POLITECNICO DI TORINO

Master degree course in Mathematical Engineering

Master Thesis

Time series modelling and analysis

Supervisor: Prof. Franco Pellery

Candidate: Kamand Taghaboni

December 2018

Abstract

The purpose of this thesis is to evaluate the GARCH model and its ability to forecast Value at Risk of financial data. To assess the forecasting performance, it has been used three different distributions on error term: Normal distribution, Student-t distribution and skewed Student-t distribution. It has been selected four stock market indexes to test : NASDAQ's daily index, Standard and Poor's 500 daily index, NIKKEI's daily index and Dow Jones's daily index. The data has been collected from January 1, 2010 to January 01, 2017.

Contents

1 Introduction

Many phenomena, like the movement of stock prices, are measured in intervals over a period of time. Time series analysis methods are very useful for analysing these types of data.
Financial time series data, like the relative return of a stock or a portfolio of stocks, often contain of periods of "calm" behaviour alternating with periods of very wild variations. One way to express this is the following quotation, taking from Mandelbrot (see [1])

"large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes"

Generally, the difficulty to predict a future value of a stock or some other asset is a measure of how risky the asset is. In financial terms this is called the volatility of the asset. There are some characteristics of financial volatility data like, long memory, fat tails and excess kurtosis, clustering volatility and leverage effect which are introduced by Baillie, Bollerslev and Mikkelsen, see [2]. The ARMA model is used to present the stationary time series based on autoregressive process and moving average of noises. For time series which needs to be differenced to be made stationary, the ARMA become ARIMA(*autoregressive integrated moving average)*. On the other hand, the autoregressive conditional heteroscedasticity (ARCH) model which was early introduced in the Engle's paper (see [3]) and it has been focused on time varying conditional variance.
Practically, high ARCH order has to be selected. For solving this problem, Bollerslev (see [4]) extended this model to Generalized ARCH (GARCH) model. GARCH model describe variance at a certain time with both past values and past variances. Most time series is sufficiently modelled using GARCH (1,1) that only includes three parameters. GARCH model have been certificated not only to catch volatility clustering but also to contain fat tails from the volatility data. The main point is that the GARCH model is symmetric, so it has a poor performance in reflecting the asymmetry. In other words, good and bad news have the same effect on the volatility in this model. This asymmetric phenomenon is leverage effect.

Today, many have designed modifications of the GARCH model, which has given rise to the expression of an ARCH/GARCH family of models, see [5].

Volatility modelling is very important in market risk applications, such as value at risk (VaR). Jorion (see [6]) defined the VaR as "*the worst loss over a target horizon that will not be exceeded with a given level of confidence (or under normal market conditions)*".

The aim of the thesis is analysing the historical data. It is mainly focused on the selection of the suitable model to estimate the financial volatility. After developing the model, we would like to know how well the model forecasts and for this aim it has been performed rolling forecast method which is used for comparing out-of-sample of value at risk (VaR) by backtesting method, see [7].

The research performs the statistical analysis of the log-returns of different stock prices by testing different GARCH models and ARMA-GARCH models with three types of residuals distributions: normal, Student-t and skewed Student-t, to find the good models. There are presented theoretical background in section 2 and data description and analysis in section 3. In section 4, it includes the application and results of models and forecasting. The conclusion is presented in section 5. It has been used software R for programming.

2 Theoretical Background

The base of analysis of a time series is stationarity. In a stationary process, means and variances do not change by time, and covariance only depends on the difference of the time subscripts, otherwise we can say a process is not a stationary.

There are two types of stationary processes: strictly stationary and weakly stationary. Strictly stationary is hard to reach empirically. A time series in discrete time $\{X_t\}_{t=-\infty}^{\infty}$ is called stationary (or weakly stationary) when both mean of X_t and covariance between X_t and X_{t-s}, where s is an integer, are invariant through the time. Before introducing the time series models, we also need to define white noise processes. There are two types of white noise processes: weak sense white noise and strict sense white noise. The time series $\{\varepsilon_t\}$ is said to be a white noise (or weak sense white noise) with mean zero and variance σ_ε^2, written as

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2),$$

if and only if ε_t has zero mean and

$$\gamma_\varepsilon(h) = \begin{cases} \sigma_\varepsilon^2 & if \quad h = 0 \\ 0 & if \quad h \neq 0 \end{cases},$$

where function $\gamma_\varepsilon: \mathbb{Z} \to \mathbb{R}$ defined by $\gamma_\varepsilon(h) = \mathrm{COV}(\varepsilon_t, \varepsilon_{t-h})$ is called autocovariance function of stationary process (weakly stationary process). It is clear that a white noise process is stationary.

A process $\{\varepsilon_t\}$ is strong sense white noise, if ε_t is iid with mean 0 and finite variance σ_ε^2, which it means in addition ε_t are not just uncorrelated but also independent.

A particularly useful white noise series is *Gaussian white noise*, where ε_t are independent normal random variables, with mean 0 and variance σ_ε^2 or more briefly,

$$\varepsilon_t \sim IIDN(O, \sigma_\varepsilon^2).$$

In this thesis, it has been used financial data, the log-return process of closing prices,$\{X_t = 100[\log(p_t) - \log(p_{t-1})], t \in \mathbb{Z}\}$ is a time series. Generally, this series can be decomposed into two elements:

$$X_t = \mu_t + \epsilon_t$$
$$\epsilon_t = \sigma_t \varepsilon_t,$$

where μ_t is a predictable process and ϵ_t is a nondeterministic process driven by a noise random variable ε_t which is iid with mean zero and unit variance. The process $\{\sigma_t\}$ called the volatility process (standard deviation process). Considering the filtration associated with the model, $\mathcal{F}_t$ is a sequence of increasing σ-algebras of $\mathcal{F}$ representing all market information up to time t. Hence, μ_t and σ_t^2 represent the conditional mean and variance of X_t:

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

2.1 Moving Average Process

A first generalization of the white noise is the moving average process (MA). Moving average models are widely used in the financial world to predict the start of trends which is important as trends are considered the best opportunity to make profits from the markets.

According to [8] moving average of order q, abbreviated $MA(q)$ is

$$X_t = \theta(B)\varepsilon_t =$$
$$= \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2) \qquad (1.1)$$

where $\theta_1, \dots, \theta_q$ are parameters and the moving average operator is

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q,$$

where B is the backshift operator (or lag operator) such as $B^k\varepsilon_t = \varepsilon_{t-k}$.

Moving average models are always in the weakly stationary form, as they are finite linear arrangements of a white noise sequence for which the first two moments are time invariant.

2.2 Autoregressive Models

Autoregressive models are based on the idea that the current value of the series,

X_t, can be explained as a function of p past values, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$, where p determines the number of steps into the past needed to forecast the current value.

According to [8] simple Autoregressive $AR(1)$ model is:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t, \qquad (1.2)$$

where $\{\varepsilon_t\}$ is defined as a white noise series with mean zero and variance σ_ε^2. This model follows the same rules of a simple linear regression model in which the dependent variable is X_t and the explanatory variable is X_{t-1}.

An autoregressive model of order p, abbreviated $AR(p)$, is of the form

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t \qquad (1.3)$$

where p is assumed to be a nonnegative integer and $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ and $\phi_1, \dots, \phi_p$ are constants $(\phi_p \neq 0)$.

The mean of X_t in (1.3) is zero. If the mean μ of X_t is not zero, replace X_t by $X_t - \mu$ in (1.3),

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \varepsilon_t \,,$$

or write

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t \qquad (1.4)$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

A useful form follows by using the backshift operator to write the $AR(p)$ model, (1.3), as

$$(1 - \phi_1 B - \phi_2\, B^2 - \cdots + \phi_p\, B^p) X_t = \varepsilon_t \,, \qquad (1.5)$$

or can be written the model as

$$\phi(B) X_t = \varepsilon_t. \qquad (1.6)$$

where B is the backshift operator such as $B^k X_t = X_{t-k}$.

2.3 ARMA Process

The general autoregressive and moving average (ARMA) statistical model is used to describe a time series that evolves over time. In this process there is a linear relationship between the values at a certain time point and past values, noise as well.

According to [8] time series $\{X_t; t \in \mathbb{Z}\}$ is $ARMA(p, q)$ if it is stationary and

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad (1.7)$$

where p is the order of the autoregressive part and q is the order of the moving average part, with $\phi_p \neq 0$, $\theta_q \neq 0$ and $\sigma_\varepsilon^2 > 0$. If X_t has a nonzero mean μ, we set $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ and write the model as

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad (1.8)$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

As previously noted, when q = 0, the model is called an autoregressive model of order p, AR(p), and when p = 0, the model is called a moving average model of order q, MA(q).

In particular, the $ARMA(p, q)$ model in (1.7) can then be written in short form as

$$\phi(B) X_t = \theta(B) \varepsilon_t, \qquad (1.9)$$

where $\phi(B)$ and $\theta(B)$ are defined as

$\phi(B) = 1 - \phi_1 B - \phi_2\, B^2 - \cdots + \phi_p\, B^p$, where B is the backshift operator such as $B^k X_t = X_{t-k}$,

$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$, where B is the backshift operator such as $B^k \varepsilon_t = \varepsilon_{t-k}$.

The autoregressive moving average (ARMA) process was introduced by Box, Jenkins, and Reinsel (see [9]) that combines the autoregressive and moving average concepts, is a way to keep the number of parameters small.

2.4 ARCH processes

The first model of stochastic volatility proposed by Engle (see [3]) is the $ARCH(q)$ process (Autoregressive conditional heteroscedasticity). In financial time-series a problem known as heteroskedasticity might occur, which explains that the variance error term is not constant over time. Working with a model that assumes constant variance would make worse the approximations and hence, the ARCH-model, that does not assume constant variance over time, might be a more suitable model to use.

According to [3] the main model can be an AR model, an ARMA model, or a standard regression model, i.e.

$$X_t = \mu_t + \epsilon_t,$$

where ϵ_t is conditionally heteroskedastic in the form of

$$\epsilon_t = \sigma_t \varepsilon_t,$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent, identically distributed (iid) random variables with zero mean and unit variance. This implied:

$$\epsilon_t \sim \mathrm{D}(0, \sigma_t^2),$$

and $ARCH(q)$ is

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 \qquad (1.10)$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$ for $j = 1, \dots, q$. The process is weakly stationary if and only if $\sum_{i=1}^{q} \alpha_i < 1$. It is easy to verify that

$$E(\epsilon_t) = 0,$$

$$var(\epsilon_t) = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_q},$$

$$E(\epsilon_t \epsilon_{t-s}) = 0, \qquad \forall s \neq 0$$

and

$$E[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = 0,$$

$$\sigma_t^2 = E(\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2.$$

The ARCH model can capture periods of tranquility and volatility in the $\{X_t\}$ series. The conditional variance σ_t^2 has two parts: a constant term α_0 and the linear combination of the information about the squared errors $\epsilon_{t-1}^2, \dots, \epsilon_{t-q}^2$ (i.e. an ARCH term).

ARCH models are simple and easy to handle. ARCH models also take care of clustered errors, nonlinearities and changes in the econometrician's ability to forecast, see [22].

2.5 GARCH processes

The most important extension of the ARCH process is certainly the *generalized* ARCH, or GARCH process (Generalized Autoregressive Conditional Heteroskedastic) which was introduced by Bollerslev (see [4]) and Taylor (see [21]).

According to [10] the error term in the main model $X_t = \mu_t + \epsilon_t$ satisfies $\epsilon_t = \sigma_t \varepsilon_t$, where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent, identically distributed (iid) random variables with zero mean and unit variance. ε_t is independent of σ_t and

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2. \qquad (1.11)$$

where $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \dots, q$, $\beta_k \geq 0$ for $k = 1, \dots, p$ and μ_t is constant.

The GARCH (p, q) model is strictly stationary with finite variance when the conditions ω > 0, and $\sum_{j=1}^{q} \alpha_j + \sum_{k=1}^{p} \beta_k < 1$ are required where the proof could be found in [4].

We can see the GARCH model has similar pattern with ARMA model, which shows we can derive GARCH process using similar theory and method with ARMA.

Particularly, in most cases structure p = q = 1 is sufficient and it is sufficient for our purposes. $\mathrm{GARCH}(1,1)$ model is the most widely used, which is given by

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

To obtain strictly stationary solution, the conditions ω > 0, $\alpha + \beta < 1$ are required and the proof could be found in Appendix. We see that $\mathrm{GARCH}(1,1)$ explains that the present volatility depends only on previous one. It is easy to calculate and simulate since there are only three parameters in $\mathrm{GARCH}(1,1)$ model. GARCH model successfully explains the volatility clustering, but it does not capture the leverage effect.

2.6 ARMA-GARCH process

According to $[10]$, one of the important extension of GARCH model comes from the dynamic frame of the conditional mean. The ARMA-GARCH model combines an ARMA model for modelling the dynamic conditional mean and a GARCH model for modelling the dynamic conditional volatility. The conditional mean of an ARMA(P, Q)-GARCH(p, q) is of the form

$$X_t = c + \sum_{i=1}^{P} \phi_i (X_{t-i} - c) - \sum_{j=1}^{Q} \theta_j \epsilon_{t-j} + \epsilon_t,$$
$$\epsilon_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \ldots, q$, $\beta_k \geq 0$ for $k = 1, \ldots, p$. we also assume that $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent, identically distributed (iid) random variables with zero mean and unit variance and if there exist real coefficients c, $\phi_1, \ldots, \phi_P$ and $\theta_1, \ldots, \theta_q$, where P and Q are integers. Stationary and invertible assumptions of the ARMA model are considered here as well.

2.7 Maximum-Likelihood Estimation (MLE)

The most common method for estimating the GARCH-type models is the *maximum likelihood estimation* (*MLE*). According to [8] the maximum likelihood estimation depends on the assumption of a particular distributional form for the observations, known apart from the values of parameters $\theta_1, \dots, \theta_m$. We can regard the estimation problem as that of selecting the most appropriate value of a parameter vector θ, taking values in a subset Θ of $\mathbb{R}^m$. We suppose that these distributions have probability densities $p(x;\theta)$, $\theta \epsilon \ \Theta$. For a fixed vector of observations x , the function $L(\theta) = p(x;\theta)$ on Θ is called the *likelihood function*. A maximum likelihood estimate $\hat{\theta}(x)$ of θ is a value of $\theta \epsilon \ \Theta$ that maximizes the value of $L(\theta)$ for the given observed value x, i.e.,

$$L(\hat{\theta}) = P\left(x;\hat{\theta}(x)\right) = \max_{\theta \epsilon \ \Theta} p(x;\theta)$$

In this thesis, the GARCH-models are to be maximized under the Normal, Student-t, and Skewed Student-t distribution for the residuals which was discovered by Fernandez and Steel, see [11].

2.8 Distribution of error term

In this thesis, it has been introduced three distributions : Normal, Student-t, and Skewed Student-t distribution. The normal distribution is the distribution that allows for less kurtosis. The Student t-distribution converges to the normal distribution as the degrees of freedom increase. The Student-t distribution is more appropriate than normal distribution to express the fat tails and a reasonable amount of excess kurtosis. The Skewed Student-t distribution can describe skewness and kurtosis appropriately, which are important characteristics in financial time series.

2.8.1 Normal Distribution

According to [29], the probability density function of ε_t is given as normal distribution,

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_t - \mu}{\sigma}\right)^2\right\},$$

where μ is mean and σ is standard deviation.

2.8.2 Student-t Distribution

According to [30], the probability density function of ε_t is given as Student-t distribution,

$$f(\varepsilon_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\left(1+\frac{\varepsilon_t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

where ν is the number of degrees of freedom and Γ denotes the Gamma function.

2.8.3 Skewed Student-t Distribution

According to [31], the probability density function of ε_t is given as skewed Student-t distribution,

$$f(\varepsilon_t;\mu,\sigma,\nu,\lambda) = \begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{b\left(\frac{\varepsilon_t-\mu}{\sigma}\right)+a}{1-\lambda}\right)^2\right)^{-\frac{\nu+1}{2}}, if\ \ \varepsilon_t < -\frac{a}{b} \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{b\left(\frac{\varepsilon_t-\mu}{\sigma}\right)+a}{1+\lambda}\right)^2\right)^{-\frac{\nu+1}{2}}, if\ \ \varepsilon_t \geq -\frac{a}{b} \end{cases},$$

where μ and σ^2 are the mean and variance of the skewed Student-t distribution.

ν is a shape parameter with $2 < \nu < \infty$ and λ is a skewness parameter with $-1 < \lambda < 1$. The constants a, b and c are given as

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), b = 1 + 3\lambda^2 - a^2 \text{ and } c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)\Gamma\left(\frac{\nu}{2}\right)}}.$$

2.9 Autocorrelation function (ACF)

According to [12], by assuming that we have a stationary time series $\{X_t\}$ with constant expectation and time independent covariance. The autocorrelation function (ACF) is

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} = \frac{\gamma_k}{\gamma_0},$$

for $k \geq 0$, where k denotes the lag and

$$\rho_{-k} = \rho_k.$$

By graph of the autocorrelation function as a function of k, we can verify if the autocorrelation decreases as the lag gets larger, or if there is any particular lag for which the autocorrelation is large.

2.10 Partial Autocorrelation function (PACF)

According to [12], the PACF of a stationary time series is a function of its ACF and could determine the appropriate lags p in an AR(p) model. PACF is the correlation between $\{X_t\}$ and $\{X_{t-k}\}$ minus the part explained by the intervening lags

$$\rho_k^* = Corr[X_t - E^*(X_t|X_{t-1},, \dots, X_{t-k+1}), X_{t-k}]$$

Where $E^*(X_t|X_{t-1},, \dots, X_{t-k+1})$ is the minimum mean-squared error predictor of X_t by $X_{t-1}, \dots, X_{t-k}$.

2.11 Information Criteria

According to [12], several information criteria are used to select order of ARMA process. All of them are based on likelihood. The well-known *Akaike information criterion* (AIC) as Akaike (see [23]) defines is

$$\text{AIC} = \frac{-2}{N}\ln(likelihood) + \frac{2}{N} \times (number\ of\ parameters), \qquad (1.12)$$

where the likelihood function is evaluated at the maximum likelihood estimates and N is the sample size.

For a Gaussian $\text{AR}(\ell)$ model, AIC reduces to

$$\text{AIC}(\ell) = \ln(\hat{\sigma}_\ell^2) + \frac{2\ell}{N},$$

where $\hat{\sigma}_\ell^2$ is the maximum-likelihood estimate of $\hat{\sigma}_\varepsilon^2$, which is the variance of the white noise $\{\varepsilon_t\}$, and N is the sample size. The first term of the AIC considering (1.12) measures the goodness of fit of the $\text{AR}(\ell)$ model to the data, and also the second term is considered as the penalty function of the criterion because it uses the number of parameters in a candidate model to penalizes it. Different penalty follow in different information criteria.

Another commonly used criterion function is the Schwarz–Bayesian information criterion (BIC), see [24]. For a Gaussian $\text{AR}(\ell)$ model, the criterion is

$$\text{BIC}(\ell) = \ln(\hat{\sigma}_\ell^2) + \frac{\ell \ln(N)}{N},$$

the penalty function for each parameter used is 2 for AIC and $\ln(N)$ for BIC. Thus, compared with AIC, BIC tends to select a lower AR model when the sample size is moderate or large. This makes BIC the main criterion to select the best fit.

Selection Process

In order to use AIC to select an AR model, we should calculate $\text{AIC}(\ell)$ for $\ell = 0, \dots, p$, where p is a pre-specified positive integer and selects the order k that has the minimum AIC value. The same steps are involved when working with BIC.

2.12 ARCH-LM Test

ARCH-LM test was introduced by Engle, see [3]. ARCH-LM test is a Lagrange multiplier test to assess the significance of *autoregressive conditional heteroscedastic* (ARCH) effects. The alternative hypothesis for Engle's ARCH test is autocorrelation in the squared residuals, given by the regression

$$H_a\colon\ \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + u_t \qquad (1.13)$$

where u_t is a white noise error process and $\text{lag}\, m$ is a pre-specified positive integer. The null hypothesis is

$$H_0\colon \alpha_0 = \alpha_1 = \cdots = \alpha_m = 0$$

The null hypothesis or H_0 is no ARCH effect, otherwise someone can reject the null and conclude that there is an ARCH effect in the time series. The test statistic is defined as $\text{LM} = N \cdot R^2$ and is distributed as a chi-square with m degree of freedom. Where N is the sample size and R^2 is computed from the regression (1.13) using estimated residuals.

2.13 Ljung-Box test

To test for stationarity of residuals, Ljung-Box test is largely used. The null hypothesis of this test is that the residuals are independently distributed. This test examines the sample autocorrelation functions simultaneously by test statistics $Q(K)$ which defined as

$$Q(K) = N(N+2)\sum_{i=1}^{K} \hat{\rho}^2(i)/(N-i),$$

where N is the sample size, $\hat{\rho}^2(i)$ is the sample autocorrelation at lag i, and K is the number of lags being tested. For large N, $Q(K)$ can be approximately the chi-squared distribution with degrees of freedom K.

The assumption of α an i.i.d. sequence is rejected at significance level if $Q(K) > \chi^2_{1-\alpha}(K)$, see [8].

2.14 Standardized Residuals

In the GARCH model we impose the normal, Student-t, and Skewed Student-t distribution on the white noise term ε_t . These assumptions are tested by plotting the standardized residuals (see [20])

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t},$$

where $\hat{\epsilon}_t = X_t - \hat{\mu}_t$. By comparing the $\tilde{\epsilon}_t$ and the white noise term $\varepsilon_t = \frac{\epsilon_t}{\sigma_t}$ we would see that it is logical to assess if the assumed distribution of ε_t is an appropriate assumption by plotting $\tilde{\epsilon}_t$.

2.15 Evaluation of Estimated GARCH models

After fitting a GARCH model, for assessing the adequacy of the fitted model, it could be used some graphical and statistical diagnostic checks. If the GARCH model is correctly specified, then the estimated standardized residuals $\tilde{\epsilon}_t$ should behave like a white noise process. To assess whether the standardized residuals seem to be white noise, they should not display autocorrelation, conditional heteroskedasticity or any type of nonlinear dependence. The Ljung-Box statistic could be used to test the null of no autocorrelation up to a specific lag and ARCH-LM statistic could be used to test the null of no remaining ARCH effects. In addition, the Q-Q plot should look like a straight line, which means the standardized residuals should follow the assumed distribution. See [20].

2.16 Forecasting

Once we have fitted the GARCH model which modelled the risk, we would like to use it for forecasting and also we would like to check how well the model forecasts. Out-of-sample forecast performance are used to determine if estimated models are potentially useful for forecasting and commonly directed by dividing a given data set into an in-sample period, used for the initial parameter estimation and fitting the model, and an out-of-sample period, used to evaluate forecasting performance. The common approach used for out-of-sample forecast is known as rolling window procedure which has a fixed length of the in-sample period and both the start and the end of the estimation dates should increase by one and the model re-estimated at each time. For the m-step ahead forecasts, this process is continued until no more m-step ahead forecast can be computed. By rolling window procedure we divide the data sets into two parts, estimation window (w_E) and forecasting window (w_F) and $w_E + w_F = N$ where N is number of our observations and $n = 1, \dots, N$. Or equivalently we could consider $n^* = 1 - w_E, \dots, w_F$.

Forecasting window and estimation window

n	n^*	Window name
1	$1 - w_E$	in-sample
…	…	in-sample
w_E	0	in-sample
$w_E + 1$	1	out-of-sample
…	…	out-of-sample
$w_E + w_F$	w_F	out-of-sample

So the first estimation window is $n = (1, \dots, w_E)$, the second estimation window is $n = (1 + 1, \dots, w_E + 1)$, the third estimation window is $n = (1 + 2, \dots, w_E + 2)$ and so on, until the last estimation window is $n = (w_E, \dots, w_E + w_F - 1)$. The first estimation window is used to forecast $\hat{\sigma}^2_{t=1}$. The second estimation window is used to forecast $\hat{\sigma}^2_{t=2}$. The last estimation window is used to forecast $\hat{\sigma}^2_{t=w_F}$. By one-day-ahead forecasting we update every day the parameters and then we will have w_F number of estimated parameters but for economic reasons the true parameters do not shift a lot from day to day. So in order to not produce bad forecasts, we will do a re-estimation of parameters every 50 days instead of every day, see [20].

2.17 Value at risk (VaR)

The most well known risk measure is value at risk (VaR). According to [6], considering a log return series at a moment in time t (r_t). The random variable of loss over the period $[t,\ t+h]$ is indicated as

$$L_{t+h} = -(r_{t+h} - r_t) = \Delta r(h).$$

Then, F_L is the cumulative function of loss distribution and it holds that $F_L(x) = P(L \leq x)$. VaR at significance level α (most often 1% and 5%, in this thesis it has been used 1%) is actually an α-quantile of the distribution function F_L ,or in other word, VaR presents the smallest real number satisfying the inequality $F_L(x) \geq \alpha$, i.e.:

$$VaR_\alpha = \inf(x | F_L(x) \geq \alpha).$$

2.18 Backtesting VaR

VaR is a good measures of risk which has several backtesting procedures for validating a set of VaR forecasts, see [7]. The aim of backtesting is to estimate whether the amount of losses predicted by VaR is correct. This process applies unconditional and conditional coverage tests for the correct number of exceedances. The unconditional tests check whether the frequency exceptions, during the selected time interval, are in accordance with the chosen confidence level and for testing, it has been used the Kupiec test, see [13]. On the other hand, conditional coverage tests examine conditionality and changes in data over time, and the most famous test is the Christoffersen independence test, see [7].

2.19 The Kupiec test

According to [7], consider N be the observed number of exceedances in the sample, in the other words, $N = \sum_{t=1}^{T} I_t$ is the total number of violations over a T period of time, where

$$I_t = \begin{cases} 0, & if\ \ r_{t,t+1} \geq VaR_t \quad (Hit) \\ 1, & if\ \ r_{t,t+1} < VaR_t \quad (violation) \end{cases}$$

The Kupiec likelihood-ratio (LR) statistic examines that the expected number of violations is indeed the stated p (In this thesis we use a value of p = 0.01). In other words, we test the null hypothesis $H_0: p = \frac{N}{T}$ against $H_a: p \neq \frac{N}{T}$. The test statistic is

$$LR_{uc} = 2\ln\frac{(\frac{N}{T})^N(1-\frac{N}{T})^{T-N}}{p^N(1-p)^{T-N}}$$

The Kupiec test has a the chi-square distribution with one degree of freedom ($\chi^2(1)$) under the null hypothesis, see [25]. If the LR_{uc} value is high, the null hypothesis will be rejected. If the null hypothesis is rejected, the specific model is not a suitable specification to estimate the VaR. The Kupiec test's power is generally poor, so conditional coverage tests, such as the Christoffersen test, can be used for the further examination of VaR model reliability, see [13].

2.20 The Christoffersen test

This test evaluates actual number of violations are the same as the stated number and if violations are clustered. According to [26], Christoffersen assumes that the violation process I_t can be represented as a Markov chaine with two states:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix}$$

and

$$\pi_{ij} = \Pr(I_t = j | I_{t-1} = i),$$

where π_{01} is the probability of a non-exception being followed by an exception, and π_{11} is the probability of an exception being followed by an exception.

We want to test if violations are clustered or not, and it is possible to test this at the same time as we test if the stated p is the correct one. The null hypothesis is

$$H_0 : \Pi = \Pi_p = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}$$

and in this thesis the coverage rate is p = 0.01. The test statistic testing independence is (see [27]):

$$LR_{ind} = -2\ln[(1-p)^{T-N}p^N] + 2\ln[(1-\pi_{01})^{\pi_{00}}{\pi_{01}}^{\pi_{00}}(1-\pi_{11})^{\pi_{10}}{\pi_{11}}^{\pi_{11}}],$$

The test statistic is distributed as chi-square distribution with two degrees of freedom ($\chi^2(2)$).By combining these two likelihood ratio tests (LR_{uc} and LR_{ind}) could create complete test for coverage and independence, which is also distributed as $\chi^2(2)$:

$$LR_{cc} = LR_{uc} + LR_{ind}.$$

This is the Christoffersen approach to check the predictive ability and accuracy of a VaR model.

A full specification of null and alternative hypotheses for these tests are:

	LR_{uc}	LR_{ind}	LR_{cc}
H_0	Correct unconditional coverage	Exceedances are independent	Correct conditional coverage
H_a	Incorrect unconditional coverage	Exceedances are not independent	Incorrect conditional coverage

We would like to accept the null hypothesis and a high p-value is a sign of a good model.

3 DATA

3.1 Data description

3.1.1 NASDAQ Stock Market Daily Closing Price Index

The NASDAQ Stock Market is an American stock exchange and was founded in 1971 by the National Association of Securities Dealers (NASD), which divested itself of NASDAQ in a series of sales in 2000 and 2001. "NASDAQ" originally represented "National Association of Securities Dealers Automated Quotations". Except the New York Stock Exchange, it already becomes the second-largest stock exchange in the world's stock market. The NASDAQ also has large trading volume than any other electronic stock exchange market. It has been chosen the data from Yahoo Finance between 2010-01-01 and 2017-01-01 because, NASDAQ daily closing price index is very significant and representative, see [14].

Figure 1. Nasdaq daily close prices in 7 years from 2010 to 2017

3.1.2 Standard & Poor 500 Stock Market Daily Closing Price Index

The Standard & Poor's 500, which often called as the S&P 500 (or the S&P), is also an American stock market index . The S&P 500 introduced its first stock index in 1923, began tracking a small number of stocks and then in 1957 it expanded to its current 500. It is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market. It has been chosen the data from Yahoo Finance between 2010-01-01 and 2017-01-01, see [15].

Figure 2. S&P500 daily close prices in 7 years from 2010 to 2017

3.1.3 NIKKEI Stock Market Daily Closing Price Index

The Nikkei (Nikkei heikin kabuki, Nikkei 225), which often called the Nikkei, the Nikkei index, or the Nikkei Stock Average, is a stock market index for the Tokyo Stock Exchange (TSE). It has been calculated daily by the Nihon Keizai Shimbun (Nikkei) newspaper since 1950. This index with longer duration and good comparability has already become the most common and reliable indicators to study the changes in the Japanese's stock market. It has been chosen the index from Yahoo Finance between 2010-01-01 and 2017-01-01, see [16].

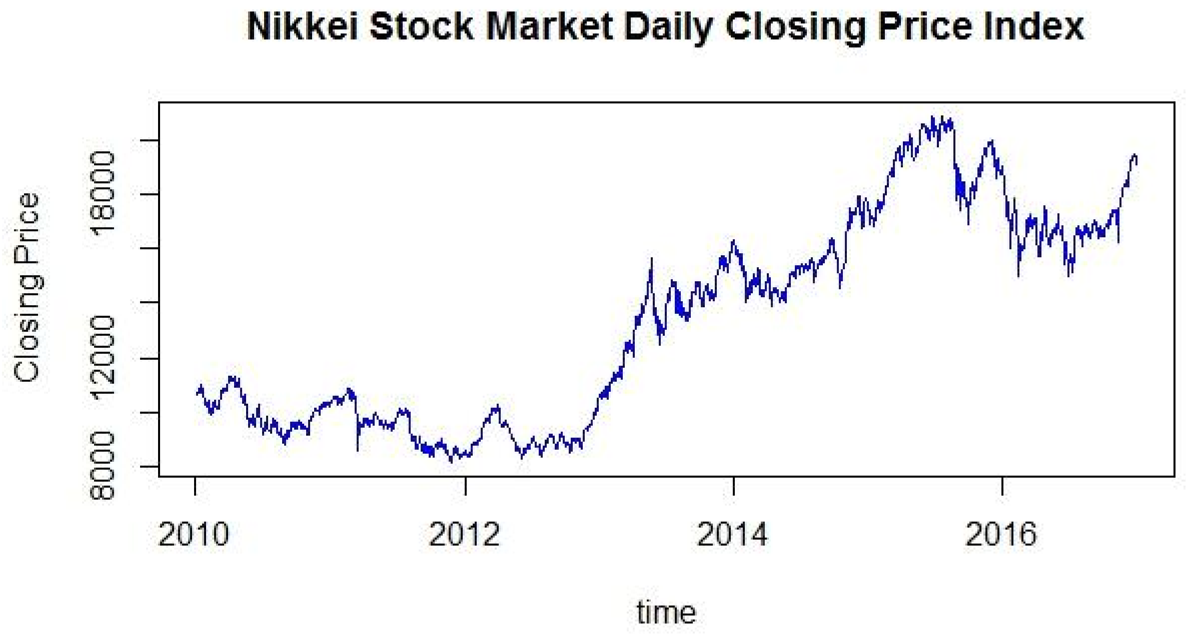


Figure 3. Nikkei daily close prices in 7 years from 2010 to 2017

3.1.4 Dow Jones Stock Market Daily Closing Price Index

The Dow Jones Industrial Average also called DJIA, the Industrial Average, the Dow Jones, the Dow Jones Industrial, the Dow 30 or simply the Dow, is a stock market index which is one of indices created by *Wall Street Journal* editor and Dow Jones & Company co-founder Charles Dow. The Dow Jones was first calculated on May 26, 1896. It is the second-oldest U.S. market index after the Dow Jones Transportation Average, which was also created by Dow. It has been chosen the index from Yahoo Finance between 2010-01-01 and 2017-01-01, see [17].

Figure 4. Dow Jones daily close prices in 7 years from 2010 to 2017

3.2 Data Analysis

The stationarity is one of the essential conditions in time series. Hence, to better examine the close returns, the continuously compounded returns, which is called as log-returns will be used instead of closing prices in this research. Log-returns simply eliminate the non-stationary properties of the data set, making the financial data more stable. The distribution of log-returns over larger periods of time (such as a month, half a year, a year) is closer to the normal distribution than for hourly or daily log-returns (or returns).

We calculate the return in percent:

$$r_t = 100[\log(p_t) - \log(p_{t-1})] \qquad (2.1)$$

Where p_t and p_{t-1} are the closing prices of current and previous date respectively.
After getting the returns of the stock price, we need to summary and list the features of these data, including sample size, mean, standard deviation, minimum, maximum, skewness, kurtosis, Shapiro–Wilk test which is a test of normality, see [18] and Quantile-Quantile plot.

3.2.1 NASDAQ Stock Prices analysis

The NASDAQ stock market includes exactly 1763 observations from the 2010-01-01 to 2017-01-01 in NASDAQ market. According to the formula (2.1), we get the 1762 returns data.

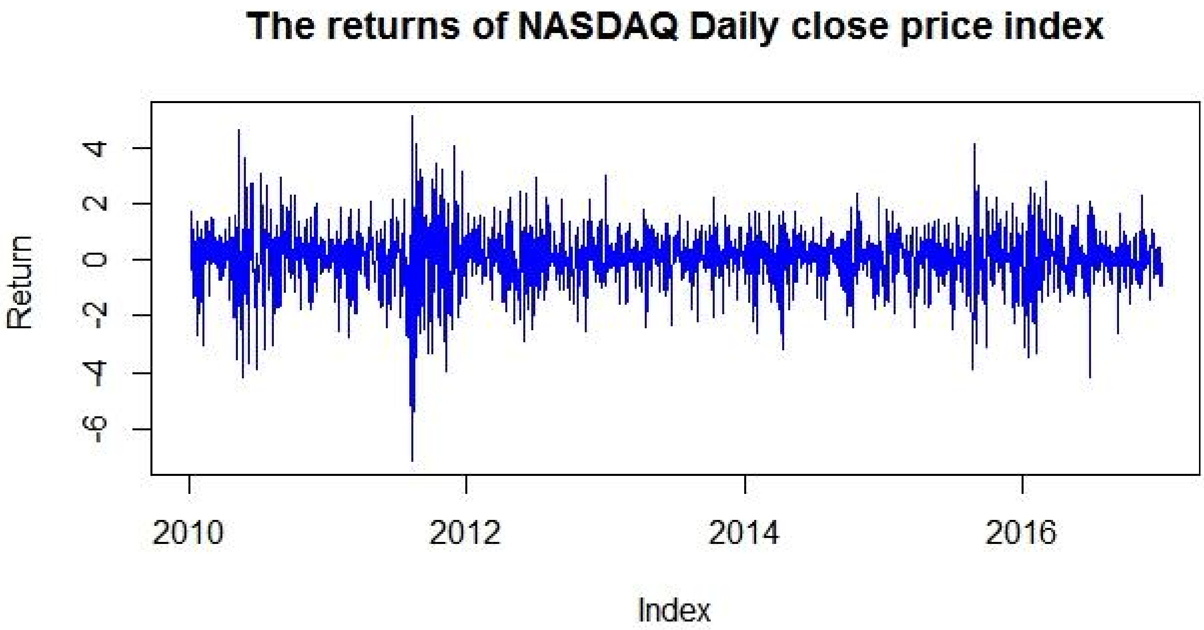


Figure 5. Nasdaq daily returns

The returns appear to oscillate around a constant level, but present volatility clustering. Large changes in the returns tend to cluster together, and small changes tend to cluster together. That is, the series present conditional heteroscedasticity.

The descriptions of Nasdaq data

	Sample size	Min	Max	Mean	Sd	Skewness	Kurtosis
NASDAQ	1762	-7.15	5.16	0.05	11.016	-0.417	3.237

The kurtosis of the NASDAQ daily returns is 3.237 which is a bit higher than the value of normal distribution (kurtosis=3). This value shows the financial time series has heavier tails and is called a leptokurtic distribution. The skewness is -0.417, not zero, which means it is not symmetric. It also has a negative value, so the left-hand tail will be longer than the right-hand tail.

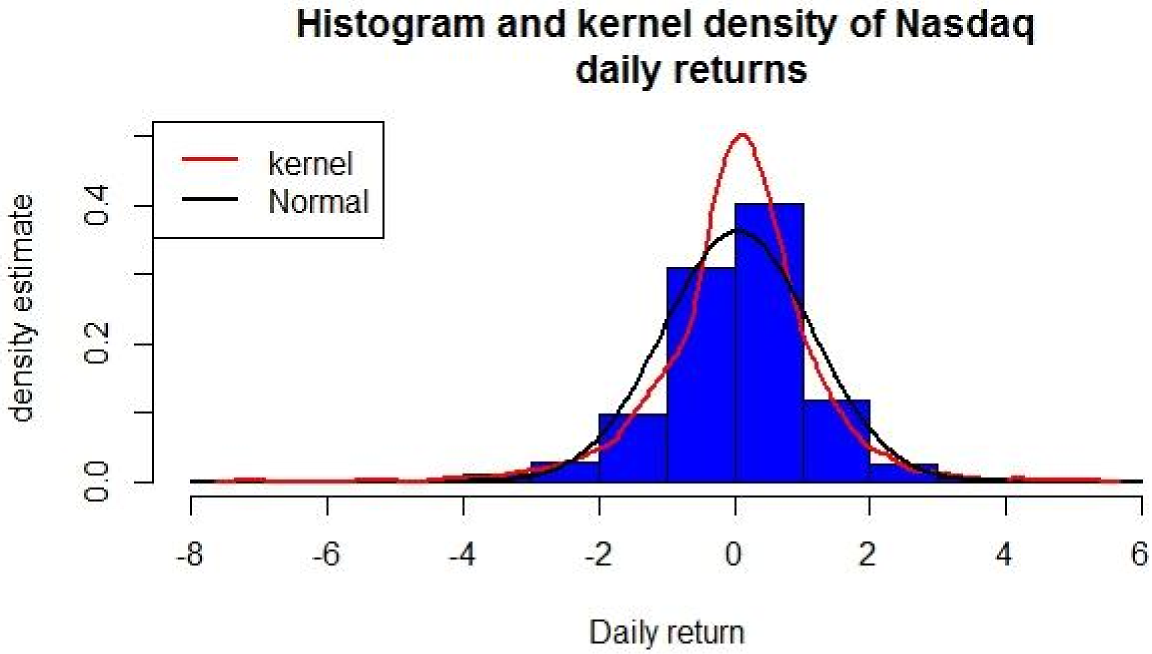


Figure 6. Nasdaq histogram and kernel density estimate (see [19])

At the same time, the Shapiro–Wilk normality test could tell us another feature. By the assumption of 5% significance level, the p-value < 0.05 represents the non-normality of the series data (when the p-value is less than 0.05 then the null hypothesis that the data are normally distributed is rejected). In our case by the Shapiro-Wilk normality test, the p-value is less than 2.2e-16, so the distribution of NASDAQ daily return is not normal. We perform a Quantile-Quantile plot (*QQ-plot*) as well. It can check whether or not a time series comes from a certain distribution, which is a visual method of analysis where the analyser can get a better knowledge of the empirical distribution and its deviations from a theoretical distribution. In our case, we can see empirical quantiles don't match normal quantiles in the tails

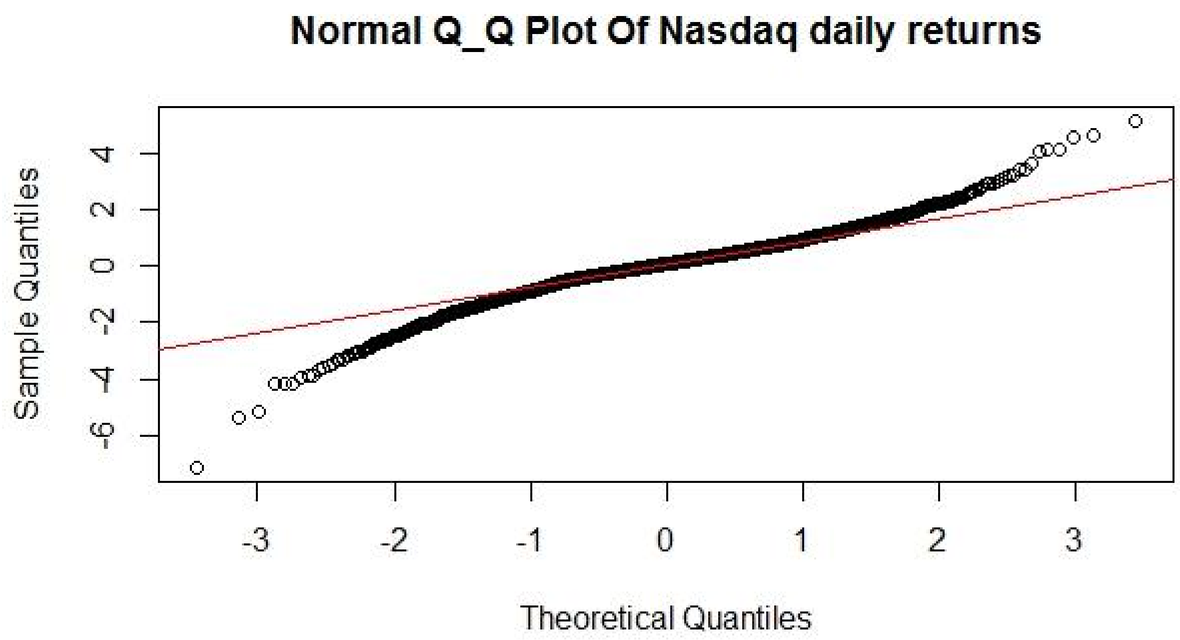


Figure 7. Quantile-Quantile plot for Nasdaq daily returns

3.2.2 Standard & Poor 500 Stock Prices analysis

The Standard & poor 500 stock market also includes exactly 1763 observations from the 2010-01-01 to 2017-01-01. According to the formula (2.1), We get the 1762 return data. Also in this case, returns exhibit the volatility clustering property.

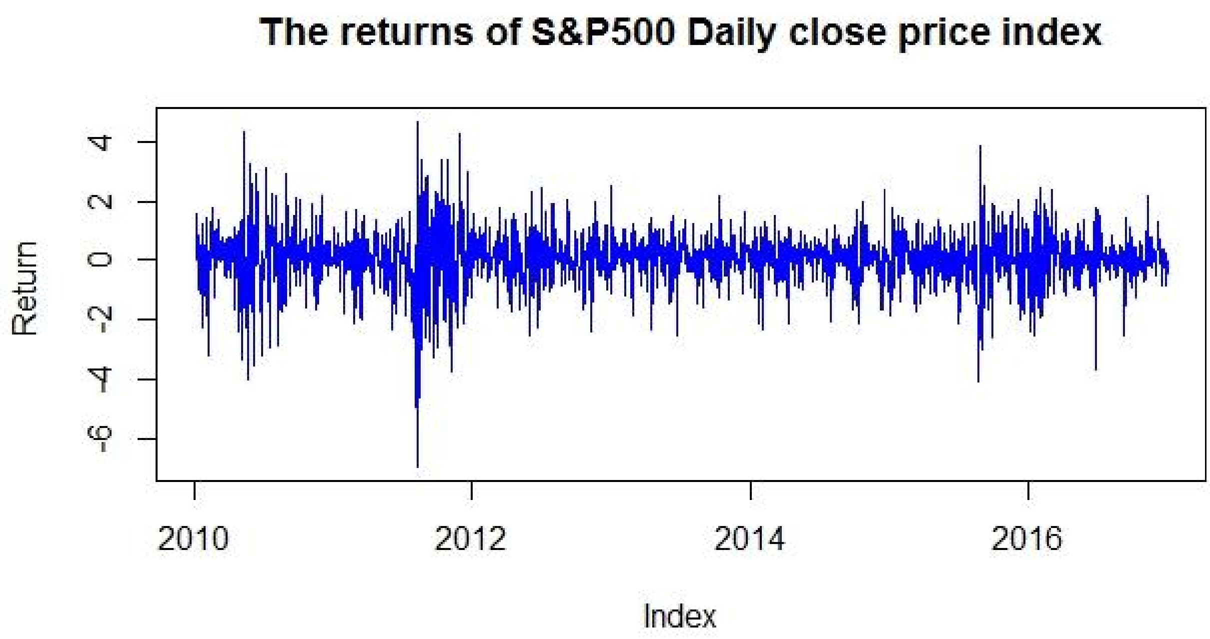


Figure 8. S&P500 daily returns

The descriptions of S&P 500 data

	Sample size	Min	Max	Mean	Sd	Skewness	Kurtosis
S&P 500	1762	-6.89	4.63	0.04	0.98	-0.438	4.19

The kurtosis of the S&P500 daily returns is 4.19 which is higher than the value of normal distribution (kurtosis=3), so S&P500 daily returns as Nasdaq daily returns also has a leptokurtic distribution. The skewness is -0.438, which means it is not symmetric and the left-hand tail will be longer than the right-hand tail. The p-value of Shapiro–Wilk normality test is less than 2.2e-16, so the distribution of S&P 500 daily returns is not normal. Also Q-Q plot shows empirical quantiles don't match normal quantiles in the tails.

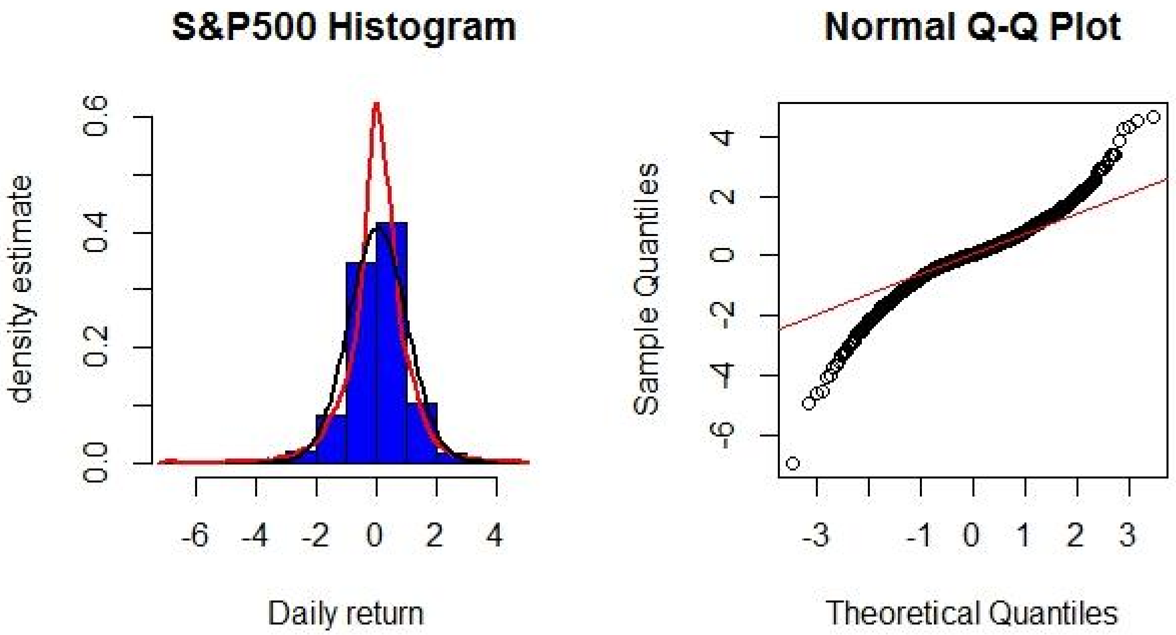


Figure 9. Histograms and Q-Q plots of the S&P500 daily returns

3.2.3 NIKKEI Stock Prices analysis

The Nikkei stock market contains exactly 1714 observations from the 2010-01-01 to 2017-01-01. According to the formula (2.1), we get the 1713 returns data. Also returns of Nikkei stock market present the volatility clustering property.

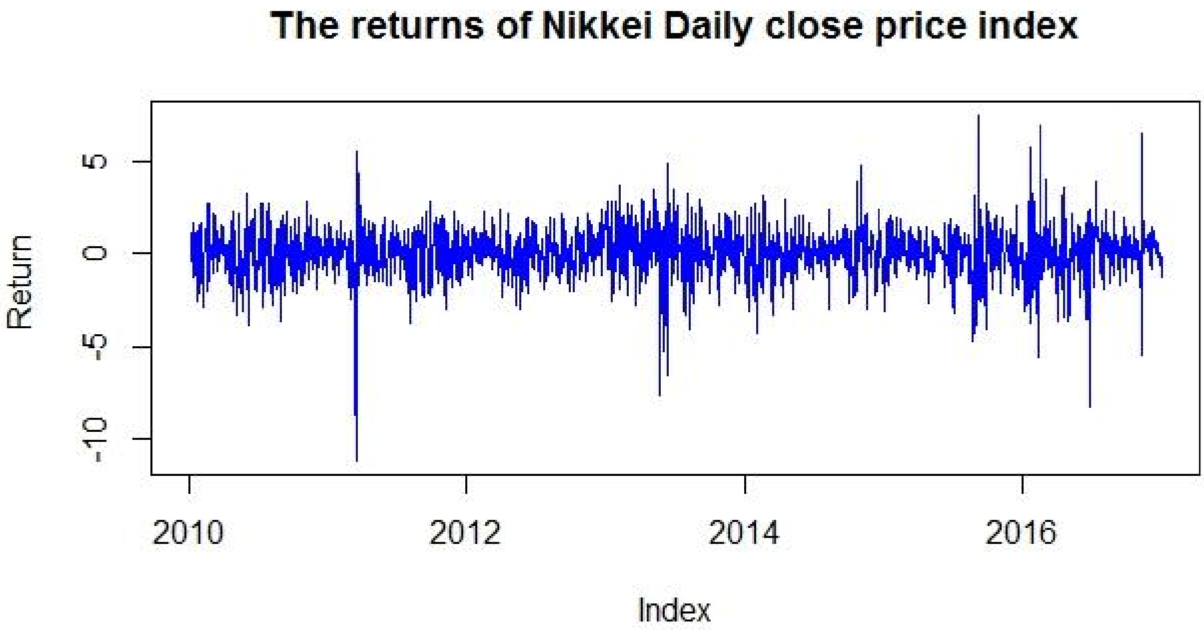
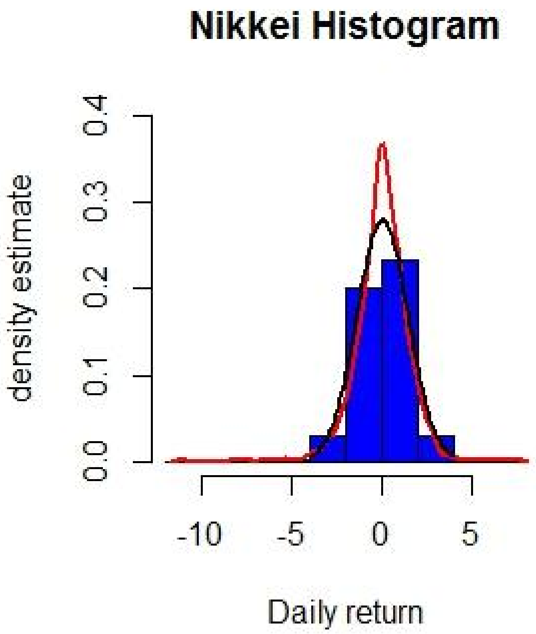


Figure 10. Nikkei daily returns

The descriptions of Nikkei data

	Sample size	Min	Max	Mean	Sd	Skewness	Kurtosis
Nikkei	1713	-11.15	7.43	0.034	1.43	-0.54	4.97

The kurtosis of the Nikkei daily returns is 4.97 which is higher than the value of normal distribution (kurtosis=3), so it has a leptokurtic distribution. The skewness is -0.54, so it is not symmetric and the left-hand tail will be longer than the right-hand tail. The p-value of Shapiro–Wilk normality test is also less than 2.2e-16, so the distribution of Nikkei daily returns is not normal. Also Q-Q plot shows empirical quantiles don't match normal quantiles in the tails.



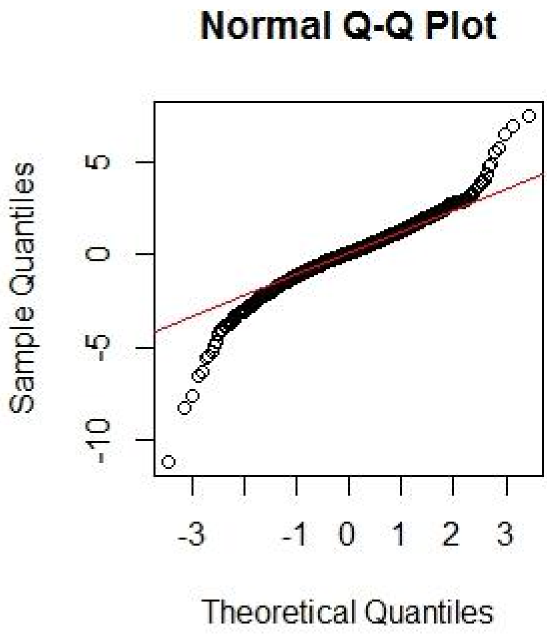


Figure 11. Histograms and Q-Q plots of the Nikkei daily returns

3.2.4 Dow Jones Stock Prices analysis

The Dow Jones stock market contains exactly 1763 observations from the 2010-01-01 to 2017-01-01. According to the formula (2.1), We get the 1762 returns data. Also here the presence of the volatility clustering property is evidence.

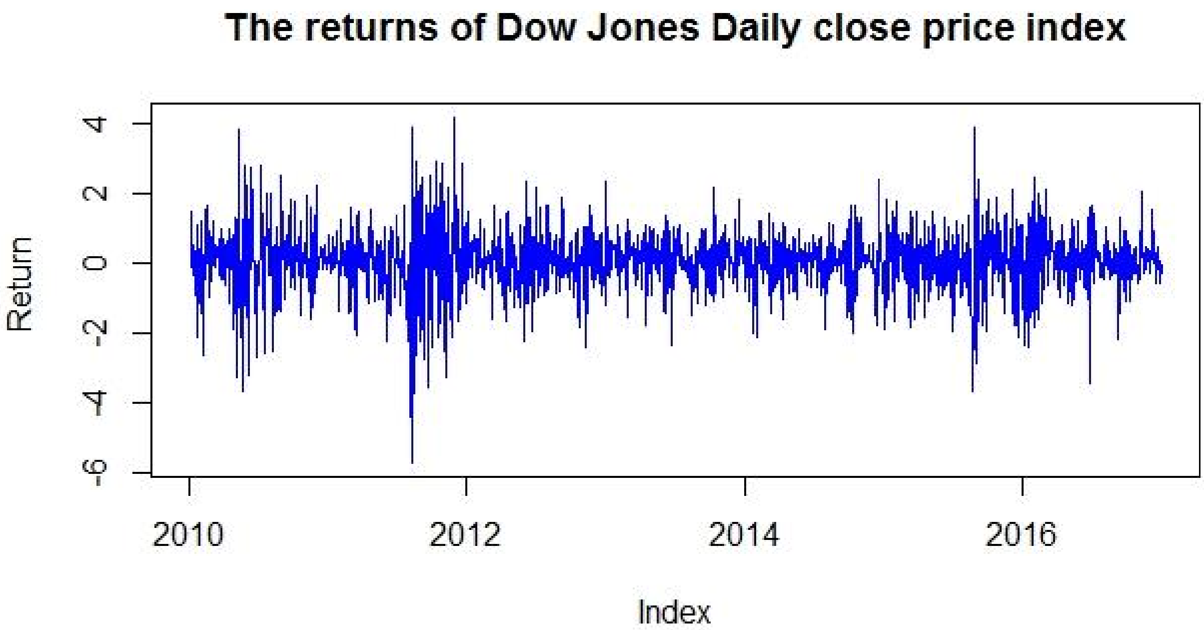


Figure 12. Dow Jones daily returns

The descriptions of Dow Jones data

	Sample size	Min	Max	Mean	Sd	Skewness	Kurtosis
Dow Jones	1762	-5.7	4.15	0.036	0.91	-0.38	3.516

Based on the kurtosis and skewness's values, the distribution of Dow Jones daily returns value is not symmetric and has the fat-tail characteristic (leptokurtic distribution) as Nasdaq, S&P 500 and Nikkei daily returns. Also the p-value of Shapiro–Wilk normality test is less than 2.2e-16, so the distribution of Dow Jones daily returns is not normal and Q-Q plot shows empirical quantiles don't match normal quantiles in the tails.

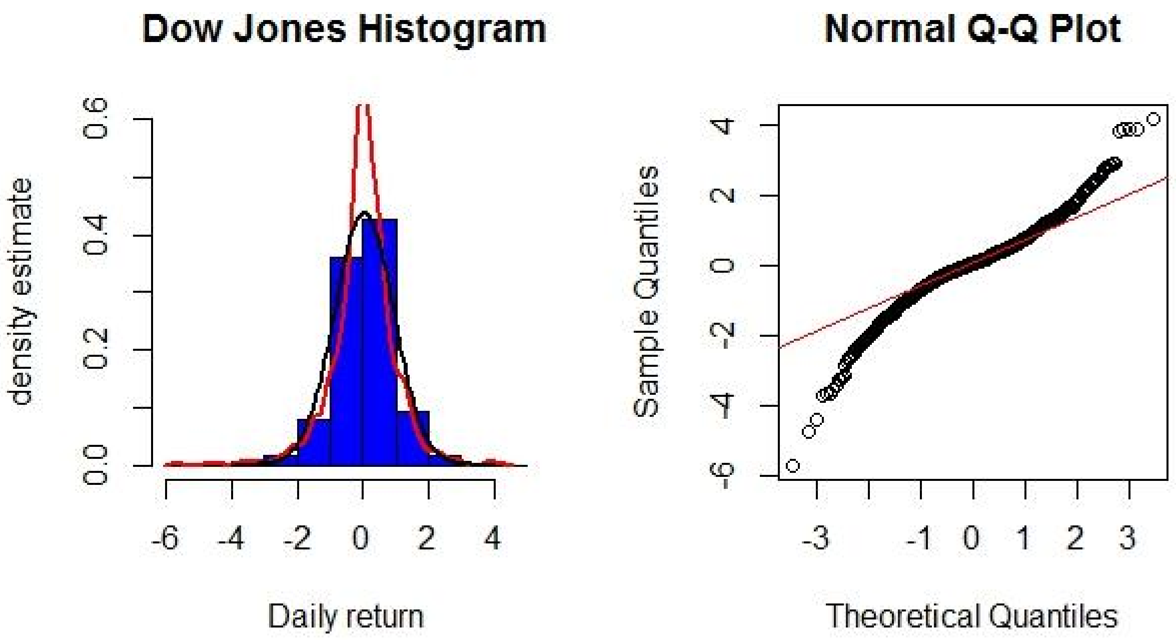


Figure 13. Histograms and Q-Q plots of the Dow Jones daily returns

3.3 Estimation

The autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of Nasdaq daily returns and its squared are presented in Figure 14 and Figure 15 respectively. The reason of looking at the ACF and PACF of squared returns to determine whether there exists ARCH effects is that the true variance of r_t is

$$\mathrm{var}(r_t|\mathcal{F}_{t-1}) = E((r_t - E[r_t])^2|\mathcal{F}_{t-1}) = E(\varepsilon_t^2|\mathcal{F}_{t-1})$$

So squaring returns is a good process to study volatility.

The results in ACF of Figure 14 represents almost all pikes within the 95% confidence band, that is, ACF decay rapidly to zero, although ACF approves significant autocorrelation at lag 3, 5,...
On the other hand squared Nasdaq returns have high correlation as we can see in ACF of Figure 15. Thus, we may conclude that the returns process has a strong non-linear dependence.
In addition, PACF of squared return data does not show a clear sign of the exponential decay of lags towards to zero mean in this study. That's why various ARMA models will be adopted in next section to search for a fit model and to see whether autocorrelation occurs at some lags.

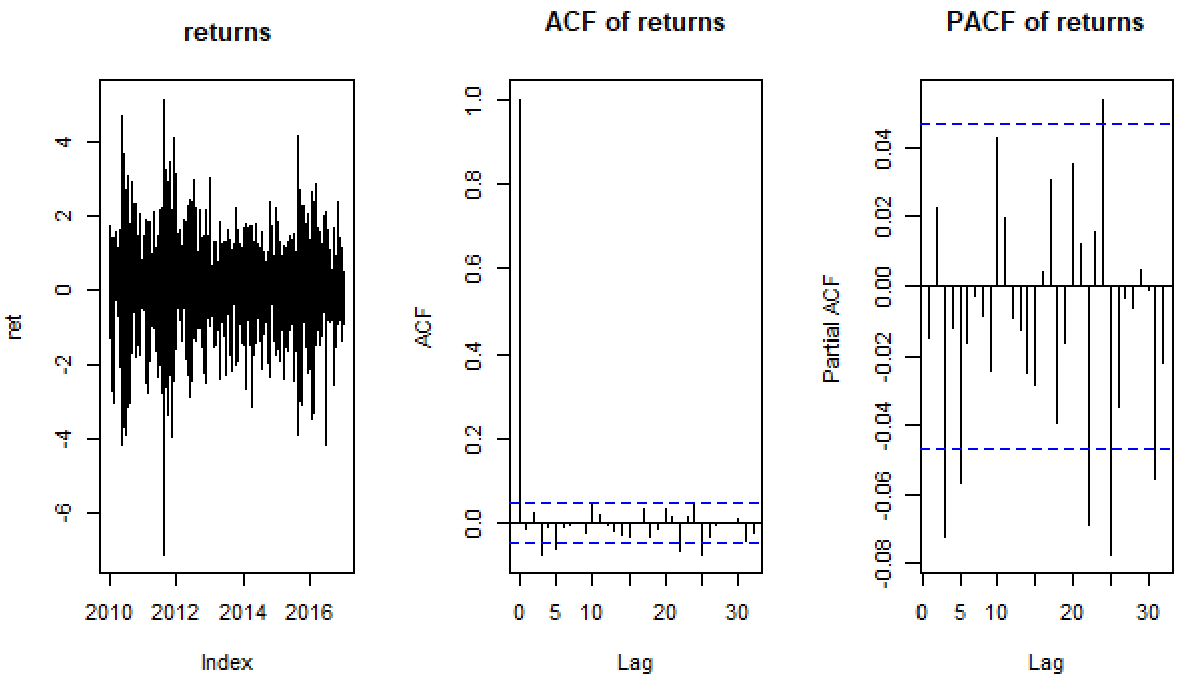


Figure 14. ACF and PACF of Nasdaq daily returns

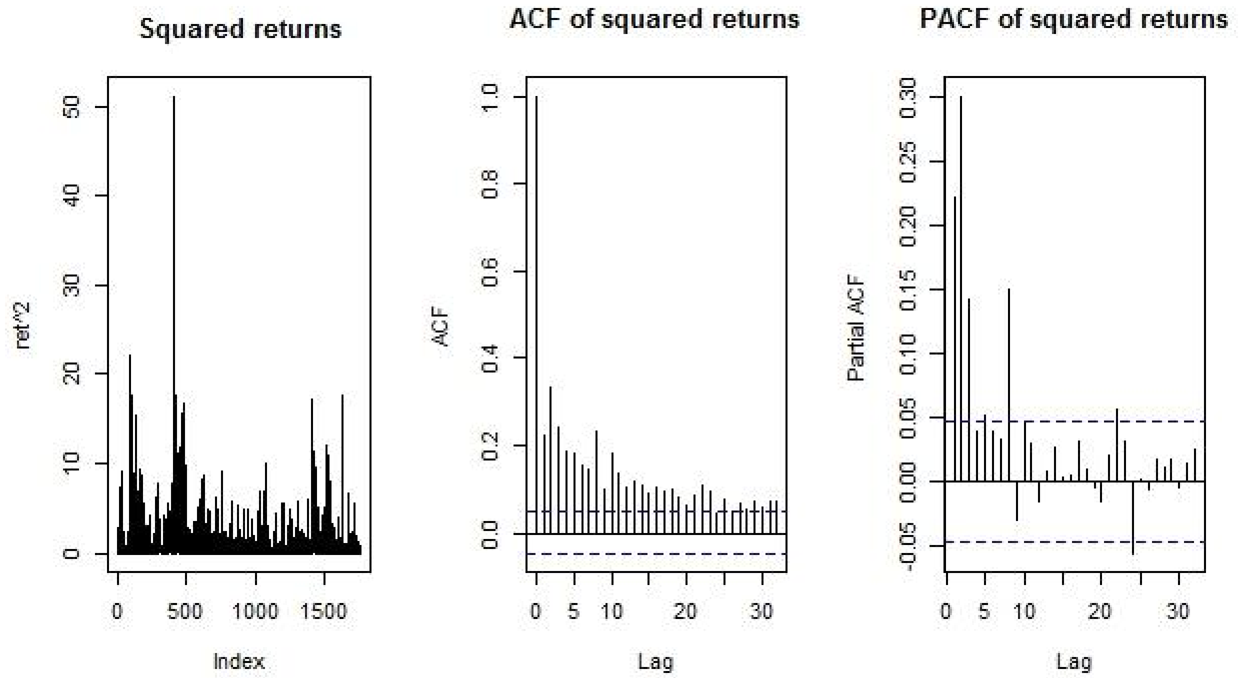


Figure 15. ACF and PACF of squared Nasdaq daily returns

ACF and PACF of daily returns and squared daily returns of S&P 500 , Nikkei and Dow Jones stock markets are also presented in Figure 16, Figure 17 and Figure 18 respectively which we can interpret almost the same as Nasdaq.

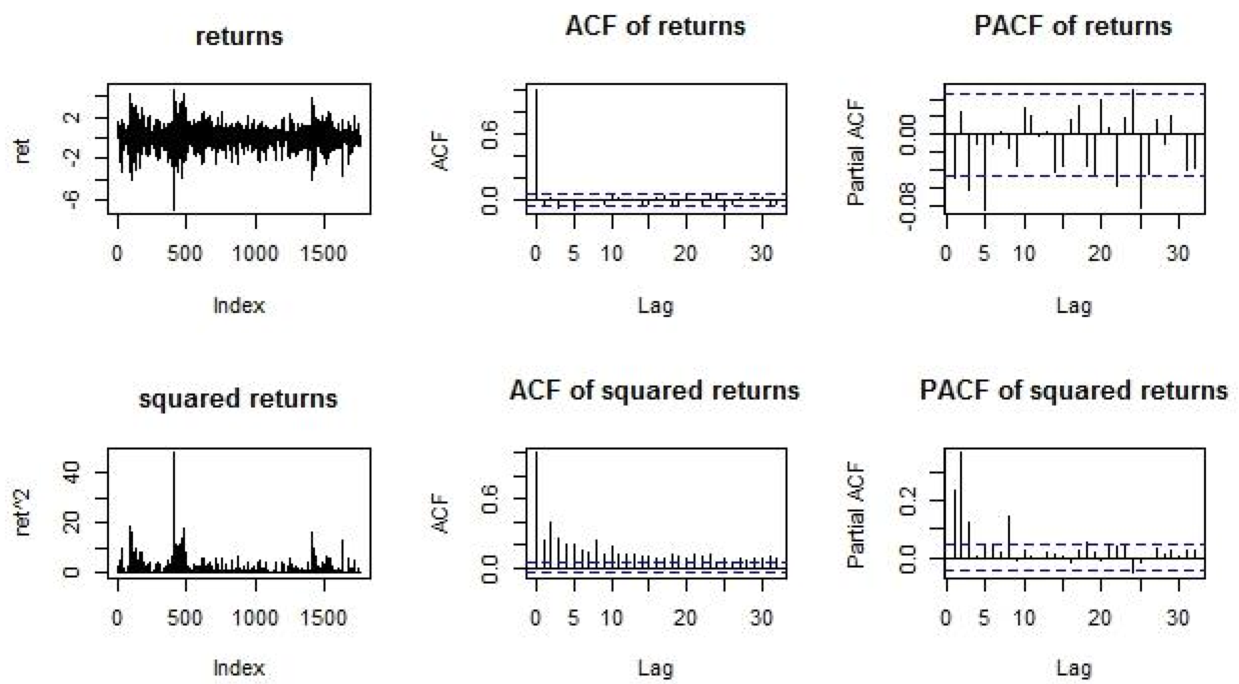


Figure 16. ACF and PACF of S&P 500 daily returns and its squared

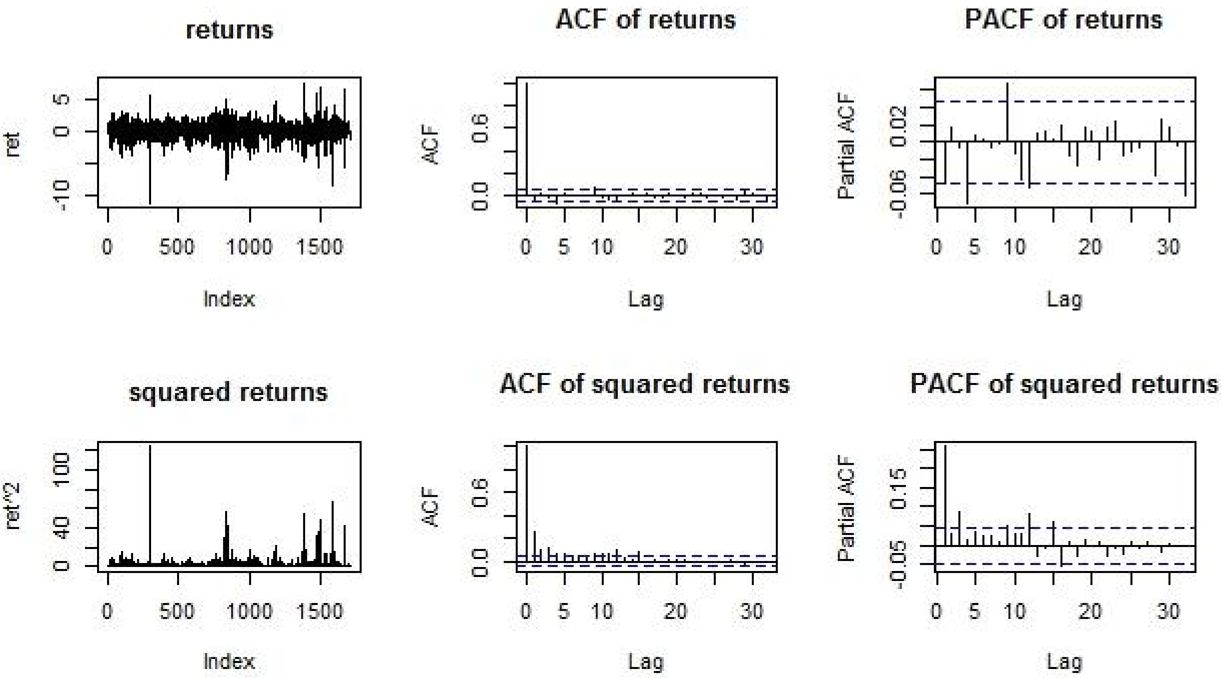


Figure 17. ACF and PACF of Nikkei daily returns and its squared

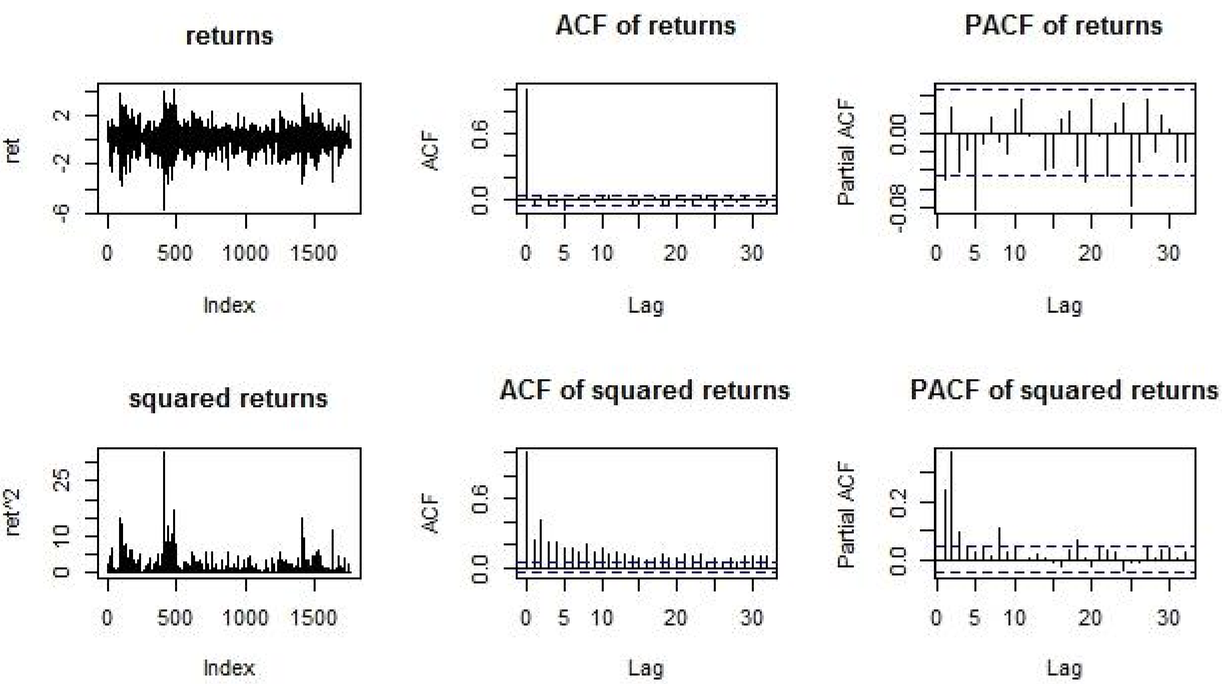


Figure 18. ACF and PACF of Dow Jones daily returns and its squared

Has been performed the Box-Ljung test (see[8]) to test the independence of daily returns and squared values of daily returns . By the assumption of 5% significance level, all of the results on Nasdaq's daily returns are significant so the null hypothesis has to be rejected (instead of Box-

Ljung test of returns at lag 2 that are not correlated as the p-values>0.05). However, it shows signs of ARCH effect on the Nasdaq's daily returns which leads us to proceed ahead to GARCH model.

The Box-Ljung test results on Nasdaq daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	1.3118	10.82	17.169
	p-value	0.519	0.02867	0.008681

The Box-Ljung test results on squared values of Nasdaq daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	284.83	447.74	548.13
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16

All of the results of S&P 500's daily returns are also significant so the null hypothesis has to be rejected (instead of Box-Ljung test of returns at lag 2 that are not correlated as the p-values>0.05). However, it shows signs of ARCH effect on the S&P 500's returns which leads us to proceed ahead to GARCH model.

The Box-Ljung test results on S&P 500 daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	5.5227	13.049	26.347
	p-value	0.06321	0.01104	0.0001918

The Box-Ljung test results on squared values of S&P 500 daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	387.66	568.36	681.64
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16

All of the results of Nikkei's daily returns are also significant so the null hypothesis has to be rejected (instead of Box-Ljung test of returns at lag 2 that are not correlated as the p-values>0.05). However, it shows signs of ARCH effect on the Nikkei's returns which leads us to proceed ahead to GARCH model.

The Box-Ljung test results on Nikkei daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	4.5328	12.84	13.139
	p-value	0.1037	0.01208	0.04088

The Box-Ljung test results on squared values of Nikkei daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	128.51	156.92	167.83
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16

All of the results of Dow Jones's daily returns are also significant so the null hypothesis has to be rejected (instead of Box-Ljung test of returns at lag 2 that are not correlated as the p-values>0.05). However, it shows signs of ARCH effect on the Dow Jones's returns which leads us to proceed ahead to GARCH model.

The Box-Ljung test results on Dow Jones daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	5.8687	9.5817	21.673
	p-value	0.05317	0.04809	0.001388

The Box-Ljung test results on squared values of Dow Jones daily returns				
Box-Ljung test		**lag=2**	**lag=4**	**lag=6**
	test value	396.76	579.27	681.57
	p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16

4 Application and Results

4.1 Application in NASDAQ daily return

4.1.1 Selection of ARMA (p, q) model

First step is selection of suitable ARMA (p, q) model for NASDAQ daily return. Based on ACF and PACF in Figure 14 and Figure 15, we could select different parameter p and q for ARMA models. We compare the value of Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). A lower AIC or BIC value indicates a better fit (more parsimonious model).

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	5345.98	5347.08	5346	5347.21	5342.44	5343.19	5343.28	5344.76
BIC	5362.4	5368.97	5362.42	5369.1	5364.34	5370.55	5370.65	5377.6

The lowest AIC comes at 5342.44 of ARMA(1,1) model in comparison to others. As a result, the study will mainly focus on ARMA(1,1) model. So the conditional mean equation is as follows

$$r_t = \alpha + \phi_1 r_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad t = 1, \dots, T.$$

The estimated parameters are $\alpha = 0.0490$, $\phi_1 = -0.8534$ and $\theta_1 = 0.8232$.
Then, we could calculate

$$\hat{\varepsilon}_t = r_t - \left(\hat{\alpha} + \hat{\phi}_1 r_{t-1} + \hat{\theta}_1 \varepsilon_{t-1}\right).$$

4.1.2 Testing estimation of residuals for ARMA model

To examine independence of residuals, it has been checked the ACF of estimated residuals of ARMA(1,1) model. According to the results in ACF of Figure 19 almost all autocorrelations stay within the 95% confidence band. Although, there are still some significant autocorrelation at lags 22 and 25.

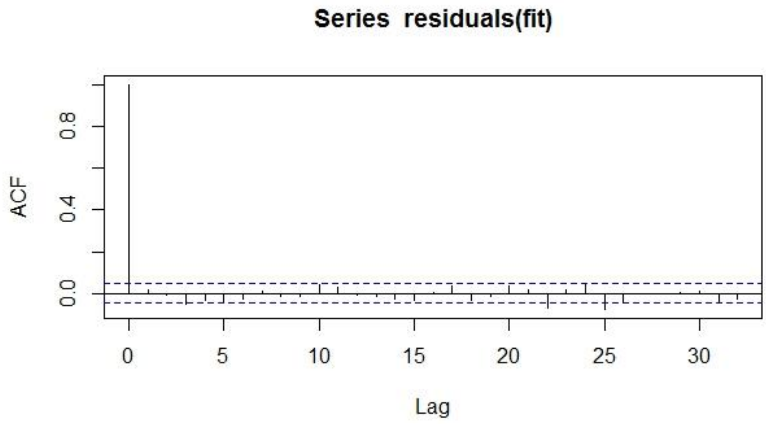


Figure 19. ACF of the estimated residuals for ARMA(1,1) models

To check the signal of autocorrelation in ARMA(1,1) model, the Ljung – Box test has been used as well. By the assumption of 5% significance level, according to the p-value of mentioned lags, the null hypothesis has to be rejected and indicates that the independence assumption of residuals can be eliminated.

		lag=22	lag=25
Box-Ljung test	test value	36.003	50.658
	p-value	0.03034	0.001767

In addition, the histogram of estimated residuals of ARMA(1,1) model and QQ–plot in Figure 20 have been used to test normality assumption of residuals. The charts approve that residuals are not normal distributed, so the GARCH models can be applied to Nasdaq's returns.

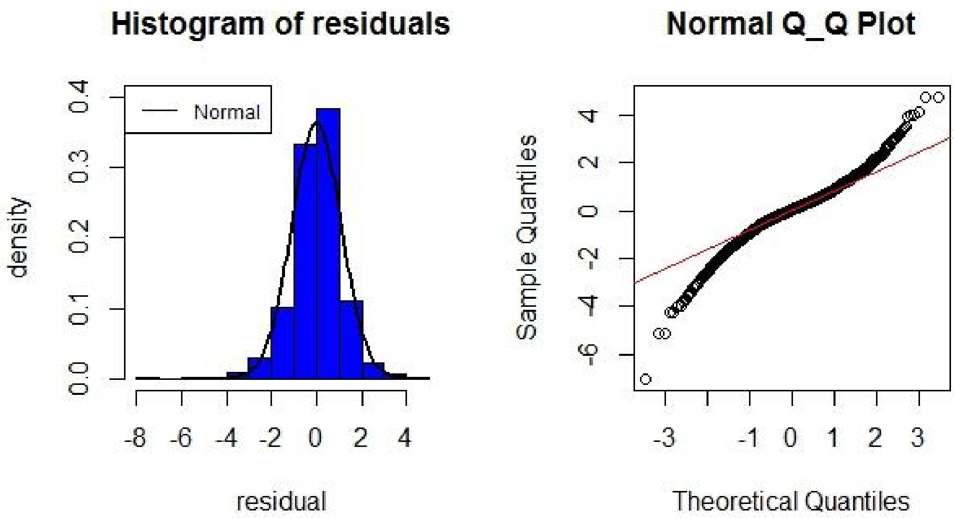


Figure 20. Histogram and QQ-plot of estimated residuals

4.1.3 GARCH Results

We estimate GARCH(1,1), GARCH(2,1), ARMA(1,1)-GARCH(1,1) and ARMA(1,1)-GARCH(2,1) with normal distribution, Student-t distribution and Skew Student-t distribution for the residuals. We can see the estimation of the models parameters for different distribution in Figure 21, Figure 22, Figure 23 and Figure 24.

Parameters estimation for GARCH(1,1)					
Normal		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.08442	0.02145	3.935	8.31e-05
	Omega	0.05520	0.01144	4.825	1.40e-06
	alpha1	0.12059	0.01751	6.888	5.66e-12
	beta1	0.83159	0.02203	37.750	< 2e-16
Student-t		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.10968	0.02026	5.413	6.19e-08
	Omega	0.04988	0.01330	3.752	0.000176
	alpha1	0.12893	0.02233	5.774	7.73e-09
	beta1	0.83385	0.02539	32.841	< 2e-16
	Shape	620.501	0.98879	6.275	3.49e-10
Skew Student-t		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.08217	0.02107	3.900	9.62e-05
	Omega	0.04557	0.01232	3.698	0.000217
	alpha1	0.12491	0.02089	5.978	2.26e-09
	beta1	0.83809	0.02441	34.336	< 2e-16
	Skew	0.87749	0.02801	31.329	< 2e-16
	Shape	703.908	127.165	5.535	3.11e-08

Figure 21. Parameters estimation for GARCH(1,1) of Nasdaq returns

By assuming the significance level of 5%, in GARCH(1,1) model, all of the estimated parameters are significant. In GARCH(2,1) model almost all of the estimated parameters are significant except α_2 for all the three different distribution (normal, Student-t and Skew Student-t distribution).

Parameters estimation for GARCH(2,1)					
Normal		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.08308	0.02149	3.866	0.000111
	Omega	0.06178	0.01433	4.310	1.63e-05
	alpha1	0.09629	0.02887	3.335	0.000852
	alpha2	0.03386	0.03337	1.015	0.310220
	beta1	0.81618	0.02841	28.730	<2e-16

		Estimate	Std.Error	t value	Pr(>|t|)
Student-t	**Mu**	0.10811	0.02029	5.327	9.96e-08
	Omega	0.06023	0.01718	3.506	0.000454
	alpha1	0.08652	0.03361	2.574	0.010043
	alpha2	0.06237	0.04165	1.497	0.134306
	beta1	0.80641	0.03431	23.503	<2e-16
	Shape	616.772	0.97708	6.312	2.75e-10
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.07963	0.02116	3.764	0.000167
	Omega	0.05680	0.01628	3.489	0.000485
	alpha1	0.07958	0.03129	2.543	0.010977
	alpha2	0.06700	0.03943	1.699	0.089273
	beta1	0.80772	0.03339	24.189	<2e-16
	Skew	0.87517	0.02800	31.261	<2e-16
	Shape	695.358	124.280	5.595	2.2e-08

Figure 22. Parameters estimation for GARCH(2,1) of Nasdaq returns

In ARMA(1,1)-GARCH(1,1) model, the parameters estimated for Skew Student-t distribution are all significant, but not under normal and Student-t distribution. As we can see in Figure 23, the coefficient of the first term of the autoregressive process, the first term of the moving average process and μ under normal distribution and Student-t distribution are not significant. On the other hand, the Skew Student-t distribution has the better estimated parameters and all are significant.

Parameters estimation for ARMA(1,1)- GARCH(1,1)					
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
Normal	**Mu**	0.08259	0.05474	1.509	0.131
	ar1	0.01316	0.60011	0.022	0.982
	ma1	-0.02178	0.60902	-0.036	0.971
	omega	0.05568	0.01154	4.826	1.40e-06
	alpha1	0.12148	0.01765	6.882	5.91e-12
	beta1	0.83045	0.02221	37.399	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.109673	0.056182	1.952	0.050925
	ar1	0.001539	0.476708	0.003	0.997425
	ma1	-0.023730	0.480911	-0.049	0.960645
	omega	0.050511	0.013482	3.747	0.000179
	alpha1	0.130804	0.022719	5.758	8.53e-09
	beta1	0.832250	0.025658	32.436	<2e-16
	shape	6.086.335	0.961863	6.328	2.49e-10

		Estimate	Std.Error	t value	Pr(>|t|)
Skew Student-t	Mu	0.024652	0.009304	2.650	0.008059
	ar1	0.695278	0.100433	6.923	4.43e-12
	ma1	-0.761835	0.096364	-7.906	2.66e-15
	omega	0.040700	0.011416	3.565	0.000363
	alpha1	0.115839	0.019655	5.894	3.78e-09
	beta1	0.849638	0.023306	36.456	<2e-16
	Skew	0.835666	0.030098	27.764	<2e-16
	shape	7.317.389	1.354.971	5.400	6.65e-08

Figure 23. Parameters estimation for ARMA(1,1)- GARCH(1,1) of Nasdaq returns

In ARMA(1,1)-GARCH(2,1) model, the parameters estimated for three distributions are not all significant. As we can see in Figure 24, the coefficient of the first term of the autoregressive process, the first term of the moving average process, µ and α_2 under normal distribution and Student-t distribution are not significant. The parameter α_2 is not significant also under Skew Student-t distribution.

Parameters estimation for ARMA(1,1)- GARCH(2,1)					
		Estimate	Std.Error	t value	Pr(>|t|)
Normal	mu	0.08043	0.05736	1.402	0.160883
	ar1	0.02356	0.64502	0.037	0.970860
	ma1	-0.03174	0.65541	-0.048	0.961373
	omega	0.06212	0.01448	4.290	1.79e-05
	alpha1	0.09553	0.02866	3.334	0.000856
	alpha2	0.03503	0.03347	1.047	0.295327
	beta1	0.81546	0.02872	28.398	<2e-16
		Estimate	Std.Error	t value	Pr(>|t|)
Student-t	mu	0.10606	0.05861	1.810	0.070355
	ar1	0.02015	0.50926	0.040	0.968446
	ma1	-0.04193	0.51414	-0.082	0.935000
	omega	0.06067	0.01743	3.480	0.000501
	alpha1	0.08695	0.03357	2.590	0.009592
	alpha2	0.06345	0.04207	1.508	0.131560
	beta1	0.80530	0.03479	23.151	<2e-16
	shape	604.840	0.94930	6.371	1.87e-10
		Estimate	Std.Error	t value	Pr(>|t|)
Skew Student-t	mu	0.024890	0.009939	2.504	0.012272
	ar1	0.683048	0.112113	6.092	1.11e-09
	ma1	-0.747956	0.108297	-6.907	4.97e-12
	omega	0.050296	0.015000	3.353	0.000799
	alpha1	0.075453	0.029555	2.553	0.010680
	alpha2	0.059321	0.037552	1.580	0.114180

	beta1	0.823159	0.031770	25.910	<2e-16
	skew	0.835337	0.029839	27.995	<2e-16
	shape	7.177.019	1.309.113	5.482	4.20e-08

Figure 24. Parameters estimation for ARMA(1,1)- GARCH(2,1) of Nasdaq returns

Figure 25 shows the Akaike Information Criterion (AIC) and Bayes information Criterion (BIC) under the three distributions for the models. As we know, a lower AIC or BIC value indicates a better fit. In our case ARMA(1,1)-GARCH(2,1) model with Skew Student-t distribution has the lowest AIC which is 2.785 and then the lowest comes at 2.786 of ARMA(1,1)-GARCH(1,1) model with Skew Student-t distribution. Obviously, the ARMA(1,1)-GARCH(1,1) model with Skew Student-t distribution is the best model, because all of the estimated parameters are significant and the fitted model is as follow

$$r_t = \mu + \phi_1(r_{t-1} - \mu) + \theta_1\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t = \sigma_t\varepsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

where estimated parameters according to Figure 23 are

$\mu = 0.024652$, $\phi_1 = 0.695278, \theta_1 = -0.761835$, $\omega = 0.0407, \alpha_1 = 0.115839$ and $\beta_1 = 0.849638$, also we could see the skewness value which has p-value < 0.05 and is significant. Since, the skew value $= 0.835666$, it indicates that the student-t distribution is skewed positively. The shape parameter is the estimated degrees of freedom in the skewed Student-t distribution of ε_t and has p-value < 0.05, so is significant.

Information Criterion Statistics		
	AIC	**BIC**
GARCH(1,1)-Normal	2.831	2.843
GARCH(1,1)-Student-t	2.798	2.814
GARCH(1,1)-Skew Student-t	2.790	2.808
GARCH(2,1)-Normal	2.831	2.847
GARCH(2,1)-Student-t	2.798	2.817
GARCH(2,1)-Skew Student-t	2.789	2.811
ARMA(1,1)-GARCH(1,1)-Normal	2.832	2.850
ARMA(1,1)-GARCH(1,1)-Student-t	2.798	2.820
ARMA(1,1)-GARCH(1,1)-Skew Student-t	2.786	2.811
ARMA(1,1)-GARCH(2,1)-Normal	2.832	2.854
ARMA(1,1)-GARCH(2,1)-Student-t	2.798	2.823
ARMA(1,1)-GARCH(2,1)-Skew Student-t	2.785	2.813

Figure 25. Information Criterion Statistics for different models of Nasdaq returns

4.1.4 Residuals Diagnostics of GARCH model

In order to test for the validity of analysis of GARCH models, we should make sure that standardised residuals and squared standardised residuals are free from serial autocorrelation. In addition, we have to make sure that our model capture all ARCH effect, which means the ARCH effect should not be exist anymore, see [20]. It has been performed the Standardised Residuals Tests of ARMA(1,1)-GARCH(1,1) with Skew Student-t distribution in Figure 26. By looking at Ljung-Box test on residuals and squared residuals which have p-values>0.05, it has been failed to reject the null hypothesis and there is no evidence of autocorrelation in the residuals and squared residuals. As a result, we can conclude that the residuals behave as white noise.

Standardised Residuals Tests:				
			Statistic	**p-value**
Jarque-Bera Test	R	Chi^2	180.4154	0
Shapiro-Wilk Test	R	W	0.9797459	4.357308e-15
Ljung-Box Test	R	Q(10)	12.86835	0.2311257
Ljung-Box Test	R	Q(15)	17.5842	0.2851557
Ljung-Box Test	R	Q(20)	21.10254	0.3911193
Ljung-Box Test	R^2	Q(10)	4.265787	0.9345613
Ljung-Box Test	R^2	Q(15)	13.17962	0.5884266
Ljung-Box Test	R^2	Q(20)	18.89043	0.5289581
LM Arch Test	R	NR^2	4.383378	0.9754836

Figure 26. Standardised Residuals Tests of ARMA(1,1)-GARCH(1,1) with skew student-t distribution

Looking at the ARCH LM Tests, the p-values>0.05 and it has been failed to reject the null hypothesis and there in no ARCH effect. This confirms that the residuals behave as a white noise process. So we could conclude that the ARMA(1,1)-GARCH(1,1) model with Skew Student-t distribution is appropriate.

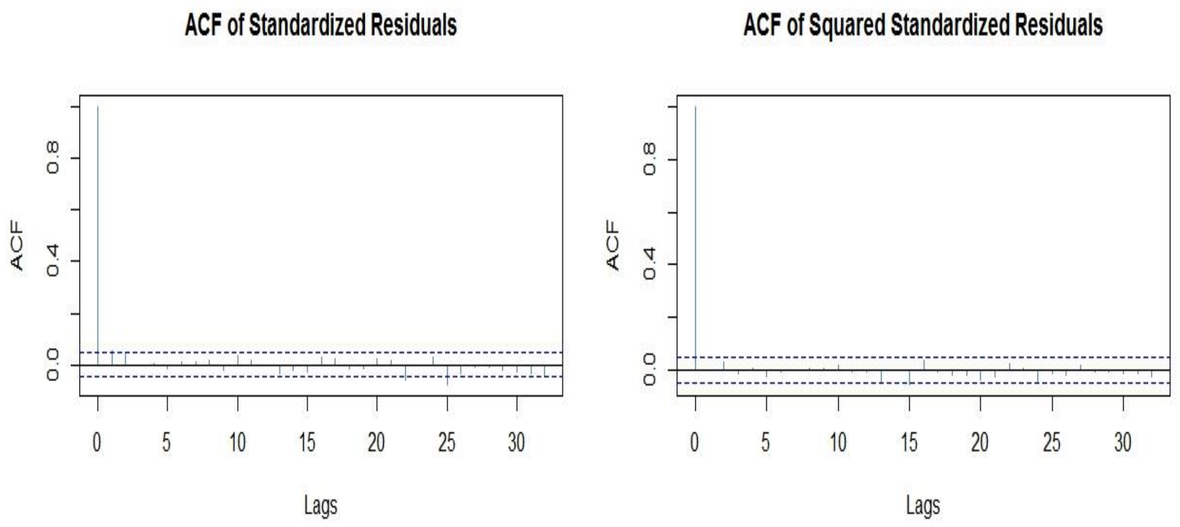


Figure 27.ACF of residuals of ARMA(1,1)- GARCH (1,1) model

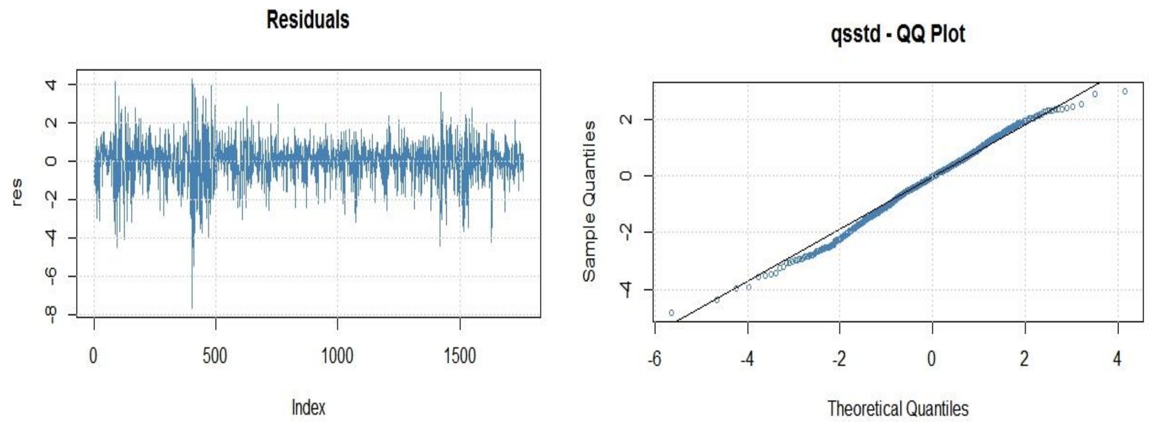


Figure 28. Residuals and Q-Q plot of residuals of ARMA(1,1)- GARCH (1,1) model

4.2 Application in the S&P 500 daily return

4.2.1 Selection of ARMA (p, q) model

As before, the first step is selection of suitable ARMA (p, q) model for S&P 500 daily return. Based on ACF and PACF in Figure 16, we could select different parameter p and q for ARMA models. By comparing the value of Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), it has been selected a model with a lower AIC or BIC value which indicates a better fit.

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	4931.62	4932.5	4931.81	4932.98	4925.77	4927.7	4927.74	4920.91
BIC	4948.04	4954.41	4948.23	4954.88	4947.66	4955.11	4955.11	4953.75

The lowest AIC and BIC come at 4925.77 and 4947.66 of ARMA(1,1) model in comparison to others with changing model orders. As a result, the study will mainly focus on ARMA(1,1) model. So the conditional mean equation is as follows

$$r_t = \alpha + \phi_1 r_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad t = 1, \dots, T.$$

The estimated parameters are $\alpha = 0.0396$, $\phi_1 = -0.8313$ and $\theta_1 = 0.7853$.
Then, we could calculate

$$\hat{\varepsilon}_t = r_t - \left(\hat{\alpha} + \hat{\phi}_1 r_{t-1} + \hat{\theta}_1 \varepsilon_{t-1}\right).$$

4.2.2 Testing estimation of residuals for ARMA model

To examine independence of residuals, it has been checked the ACF of estimated residuals of ARMA(1,1) model. According to the results in ACF of Figure 29 almost all autocorrelations stay within the 95% confidence band. Although, there are still some significant autocorrelation at lags 5 and 25.

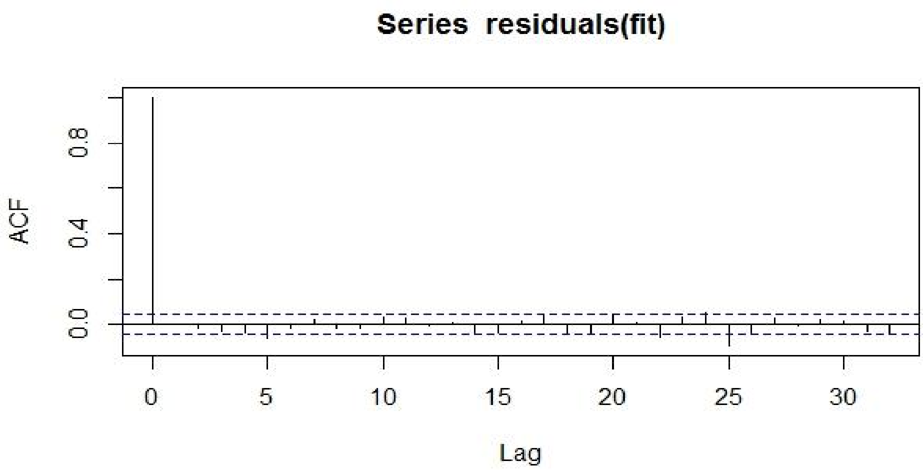


Figure 29. ACF of the estimated residuals for ARMA(1,1) model

To check the signal of autocorrelation in ARMA(1,1) model, the Ljung – Box test has been used as well. By the assumption of 5% significance level, according to the p-value of mentioned lags, the null hypothesis has to be rejected and indicates that the independence assumption of residuals can be eliminated, so the GARCH models can be applied to S&P 500's returns.

		lag=5	lag=25
Box-Ljung test	test value	11.92	62.514
	p-value	0.0359	4.668e-05

In addition, the histogram of estimated residuals of ARMA(1,1) model and QQ–plot in Figure 30 have been used to test normality assumption of residuals. The charts approve that residuals are not normal distributed.

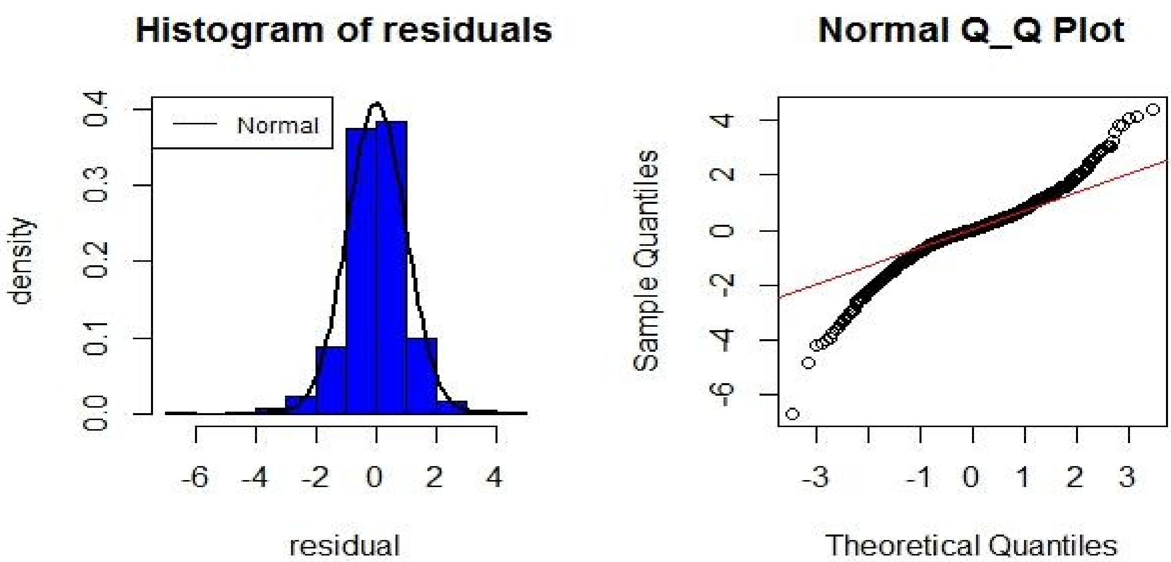


Figure 30. Histogram and QQ-plot of estimated residuals

4.2.3 GARCH Results

We estimate GARCH(1,1), GARCH(2,1), ARMA(1,1)-GARCH(1,1) and ARMA(1,1)-GARCH(2,1) with normal distribution, Student t distribution and Skew Student t distribution. We can see the estimation of the models parameters for different distribution in Figure 31, Figure 32, Figure 33 and Figure 34.

Parameters estimation for GARCH(1,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.067444	0.017925	3.763	0.000168
	omega	0.046361	0.008154	5.686	1.30e-08
	alpha1	0.150323	0.020253	7.422	1.15e-13
	beta1	0.799864	0.022576	35.430	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.08332	0.01660	5.019	5.19e-07
	omega	0.03873	0.00959	4.039	5.37e-05
	alpha1	0.16205	0.02647	6.121	9.30e-10
	beta1	0.80721	0.02622	30.790	< 2e-16
	shape	5.44624	0.75969	7.169	7.55e-13
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.064585	0.017567	3.677	0.000236
	omega	0.036284	0.009064	4.003	6.25e-05
	alpha1	0.157836	0.025112	6.285	3.27e-10
	beta1	0.810325	0.025557	31.707	< 2e-16
	skew	0.905846	0.028964	31.275	< 2e-16
	shape	5.915279	0.892840	6.625	3.47e-11

Figure 31. Parameters estimation for GARCH(1,1) of S&P500 returns

By assuming the significance level of 5%, in GARCH(1,1) model, all of the estimated parameters are significant. In GARCH(2,1) model also all of the estimated parameters are significant for all the three different distribution (normal, Student-t and Skew Student-t distribution).

Parameters estimation for GARCH(2,1)					
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
Normal	**mu**	0.06601	0.01788	3.692	0.000223
	omega	0.05885	0.01151	5.113	3.18e-07
	alpha1	0.09438	0.02768	3.410	0.000650
	alpha2	0.08523	0.03506	2.431	0.015066
	beta1	0.75671	0.03228	23.444	< 2e-16
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
Student-t	**mu**	0.08313	0.01654	5.025	5.02e-07
	omega	0.05363	0.01365	3.929	8.54e-05
	alpha1	0.07816	0.03356	2.329	0.01985
	alpha2	0.12941	0.04571	2.831	0.00464
	beta1	0.74965	0.03802	19.719	< 2e-16
	shape	5.44116	0.75702	7.188	6.59e-13
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
Skew Student-t	**mu**	0.06393	0.01752	3.650	0.000262
	omega	0.05082	0.01288	3.945	7.97e-05
	alpha1	0.07610	0.03189	2.387	0.017008
	alpha2	0.12677	0.04354	2.912	0.003596
	beta1	0.75245	0.03686	20.413	< 2e-16
	skew	0.90333	0.02897	31.182	< 2e-16
	shape	5.92282	0.89379	6.627	3.43e-11

Figure 32. Parameters estimation for GARCH(2,1) of S&P500 returns

In ARMA(1,1)-GARCH(1,1) model, the parameters estimated for Skew Student-t distribution are all significant, but not under normal and Student-t distribution. As we can see in Figure 33, the coefficient of the first term of the autoregressive process, the first term of the moving average process and μ under Student-t distribution are not significant. Also μ is not significant under Normal distribution.

Parameters estimation for ARMA(1,1)- GARCH(1,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.030996	0.016480	1.881	0.05999
	ar1	0.542647	0.222417	2.440	0.01470
	ma1	-0.591449	0.220786	-2.679	0.00739
	omega	0.046411	0.008162	5.686	1.30e-08
	alpha1	0.151030	0.020428	7.393	1.43e-13
	beta1	0.799215	0.022755	35.123	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.046513	0.032562	1.428	0.153
	ar1	0.451083	0.376584	1.198	0.231
	ma1	-0.504515	0.374120	-1.349	0.177
	omega	0.038576	0.009597	4.019	5.83e-05
	alpha1	0.163406	0.026951	6.063	1.34e-09
	beta1	0.807012	0.026378	30.595	< 2e-16
	shape	5.355260	0.742062	7.217	5.32e-13
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.021874	0.008499	2.574	0.010060
	ar1	0.659843	0.117815	5.601	2.14e-08
	ma1	-0.739179	0.112310	-6.582	4.65e-11
	omega	0.032583	0.008503	3.832	0.000127
	alpha1	0.147080	0.024090	6.106	1.02e-09
	beta1	0.822473	0.025070	32.807	< 2e-16
	skew	0.860590	0.031805	27.058	< 2e-16
	shape	6.123877	0.940312	6.513	7.39e-11

Figure 33. Parameters estimation for ARMA(1,1)- GARCH(1,1) of S&P500 returns

In ARMA(1,1)-GARCH(2,1) model, the parameters estimated for Skew Student-t distribution are all significant, but not under normal and Student-t distribution as in ARMA(1,1)-GARCH(1,1) model . As we can see in Figure 34, the coefficient of the first term of the autoregressive process, the first term of the moving average process and μ under student-t distribution are not significant. Also μ is not significant under Normal distribution.

Parameters estimation for ARMA(1,1)- GARCH(2,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.03063	0.01637	1.872	0.061275
	ar1	0.53695	0.22776	2.358	0.018398
	ma1	-0.58530	0.22566	-2.594	0.009495
	omega	0.05936	0.01166	5.091	3.55e-07
	alpha1	0.09262	0.02691	3.442	0.000577
	alpha2	0.08869	0.03485	2.545	0.010929
	beta1	0.75440	0.03296	22.888	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.04455	0.02907	1.533	0.12537
	ar1	0.47181	0.33726	1.399	0.16183
	ma1	-0.52563	0.33437	-1.572	0.11594
	omega	0.05413	0.01384	3.911	9.19e-05
	alpha1	0.07469	0.03248	2.300	0.02146
	alpha2	0.13615	0.04558	2.987	0.00282
	beta1	0.74710	0.03865	19.329	< 2e-16
	shape	5.34664	0.73536	7.271	3.57e-13
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.022676	0.009053	2.505	0.012251
	ar1	0.641594	0.128836	4.980	6.36e-07
	ma1	-0.720690	0.122946	-5.862	4.58e-09
	omega	0.047335	0.012355	3.831	0.000128
	alpha1	0.065672	0.028974	2.267	0.023417
	alpha2	0.126639	0.041042	3.086	0.002032
	beta1	0.763654	0.036495	20.925	< 2e-16
	skew	0.859435	0.031713	27.100	< 2e-16
	shape	6.082555	0.926129	6.568	5.11e-11

Figure 34. Parameters estimation for ARMA(1,1)- GARCH(2,1) of S&P500 returns

Figure 35 shows the Akaike Information Criterion (AIC) and Bayes information Criterion (BIC) under the three distributions for the models. As we know, a lower AIC or BIC value indicates a better fit. In our case ARMA(1,1)-GARCH(2,1) model with Skew Student-t distribution has the lowest AIC which is 2.476 and BIC which is 2.504. So the ARMA(1,1)-GARCH(2,1) model with Skew Student-t distribution is the best model, because all of the estimated parameters are significant and the fitted model is as follow

$$r_t = \mu + \phi_1(r_{t-1} - \mu) + \theta_1\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t = \sigma_t\varepsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \beta_1\sigma_{t-1}^2,$$

where estimated parameters according to Figure 34 are $\mu = 0.022676$, $\phi_1 = 0.641594, \theta_1 = -0.720690$, $\omega = 0.047335$, $\alpha_1 = 0.065672$, $\alpha_2 = 0.126639$ and $\beta_1 = 0.763654$, also we could the skew value $= 0.859435$ which indicates that the student-t distribution is skewed positively.

Information Criterion Statistics		
	AIC	**BIC**
GARCH(1,1)-Normal	2.539	2.551
GARCH(1,1)-Student-t	2.492	2.508
GARCH(1,1)-Skew Student-t	2.488	2.507
GARCH(2,1)-Normal	2.536	2.552
GARCH(2,1)-Student-t	2.489	2.507
GARCH(2,1)-Skew Student-t	2.484	2.506
ARMA(1,1)-GARCH(1,1)-Normal	2.537	2.555
ARMA(1,1)-GARCH(1,1)-Student-t	2.489	2.511
ARMA(1,1)-GARCH(1,1)-Skew Student-t	2.480	2.505
ARMA(1,1)-GARCH(2,1)-Normal	2.534	2.556
ARMA(1,1)-GARCH(2,1)-Student-t	2.485	2.510
ARMA(1,1)-GARCH(2,1)-Skew Student-t	2.476	2.504

Figure 35. Information Criterion Statistics for different models of S&P500 returns

4.2.4 Residuals Diagnostics of GARCH model

It has been performed the Standardised Residuals Tests of ARMA(1,1)-GARCH(2,1) with skew Student-t distribution in Figure 36. By looking at Ljung-Box test on residuals and squared residuals which have p-values>0.05, it has been failed to reject the null hypothesis and there is no evidence of autocorrelation in the residuals and squared residuals. As a result, we can conclude that the residuals behave as white noise.

Standardised Residuals Tests:				
			Statistic	**p-value**
Jarque-Bera Test	R	Chi^2	307.5923	0
Shapiro-Wilk Test	R	W	0.9761527	0
Ljung-Box Test	R	Q(10)	13.7967	0.1824684
Ljung-Box Test	R	Q(15)	19.5047	0.191767
Ljung-Box Test	R	Q(20)	24.45178	0.2232144
Ljung-Box Test	R^2	Q(10)	10.06548	0.4347671
Ljung-Box Test	R^2	Q(15)	17.97802	0.263823
Ljung-Box Test	R^2	Q(20)	19.88817	0.4649444
LM Arch Test	R	NR^2	10.21848	0.5968004

Figure 36. Standardised Residuals Tests of ARMA(1,1)-GARCH(2,1) with Skew Student-t distribution

Looking at the ARCH LM Tests, the p-values>0.05 and it has been failed to reject the null hypothesis and there is no ARCH effect. This confirms that the residuals behave as a white noise process. So we could conclude that the ARMA(1,1)-GARCH(2,1) model with Skew Student-t distribution is appropriate.

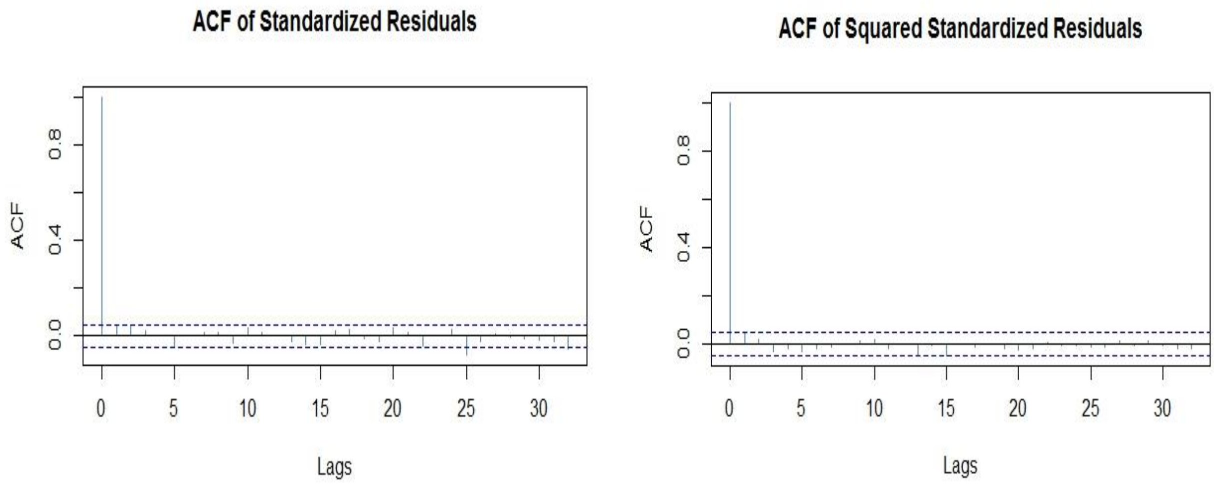


Figure 37.ACF of residuals of ARMA(1,1)- GARCH (2,1) model

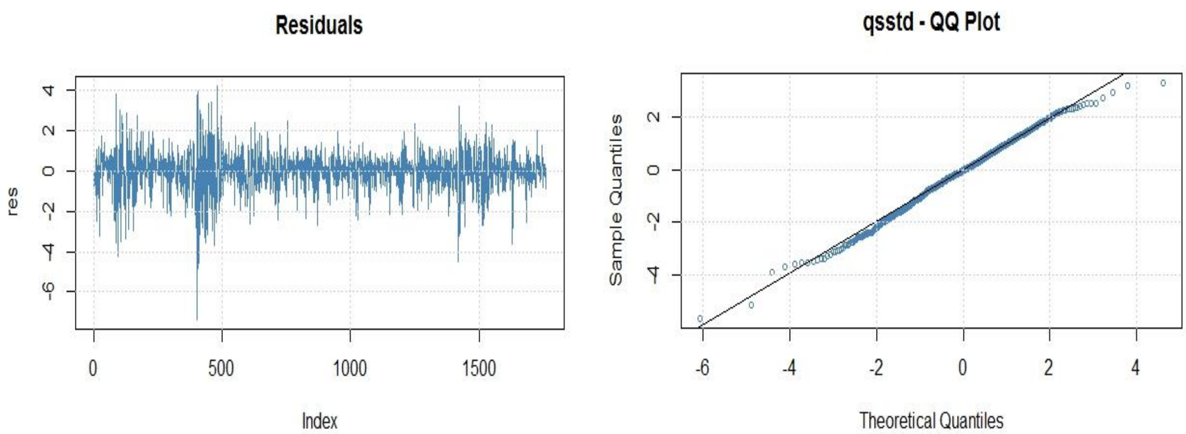


Figure 38. Residuals and Q-Q plot of residuals of ARMA(1,1)- GARCH (2,1) model

4.3 Application in the Nikkei daily return

4.3.1 Selection of ARMA (p, q) model

Based on ACF and PACF in Figure 17, we select different parameter p and q for ARMA models for Nikkei daily returns.

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	6085.91	6087.38	6086.05	6087.35	6087.55	6089.3	6089.37	6073.79
BIC	6102.25	6109.16	6102.38	6109.13	6109.33	6116.53	6116.6	6106.46

The lowest AIC and BIC come at 6073.79 and 6106.46 of ARMA(2,2) model in comparison to others with changing model orders. As a result, the study will mainly focus on ARMA(2,2) model. So the conditional mean equation is as follows

$$r_t = \alpha + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \qquad t = 1, \ldots, T.$$

The estimated parameters are

$\alpha = 0.0335, \phi_1 = -0.4707, \phi_2 = -0.9037, \theta_1 = 0.4502$ and $\theta_2 = 0.9418.$

Then, we could calculate

$$\hat{\varepsilon}_t = r_t - \left(\hat{\alpha} + \hat{\phi}_1 r_{t-1} + \hat{\phi}_2 r_{t-2} + \hat{\theta}_1 \varepsilon_{t-1} + \hat{\theta}_2 \varepsilon_{t-2}\right).$$

4.3.2 Testing estimation of residuals for ARMA model

It has been checked the ACF of estimated residuals of ARMA(2,2) model to examine independence of residuals. According to the results in ACF of Figure 39 all autocorrelations stay within the 95% confidence band.

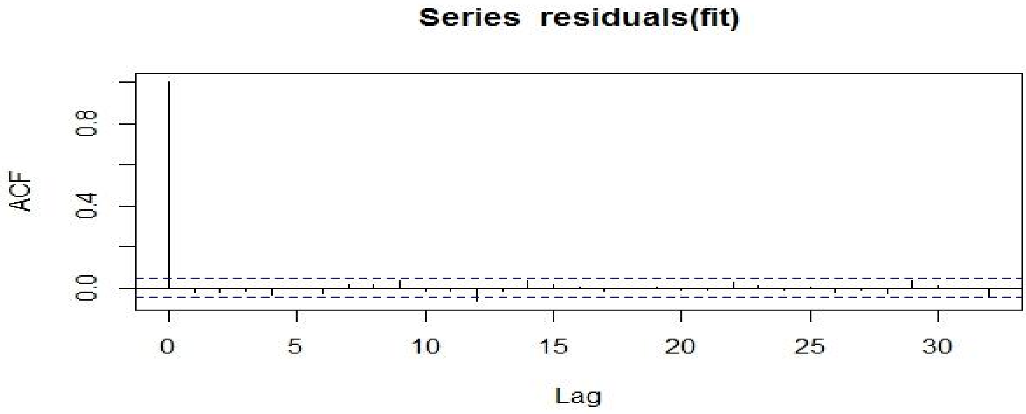


Figure 39. ACF of the estimated residuals for ARMA(2,2) model

In addition, the histogram of estimated residuals of ARMA(2,2) model and QQ–plot in Figure 40 have been used to test normality assumption of residuals. The charts approve that residuals are not normal distributed. So the GARCH models can be applied to Nikkie's returns.

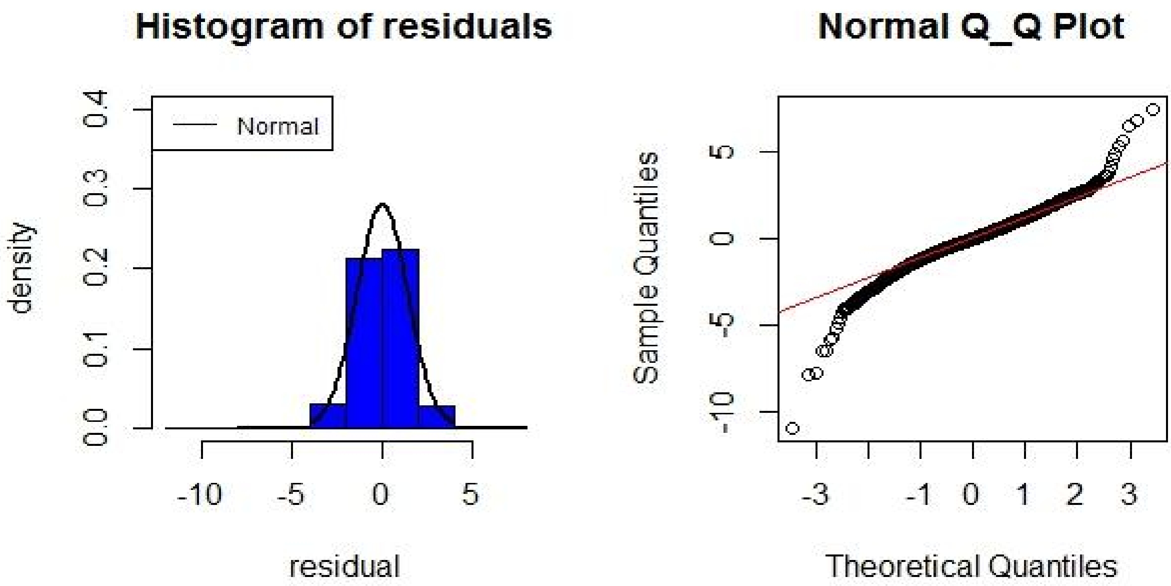


Figure 40. Histogram and QQ-plot of estimated residuals

4.3.3 GARCH Results

We estimate GARCH(1,1), GARCH(2,1), ARMA(2,2)-GARCH(1,1) and ARMA(2,2)-GARCH(2,1) with normal distribution, Student t distribution and Skew Student t distribution. We can see the estimation of the models parameters for different distribution in Figure 41, Figure 42, Figure 43 and Figure 44.

Parameters estimation for GARCH(1,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.07185	0.02878	2.497	0.0125
	omega	0.09836	0.02285	4.304	1.67e-05
	alpha1	0.14428	0.01913	7.543	4.60e-14
	beta1	0.81419	0.02170	37.519	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.09672	0.02806	3.447	0.000568
	omega	0.10090	0.02931	3.442	0.000577
	alpha1	0.12497	0.02259	5.532	3.17e-08
	beta1	0.82796	0.02807	29.501	< 2e-16
	shape	7.21153	1.19929	6.013	1.82e-09
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	mu	0.07106	0.02903	2.448	0.014363
	omega	0.09431	0.02714	3.474	0.000512
	alpha1	0.12308	0.02152	5.720	1.07e-08
	beta1	0.83123	0.02667	31.169	< 2e-16
	skew	0.89661	0.03068	29.227	< 2e-16
	shape	8.02206	1.47767	5.429	5.67e-08

Figure 41. Parameters estimation for GARCH(1,1) of Nikkei returns

By assuming the significance level of 5%, in GARCH(1,1) model, all of the estimated parameters are significant. In GARCH(2,1) model almost all of the estimated parameters are significant except α_2 for all the three different distribution (Normal, Student-t and Skew Student-t distribution).

Parameters estimation for GARCH(2,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.07261	0.02881	2.521	0.0117
	omega	0.10027	0.02397	4.183	2.88e-05
	alpha1	0.13621	0.02722	5.004	5.61e-07
	alpha2	0.01284	0.03341	0.384	0.7008
	beta1	0.80902	0.02634	30.712	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.09792	0.02798	3.499	0.000467

	omega	0.11074	0.03400	3.257	0.001126
	alpha1	0.10088	0.03179	3.174	0.001505
	alpha2	0.03859	0.04007	0.963	0.335530
	beta1	0.80993	0.03615	22.405	< 2e-16
	shape	7.15724	1.18154	6.058	1.38e-09
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.07246	0.02901	2.498	0.012486
	omega	0.10091	0.03061	3.297	0.000979
	alpha1	0.10490	0.03116	3.367	0.000760
	alpha2	0.02866	0.03871	0.740	0.459072
	beta1	0.81828	0.03368	24.297	< 2e-16
	Skew	0.89828	0.03076	29.206	< 2e-16
	shape	7.94212	1.45075	5.474	4.39e-08

Figure 42. Parameters estimation for GARCH(2,1) of Nikkei returns

As we can see in Figure 43, in ARMA(2,2)-GARCH(1,1) model, the parameters estimated for normal, Student-t and Skew Student-t distribution are all significant.

Parameters estimation for ARMA(2,2)- GARCH(1,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.18301	0.07078	2.586	0.00972
	ar1	-0.52989	0.01461	-36.276	< 2e-16
	ar2	-0.96064	0.01464	-65.609	< 2e-16
	ma1	0.51490	0.01020	50.495	< 2e-16
	ma2	0.97512	0.01290	75.617	< 2e-16
	Omega	0.10398	0.02424	4.290	1.78e-05
	alpha1	0.14669	0.01995	7.354	1.93e-13
	beta1	0.80827	0.02327	34.740	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.24334	0.06957	3.498	0.000470
	ar1	-0.53186	0.01617	-32.885	< 2e-16
	ar2	-0.95956	0.02417	-39.706	< 2e-16
	ma1	0.51368	0.01632	31.475	< 2e-16
	ma2	0.96936	0.02197	44.115	< 2e-16
	Omega	0.10646	0.03089	3.446	0.000569
	alpha1	0.12759	0.02324	5.490	4.03e-08
	beta1	0.82214	0.02947	27.896	< 2e-16
	Shape	7.21814	1.20782	5.976	2.28e-09
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.18668	0.07149	2.611	0.009018
	ar1	-0.53044	0.01541	-34.428	< 2e-16
	ar2	-0.96256	0.01883	-51.108	< 2e-16
	ma1	0.51308	0.01342	38.225	< 2e-16
	ma2	0.97157	0.01792	54.208	< 2e-16
	Omega	0.09909	0.02843	3.486	0.000491

	alpha1	0.12613	0.02213	5.698	1.21e-08
	beta1	0.82538	0.02782	29.668	< 2e-16
	Skew	0.90056	0.03125	28.820	< 2e-16
	Shape	8.04971	1.49621	5.380	7.45e-08

Figure 43. Parameters estimation for ARMA(2,2)- GARCH(1,1) of Nikkei returns

In ARMA(2,2)-GARCH(2,1) model, the parameters estimated for normal, Student-t and Skew Student-t distributions are almost all significant except the parameter α_2 that is not significant under all three distributions.

Parameters estimation for ARMA(2,2)- GARCH(2,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	1.828e-01	7.088e-02	2.579	0.00991
	ar1	-5.298e-01	1.461e-02	-36.263	< 2e-16
	ar2	-9.606e-01	1.466e-02	-65.540	< 2e-16
	ma1	5.149e-01	1.018e-02	50.584	< 2e-16
	ma2	9.752e-01	1.298e-02	75.134	< 2e-16
	Omega	1.037e-01	2.491e-02	4.161	3.17e-05
	alpha1	1.464e-01	2.930e-02	4.996	5.87e-07
	alpha2	1.000e-08	3.486e-02	0.000	1.00000
	beta1	8.087e-01	2.749e-02	29.418	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.24441	0.06949	3.517	0.000436
	ar1	-0.53139	0.01756	-30.260	< 2e-16
	ar2	-0.95774	0.02955	-32.415	< 2e-16
	ma1	0.51293	0.01944	26.390	< 2e-16
	ma2	0.96737	0.02695	35.894	< 2e-16
	omega	0.11584	0.03592	3.224	0.001262
	alpha1	0.10582	0.03330	3.177	0.001486
	alpha2	0.03422	0.04135	0.827	0.407959
	beta1	0.80611	0.03785	21.296	< 2e-16
	Shape	7.14198	1.18660	6.019	1.76e-09
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.18828	0.07142	2.636	0.00838
	ar1	-0.53023	0.01586	-33.431	< 2e-16
	ar2	-0.96156	0.02084	-46.138	< 2e-16
	ma1	0.51272	0.01467	34.959	< 2e-16
	ma2	0.97035	0.01979	49.028	< 2e-16
	omega	0.10500	0.03204	3.277	0.00105
	alpha1	0.11051	0.03264	3.386	0.00071
	alpha2	0.02415	0.03995	0.605	0.54543
	beta1	0.81452	0.03494	23.310	< 2e-16
	Skew	0.90195	0.03138	28.746	< 2e-16
	Shape	7.95517	1.46915	5.415	6.14e-08

Figure 44. Parameters estimation for ARMA(2,2)- GARCH(2,1) of Nikkei returns

Figure 45 shows the Akaike Information Criterion (AIC) and Bayes information Criterion (BIC) under the three distributions for the models.

In our case ARMA(2,2)-GARCH(1,1) model with Skew Student-t distribution has the lowest AIC which is 3.358 .Obviously, the ARMA(2,2)-GARCH(1,1) model with Skew Student-t distribution is the best model, because all of the estimated parameters are significant and has the lowest AIC. The fitted model is as follow

$$r_t = \mu + \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \epsilon_t, \quad \epsilon_t = \sigma_t\varepsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

where estimated parameters according to Figure 43 are

$\mu = 0.18668$, $\phi_1 = -0.53044, \phi_2 = -0.96256$, $\theta_1 = 0.51308$, $\theta_2 = 0.97157$, $\omega = 0.09909$, $\alpha_1 = 0.12613$, and $\beta_1 = 0.82538$. Also we could see the skew value = 0.90056 and it indicates that the student-t distribution is skewed positively.

Information Criterion Statistics		
	AIC	**BIC**
GARCH(1,1)-Normal	3.404	3.416
GARCH(1,1)-Student-t	3.366	3.382
GARCH(1,1)-Skew Student-t	3.360	3.380
GARCH(2,1)-Normal	3.405	3.421
GARCH(2,1)-Student-t	3.366	3.385
GARCH(2,1)-Skew Student-t	3.362	3.384
ARMA(2,2)-GARCH(1,1)-Normal	3.399	3.425
ARMA(2,2)-GARCH(1,1)-Student-t	3.363	3.391
ARMA(2,2)-GARCH(1,1)-Skew Student-t	3.358	3.390
ARMA(2,2)-GARCH(2,1)-Normal	3.401	3.429
ARMA(2,2)-GARCH(2,1)-Student-t	3.363	3.395
ARMA(2,2)-GARCH(2,1)-Skew Student-t	3.359	3.394

Figure 45. Information Criterion Statistics for different models of Nikkei returns

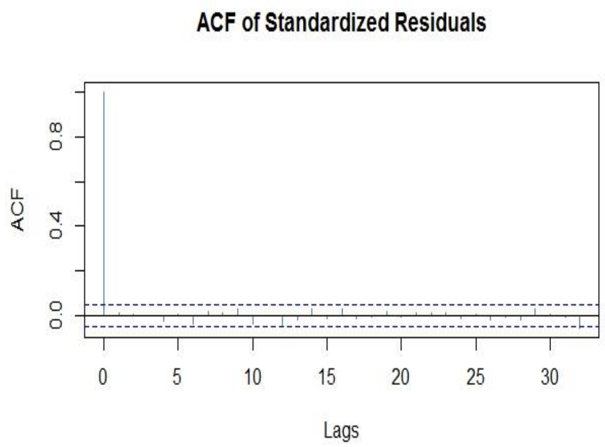
4.3.4 Residuals Diagnostics of GARCH model

It has been performed the Standardised Residuals Tests of ARMA(2,2)-GARCH(1,1) with skew student-t distribution in Figure 46. By looking at Ljung-Box test on residuals and squared residuals which have p-values>0.05, it has been failed to reject the null hypothesis and there is no evidence of autocorrelation in the residuals and squared residuals. As a result, we can conclude that the residuals behave as white noise.

Standardised Residuals Tests:				
			Statistic	**p-value**
Jarque-Bera Test	R	Chi^2	286.1826	0
Shapiro-Wilk Test	R	W	0.9818964	6.825436e-14
Ljung-Box Test	R	Q(10)	8.820061	0.5492615
Ljung-Box Test	R	Q(15)	15.19217	0.4376655
Ljung-Box Test	R	Q(20)	17.48503	0.6212913
Ljung-Box Test	R^2	Q(10)	14.20882	0.163678
Ljung-Box Test	R^2	Q(15)	22.90597	0.08615685
Ljung-Box Test	R^2	Q(20)	27.68497	0.1170691
LM Arch Test	R	NR^2	21.98413	0.05769804

Figure 46. Standardised Residuals Tests of ARMA(2,2)-GARCH(1,1) with Skew Student-t distribution

Looking at the ARCH LM Tests, the p-values>0.05 and it has been failed to reject the null hypothesis and there in no ARCH effect. So we could conclude that the ARMA(2,2)-GARCH(1,1) model with Skew Student-t distribution is appropriate.



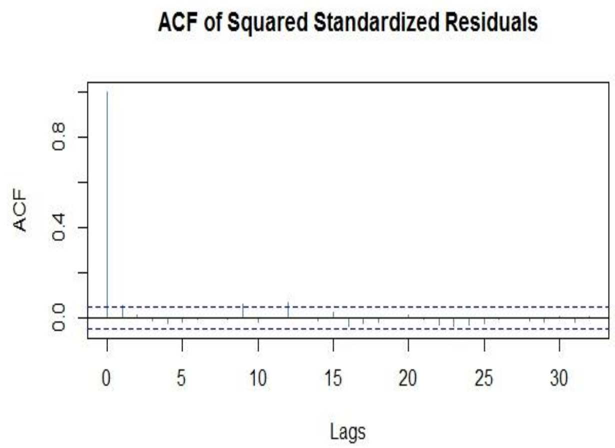


Figure 47.ACF of residuals of ARMA(1,1)- GARCH (1,1) model

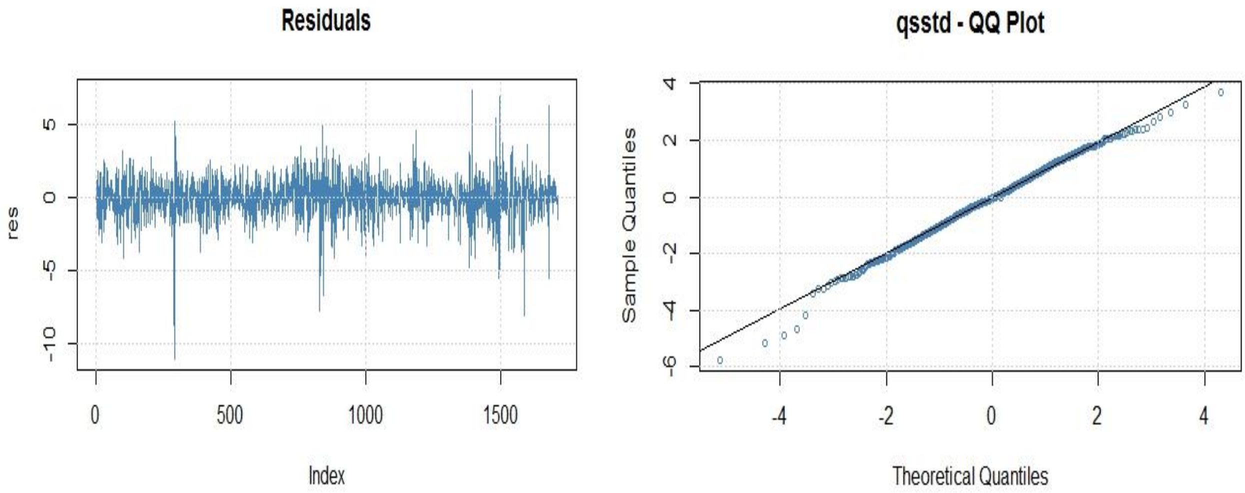


Figure 48. Residuals and Q-Q plot of residuals of ARMA(1,1)- GARCH (1,1) model

4.4 Application in the Dow Jones daily return

4.4.1 Selection of ARMA (p, q) model

As before, the first step is selection of suitable ARMA (p, q) model for Dow Jones daily return. Based on ACF and PACF in Figure 16, we could select different parameter p and q for ARMA models.

	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	4670.32	4670.96	4670.54	4671.28	4667.71	4669.71	4669.71	4662
BIC	4686.74	4692.85	4686.96	4693.18	4689.61	4697.08	4697.08	4694.85

Many models have been tried, such as AR (1), AR(2), MA (1), MA(2) , ARMA (1, 1) and so on. The lowest AIC and BIC come at 4662 and 4694.85 of ARMA(2,2) model in comparison to others with changing model orders. As a result, the study will mainly focus on ARMA(2,2) model. So the conditional mean equation is as follows

$$r_t = \alpha + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \qquad t = 1, \dots, T.$$

The estimated parameters are

$\alpha = 0.0354, \phi_1 = 0.1933, \phi_2 = 0.7660\ , \theta_1 = -0.2509 \text{and}\ \ \theta_2 = -0.7372.$

Then, we could calculate

$$\hat{\varepsilon}_t = r_t - \left(\hat{\alpha} + \hat{\phi}_1 r_{t-1} + \hat{\phi}_2 r_{t-2} + \hat{\theta}_1 \varepsilon_{t-1} + \hat{\theta}_2 \varepsilon_{t-2}\right).$$

4.4.2 Testing estimation of residuals for ARMA model

To examine independence of residuals, it has been checked the ACF of estimated residuals of ARMA(2,2) model. According to the results in ACF of Figure 49 almost all autocorrelations stay within the 95% confidence band. Although, there is a significant autocorrelation at lag 25.

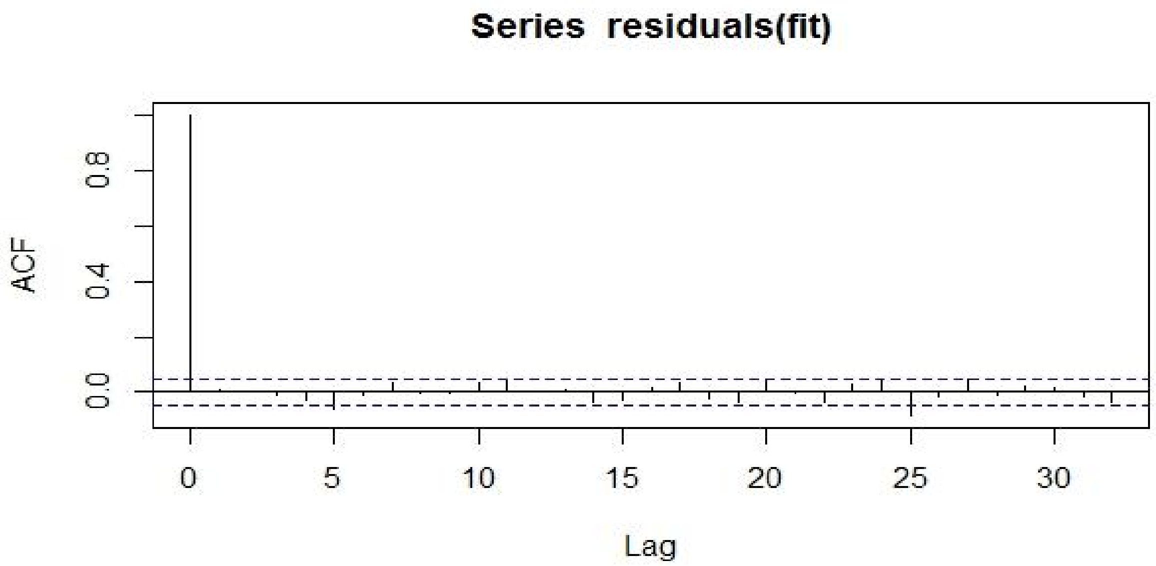


Figure 49. ACF of the estimated residuals for ARMA(2,2) model

To check the signal of autocorrelation in ARMA(2,2) model, the Ljung – Box test has been used as well. By the assumption of 5% significance level, according to the p-value of mentioned lag, the null hypothesis has to be rejected and indicates that the independence assumption of residuals can be eliminated.

		lag=25
Box-Ljung test	test value	48.067
	p-value	0.003661

In addition, the histogram of estimated residuals of ARMA(2,2) model and QQ–plot in Figure 30 have been used to test normality assumption of residuals. The charts approve that residuals are not normal distributed. So the GARCH models can be applied to Dow jones's returns

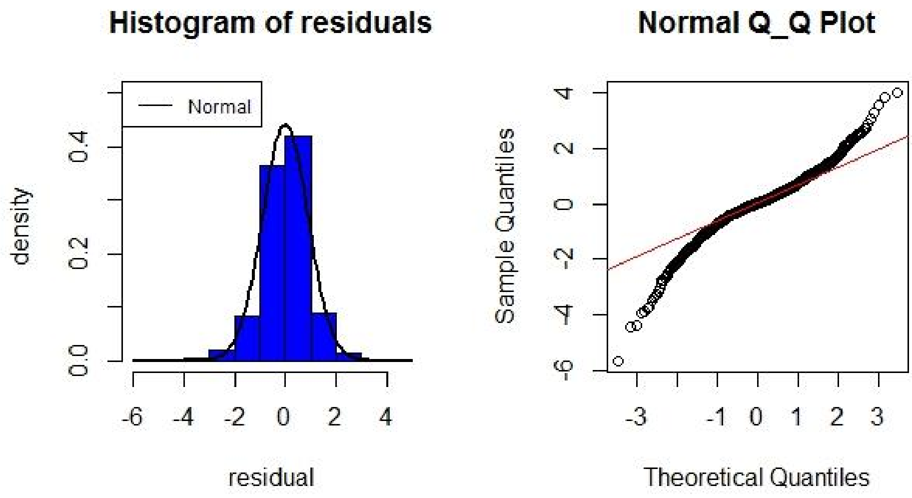


Figure 50. Histogram and QQ-plot of estimated residuals

4.4.3 GARCH Results

We estimate GARCH(1,1), GARCH(2,1), ARMA(1,1)-GARCH(1,1) and ARMA(1,1)-GARCH(2,1) with normal distribution, Student t distribution and Skew Student t distribution. We can see the estimation of the models parameters for different distribution in Figure 51, Figure 52, Figure 53 and Figure 54.

Parameters estimation for GARCH(1,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.06681	0.01672	3.997	6.41e-05
	Omega	0.04183	0.00739	5.660	1.52e-08
	alpha1	0.16445	0.02153	7.636	2.24e-14
	beta1	0.78608	0.02326	33.798	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.076583	0.015580	4.916	8.85e-07
	Omega	0.034382	0.008353	4.116	3.85e-05
	alpha1	0.179157	0.028053	6.386	1.70e-10
	beta1	0.792700	0.026411	30.014	< 2e-16
	Shape	5.691578	0.828895	6.866	6.58e-12
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.061295	0.016510	3.713	0.000205

	Omega	0.032897	0.008009	4.107	4.00e-05
	alpha1	0.174832	0.026821	6.518	7.10e-11
	beta1	0.795065	0.025951	30.638	< 2e-16
	Skew	0.916932	0.029408	31.180	< 2e-16
	Shape	6.111587	0.947633	6.449	1.12e-10

Figure 51. Parameters estimation for GARCH(1,1) of Dow Jones returns

By assuming the significance level of 5%, in GARCH(1,1) model, all of the estimated parameters are significant. In GARCH(2,1) model also all of the estimated parameters are significant for all the three different distribution (normal, Student-t and Skew Student-t distribution).

Parameters estimation for GARCH(2,1)					
Normal		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.065832	0.016715	3.938	8.20e-05
	omega	0.050700	0.009943	5.099	3.42e-07
	alpha1	0.113954	0.030539	3.731	0.00019
	alpha2	0.075395	0.036525	2.064	0.03900
	beta1	0.750083	0.032022	23.424	< 2e-16
Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.07643	0.01557	4.908	9.21e-07
	omega	0.04451	0.01133	3.929	8.54e-05
	alpha1	0.10181	0.03807	2.674	0.0075
	alpha2	0.11491	0.04785	2.401	0.0163
	beta1	0.74596	0.03678	20.281	< 2e-16
	Shape	5.69978	0.82811	6.883	5.86e-12
Skew Student-t		**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.06054	0.01651	3.667	0.000245
	omega	0.04314	0.01093	3.948	7.87e-05
	alpha1	0.09786	0.03610	2.711	0.006714
	alpha2	0.11519	0.04578	2.516	0.011871
	beta1	0.74662	0.03625	20.596	< 2e-16
	Skew	0.91354	0.02954	30.924	< 2e-16
	Shape	6.15487	0.95750	6.428	1.29e-10

Figure 52. Parameters estimation for GARCH(2,1) of Dow Jones returns

In ARMA(2,2)-GARCH(1,1) model, the parameters estimated are not all significant under normal, Student-t and Skew Student-t distribution. As we can see in Figure 53, the coefficients of the first and the second term of the autoregressive process and the first term of the moving average process under all three distributions are not significant. On the other hand, the second term of the moving average process is not significant under Student-t and Skew Student-t distributions.

D					
Normal		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.093096	0.038951	2.390	0.0168
	ar1	0.102368	0.402026	0.255	0.7990
	ar2	-0.491642	0.255056	-1.928	0.0539
	ma1	-0.149328	0.393795	-0.379	0.7045
	ma2	0.515271	0.260363	1.979	0.0478
	Omega	0.041957	0.007386	5.680	1.34e-08
	alpha1	0.165251	0.021710	7.612	2.71e-14
	beta1	0.785263	0.023347	33.635	< 2e-16
Student-t		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.100611	0.034867	2.886	0.00391
	ar1	0.117468	0.331337	0.355	0.72294
	ar2	-0.423896	0.312245	-1.358	0.17460
	ma1	-0.173294	0.326809	-0.530	0.59593
	ma2	0.450020	0.321826	1.398	0.16201
	Omega	0.034611	0.008421	4.110	3.95e-05
	alpha1	0.181826	0.028629	6.351	2.14e-10
	beta1	0.791249	0.026563	29.788	< 2e-16
	Shape	5.535894	0.793990	6.972	3.12e-12
Skew Student-t		Estimate	Std.Error	t value	Pr(>|t|)
	Mu	0.085090	0.038561	2.207	0.0273
	ar1	-0.117849	0.670425	-0.176	0.8605
	ar2	-0.296551	0.395200	-0.750	0.4530
	ma1	0.051208	0.668907	0.077	0.9390
	ma2	0.290676	0.437439	0.664	0.5064
	Omega	0.032890	0.008011	4.106	4.03e-05
	alpha1	0.176919	0.027154	6.515	7.25e-11
	beta1	0.794061	0.025987	30.556	< 2e-16
	Skew	0.904795	0.030376	29.787	< 2e-16
	Shape	5.954617	0.907684	6.560	5.37e-11

Figure 53. Parameters estimation for ARMA(2,2)- GARCH(1,1) of Dow Jones returns

In ARMA(2,2)-GARCH(2,1) model, the parameters estimated for three distributions are not all significant. As we can see in Figure 54, the coefficient of the first term of the autoregressive process and the first term of the moving average process, under normal distribution and Student-t distribution are not significant. On the other hand, the coefficient of the first and second term of the autoregressive process and the first and second term of the moving average process are not significant also under Skew Student-t distribution.

Parameters estimation for ARMA(2,2)- GARCH(2,1)					
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.08925	0.03028	2.948	0.003203
	ar1	0.19261	0.31083	0.620	0.535467
	ar2	-0.55902	0.19507	-2.866	0.004159
	ma1	-0.24016	0.29738	-0.808	0.419337
Normal	**ma2**	0.59281	0.20724	2.860	0.004230
	Omega	0.05188	0.01014	5.119	3.07e-07
	alpha1	0.10433	0.02995	3.484	0.000495
	alpha2	0.08737	0.03628	2.408	0.016043
	beta1	0.74582	0.03248	22.960	< 2e-16
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.10486	0.02657	3.946	7.93e-05
	ar1	0.21380	0.22638	0.944	0.34494
	ar2	-0.58090	0.24026	-2.418	0.01562
	ma1	-0.26952	0.21527	-1.252	0.21057
Student-t	**ma2**	0.61650	0.25289	2.438	0.01478
	Omega	0.04601	0.01165	3.949	7.86e-05
	alpha1	0.08989	0.03667	2.451	0.01424
	alpha2	0.13337	0.04768	2.797	0.00515
	beta1	0.73897	0.03737	19.776	< 2e-16
	Shape	5.57064	0.79469	7.010	2.39e-12
		Estimate	**Std.Error**	**t value**	**Pr(>|t|)**
	Mu	0.08345	0.03238	2.577	0.00997
	ar1	-0.05424	0.62263	-0.087	0.93059
	ar2	-0.35951	0.46842	-0.767	0.44279
	ma1	-0.01599	0.61884	-0.026	0.97938
	ma2	0.35805	0.51905	0.690	0.49030
Skew Student-t	**Omega**	0.04497	0.01129	3.983	6.81e-05
	alpha1	0.08675	0.03380	2.566	0.01028
	alpha2	0.13468	0.04525	2.977	0.00291
	beta1	0.73738	0.03718	19.833	< 2e-16
	Skew	0.89862	0.03093	29.053	< 2e-16
	Shape	6.02463	0.91979	6.550	5.75e-11

Figure 54. Parameters estimation for ARMA(2,2)- GARCH(2,1) of Dow Jones returns

Figure 55 shows the Akaike Information Criterion (AIC) and Bayes information Criterion (BIC) under the three distributions for the models.

In our case ARMA(2,2)-GARCH(2,1) model with Skew Student-t distribution has the lowest AIC which is 2.354 and then the lowest comes at 2.357 of GARCH(2,1) model with Skew Student-t distribution. Obviously, the GARCH(2,1) model with Skew Student-t distribution is the best model, because all of the estimated parameters are significant and the fitted model is as follow

$$r_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t \varepsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2,$$

where estimated parameters according to Figure 52 are $\mu = 0.06054, \omega = 0.04314$, $\alpha_1 = 0.09786$, $\alpha_2 = 0.11519$ and $\beta_1 = 0.74662$ also we could the skew value $= 0.91354$which indicates that the student-t distribution is skewed positively.

Information Criterion Statistics		
	AIC	**BIC**
GARCH(1,1)-Normal	2.403	2.416
GARCH(1,1)-Student-t	2.363	2.378
GARCH(1,1)-Skew Student-t	2.359	2.378
GARCH(2,1)-Normal	2.402	2.417
GARCH(2,1)-Student-t	2.360	2.378
GARCH(2,1)-Skew Student-t	2.357	2.378
ARMA(2,2)-GARCH(1,1)-Normal	2.403	2.428
ARMA(2,2)-GARCH(1,1)-Student-t	2.361	2.389
ARMA(2,2)-GARCH(1,1)-Skew Student-t	2.357	2.388
ARMA(2,2)-GARCH(2,1)-Normal	2.402	2.430
ARMA(2,2)-GARCH(2,1)-Student-t	2.358	2.389
ARMA(2,2)-GARCH(2,1)-Skew Student-t	2.354	2.387

Figure 55. Information Criterion Statistics for different models of Dow Jones returns

4.4.4 Residuals Diagnostics of GARCH model

It has been performed the Standardised Residuals Tests of GARCH(2,1) with Skew Student-t distribution in Figure 56. By looking at Ljung-Box test on residuals and squared residuals which have p-values>0.05, it has been failed to reject the null hypothesis and there is no evidence of autocorrelation in the residuals and squared residuals. As a result, we can conclude that the residuals behave as white noise. Looking at the ARCH LM Tests, the p-values>0.05 and it has been failed to reject the null hypothesis and there in no ARCH effect. So we could conclude that the GARCH(2,1) model with Skew Student-t distribution is appropriate.

Standardised Residuals Tests:				
			Statistic	**p-value**
Jarque-Bera Test	R	Chi^2	192.8589	0
Shapiro-Wilk Test	R	W	0.9826086	8.47313e-14
Ljung-Box Test	R	Q(10)	12.67659	0.2423202
Ljung-Box Test	R	Q(15)	17.97563	0.2639489
Ljung-Box Test	R	Q(20)	22.86617	0.2954138
Ljung-Box Test	R^2	Q(10)	5.922708	0.8217147
Ljung-Box Test	R^2	Q(15)	12.31793	0.6548226
Ljung-Box Test	R^2	Q(20)	16.51499	0.6842053
LM Arch Test	R	NR^2	6.954667	0.8605947

Figure 56. Standardised Residuals Tests of GARCH(2,1) with skew student-t distribution

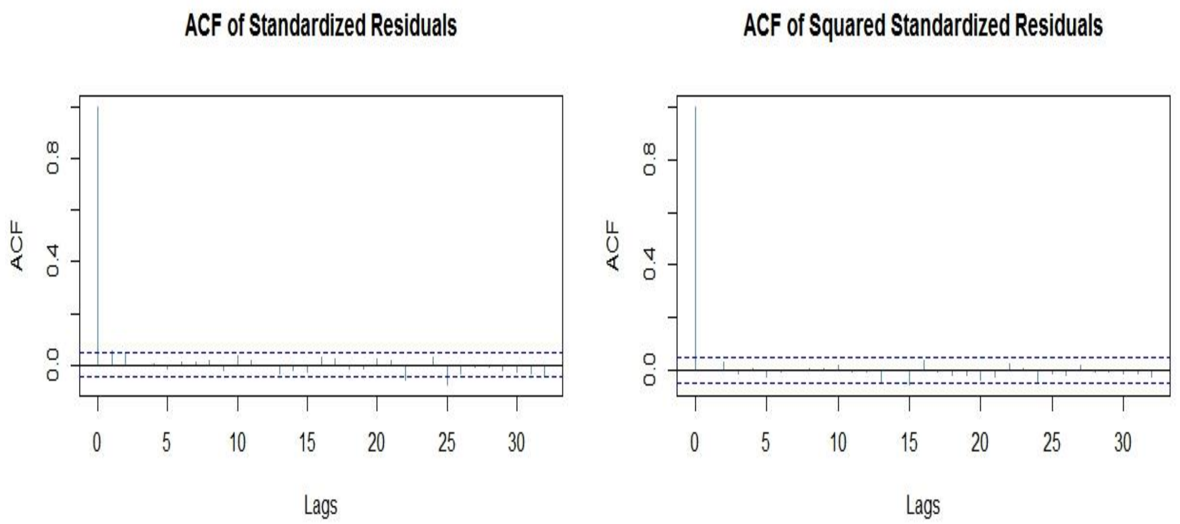


Figure 57.ACF of residuals of GARCH (2,1) model

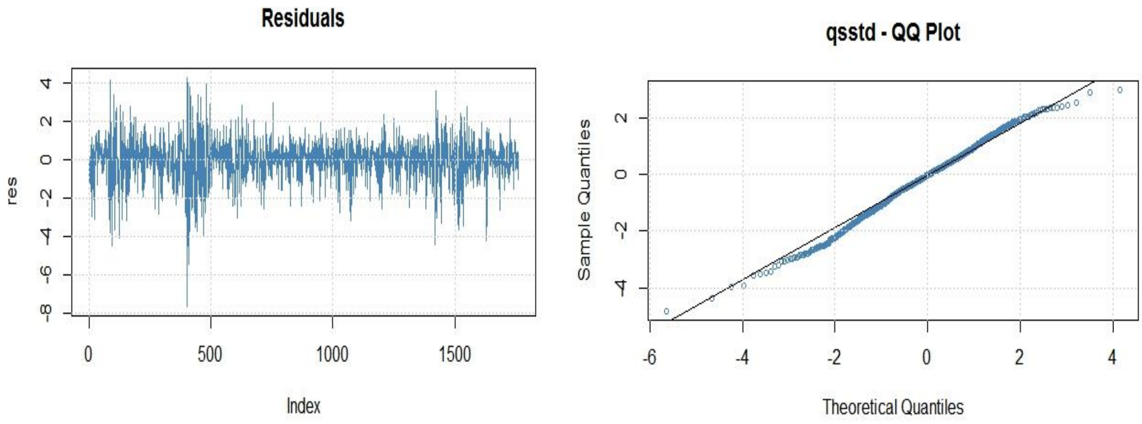


Figure 58. Residuals and Q-Q plot of residuals of GARCH (2,1) model

4.5 Out-of-sample forecast performance of Nasdaq's GARCH Model

For Nasdaq daily returns, it has been chosen ARMA(1,1)-GARCH(1,1) model with skew Student-t distribution as best model. The NASDAQ daily return includes 1762 observations of 7 years. It has been taken the last 10 percent of data as out-of-sample which includes 175 observations ($w_F = 175$). By rolling forecast approach, it has been fixed the length of the in-sample period which is 1587 observations ($w_E = 1587$). In addition for not producing bad forecasts, we will do a re-estimation of these parameters every 50 days (Refit Horizon= 50, No.Refits =$\frac{175}{50} \sim 3.5$ which is rounded up to 4). It has been performed the forecast density in the figure 59.

According to our model ARMA(1,1)-GARCH(1,1), by first estimation window we forecast the variance for T_1:

$$\sigma_{T_1}^2 = \widehat{\omega}^{(1)} + \hat{\alpha}_1^{(1)} \epsilon_{T_0}^2 + \hat{\beta}_1^{(1)} \sigma_{T_0}^2,$$

where the estimate of $\sigma_{T_0}^2$ is set equal to the sample variance of the estimated window and $\epsilon_{T_0}^2$ is the last squared innovation in the estimated window.

We forecast the return series using the mean model:

$$r_{T_1} = \hat{\mu}^{(1)} + \hat{\phi}_1^{(1)} \left(r_{T_0} - \hat{\mu}^{(1)} \right) + \hat{\theta}_1^{(1)} \epsilon_{T_0}.$$

It can be calculated new residual ϵ_{T_1} by comparing the predicted return r_{T_1} to the observed r_1 as $\epsilon_{T_1} = r_1 - r_{T_1}$. So for the first 50 days of our out of sample we use these parameters $(\widehat{\omega}^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\beta}_1^{(1)}, \hat{\mu}^{(1)}, \hat{\phi}_1^{(1)}, \hat{\theta}_1^{(1)})$, but for the next 50 days (i.e. day 51 to 100) we will do a re-estimation and use $\widehat{\omega}^{(2)}, \hat{\alpha}_1^{(2)}, \hat{\beta}_1^{(2)}, \hat{\mu}^{(2)}, \hat{\phi}_1^{(2)}, \hat{\theta}_1^{(2)}$ and so on.

GARCH Roll				
No.Refits : 4 Refit Horizon : 50 No.Forecasts : 175 Model : ARMA(1,1)-GARCH(1,1) Distribution : sstd				
Forecast Density :				
	Mu	**Sigma**	**Skew**	**Shape**
25/04/2016	0.1199	0.7984	0.8331	7.79
26/04/2016	0.1336	0.7747	0.8331	7.79
27/04/2016	0.1425	0.7517	0.8331	7.79
28/04/2016	0.1708	0.7557	0.8331	7.79
29/04/2016	0.2346	0.8520	0.8331	7.79
02/05/2016	0.2592	0.8613	0.8331	7.79
....				
	Mu	**Sigma**	**Skew**	**Shape**
22/12/2016	0.0669	0.7069	0.8806	6.8557
23/12/2016	0.1076	0.7048	0.8806	6.8557
27/12/2016	0.0612	0.6823	0.8806	6.8557
28/12/2016	0.1017	0.6745	0.8806	6.8557
29/12/2016	0.0811	0.7433	0.8806	6.8557
30/12/2016	0.0905	0.7165	0.8806	6.8557

Figure 59. Forecast density of GARCH ROLL

It has been performed the backtesting of VaR in the Figure 60. The size of the forecasting window is $w_F = 175$ and $p = 0.01$ so the expected number of violations is given by

$$\text{expected} = w_F * p = 175 * 0.01 = 1{,}75 \sim 1{,}8$$

but the actual violations is given by

$$actual = \sum_{t=1}^{w_F} I_t$$

Where

$$I_t = \begin{cases} 0, & if \ r_{t,t+1} \geq VaR_t \quad (Hit) \\ 1, & if \ r_{t,t+1} < VaR_t \ (violation) \end{cases}$$

Let $\hat{\mu}_t$ be the estimated mean in the rolling forecasting. Let $skst_{0.99}(\hat{v})$ be a 99 percent quantile from the Skewed Student-t distribution with $\hat{v}$ degrees of freedom. Let $\hat{\sigma}_t$ be the estimated volatility from the GARCH model. The formula for calculating VaR is

$$VaR_t = \hat{\mu}_t - t_{0.99}(\hat{v})\hat{\sigma}_t$$

If actual < expected the model is said to overestimate risk. If actual > expected the model is said to underestimate risk. A good model will have actual ≈ expected.

VaR Backtest Report	
Model:ARMA(1,1)-GARCH(1,1)-sstd	
Backtest Length:	175
Data:	
alpha: 1%	
Expected Exceed: 1.8	
Actual VaR Exceed: 2	
Actual %: 1.1%	
Unconditional Coverage (Kupiec)	
Null-Hypothesis:	Correct Exceedances
LR.uc Statistic:	0.034
LR.uc Critical:	6.635
LR.uc p-value:	0.853
Reject Null:	NO
Conditional Coverage (Christoffersen)	
Null-Hypothesis:	Correct Exceedances and Independence of Failures
LR.cc Statistic:	0.081
LR.cc Critical:	9.21
LR.cc p-value:	0.96
Reject Null:	NO

Figure 60. Backtesting of VaR of Nasdaq daily returns

We performed the test for the 99% confidence region. The higher the p-value of the unconditional coverage (Kupiec) test and conditional coverage (Christoffersen) test the better the model is, because a high p-value indicates that the null hypothesis is correct. So we can conclude the selected ARMA(1,1)-GARCH(1,1) model with skewed Student-t distribution of residuals is good model for forecasting risk for Nasdaq Daily returns.

In Figure 61, it has been performed the forecasted conditional mean, the forecasted conditional volatility and value at risk (VaR) together with the realized return. The VaR forecast varies over time, and when returns are more volatile the VaR forecast increases. The plot shows that the violations occur whenever the realized return (red) is lower than the VaR forecast (black).

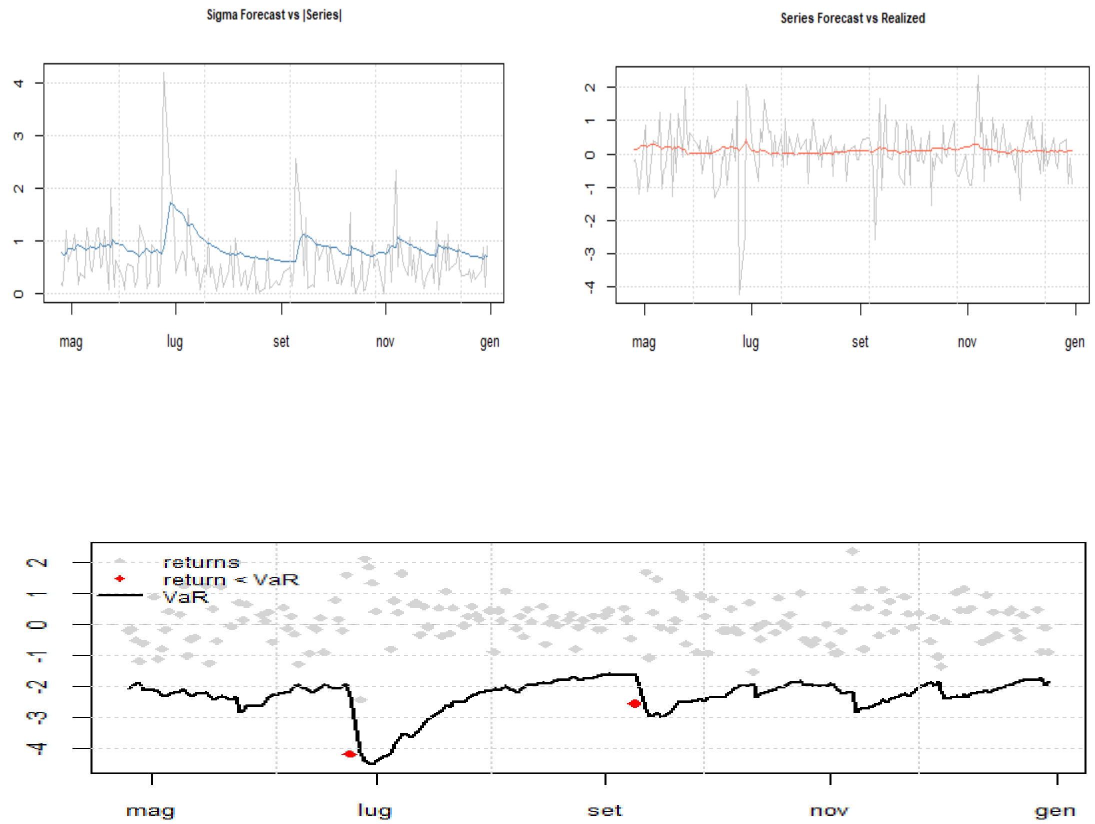


Figure 61. ARMA-GARCH rolling forecast plots of Nasdaq's daily returns

4.6 Out-of-sample forecast performance of S&P 500's GARCH Model

For S&P 500 daily returns, it has been chosen ARMA(1,1)-GARCH(2,1) model with skew Student-t distribution as best model. The S&P 500 daily return includes 1762 observations of 7 years. It has been taken the last 10 percent of data as out-of-sample which includes 175 observations ($w_F = 175$). By rolling forecast approach, it has been fixed the length of the in-sample period which is 1587 observations ($w_E = 1587$). For not producing bad forecasts, we will do a re-estimation of these parameters every 50 days (Refit Horizon= 50, No.Refits =$\frac{175}{50}\sim 3.5$ which is rounded up to 4). It has been performed the forecast density in the figure 62.

According to our model ARMA(1,1)-GARCH(2,1), by first estimation window we forecast the variance for T_1:

$$\sigma_{T_1}^2 = \widehat{\omega}^{(1)} + \hat{\alpha}_1^{(1)}\epsilon_{T_0}^2 + \hat{\alpha}_2^{(1)}\epsilon_{T_{-1}}^2 + \hat{\beta}_1^{(1)}\sigma_{T_0}^2,$$

where the estimate of $\sigma_{T_0}^2$ is set equal to the sample variance of the estimated window and $\epsilon_{T_0}^2$ is the last squared innovation in the estimated window.

We forecast the return series using the mean model:

$$r_{T_1} = \hat{\mu}^{(1)} + \hat{\phi}_1^{(1)}\left(r_{T_0} - \hat{\mu}^{(1)}\right) + \hat{\theta}_1^{(1)}\epsilon_{T_0}.$$

It can be calculated new residual ϵ_{T_1} by comparing the predicted return r_{T_1} to the observed r_1 as $\epsilon_{T_1} = r_1 - r_{T_1}$. So for the first 50 days of our out of sample we use these parameters ($\widehat{\omega}^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\alpha}_2^{(1)}, \hat{\beta}_1^{(1)}, \hat{\mu}^{(1)}, \hat{\phi}_1^{(1)}, \hat{\theta}_1^{(1)}$), but for the next 50 days (i.e. day 51 to 100) we will do a re-estimation and use $\widehat{\omega}^{(2)}, \hat{\alpha}_1^{(2)}, \hat{\alpha}_2^{(2)}, \hat{\beta}_1^{(2)}, \hat{\mu}^{(2)}, \hat{\phi}_1^{(2)}, \hat{\theta}_1^{(2)}$ and so on.

GARCH Roll				
No.Refits : 4 Refit Horizon : 50 No.Forecasts : 175 Model : ARMA(1,1)-GARCH(2,1) Distribution : sstd				
Forecast Density :				
	Mu	**Sigma**	**Skew**	**Shape**
25/04/2016	0.0558	0.6338	0.8417	6.9213
26/04/2016	0.0752	0.6026	0.8417	6.9213
27/04/2016	0.0671	0.5815	0.8417	6.9213
28/04/2016	0.0615	0.5593	0.8417	6.9213
29/04/2016	0.1309	0.5770	0.8417	6.9213

02/05/2016	0.1631	0.6713	0.8417	6.9213
....				
	Mu	**Sigma**	**Skew**	**Shape**
22/12/2016	0.0304	0.5921	0.8527	6.1395
23/12/2016	0.0501	0.5729	0.8527	6.1395
27/12/2016	0.0485	0.5536	0.8527	6.1395
28/12/2016	0.0409	0.5354	0.8527	6.1395
29/12/2016	0.0996	0.5705	0.8527	6.1395
30/12/2016	0.1021	0.6231	0.8527	6-1395

Figure 62. Forecast density of GARCH Roll

It has been performed the backtesting of VaR in the Figure 63.

VaR Backtest Report
Model: ARMA-GARCH-sstd Backtest Length: 175 Data:
alpha: 1% Expected Exceed: 1.8 Actual VaR Exceed: 2 Actual %: 1.1% Unconditional Coverage (Kupiec) Null-Hypothesis: Correct Exceedances LR.uc Statistic: 0.034 LR.uc Critical: 6.635 LR.uc p-value: 0.853 Reject Null: NO Conditional Coverage (Christoffersen) Null-Hypothesis: Correct Exceedances and Independence of Failures LR.cc Statistic: 0.081 LR.cc Critical: 9.21 LR.cc p-value: 0.96 Reject Null: NO

Figure 63. Backtesting of VaR of S&P 500's daily returns

We performed the test for the 99% confidence region. The higher the p-value of the unconditional coverage (Kupiec) test and conditional coverage (Christoffersen) test the better the model is, because a high p-value indicates that the null hypothesis is correct. So we can conclude the selected ARMA(1,1)-GARCH(2,1) model with skewed Student-t distribution of residuals is good model for forecasting risk for S&P 500's Daily returns.

In Figure 64, it has been performed the forecasted conditional mean, the forecasted conditional volatility and value at risk (VaR) together with the realized return.

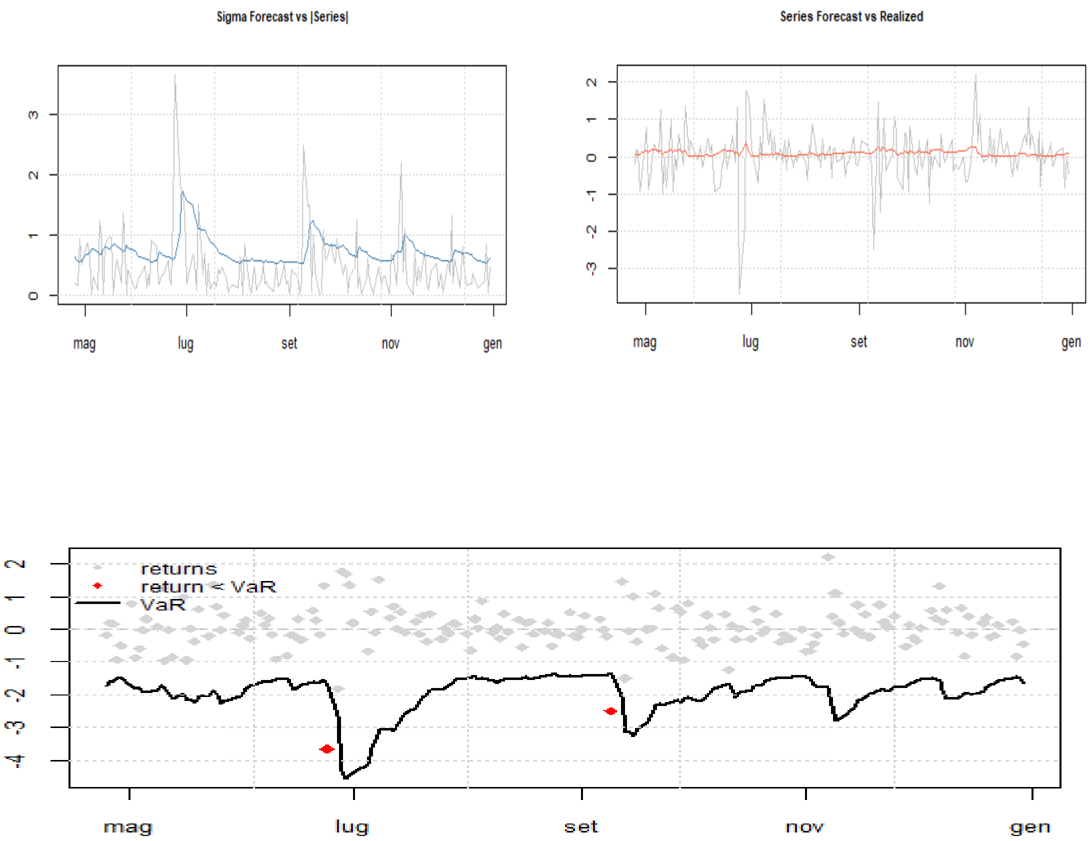


Figure 64. ARMA-GARCH rolling forecast plots of S&P 500's daily returns

4.7 Out-of-sample forecast performance of Nikkei's GARCH Model

For Nikkei daily returns, it has been chosen ARMA(2,2)-GARCH(1,1) model with skew Student-t distribution as best model. The Nikkei daily return includes 1713 observations of 7 years. It has been taken the last 10 percent of data as out-of-sample which includes 171 observations ($w_F =$

171). By rolling forecast approach, it has been fixed the length of the in-sample period which is 1542 observations ($w_E = 1542$).

For not producing bad forecasts, we will do a re-estimation of these parameters every 50 days (Refit Horizon$= 50$, No.Refits $=\frac{171}{50} \sim 3.5$ which is rounded up to 4). It has been performed the forecast density in the figure 65.

According to our model ARMA(2,2)-GARCH(1,1), by first estimation window we forecast the variance for T_1:

$$\sigma_{T_1}^2 = \widehat{\omega}^{(1)} + \hat{\alpha}_1^{(1)} \epsilon_{T_0}^2 + \hat{\beta}_1^{(1)} \sigma_{T_0}^2,$$

where the estimate of $\sigma_{T_0}^2$ is set equal to the sample variance of the estimated window and $\epsilon_{T_0}^2$ is the last squared innovation in the estimated window.

We forecast the return series using the mean model:

$$r_{T_1} = \hat{\mu}^{(1)} + \hat{\phi}_1^{(1)}\left(r_{T_0} - \hat{\mu}^{(1)}\right) + \hat{\phi}_2^{(1)}\left(r_{T_{-1}} - \hat{\mu}^{(1)}\right) + \hat{\theta}_1^{(1)} \epsilon_{T_0} + \hat{\theta}_2^{(1)} \epsilon_{T_{-1}}.$$

It can be calculated new residual $\epsilon_{T_1}$by comparing the predicted return r_{T_1} to the observed r_1 as $\epsilon_{T_1} = r_1 - r_{T_1}$. So for the first 50 days of our out of sample we use these parameters ($\widehat{\omega}^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\beta}_1^{(1)}, \hat{\mu}^{(1)}, \hat{\phi}_1^{(1)}, \hat{\phi}_2^{(1)}, \hat{\theta}_1^{(1)}, \hat{\theta}_2^{(1)}$), but for the next 50 days (i.e. day 51 to 100) we will do a re-estimation and use $\widehat{\omega}^{(2)}, \hat{\alpha}_1^{(2)}, \hat{\beta}_1^{(2)}, \hat{\mu}^{(2)}, \hat{\phi}_1^{(2)}, \hat{\phi}_2^{(2)}, \hat{\theta}_1^{(2)}, \hat{\theta}_2^{(2)}$ and so on.

GARCH Roll				
No.Refits : 4 Refit Horizon : 50 No.Forecasts : 171 Model : ARMA(2,2)-GARCH(1,1) Distribution : sstd				
Forecast Density :				
	Mu	**Sigma**	**Skew**	**Shape**
20/04/2016	0.1379	2.2100	0.9002	9.7143
21/04/2016	0.2329	2.0559	0.9002	9.7143
22/04/2016	-0.0483	2.0857	0.9002	9.7143
25/04/2016	0.0196	1.9876	0.9002	9.7143
26/04/2016	0.2323	1.8716	0.9002	9.7143
27/04/2016	0.0373	1.7643	0.9002	9.7143
....				

	Mu	Sigma	Skew	Shape
22/12/2016	-0.1951	0.9394	0.8938	7.6451
26/12/2016	0.3084	0.9131	0.8938	7.6451
27/12/2016	0.2246	0.9040	0.8938	7.6451
28/12/2016	-0.2415	0.8843	0.8938	7.6451
29/12/2016	0.0372	0.8687	0.8938	7.6451
30/12/2016	0.3818	0.9721	0.8938	7.6451

Figure 65. Forecast density of GARCH Roll

It has been performed the backtesting of VaR in the Figure 66. The size of the forecasting window is $w_F = 171$ and $p = 0.01$ so the expected number of violations is given by

$$\text{expected} = w_F * p = 171 * 0.01 = 1{,}71 \sim 1{,}7$$

but the actual violations is given by $actual = \sum_{t=1}^{W_F} I_t$

VaR Backtest Report
Model: ARMA-GARCH-sstd Backtest Length: 171 Data:
alpha: 1% Expected Exceed: 1.7 Actual VaR Exceed: 3 Actual %: 1.8% Unconditional Coverage (Kupiec) Null-Hypothesis: Correct Exceedances LR.uc Statistic: 0.083 LR.uc Critical: 6.635 LR.uc p-value: 0.37 Reject Null: NO Conditional Coverage (Christoffersen) Null-Hypothesis: Correct Exceedances and Independence of Failures LR.cc Statistic: 0.91 LR.cc Critical: 9.21 LR.cc p-value: 0.634 Reject Null: NO

Figure 66. Backtesting of VaR of Nikkei's daily returns

We performed the test for the 99% confidence region. The higher the p-value of the unconditional coverage (Kupiec) test and conditional coverage (Christoffersen) test the better the model is, because a high p-value indicates that the null hypothesis is correct. So we can conclude the selected ARMA(2,2)-GARCH(1,1) model with skewed Student-t distribution of residuals is good model for forecasting risk for Nikkei's Daily returns.

In Figure 67, it has been performed the forecasted conditional mean, the forecasted conditional volatility and value at risk (VaR) together with the realized return.

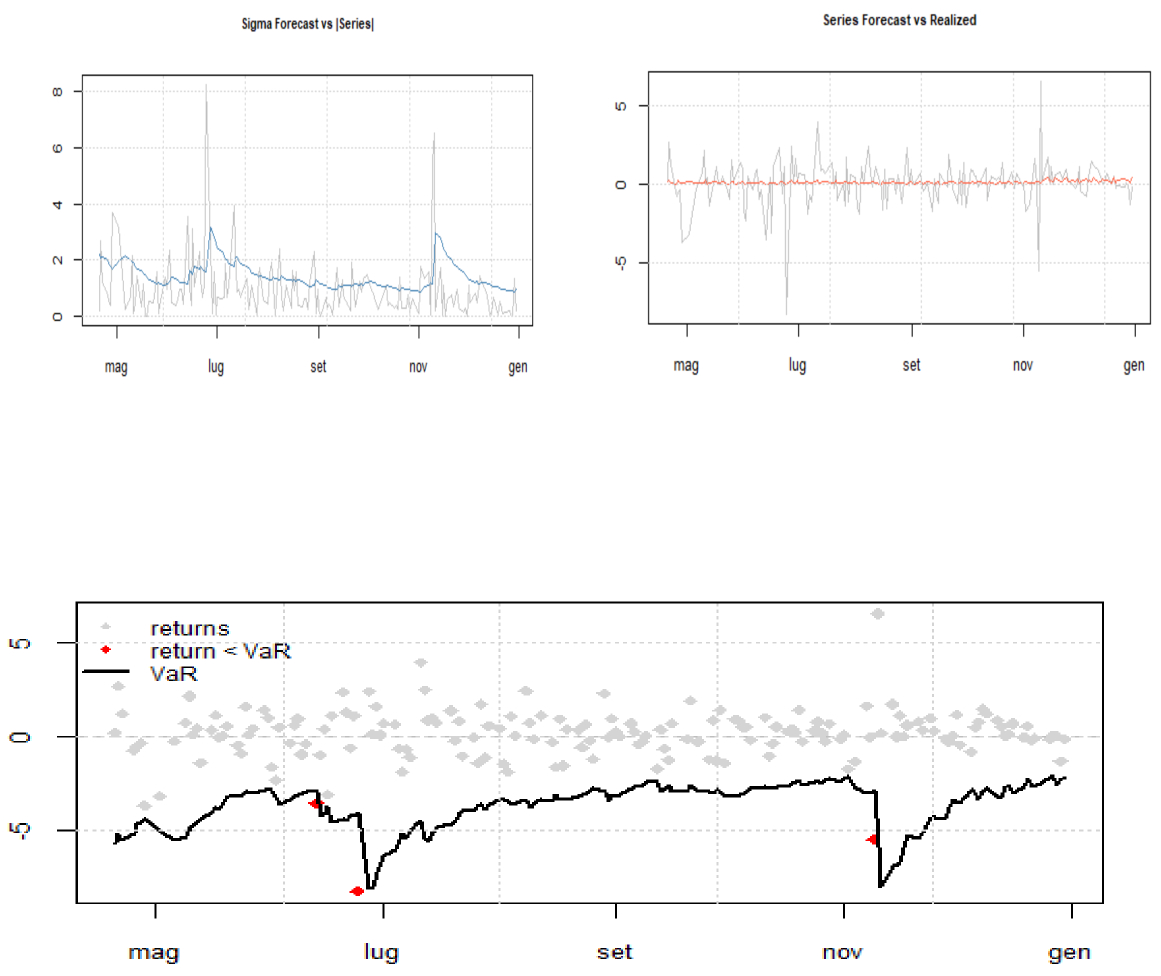


Figure 67. ARMA-GARCH rolling forecast plots of Nikkei's daily returns

4.8 Out-of-sample forecast performance of Dow Jones's GARCH Model

For Dow Jones daily returns, it has been chosen GARCH(2,1) model with skew Student-t distribution as best model. The Dow Jones daily return includes 1762 observations of 7 years. It has been taken the last 10 percent of data as out-of-sample which includes 175 observations ($w_F = 175$).

By rolling forecast approach, it has been fixed the length of the in-sample period which is 1587 observations ($w_E = 1587$). For not producing bad forecasts, we will do a re-estimation of these parameters every 50 days (Refit Horizon$= 50$, No.Refits =$\frac{175}{50} \sim 3.5$ which is rounded up to 4). It has been performed the forecast density in the figure 68.

According to our model GARCH(2,1), by first estimation window we forecast the variance for T_1:

$$\sigma_{T_1}^2 = \widehat{\omega}^{(1)} + \hat{\alpha}_1^{(1)} \epsilon_{T_0}^2 + \hat{\alpha}_2^{(1)} \epsilon_{T_{-1}}^2 + \hat{\beta}_1^{(1)} \sigma_{T_0}^2,$$

where the estimate of $\sigma_{T_0}^2$ is set equal to the sample variance of the estimated window and $\epsilon_{T_0}^2$ is the last squared innovation in the estimated window.

We forecast the return series using the mean model:

$$r_{T_1} = \hat{\mu}^{(1)} + \epsilon_{T_0}.$$

It can be calculated new residual $\epsilon_{T_1}$by comparing the predicted return r_{T_1} to the observed r_1 as $\epsilon_{T_1} = r_1 - r_{T_1}$. So for the first 50 days of our out of sample we use these parameters ($\widehat{\omega}^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\alpha}_2^{(1)}, \hat{\beta}_1^{(1)}, \hat{\mu}^{(1)}$), but for the next 50 days (i.e. day 51 to 100) we will do a re-estimation and use $\widehat{\omega}^{(2)}, \hat{\alpha}_1^{(2)}, \hat{\alpha}_2^{(2)}, \hat{\beta}_1^{(2)}, \hat{\mu}^{(2)}$ and so on.

GARCH Roll				
No.Refits : 4 Refit Horizon : 50 No.Forecasts : 175 Model : GARCH(2,1) Distribution : sstd				
Forecast Density :				
	Mu	**Sigma**	**Skew**	**Shape**
25/04/2016	0.0646	0.6110	0.898	7.5597
26/04/2016	0.0646	0.5739	0.898	7.5597
27/04/2016	0.0646	0.5457	0.898	7.5597
28/04/2016	0.0646	0.5211	0.898	7.5597
29/04/2016	0.0646	0.6024	0.898	7.5597
02/05/2016	0.0646	0.7208	0.898	7.5597
…………….				
	Mu	**Sigma**	**Skew**	**Shape**
22/12/2016	0.059	0.5462	0.9097	6.416
23/12/2016	0.059	0.5251	0.9097	6.416
27/12/2016	0.059	0.5037	0.9097	6.416
28/12/2016	0.059	0.4835	0.9097	6.416
29/12/2016	0.059	0.5063	0.9097	6.416
30/12/2016	0.059	0.5293	0.9097	6.416

Figure 68. Forecast density of GARCH Roll

It has been performed the backtesting of VaR in the Figure 69. We performed the test for the 99% confidence region. The higher the p-value of the unconditional coverage (Kupiec) test and conditional coverage (Christoffersen) test the better the model is, because a high p-value indicates that the null hypothesis is correct. So we can conclude the selected GARCH(2,1) model with skewed Student-t distribution of residuals is good model for forecasting risk for Dow Jones's Daily returns.

In Figure 70, it has been performed the forecasted conditional mean, the forecasted conditional volatility and value at risk (VaR) together with the realized return.

VaR Backtest Report

Model: GARCH-sstd
Backtest Length: 175
Data:

alpha: 1%
Expected Exceed: 1.8
Actual VaR Exceed: 2
Actual %: 1.1%

Unconditional Coverage (Kupiec)
Null-Hypothesis: Correct Exceedances
LR.uc Statistic: 0.034
LR.uc Critical: 6.635
LR.uc p-value: 0.853
Reject Null: NO

Conditional Coverage (Christoffersen)
Null-Hypothesis: Correct Exceedances and Independence of Failures
LR.cc Statistic: 0.081
LR.cc Critical: 9.21
LR.cc p-value: 0.96
Reject Null: NO

Figure 69. Backtesting of VaR of Dow Jones's daily returns

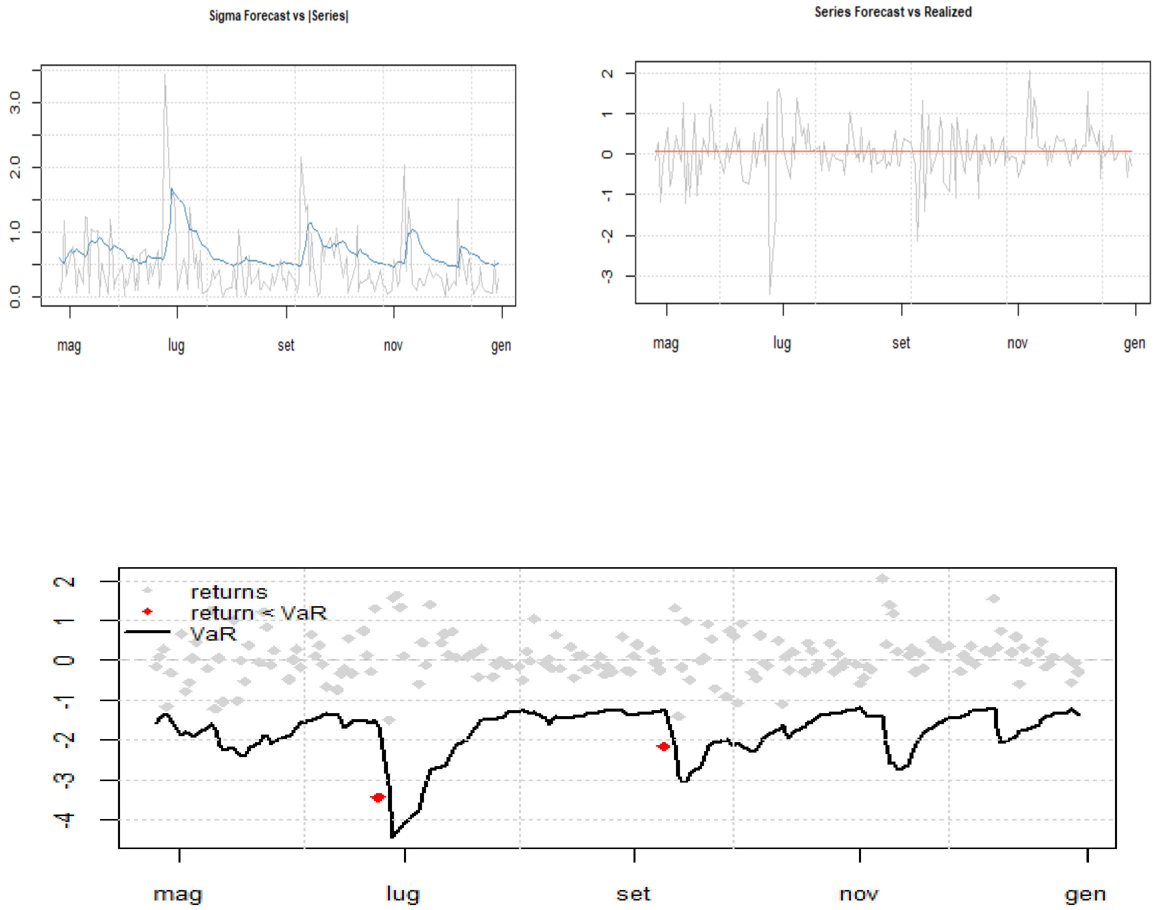


Figure 70. GARCH rolling forecast plots of Dow Jones's daily returns

5 Conclusion

Given the purpose of this thesis which is evaluating the GARCH model and its ability to forecast Value at Risk of financial data, it has been used different models to estimate separately on four different financial data (Nasdaq, S&P 500, Nikkei, Dow jones). In addition, it has been chosen the normal distribution, the Student-t distribution and the skewed Student-t distribution as the error term distribution. For each index, it has been selected the best model to forecast.

To forecast the volatility, for NASDAQ daily return, it has been chosen the ARMA(1,1)-GARCH(1,1) model with skewed Student-t distribution, for S&P500 daily return, it has been chosen the ARMA(1,1)-GARCH (2, 1) model with skewed Student-t distribution, for Nikkei daily return, it has been chosen the ARMA(2,2)-GARCH(1,1) model with skewed Student-t distribution and for Dow Jones daily return, it has been chosen the GARCH(2,1) model with skewed Student-t distribution.

As we could noticed, for all four model the skew Student-t distribution is more efficient than the normal distribution and Student-t distribution.

The volatility forecast has been used to calculate value of risk (VaR). Finally, for validating a set of VaR forecasts it has been used backtesting procedure. In this thesis, the unconditional coverage (Kupiec) test and conditional coverage (Christoffersen) test have been used for backtesting VaR. According to the results of these tests, all four model adequately captures the risk.

6 References

[1] Mandelbrot, B. B., 1963. The Variation of Certain Speculative Prices, *The Journal of Business*, volume 36, No. 4, 394-419.

[2] Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, volume 74, 3-30.

[3] Engle, R. F. , 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, volume 50, No. 4, 987-1007.

[4] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, Elsevier, volume 31, No. 3, 307-327.

[5] Bollerslev, T., Engle, R.F. and Nelson, 1994. D.B.: ARCH Models. In: Engle, R.F. and McFadden, D.L. (Eds.): *Handbook of Econometrics*, volume 4, 2959-3038. North-Holland, Amsterdam.

[6] P. Jorion, 2006. Value at risk: The New Benchmark for managing Financial Risk, 3rd Edition. McGraw Hill, New York.

[7] Campbell, S., 2005. *A review of backtesting and backtesting procedures* (Discussion Paper No. 2005-21), Washington, D.C.: Finance and Economics Discussion Series, Federal Reserve Board.

[8] P. J. Brockwell and R. A. David., 1990. Time Series: Theory and Methods, 2rd edition. Springer, New York, NY, USA.

[9] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 1994. Time Series Analysis; Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

[10] Christian Francq and Jean-Michel Zakoian, 2010. GARCH Models: Structure, Statistical Inference and Financial Applications :University Lille3,France.

[11] Fernandez, C. and Steel, M. F. J., 1998. *On Bayesian modeling of fat tails and skewness*. J. Amer. Statist. Assoc. volume 93, 359–371.

[12] RUEY S. TSAY, 2005. Analysis of Financial Time Series, 2rd Edition: University of Chicago.

[13] Kupiec, P. H., 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives,* volume 3, 73–84.

[14] Yahoo finance: NASDAQ historical prices,
https://finance.yahoo.com/quote/%5EIXIC/history?p=%5EIXIC

[15] Yahoo finance: S&P 500 historical prices,
https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC

[16] Yahoo finance: Nikkei 225 historical prices,

https://finance.yahoo.com/quote/%5EN225?p=^N225

[17] Yahoo finance: Dow 30 historical prices,

https://finance.yahoo.com/quote/%5EDJI/history?p=%5EDJI

[18] S. S. Shapiro and M. B. Wilk, 1965. An Analysis of Variance Test for Normality (Complete Samples): Biometrika.

[19] Sheather, S. J. and Jones, M. C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society series B*, volume 53, 683–690.

[20] Brooks, C., 2008. Introductory econometrics for finance. Cambridge University Press, UK

[21] Taylor, S., 1986. Modeling Financial Time Series. John Willy & Sons, New York.

[22] Bera, A. K., and Higgins, M. L., 1993. ARCH models : Properties, estimation and testing, *Journal of Economic Surveys*, volume 7, No. 4, 307-366.

[23] Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., 2rd edition. *International Symposium on Information Theory*, Armenia, 2-8 September 1971, 267-281.

[24] Schwarz, Gideon E., 1978. Estimating the dimension of a model, *Annals of Statistics*, volume 6, No. 2, 461–464.

[25] Orhan, M., Köksal, B., 2011. A comparison of GARCH models for VaR estimation. *Expert Systems with Applications,* volume 39, 3582–3592

[26] Christoffersen, P. F., 1998. Evaluating Interval Forecasts*, International Economic Review*, volume 39, No. 4, 841-862.

[27] Dowd, K., 2005. Measuring market risk, 2rd Edition. Chichester: John Wiley & Sons.

[28] Engle, R. F., 2001. GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, volume 15, No. 4, 157–168.

[29] https://en.wikipedia.org/wiki/Normal_distribution#cite_note-7

[30] Shamiri, A. and Z. Isa, 2009. Modelling and forecasting volatility of Malaysian stock markets. *Journal of Mathematics and Statistics*, volume 3, 234-240.

[31] Bali, T.G., 2007. Modelling the dynamics of interest rate volatility with skewed fat-tailed distribution. *Annals of Operations Research*, volume 151, 151-178.

Appendix :

For the GARCH (1, 1) process given by

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

and

$$\epsilon_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim IIDN(0,1),$$

where $\omega > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$, a necessary and sufficient condition for stationarity is

$$\alpha_1 + \beta_1 < 1 \tag{1}$$

Proof: we use recursive substitution to show that

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \omega + \alpha_1 \varepsilon_{t-1}^2 (\omega + \alpha_1 \varepsilon_{t-2}^2 \sigma_{t-2}^2 + \beta_1 \sigma_{t-2}^2) + \beta_1 (\omega + \alpha_1 \varepsilon_{t-2}^2 \sigma_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\ &\vdots \\ &= \omega \sum_{k=0}^{\infty} M(t,k), \qquad (2)\end{aligned}$$

where $M(t,k) = \alpha_1^a \beta_1^b \prod_{l=1}^{a} \varepsilon_{t-S_l}^2$, for $a + b = k$ and $1 \leq S_1 < S_2 < \cdots < S_a \leq k$.

Since ε_t is i.i.d., the expected values of $M(t,k)$ do not depend on t, and

$$E\big(M(t,k)\big) = E\big(M(s,k)\big) \text{ for all } k, t, s. \tag{3}$$

In addition

$$\begin{aligned} M(t,0) &= 1, \\ M(t,1) &= \alpha_1 \varepsilon_{t-1}^2 + \beta_1, \\ M(t,2) &= (\alpha_1 \varepsilon_{t-1}^2 + \beta_1)(\alpha_1 \varepsilon_{t-2}^2 + \beta_1),\end{aligned}$$

and generally,

$$M(t,k+1) = (\alpha_1 \varepsilon_{t-1}^2 + \beta_1) M(t-1,k),$$

which yields together with (3), that

$$\begin{aligned} E\big(M(t,k+1)\big) &= (\alpha_1 + \beta_1) E\big(M(t,k)\big) \\ &\vdots \\ &= (\alpha_1 + \beta_1)^{k+1} E(M(t,0)) \end{aligned}$$

$$= (\alpha_1 + \beta_1)^{k+1}. \tag{4}$$

and by (2) and (4),

$$E(\epsilon_t^2) = \omega E(\sum_{k=0}^{\infty} M(t,k))$$
$$= \omega \sum_{k=0}^{\infty} E(M(t,k))$$
$$= \omega \textstyle\sum_{k=0}^{\infty} 1/(\alpha_1 + \beta_1)^{k+1},$$

where $(\epsilon_t|\mathcal{F}_{t-1}) \sim N(0, \sigma_t^2)$ and

$$E(\epsilon_t^2) = \omega(1 - \alpha_1 - \beta_1)^{-1}$$

if and only if (1) holds and ϵ_t^2 converges almost surely. Then $E(\epsilon_t) = 0$ and $COV(\epsilon_t, \epsilon_s) = 0$ for $t \neq s$ follows by symmetry. ■