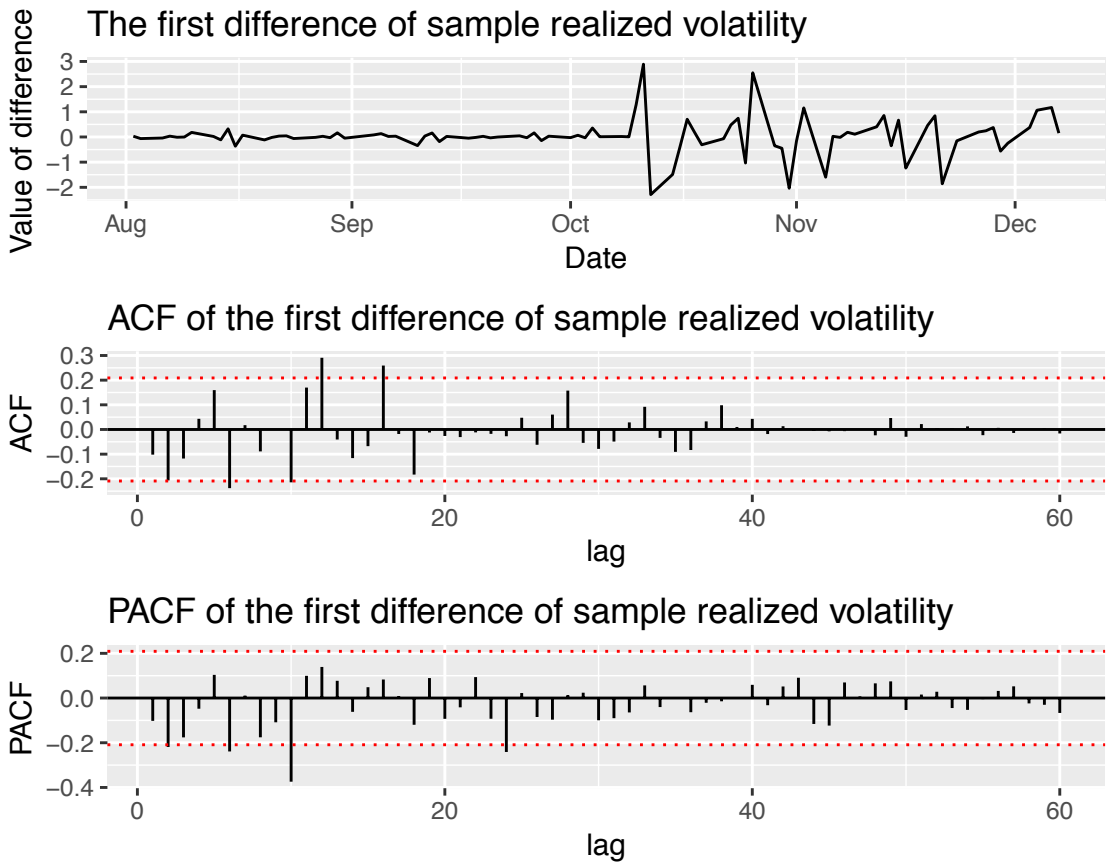



Figure 3.11: The time series, auto-correlogram and partial auto-correlogram of the first difference of the sample realized volatility: $D_t$.

According to the time series in Figure 3.11, $D_t$ does not show any linear trend or periodical pattern. In the ACF plot, the sixth, ninth, eleventh and fifteenth lags are significant; in the PACF plot, the second, sixth, tenth and twenty-fourth lags are significant, although none of them appears high value. Therefore, we first consider two models, AR(10) and MA(9) for the series $D_t$.

Table 3.9 and Table 3.10 show the summary statistics of the AR(10) and MA(9) models. After removing the insignificant coefficients, we refit the models again. The summary statistics of the refitted models are presented in Table 3.11.

Figure 3.12 presents the ACFs and PACFs of the residuals from the refitted AR(10)

| Model | Estimate | Std. Error | t value | P($>|t|$) |
|---|---|---|---|---|
| AR(10): | | | | |
| ar1 | -0.2656 | 0.0948 | -2.8014 | 0.0051 ** |
| ar2 | -0.4430 | 0.0977 | -4.5350 | 5.760e-06 *** |
| ar3 | -0.3352 | 0.1021 | -3.2828 | 0.0010 ** |
| ar4 | -0.3229 | 0.1067 | -3.0269 | 0.0025 ** |
| ar5 | -0.0722 | 0.1024 | -0.7053 | 0.4806 |
| ar6 | -0.4114 | 0.0987 | -4.1690 | 3.060e-05 *** |
| ar7 | -0.2323 | 0.1026 | -2.2630 | 0.0236 * |
| ar8 | -0.3880 | 0.0987 | -3.9313 | 8.448e-05 *** |
| ar9 | -0.2594 | 0.0955 | -2.7174 | 0.0066 ** |
| ar10 | -0.4328 | 0.0928 | -4.6634 | 3.110e-06 *** |
| intercept | 0.0160 | 0.0153 | 1.0428 | 0.2971 |
| $\sigma^2$ estimated as 0.3252: log-likelihood = -77.98, aic = 179.96 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.9: Summary statistics of the AR(10) model with sample $D_t$.

| Model | Estimate | Std. Error | t value | P($>|t|$) |
|---|---|---|---|---|
| MA(9): | | | | |
| ma1 | -0.1458 | 0.1130 | -1.2906 | 0.1969 |
| ma2 | -0.3212 | 0.1397 | -2.2998 | 0.0215 * |
| ma3 | -0.2610 | 0.1468 | -1.7781 | 0.0754 . |
| ma4 | 0.0028 | 0.1330 | 0.0209 | 0.9834 |
| ma5 | 0.5024 | 0.1460 | 3.4425 | 0.0006 *** |
| ma6 | -0.3275 | 0.1148 | -2.8522 | 0.0043 ** |
| ma7 | -0.4616 | 0.1275 | -3.6192 | 0.0003 *** |
| ma8 | -0.1640 | 0.1197 | -1.3700 | 0.1707 |
| ma9 | 0.4325 | 0.1343 | 3.2196 | 0.0013 ** |
| $\sigma^2$ estimated as 0.3311: log-likelihood = -81.37, aic = 182.73 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.10: Summary statistics of the MA(9) model with sample $D_t$.

and MA(9) model. It is obvious that those two models explain the auto-correlations in the series $D_t$ well, there is no significant correlation left.

In Table 3.11, although AR(10) model contains three more parameters, its AIC does not improve significantly compared to MA(9) model. Consequently, it seems MA(9) model is a good option for our data. The time series of the residuals from MA(9) shows in Figure 3.13.

| Model | Estimate | Std. Error | t value | P($>\|t\|$) |
|---|---|---|---|---|
| AR(10): | | | | |
| ar1 | -0.2476 | 0.0940 | -2.6332 | 0.0085 ** |
| ar2 | -0.4190 | 0.0946 | -4.4290 | 9.469e-06 *** |
| ar3 | -0.3035 | 0.0950 | -3.1933 | 0.0014 ** |
| ar4 | -0.2910 | 0.0999 | -2.9133 | 0.0036 ** |
| ar6 | -0.3829 | 0.0928 | -4.1257 | 3.697e-05 *** |
| ar7 | -0.2021 | 0.0966 | -2.0914 | 0.0365 * |
| ar8 | -0.3659 | 0.0960 | -3.8122 | 0.0001 *** |
| ar9 | -0.2444 | 0.0949 | -2.5747 | 0.0100 * |
| ar10 | -0.4253 | 0.0934 | -4.5530 | 5.289e-06 *** |
| ar5 and intercept are insignificant, whose coefficients are set to be 0 | | | | |
| $\sigma^2$ estimated as 0.3311: log-likelihood = -78.73, aic = 175.46 | | | | |
| MA(9): | | | | |
| ma2 | -0.4962 | 0.1512 | -3.2808 | 0.0010 ** |
| ma5 | 0.4377 | 0.1165 | 3.7580 | 0.0002 *** |
| ma6 | -0.3627 | 0.0974 | -3.7229 | 0.0002 *** |
| ma7 | -0.3996 | 0.1425 | -2.8039 | 0.0050 ** |
| ma8 | -0.1825 | 0.0979 | -1.8651 | 0.0622 . |
| ma9 | 0.3415 | 0.1501 | 2.2758 | 0.0229 * |
| ma1, ma3 and ma4 are insignificant, whose coefficients are set to be 0 | | | | |
| $\sigma^2$ estimated as 0.3155: log-likelihood = -83.21, aic = 178.41 | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Table 3.11: Summary statistics of the AR(10) and MA(9) model with sample $D_t$ once zero-coefficients are removed from the models.

The fitted model can be written as:

$$\begin{aligned} D_t = &- 0.4962a_{d,t-2} + 0.4377a_{d,t-5} - 0.3627a_{d,t-6} - 0.3996a_{d,t-7} \\ &- 0.1825a_{d,t-8} + 0.3415a_{d,t-9} + a_{d,t} \end{aligned} \tag{3.4}$$

Here, $a_{d,t-i} \sim N(0, \sigma_a^2)$ and $D_t = RV_t - RV_{t-1}$.

From R output, $RV_{90} = 3.0861$, $a_{d,89} = 1.1609$, $a_{d,86} = 0.2258$, $a_{d,85} = -0.3925$, $a_{d,84} = 0.4182$, $a_{d,83} = -0.1179$ and $a_{d,82} = 0.2652$. Plug them into equation (3.4), we can obtain the one-step-ahead forecast of the realized volatility of December $10^{th}$, 2018 (day 91):

$$\hat{RV}_{91} = RV_{90} - 0.4962a_{d,89} + 0.4377a_{d,86} - 0.3627a_{d,85} - 0.3996a_{d,84} - 0.1825a_{d,83} + 0.3415a_{d,82}$$

$$\Rightarrow \hat{RV}_{91} = 2.6962$$

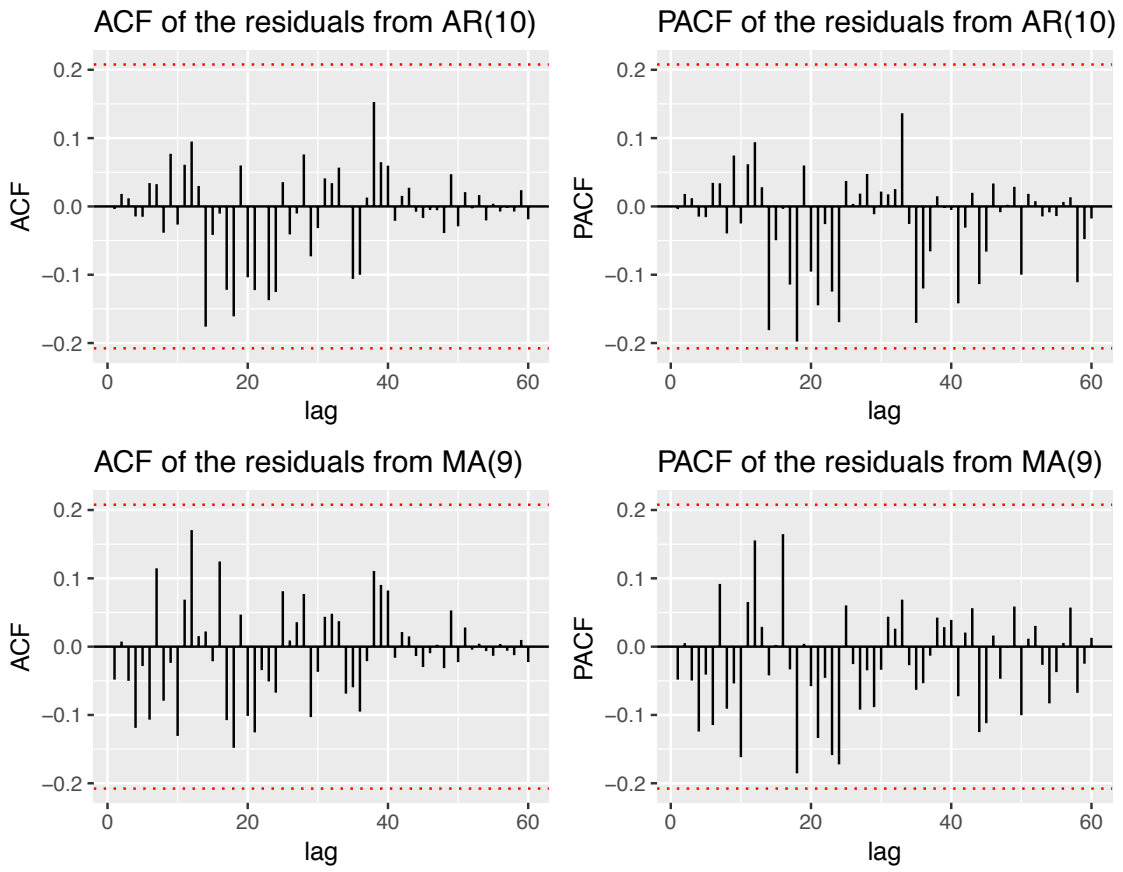
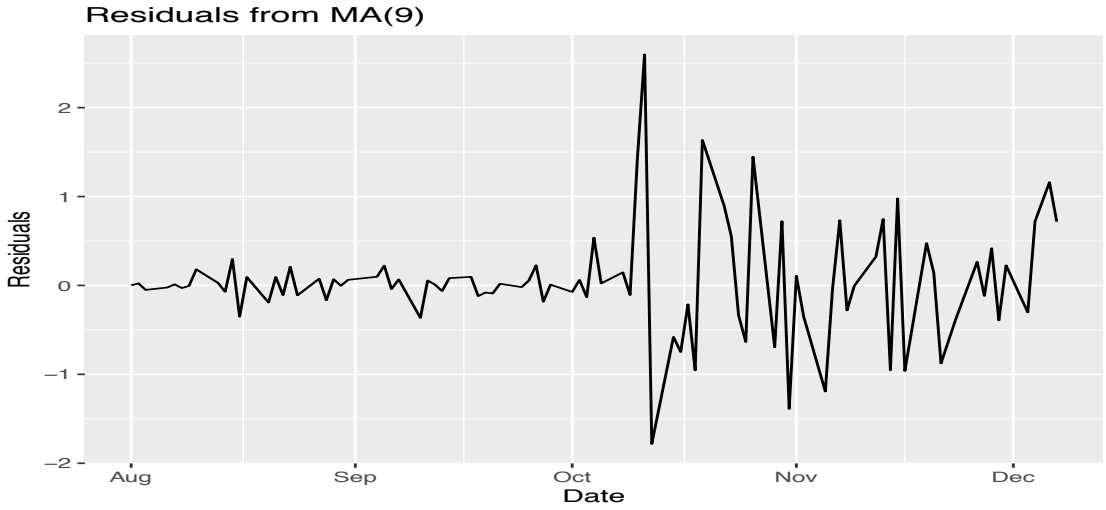



Figure 3.12: The auto-correlogram and partial auto-correlogram of the residuals from AR(10) model (first row) and MA(9) model (second row) about the sample series $D_t$.




s

Figure 3.13: The time series of the residuals from MA(9) model about sample $RV_t$.

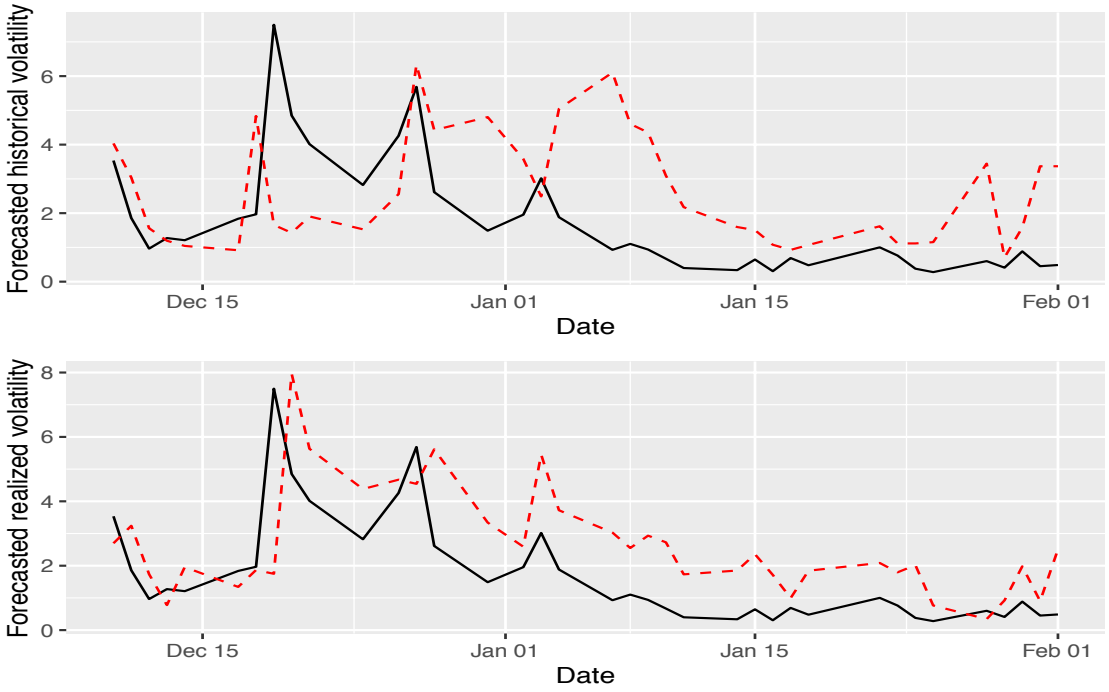
# Chapter 4

# Forecast evaluation and conclusion

As mentioned in Chapter 3, our forecasting results are constructed based on the data from the estimation period. Using the daily return and the realized volatility, we conducted one-day ahead forecasting through the GARCH family models and the ARMA model, respectively. Our evaluation period is from the $10^{th}$ of December, 2018 to the $1^{st}$ of February, 2019, which contains 37 trading days. We calculated the realized volatility of each trading day from the evaluation period to compare with our forecasting results for assessment. The forecasting results are presented in Figure 4.1 as a dashed line along with the realized volatility displayed as a solid line.

As revealed in Figure 4.1, the volatility that is forecasted based on the daily return did not accurately predict all the shocks that occurred during the evaluation period. The forecasts on dates before December $28^{th}$ 2018 appear to be lower than the actual volatility, although the forecasts since December $28^{th}$ 2018 are higher than the actuality.

The volatility that is forecasted based on the realized volatility describes well the general changes of the actual volatility from the evaluation period, although the forecasting values are generally higher than the actual volatility. This forecast based on realized volatility also fairly predict the three shocks occurred during the period of study although they seem to be one day behind the actual occurring date.




s

Figure 4.1: One-day ahead forecasting results of historical volatility (first row) and realized volatility (second row) from the evaluation period (displayed in red dash). The black solid line presents the realized volatility.

| | Evaluation criteria | |
|---|---|---|
| Data | HMAE | HRMSE |
| Daily return with GARCH(1,1) | 0.7072 | 0.9360 |
| Realized volatility with MA(9) | 0.5989 | 0.7747 |

Table 4.1: Evaluation criteria: HMAE and HRMSE with prediction results.

Table 4.1 shows two forecasting evaluaton criteria: HMAE and HRMSE, which are also used in the studies of Andersen et al. (1999) and Martens (2002). The forecasts conducted based on the realized volatility has HMAE: 0.5989 and HRMSE: 0.7747, which are lower than the forecasts result based on the daily return. It indicates that the predictions based on the realized volatility perform better compared to the forecasts based on the daily return (Martens, 2002). This result matches our interpretation of Figure 4.1 and suggests that the volatility forecasting based on the daily return is not as good as the one based on realized volatility. Some previous

studies, which are conducted by Andersen and Bollerslev (1998), Martens (2002) and Koopman et al. (2005), also show similar results.

# Bibliography


[1] T. G. Andersen and T. Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905, 1998.

[2] T. G. Andersen, T. Bollerslev, and S. Lange. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of empirical finance*, 6(5):457–477, 1999.

[3] N. M. Areal and S. J. Taylor. The realized volatility of FTSE-100 futures prices. *Journal of futures markets: futures, options, and other derivative products*, 22(7):627–648, 2002.

[4] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[5] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

[6] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: journal of the econometric society*, pages 987–1007, 1982.

[7] K. R. French, G. W. Schwert, and R. F. Stambaugh. Expected stock returns and volatility. *Journal of financial economics*, 19(1):3–29, 1987.

[8] P. Hansen and A. Lunde. Volatility estimation using high frequency data with partial availability. Technical report, Discussion Paper, Brown University Working Paper, 2002.

[9] S. J. Koopman, B. Jungbacker, and E. Hol. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3):445–475, 2005.

[10] M. Martens. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of futures markets: futures, options, and other derivative products*, 22(6):497–518, 2002.

[11] M. H. Miller, J. Muthuswamy, and R. E. Whaley. Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion? *The journal of finance*, 49(2):479–513, 1994.

[12] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: journal of the econometric society*, pages 347–370, 1991.

[13] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer, 2017.

[14] R. S. Tsay. *Analysis of financial time series*. John wiley & sons, 2005.

[15] R. S. Tsay. *An introduction to analysis of financial data with R*. John Wiley & Sons, 2014.

[16] J.-M. Zakoian. Threshold heteroskedastic models. *Journal of economic dynamics and control*, 18(5):931–955, 1994.