Stockholms universitet

# Volatility forecast of Google stock daily log returns

Viktor Törnegren

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2018:1
Matematisk statistik
Juni 2018

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm
University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2018:1**
http://www.math.su.se

# Volatility forecast of Google stock daily log returns

Viktor Törnegren*

June 2018

**Abstract**

This thesis aims to fit an appropriate general atuoreggressive heteroscedastic model to Google stock daily log returns from 2006-02-01 to 2018-01-31. The purpose is to evaluate whether or not the GARCH$(1,1)$ model can be used to predict the one step ahead volatility of the log returns. This is done by applying a back testing procedure and then compute interval forecasts respectively density forecasts. The results indicates that the GARCH$(1,1)$ model assuming a normal distribution in combination with rolling window length of 750 days respectively 1250 days yields correct conditional coverage when the actual coverage probability of the predicted interval is 0.95. It further implies that the interval forecast from the student-t distribution is to cautious, regardless of the window length used, when the actual coverage probability is 0.95. But the results from the density forecasts implies that the GARCH$(1,1)$ model, assuming a student-t distribution and using window length 250 days, yields good predictions. However, the density forecasts also indicates that a skewed student-$t$ distribution might have a better fit. Hence, for further research it is suggested to calculate interval forecast, letting the actual coverage probability, $p$, vary between 0.8 and 0.95. It is also suggested that in addition to the normal and student-$t$ distribution also include the skewed student-t distribution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: viktor.tornegren@gmail.com. Supervisor: Mathias Lindholm and Filip Lindskog.

I sincerely would like to thank my supervisors Mathias Lindholm and Filip Lindskog for their encouraging help and suggestions.

# Contents

## 1 Introduction

Investing money in the financial market means that you are exposed to risk. The daily closing price of the asset will change day by day. Hence, in the financial sector it is common to model and predict the volatility of an financial asset. The volatility is defined as the conditional standard deviation of the asset return. Thus, the volatility is a measure of risk. Since there is only one observation made on a trading day, the stock volatility is not directly observable. However, volatility has some characteristics that is frequently observed in the stock returns. For instance, volatility is often high for certain periods and low for other periods, this pattern is known as volatility clusters. When volatility clusters exists heteroscedastic models can be used to predict the one step ahead volatility. Robert F. Engle created the autoregressive conditional heteroscedastic (ARCH) model in 1982. This model uses past values of the shock of the stock returns to predict the future volatility. In 1986 Tim Bollerslev extended the ARCH model to the general autoregressive conditional heteroscedastic (GARCH) model. In addition to the ARCH model, Bollerslevs model also depends on previous values of the conditional variance of the asset return.

This thesis aims to fit an appropriate general atuoreggressive heteroscedastic model to Google stock daily log returns from 2006-02-01 to 2018-01-31 from 2006-02-01 to 2018-01-31. The purpose is to evaluate whether or not the GARCH$(1, 1)$ model can be used to predict the one step ahead volatility of the Google stock daily log return. This is done by applying a back testing procedure and then compute interval forecasts respectively density forecasts.

In section 2 the theory of asset returns and financial time series are presented. Section 3 describes the methodology and section 4 includes the data anlysis as well as the procedure of fitting an appropriate model. The results are presented in section 5. Section 6 includes discussion and conclusion of the results as well as suggestions to further research.

# 2 Theory

In this section, theories of linear time series analysis are introduced, as well as modelling and tests.

## 2.1 Return

The theory in this subsection is from Tsay (2010, chapter 1)

When studying financial time series it is common to examine returns, instead of prices, of an asset. However, the return is a function of asset prices and can be defined in several ways.

The one-period simple gross return is defined as

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

where $P_t$ is the closing price today and $P_{t-1}$ is yesterdays closing price. $R_t$ is the simple return and it is define as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1. \tag{2}$$

The return of holding the asset for k days, is called the multiperiod simple return

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \tag{3}$$

It is often assumed that the simple returns are independently and identically (iid) normally distributed. There are mainly two problems with this assumption. First, we have that $P_t, P_{t-k+1} > 0$, hence, the lower bound of $R_t$ is $-1$. Whereas the normal distribution can take any value in the interval $[-\infty, \infty]$. Second, the multiperiod simple returns are a product of normally distributed, one-period simple returns and by definition is not normally distributed.

Hence, it is common to use the simple log returns. Taking the natural logarithm of equation (1) yields the simple log return

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{4}$$

and the multiperiod log return is defined as

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) \\ &= \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) \cdots \ln(1 + R_{t-k+1}) \end{aligned} \tag{5}$$

The simple log returns can take any value in the interval $[-\infty, \infty]$ and the multiperiod log returns is a sum of normally distributed one-period returns, hence, $r_t[k]$ is also normally distributed.

## 2.2 Financial time series

The theory in section 2.2 to 2.2.10 is based on Tsay (2010, chapter 1-3)

The log returns, $r_t$ is as a collection of random variables over time. Thus, $\{r_t\}_{t=1}^{T}$, is a time series, were T is the total number of days observed.

Moreover, if $r_t$ is a linear time series, then it can be written on the form

$$r_t = \mu_t + a_t \tag{6}$$

where $\mu_t$ is the mean of $r_t$ and $\{a_t\}$ is a sequence of iid random variables, with mean zero and variance $\sigma_a^2$, thus, $\{a_t\}$ is a white noise series. It is often referred to as shocks or innovations and it will be seen in the succeeding sections that $a_t$ denotes the new information at time t.

A time series $r_t$ is weakly stationary if the following two criteria are fulfilled

1. If $\{r_t\}$ has constant mean, $E[r_t] = \mu$
2. and $Cov(r_t, r_{t-\ell}) = \gamma_\ell$ is time invariant, i.e, the covariance only depends on the lag length $\ell$.

If the sequence $\{r_t\}$ is weakly stationary it is said to be a white noise time series. This implies that the first two moments of $r_t$ are finite. In addition to this, if $r_t$ is also normally distributed with mean zero and variance $\sigma^2$ it is called a Gaussian white noise process.

### 2.2.1 Autocorrelation function (ACF)

A measurement of linear dependence (correlation) between $r_t$ and its past value $r_{t-\ell}$ is called the autocorrelation function (ACF), it is often denoted by $\rho_\ell$. Under the assumption of weak stationarity, $\rho_\ell$, is time invariant. That is, the ACF depends only of the lag length $\ell$. It is defined as

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0} \tag{7}$$

where $Var(r_t) = Var(r_{t-\ell})$ under the assumption that $\{r_t\}$ is weakly stationary. By definition, $\rho_0 = 1$, $\rho_\ell = \rho_{-\ell}$ and $-1 \leq \rho_\ell \leq 1$. Given a sample $\{r_t\}_{t=1}^{T}$, the ACF is estimated with

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^{T}(r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=\ell+1}^{T}(r_t - \bar{r})^2}, \quad 0 \leq \ell < T - 1 \tag{8}$$

where $\bar{r} = \sum_{t=\ell+1}^{T} r_t/T$ is the sample mean. If all $\hat{\rho}_\ell \approx 0$ the series $\{r_t\}$ is said to be white noise. And if $\{r_t\}$ is iid, with $E(r_t^2) < \infty$, then $\hat{\rho}_\ell$ is asymptotically normally distributed with mean zero and variance $1/T$. For any positive integer $\ell$, the former result can be used to test

$$H_0 : \rho_\ell = 0 \text{ against } H_a : \rho_\ell \neq 0$$

with the test statistic

$$\text{t ratio} = \frac{\hat{\rho}_\ell}{\sqrt{(1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_i^2)/T}} \sim t_{T-1} \tag{9}$$

### 2.2.2 Ljung-Box test

In financial time series it is of great importance to test jointly that several autocorrealtions of $r_t$ are zero. Ljung and Box (1978, citied in Tsay, p32) has modified the Portmanteau statistic to increase the power when working with finite samples

$$Q(m) = T(T+2)\sum_{\ell=1}^{m} \frac{\hat{\rho}_\ell^2}{T-\ell} \sim \chi_\alpha^2(m) \tag{10}$$

where $m$ is the number of lags. Studies suggest that the choice of $m \approx \ln(T)$. This test, tests $H_0 : \rho_1 = \cdots = \rho_m = 0$ against $H_a : \rho_\ell \neq 0$ for some $\ell$. The null hypothesis, $H_0$, is rejected if $Q(m) > \chi_\alpha^2$, where $\chi_\alpha^2$ denotes the $100(1-\alpha)th$ percentile of a chi-squared distribution with $m$ degrees of freedom.

### 2.2.3 Skewness

Let X be a random variable. The third central moment of X is called Skewness. As the the name indicates this moment measures the symmetry of X with respect to its mean. The skewness of X is defined as

$$S(X) = E\left[\frac{(X - \mu_x)^3}{\sigma_x^3}\right]$$

and the sample skewness

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^3. \tag{11}$$

It is often assumed that X has a symmetric distribution if $\hat{S}(x) \in [-0.5, 0.5]$.

### 2.2.4 Kurtosis

The fourth central moment of X measures tail thickness, it is known as kurtosis

$$K(X) = E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right],$$

the sample kurtosis is defined as

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^4. \tag{12}$$

A normal distributed variable X is known to have $K(x) = 3$ and a distribution with $K(x) > 3$ is said to be leptokurtic. If the excess kurtosis $K(x) - 3 > 0$ the distribution tends to contain more extreme values. Hence, the density function is characterized by a high, thin peak around its mean and has heavy tails.

### 2.2.5 Jarque and Bera (JB) test

Jarque and Bera (1987, citied in Tsay, p.10) created a test-statistic that combines skewness and kurtosis to test for normality

$$JB = \frac{\hat{S}^2(x)}{6/T}\frac{[\hat{K}(x)-3]^2}{24/T} \sim \chi^2_\alpha(2). \tag{13}$$

Where $6/T$ and $24/T$ is the variance of $\hat{S}(x)$ respectively $\hat{K}(x)$. $H_0$ that X is normally distributed is rejected if the p-value of the JB statistic is less than the significance level.

### 2.2.6 Autoregressive (AR)

If $r_t$ has a significant lag-$p$ autocorrelation, then lags up to $r_{t-p}$ can be used to predict $r_t$ with an autoregressive model of order $p$, AR($p$)

$$r_t = \phi_0 + \sum_{i=1}^{p}\phi_i r_{t-i} + a_t \tag{14}$$

where $|\sum_{i=1}^{p}\phi_i| < 1$ and $\{a_t\}$ is assumed to be a white noise series with mean zero and variance equal to one.

### 2.2.7 Partial Autocorrelation Function (PACF)

In time series analysis a useful tool to determine the order $p$ of an AR model is the partial autocorrelation function. The PACF is a function of its ACF, consider the AR models

$$\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t} \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t} \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t} \\
&\vdots \\
r_t &= \phi_{0,j} + \phi_{1,j} r_{t-1} + \phi_{2,j} r_{t-2} + \phi_{3,j} r_{t-3} + \ldots + \phi_{j,j} r_{t-j} + e_{jt}
\end{aligned} \tag{15}$$

where $\phi_{0,j}$ is the constant term, $\phi_{i,j}$ the coefficient of $r_{t-i}$ and $\{e_{jt}\}$ is the error term of an AR($j$) model. The models in equation (15) are recognized as multiple linear regression models. Thus, the coefficients can be estimated with the ordinary least-square (OLS) method. The estimate $\hat{\phi}_{1,1}$ from the first equation, is called the lag-1 sample PACF of $r_t$. The estimate $\hat{\phi}_{j,j}$ from the $j$:th equation, is called the lag-$j$ sample PACF of $r_t$. Where $\hat{\phi}_{2,2}$ shows the added contribution of $r_{t-2}$ to $r_t$ over the AR(1) model. And $\hat{\phi}_{3,3}$ shows the added contribution of $r_{t-3}$ to $r_t$ over the AR(2) model, etc. Hence, the lag order is chosen such that the lag-$p$ sample PACF is non zero for an AR($p$) model, while $\hat{\phi}_{j,j}$ is close to zero for all $j > p$. Thus, if the PACF cuts of after lag-$p$, then the AR model should have order $p$.

### 2.2.8 Conditional Heteroskedastic Models

It is well known that stock volatility is not directly observable, since there is only one observation made in a trading day. However, the volatility has some characteristics that are important to capture, when trying to model volatility

1. The volatility is high for certain periods and low for other periods, known as volatility clusters.
2. Volatility jumps are rare, meaning that volatility evolves contionuously over time.
3. Volatility often varies within some fixed ranges, which means that it is often stationary.
4. Volatility should react differently to a big price increase compared to a decrease.

Moreover, a common assumption made when studying volatility is that the serie $\{r_t\}$ is serially uncorrelated, but dependent. Hence, given the information set available at time $t-1$, denoted as $\mathcal{F}_{t-1}$, the conditional mean and variance of $r_t$ is

$$\mu_t = E[r_t|\mathcal{F}_{t-1}], \quad \sigma_t^2 = Var(r_t|\mathcal{F}_{t-1}) = E[(r_t-\mu_t)^2|\mathcal{F}_{t-1}]. \tag{16}$$

Combining equation (6) and (16) results in

$$\sigma_t^2 = Var(r_t|\mathcal{F}_{t-1}) = Var(a_t|\mathcal{F}_{t-1}), \tag{17}$$

which is and will be referred to as the volatility equation of $r_t$.

### 2.2.9 ARCH

Robert F. Engle (1982, cited in Tsay, p.115) created the autoregressive conditional heteroskedasticity (ARCH) model. The basic idea of the model is simple, the first assumption is that the shock $a_t$ is serially uncorrealted, that is, there should not be any linear dependence betwen $a_t$ and its previous values. Second, the shock $a_t$ of an asset return should have a non-linear dependency. This dependency is described by a quadratic function of the lagged values of $a_t$. The ARCH($m$) model is defined as

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2, \tag{18}$$

where $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0, variance 1 and $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i > 0$, due to the fact that $\sigma_t^2 \geq 0$.

Equation (18) shows that large changes of $\{a_{t-i}^2\}_{i=1}^m$ implies large change of the conditional variance $\sigma_t^2$ for the shock $a_t$. This is a characteristic behaviour of volatility clustering.

Consider the model of $\sigma_t^2$ from equation (18). Assuming that $a_t^2$ linearly depends on its lagged values and that $a_t^2$ is an unbiased estimate of $\sigma_t^2$. One can use the PACF (see section 2.2.7) of $a_t^2$ to determine the order $m$ of an ARCH($m$) model.

Moreover, the standardized residuals from the ARCH model, $\tilde{a}_t$, is a sequence of iid random variables.

$$\tilde{a}_t = \frac{a_t}{\sigma_t} \tag{19}$$

Thus, the Quantile Quantile (QQ) plot of $\tilde{a}_t$ can be used to check the distribution assumption of the model. And the Ljung-Box test (see section 2.2.2) of $\tilde{a}_t^2$ can be used to test the validity of the volatility equation.

However, there are some weaknesses of the ARCH model. For instance, since the model depends on the square values of the previous innovations, it assumes that positive and negative shocks have the same effect on $\sigma_t^2$. In addition to this, the ARCH model tends to respond slowly to large isolated shocks, thus, it is likely to overpredict the volatility.

### 2.2.10 GARCH

Bollerslev (1986, cited in Tsay(2010), p.131) created the generalized autoregressive conditional heteroscedastic (GARCH) model. In addition to ARCH($m$), this model also depends on previous values of the conditional variance, $\sigma_{t-j}^2$. Hence, a GARCH model often requires a lot fewer parameters than an ARCH model. The GARCH($m, s$) model is defined as

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{i=1}^{s} \beta_j \sigma_{t-j}^2, \tag{20}$$

where $\{\epsilon_t\}$ is a sequence of independently and identically distributed random variables, with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{\max(m,s)}(\alpha_i + \beta_i) < 1$. The simplest GARCH models occurs when $m = s = 1$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{21}$$

Using the above equation and assuming that the forecast origin is $h$. Then, $\alpha_h^2$ and $\sigma_h^2$ are known at time index $h$. Thus,

$$\sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2, \tag{22}$$

which yields that the 1-step-ahead forecast is

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2. \tag{23}$$

The calculation of the maximum likelihood estimates $(\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_m, \hat{\beta}_1, \ldots, \hat{\beta}_s)$ is described in Appendix B.1.1

### 2.2.11 Back testing

The theory in section 2.2.11 to 2.2.12 is based on Christoffersen (1998) and Christoffersen (2012, chapter 13).

Given a model, for example GARCH(1,1), rolling windows of a given length can be used to predict the 1-step-ahead volatility. More precisely, if the window length is 250 days, then the first 250 observations will be used to predict the volatility of day 251. When that is done the first observation will be left out and observation of day 2 to day 251 will be used to predict the volatility of day 252 and so on. These predictions can then be used to calculate a sequence of out-of-sample interval forecasts $\left\{L_{t|t-1}(p), U_{t|t-1}(p)\right\}_{t=251}^{T}$, where $L_{t|t-1}(p)$ and $U_{t|t-1}(p)$ are the lower and upper limits of the interval forecast. Where the interval forecast are given by

$$\begin{aligned}\text{standard normal distribution} &: r_t \pm z_{\alpha/2}\sigma_{t-1}(1)\\ \text{standardized student-t distribution} &: r_t \pm t_{\alpha/2}(df)\sigma_{t-1}(1)\end{aligned}$$

where $\sigma_{t-1}(1)$ is calculated according to equation (22) and (23). Given the sequence of out-of-sample interval forecasts let

$$I_t = \begin{cases} 1, & \text{if } r_t \in \left[L_{t|t-1}(p), U_{t|t-1}(p)\right] \\ 0, & \text{if } r_t \notin \left[L_{t|t-1}(p), U_{t|t-1}(p)\right] \end{cases}$$

$I_t$ can then be used to test if the actual degree of coverage is equal to the degree of coverage implied by the prediction interval. If the interval is a 95% prediction interval the hypothesis for unconditional coverage are

$$\begin{aligned} H_0 &: p = 0.95 \\ H_a &: p \neq 0.95 \end{aligned}$$

and under the null hypothesis $I_t \sim Be(T,p)$ and $\sum_{t=251}^{T} I_t \sim Bin(T,p)$.

However, if there are asymmetries in the tail probabilities, it is important to state whether the realizations that fell outside the predicted interval were in the upper or lower tail of the conditional distribution. Let $\tau_l$ and $\tau_u$ be the lower and upper tail probabilities, such that $1-p = \tau_l + \tau_u$. If the tail probabilities are symmetric $\tau_l = \tau_u = (1-p)/2$. Define

$$S_t = \begin{cases} 1, & \text{if } r_t \leq L_{t|t-1}(p) \\ 2, & \text{if } L_{t|t-1}(p) < r_t < U_{t|t-1}(p) \\ 3, & \text{if } r_t \geq U_{t|t-1}(p) \end{cases}$$

Under the null hypothesis that the actual degree of coverage is equal to the degree of coverage implied by the prediction interval, the transition matrix for $S_t$ is

$$\mathbf{\Pi}_0 = \begin{bmatrix} \tau_l & 1-\tau_l-\tau_u & \tau_u \\ \tau_l & 1-\tau_l-\tau_u & \tau_u \\ \tau_l & 1-\tau_l-\tau_u & \tau_u \end{bmatrix} \tag{24}$$

and the alternative hypothesis of independence, but incorrect coverage is

$$\mathbf{\Pi}_2 = \begin{bmatrix} \pi_l & 1-\pi_l-\pi_u & \pi_u \\ \pi_l & 1-\pi_l-\pi_u & \pi_u \\ \pi_l & 1-\pi_l-\pi_u & \pi_u \end{bmatrix} \tag{25}$$

The likelihood ratio (LR) test can then be used, where under the null hypothesis of unconditional coverage the unconditional likelihood is given by

$$L(\mathbf{\Pi}_0; S_1, S_2, \ldots, S_T) = \tau_l^{n_1}(1-\tau_l-\tau_u)^{n_2}\tau_u^{n_3},$$

where $n_i$, $i = 1, 2, 3$, is the observed number of times $r_t$ is in state $i$.

Under the alternative hypothesis the unconditional likelihood is given by

$$\begin{aligned} L(\hat{\mathbf{\Pi}}_2; S_1, S_2, \ldots, S_T) &= \hat{\pi}_l^{n_1}(1-\hat{\pi}_l-\hat{\pi}_u)^{n_2}\hat{\pi}_u^{n_3} \\ &\propto \text{Mult}(n_1, n_2, n_3, \pi_l, 1-\pi_l-\pi_u, \pi_u), \end{aligned}$$

where the maximum likelihood estimates are $\hat{\pi}_l = \frac{n_1}{n_1+n_2+n_3}$ respectively $\hat{\pi}_u = \frac{n_3}{n_1+n_2+n_3}$.

The distribution of the LR test of unconditional coverage is asymptotically $\chi^2$ with $s-1$ degrees of freedom, where $s = 3$ is the number of states

$$LR_{uc} = -2\log[L(\mathbf{\Pi}_0; S_1, S_1, \ldots, S_T)/L(\hat{\mathbf{\Pi}}_2; S_1, S_1, \ldots, S_T)] \overset{asym}{\sim} \chi^2(2).$$

### 2.2.12 Independence and conditional coverage

The unconditional coverage test in the section above does not take into account that the states 1,2 and 3 from $S_t$ could come clustered together in a time-dependent manner. Thus in the unconditional coverage test the order of the states does not matter. However, it is a well known fact that volatility cluster is common in financial time-series. Hence, to take this into account the first step is to test the independence assumption and the second step is to combine the test for unconditional coverage respectively independence and test for conditional coverage.

The null hypothesis of independence is given by $\mathbf{\Pi}_2$, see equation (25). The alternative hypothesis for first-order dependence and incorrect coverage is

$$\mathbf{\Pi}_1 = \begin{bmatrix} \pi_{ll} & 1-\pi_{ll}-\pi_{lu} & \pi_{lu} \\ \pi_{ml} & 1-\pi_{ml}-\pi_{mu} & \pi_{mu} \\ \pi_{ul} & 1-\pi_{ul}-\pi_{uu} & \pi_{uu} \end{bmatrix} \tag{26}$$

Under the alternative hypothesis the conditional likelihood is given by

$$L(\hat{\mathbf{\Pi}}_1; S_1, S_2, \ldots, S_T) = \hat{\pi}_{ll}^{n_{11}}(1-\hat{\pi}_{lm}-\hat{\pi}_{lu})^{n_{12}}\hat{\pi}_{lu}^{n_{13}} \ldots \hat{\pi}_{ul}^{n_{31}}(1-\hat{\pi}_{um}-\hat{\pi}_{uu})^{n_{32}}\hat{\pi}_{uu}^{n_{33}},$$

where $\pi_{ij}$ is the probability of moving from state $i$ to state $j$ and $n_{11}$ is the observed number of times $r_t$ moves from state $l$ to state $l$. Let $1 = l$, $2 = m$ and $3 = u$, then the maximum likelihood estimates are given by

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n_{il} + n_{im} + n_{iu}}, \text{ for } i = l, m, u \text{ and } j = l, u$$

The distribution of the LR test of independence is asymptotically $\chi^2$ with $(s-1)^2$ degrees of freedom, where $s = 3$ is the number of states

$$LR_{ind} = -2\log[L(\hat{\mathbf{\Pi}}_2; S_1, S_1, \ldots, S_T)/L(\hat{\mathbf{\Pi}}_1; S_1, S_1, \ldots, S_T)] \overset{a}{\sim} \chi^2(4)$$

This test is only for independence and it does not depend on the true coverage $p$. However, testing the null hypothesis, $\mathbf{\Pi}_0$, of the unconditional coverage, see equation (24) against the the alternative hypothesis, $\mathbf{\Pi}_1$, of the independence test, see equation (26), yields a complete test of conditional coverage.

The likelihood ratio test of conditional coverage is asymptotically Chi-squared distributed with $s(s-1)$ degrees of freedom, where $s = 3$

$$LR_{cc} = -2\log[L(\hat{\mathbf{\Pi}}_0; S_1, S_1, \ldots, S_T)/L(\hat{\mathbf{\Pi}}_1; S_1, S_1, \ldots, S_T)] \overset{a}{\sim} \chi^2(6).$$

### 2.2.13 Density forecast

The theory in this section is based on Christoffersen (2012, chapter 13), Tay & Wallis(2000) and Held & Sabanés Bovés(2014, p 309).

Density forecast can be used to backtest the entire distribution of a random variable. It is an estimate of the probability distribution of the possible future values of a random variable. Thus, if for instance the GARCH model (see section 2.2.10) can be used to predict the one step ahead volatility of the log returns. Then the one step ahead standardized residuals, $\frac{a_{t+1}}{\sigma_{t+1}}$, should be $U(0,1)$ distributed according to the probability integral transform. This can be visually checked with a histogram that should have a rectangular form or more formally checked with the one sample Kolmogorv Smirnov test, see section 2.2.14

Let $X$ be a continuous random variable with cumulative distribution function $F_X$ and define $Y = F_X(X)$. Then according to the probability integral transform $Y$ is uniformly distributed on $(0,1)$.

$$F_Y(y) = P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y,$$

thus, $Y$ is uniformly distributed on the interval $(0, 1)$.

### 2.2.14 Kolmogorov-Smirnov one sample test

The theory in this section is based on Bagdonavičius, V., Julius, K. and Nikulin, M. S.(2011, chapter 3).

The Kolmogorv-Smirnov test is a nonparametric test that is often used to test if a sample from an unknown distribution $F$ is equal to a reference distribution $F_0$.

$$H_0 : F = F_0, \quad H_a : F \neq F_0$$

Let $X_1, \ldots, X_n$ be a independently and identically distributed sample from some unknown distribution. Let $F(x) = P(X_1 \leq x)$ denote the true cumulative distribution function (c.d.f) of the true underlying distribution of the data. Define the empirical c.d.f.

$$F_n(x) = P_n(X \leq x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x),$$

where $I$ is an indicator variable, taking the value 1 if a sample point is less than or equal to the level $x$ and 0 otherwise. According to the law of large numbers

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \to E[I(X_1 \leq x)] = P(X_1 \leq x) = F(x) \quad \text{as } n \to \infty,$$

thus the largest distance between the the empirical c.d.f. and the sample c.d.f. converges to 0 in probability as the sample size, $n$, goes to infinity. The test statistics of the Kolmogorv-Smirnov test is given by

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|,$$

under the null hypothesis the test statistic $D_n$ does not depend on the reference distribution instead it only depends on the sample size $n$. The null hypothesis is rejected if $D_n$ exceeds a threshold $c$ on the significance level $\alpha$. Otherwise if $D_n$ is less than or equal to $c$, it is not rejected that $F = F_0$.

# 3 Methodology

The financial time series data used in this paper is daily closing prices, from 2006-01-31 to 2018-01-31, of the Google stock. This data is transformed into log returns, see section 5.1. To carry out the analysis on this data-set the software RStudio is used, with the time series packages "fGarch" and "rugarch". We have also implemented our own functions, for the back testing procedure. Moreover, since the true distribution of the log returns is not known, simulation from a student-$t$ distributed GARCH$(1, 1)$ process has been made with the function garchSim from the "fGarch" package. Thus, the results of the log return series can then be compared with the ones observed from the simulated data.

To model and predict volatility the ARCH and GARCH models are widely used. However, the ARCH model often requires many parameters, see section 5.3.1. Hence, this thesis will focus on the GARCH$(1, 1)$ model assuming a normal distribution respectively a student-$t$ distribution. Time series analysis is not really different from regression analysis. Hence, to get a sense of the underlying distribution of the log returns respectively the standardized residuals, the same methods that are used when dealing with regression analysis can be applied. This analysis is carried out in section 4.1 respectivel section 4.3.2.

To predict the one step ahead volatility, the back testing method in section 2.2.11 is applied using window length 250 days, 750 days respectively 1250 days. Since there is only one observation made on a trading day, the stock volatility is not directly observable. Hence, to evaluate how well the GARCH model predicts volatility, predictions intervals of the log returns are calculated, assuming a normal distribution respectively student-$t$ distribution

$$\text{standard normal distribution} : r_t \pm z_{\alpha/2}\sigma_{t-1}(1)$$
$$\text{standardized student-}t\text{ distribution} : r_t \pm t_{\alpha/2}(df)\sigma_{t-1}(1).$$

Where $\sigma_{t-1}(1)$ is calculated according to equation (22) and (23). To test how well the GARCH(1,1) model predicts over all possible percentiles, density forecasts are also constructed to back test the entire distribution.

# 4 Analysis

## 4.1 Data analysis

The financial time series data used in this paper is daily closing prices, from 2006-01-31 to 2018-01-31, of the Google stock. This data is downloaded from finance.yahoo.com. From Figure 1 it can be seen that the daily closing prices follows a positive trend. However, it is common to transform the data to simple log returns, $r_t$, see Equation (4).

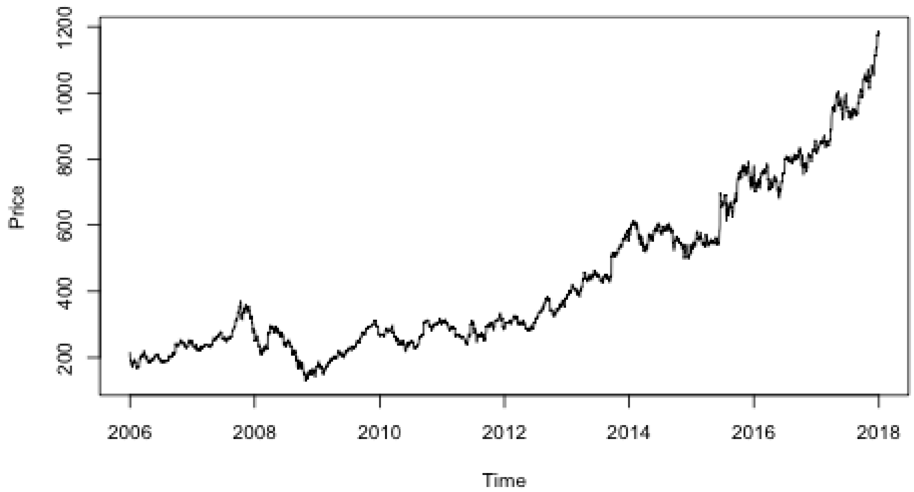



**Figure 1:** Google daily closing prices, period 2006-01-31 - 2018-01-31.

Looking at Figure 2(a) it is clear that the log returns does not contain a trend. Instead it can be seen that the fluctuation of $r_t$ varies over time and that the most volatile period is coinciding with the financial crisis (2007-2009). After the financial crisis comes a more tranquile period, this alternating pattern becomes more clear in Figure 2(b) and it is known as volatility clustering. This indicates that an conditional heteroscedastic model should be used. Hence, the data that will be analyzed throughout this paper, is the simple log returns.

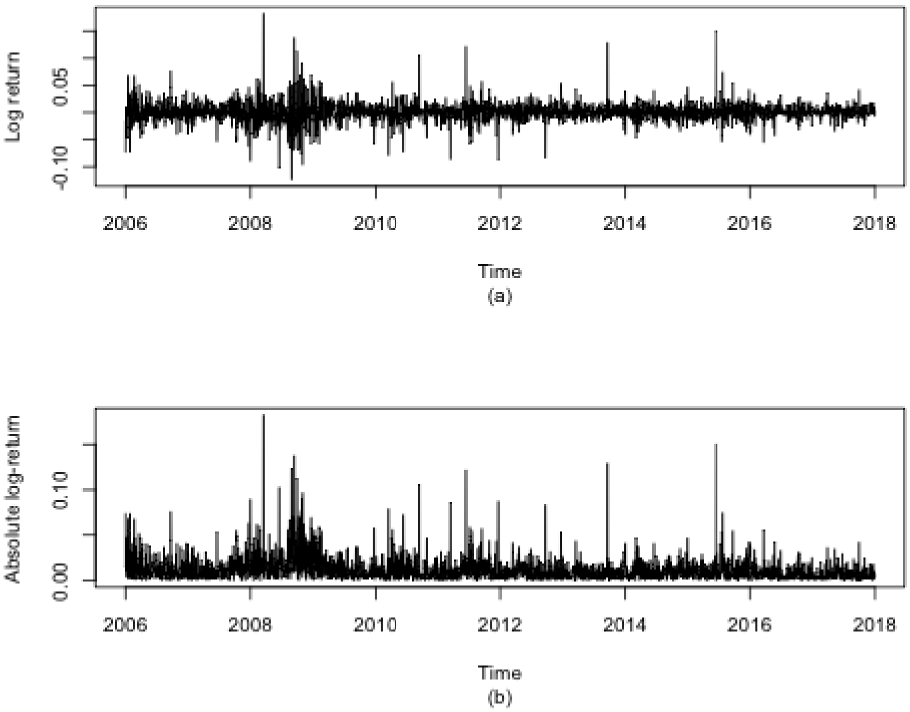



**Figure 2:** Google daily, (a) log returns and (b) absolute log returns period 2006-02-01 - 2018-01-31.

The next step is to try to understand which unconditional distribution the series $\{r_t\}$ follows and a good start is to look at some descriptive statistic. The mean and standard deviation of $r_t$ in Table 1 is equal to 0.000562 respectively 0.018162. Using a t-test it can be tested if the mean is zero or not

$$t = \frac{\bar{r}_t}{s/\sqrt{T}} = 1.70 < t_{0.025}(T-1) \approx 1.96,$$

since 1.70 < 1.96 we can't reject that the mean is zero. However, the t-test statistic assumes that the random variables are independently and identically distributed. In subsection 4.2 it will be seen that the log returns are not independently distributed.

| Mean | Standard deviation | Skewness | Kurtosis |
|---|---|---|---|
| 0.000562 | 0.018162 | 0.543235 | 14.38697 |

**Table 1:** Descriptive statistics for log returns, $r_t$.

Moreover, in subsection 2.2.3 it is explained that $r_t$ has a symmetric distribution if the skewnees lies in the interval $[-0.5, 0.5]$. And in subsection 2.2.4 it is stated that a distribution with a kurtosis greater than 3, tends to contain more extreme values, thus, the density function is likely to have heavy tails, as well as a high thin peak around its mean. In Table 1 it can be seen that the log returns has a skewness of 0.543235 which implies that the distribution might be skeewed, and $r_t$ has a quite high kurtosis, 14.38697, indicating that the distribution is leptukortic. Another way to investigate the distribution of the series $\{r_t\}$, is by plotting a histogram and a QQ-plot.

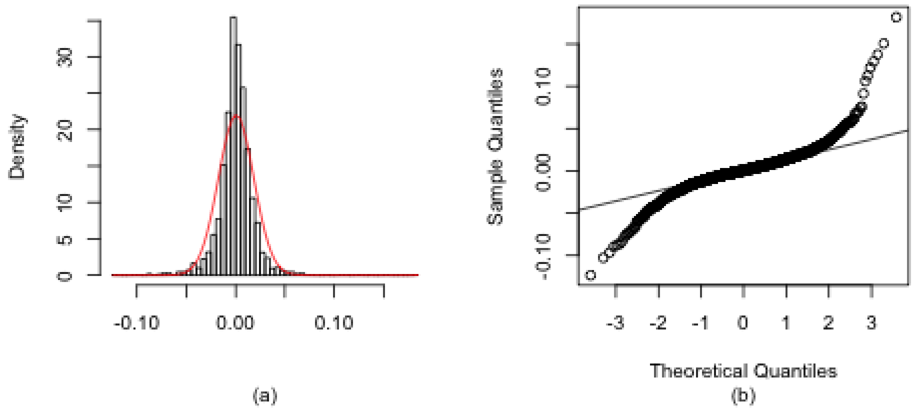



**Figure 3:** (a) histogram with normal density curve, (b) normal QQ-plot, of daily Google log returns

In Figure 3 (a) it can be seen that the histogram of $r_t$ deviates quite a lot from the normal density curve. The tails of the QQ-plot in Figure 3 (b), deviates a lot from the straight line. This implies that the normal distribution does not capture the extreme observations of the log returns. Hence, it is a good idea to check if the student-$t$ distribution has a better fit, since this distribution is known to have heavier tails.

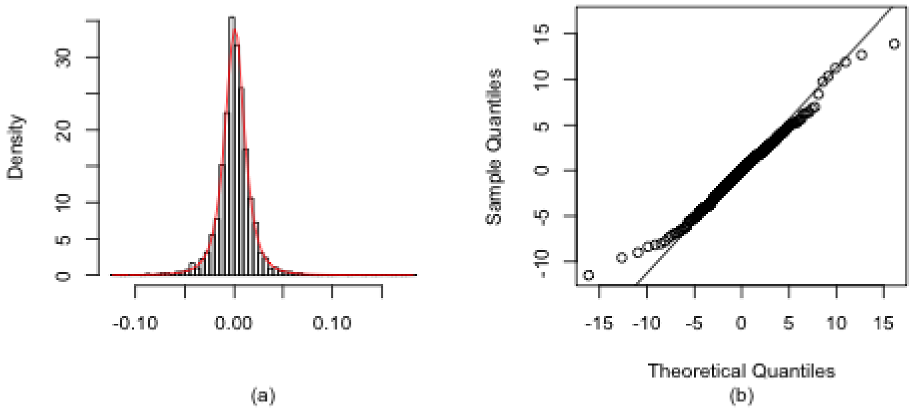



**Figure 4:** (a) histogram with student-$t$ density curve, (b) student-$t$ QQ-plot, of daily Google log returns with four degrees of freedom

From Table 1 we know that the log returns has a very high kurtosis, indicating that the data might be leptukortic. The histogram of $r_t$ has a high thin peak around its mean as well as heavy tails, this confirms what the high kurtosis has already implicated. That is, that the normal distribution is not a good fit. However, the density curve of of the student-$t$ distribution in Figure 4 (a) seems to fit well to the histogram of $r_t$. Sadly enough, the QQ-plot in Figure 4 (b) shows that the tails deviates from the straight line. Still, the conclusion will be that the student-$t$ distribution has a better fit to the data compared to a normal distribution. It should also be noted that the analysis made in this section applies to the unconditional distribution of the log returns.

## 4.2 Model identification

In subsection 2.2.8 it is stated that volatility is not directly observable. However, it has some characteristic that can be captured. For example, volatility is often high for certain periods and low for other periods, known as volatility clustering. Conditional heteroscedastic models, are often used to try and model this kind of behavioring. From subsection 4.1 we know that volatility clustering seems to exist for the Google daily log returns. The next step in our path to identify a model, is to check if the series $\{r_t\}$ is dependent, but serially uncorrelated, or at least has minor lower order serial correlations. This is often visually done with autocorrelation plots, of different functions of the series $\{r_t\}$. The ACF plot of the log returns, shows if there exists any serial correlation of different lag length, between $r_t$ and its previous values. While the ones for the squared respectively absolute-value of the log returns, reveals if there exists any non-linear dependency in the time series. The partial autocorrelation function plot, of the squared log returns, reveals if the series is serially independent or not. If not, the serie is said to contain ARCH effects.

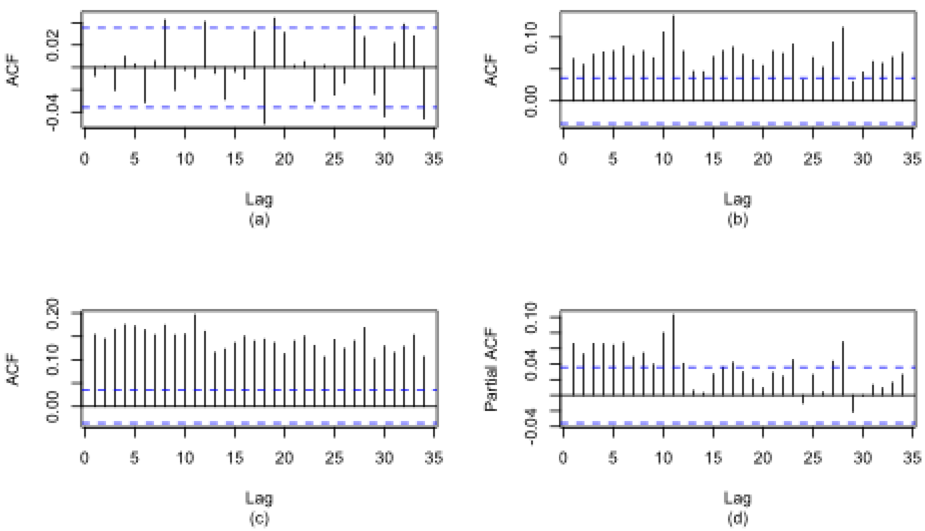



**Figure 5:** (a) ACF of log returns, (b) ACF of squared log returns, (c) ACF of absolute log returns, (d) PACF of squared log returns

It is clear from Figure 5 (a) that the daily log returns of the Google stock exhibit no significant autocorrelation, supporting the hypothesis that the returns of a financial asset are uncorrelated across time. However, in Figure 5 (b),(c) there are significant autocorrelations in the squared log returns and more in the absolute log returns, this confirm the presence of volatility clustering. The absolute log returns exhibit significant and persistent autocorrelations, a characteristic known as long memory process. Looking at the PACF of the squared log returns (Figure 5 (d)), it is seen that there are several big spikes. This means that the series is serially dependent, which indicates that it contains autoregressive conditional heteroscedastic effects. Moreover, the conclusions drawn from the ACF and PACF plots in Figure 5 also applies to the data simulated from a student-$t$ distributed GARCH$(1,1)$ process.

### 4.3 Fitting a model

To summarize what has been established so far, the Google daily log returns seems to be serially uncorrelated, but they have a non-linear dependency. Which is a sign of volatility clustering. We have also concluded that the mean is approximately zero, see 1 and from the ACF plot of the log returns in Figure 5 there are no signs of a trend. Hence, we do not need to specify a mean Equation for the conditional heteroscedastic models, when trying to model the volatility.

### 4.3.1 ARCH($m$) models

To decide how many lags, ($m$), that should be used in the ARCH model, the PACF plot in Figure 5 (D) can be used. This plot indicates that lags up to order 12 of the squared innovations should be included in the model. As noted the ARCH model requires many parameters to describe the volatility of the log returns. Hence, it is be better to try and fit a simpler model such as the genrealized autoregressive conditional heteroscedastic, GARCH$(1,1)$, model.

### 4.3.2 GARCH$(1,1)$

In this section we will fit a GARCH$(1,1)$ model, using the normal distribution as well as the student-$t$ distribution. Due to previous result it is assumed that $\mu_t = 0$, hence, $r_t = a_t$. Thus, the GARCH(1,1) model can be written on the form

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{27}$$

Dividing the innovations $a_t$ with the standard deviation $\sigma_t$ yields the standardized residuals, see equation (19). They form a sequence of independently and identically distributed random variables. Hence, the series $\{\tilde{a}_t\}$ can be used to check if the model is correctly specified. In Figure 6 it can be seen that the standardized residuals does not satisfy the normality assumption, when fitting a GARCH(1,1) model, using the normal distribution. Rather it seems as the student-$t$ distribution has a better fit. However, as can be seen in the right plot in Figure 6, the right tail of the standardized residuals, deviates quite a lot from the theoretical line. Still, the conclusion will be that the student-$t$ distribution seems to be the most appropriate choice.

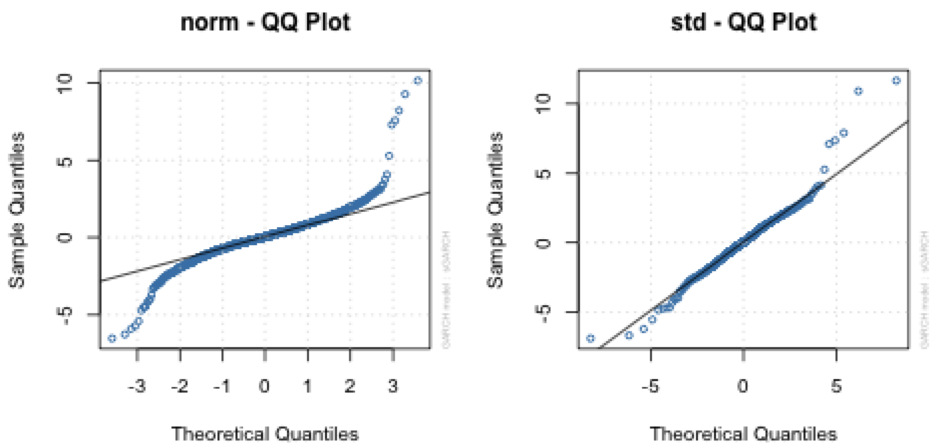



**Figure 6:** QQ-plots of the standardized residuals for the normal distribution respectively the student-$t$ distribution with 3.94 degrees of freedom

A plot of the squared standardized residuals can be used to check if they are a white noise process. Looking at Figure 7, this seems to be the case, since there are no significant spikes. This is further confirmed by the Ljung-Box test (see equation (10)), which states that we can't reject at the 5% significance level that the autocorrelations are jointly zero.

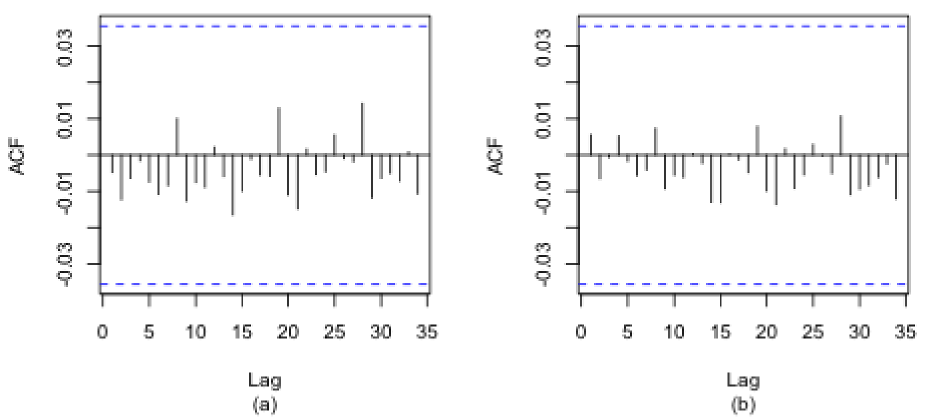



**Figure 7:** ACF plots for the squared standardized residuals, for (a) the normal distribution and (b) the student-$t$ distribution

Moreover, it is also assumed that the standardized residuals has a mean equal to zero and a constant variance. According to Figure 8 the mean seems to be zero, it looks like the variance is constant and the observations seems to be spread out randomly. Hence, the standardized residuals are a white noise process, with mean zero and variance 1, when both fitting a GARCH$(1,1)$ model using the normal distribution as well as the student-$t$ distribution.

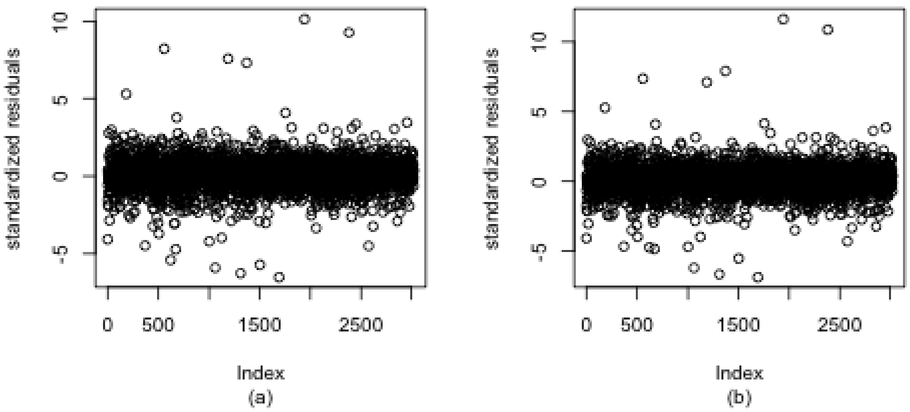



**Figure 8:** Plots of the standardized residuals, for (a) the normal distribution and (b) the student-$t$ distribution

Table 2 shows the parameter estimates for the GARCH(1,1) model using both the normal distribution as well as the student-$t$ distribution. Both gives similar estimates, however, the intercept, $\hat{\alpha}_0$, is significantly different from zero when using the normal distribution but insignificant when using the student-$t$ distribution. Still we will keep the intercept in the GARCH(1,1) model when using the student-$t$ distribution. What's more interesting is that $\hat{\beta}_1$ is quite high for both distributions (0.918985 n-distribution, 0.965305 t-distribution) while $\hat{\alpha}_1$ is quite low (0.058733 n-distribution, 0.028865 t-distribution). This can be interpreted as that the squared volatility at time t depends heavily on the past squared volatility from time period t-1 but it has a quite low dependency with the previous shock at time period t-1. Moreover, the sum of $\hat{\alpha}_1$ and $\hat{\beta}_1$ is 0.977718 using the normal distribution and 0.99417 using the student-$t$ distribution. Both of them are less then 1 which satisfies the criteria stated in section 2.2.10. Thus the conclusion will be that the GARCH(1,1) model seems to be the appropriate choice.

| | Normal | | Student-$t$ | |
|---|---|---|---|---|
| | **Estimate** | **P-value** | **Estimate** | **P-value** |
| $\alpha_0$ | 0.000007<br>(0.000002) | 0.00000 | 0.000002<br>(0.000001) | 0.11135 |
| $\alpha_1$ | 0.058733<br>(0.008690) | 0.00000 | 0.028865<br>(0.008678) | 0.000880 |
| $\beta_1$ | 0.918985<br>(0.011004) | 0.00000 | 0.965305<br>(0.005519) | 0.00000 |

Note: Std. Error are given within parenthesis

**Table 2:** Parameterestimates for GARCH(1,1), to the left normal distribution and to the right student-$t$ distribution. Data used is the Google stock daily log returns.

# 5 Results

## 5.1 Back testing

To evaluate how well the GARCH$(1, 1)$ model predicts the one step ahead volatility of the Google stock log-returns, back testing with rolling window of length 250, 750 and 1250 days has been constructed for both the normal distribution as well as the student-t distribution. Figure 9, 10 and 11 shows 95 % prediction intervals using all of the rolling windows length. Not surprisingly it can be seen that the intervals assuming a student-t distribution is much more wider than the ones assuming a normal distribution. This result is expected since the student-t distribution is known to have heavier tails, thus it will capture more of the extreme observations. And from Table 1 in section 5.1 we know that the log-returns have a quite high kurtosis indicating that the distribution is likely to have heavy tails. Moreover, regardless of the distribution assumption, the intervals are having trouble to capture the extreme and sudden change of the log-return when using windows longer than 250 days. Indicating that using more recent data makes the GARCH$(1, 1)$ model more sensitive to changes of the log-returns. This is also quite expected, since we only are interested in prediction of the one step ahead volatility. Prediction intervals are also constructed for the simulated data. As can be seen from Figure 14, 15 and 16 in Appendix C, they show a similar pattern as the ones using the Google stock daily log return.

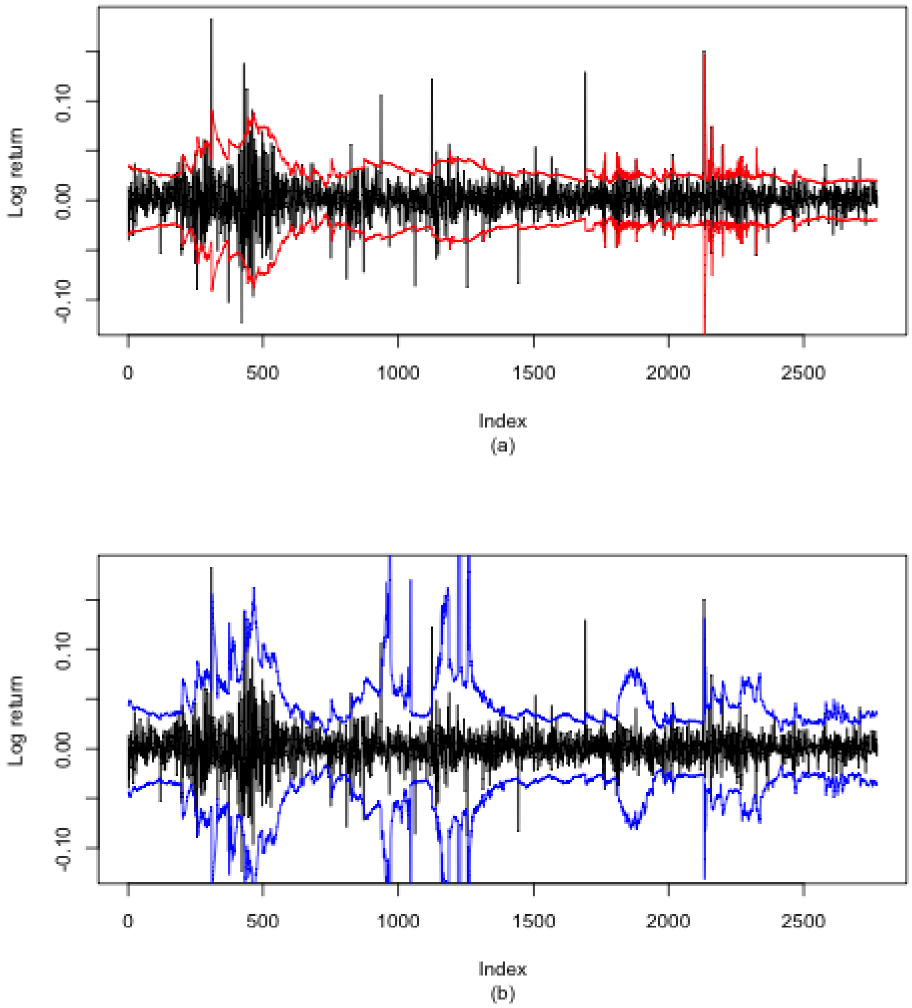



**Figure 9:** Log-returns with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 250 days,
(b) GARCH(1,1) student-t distribution, rolling windows length 250 days

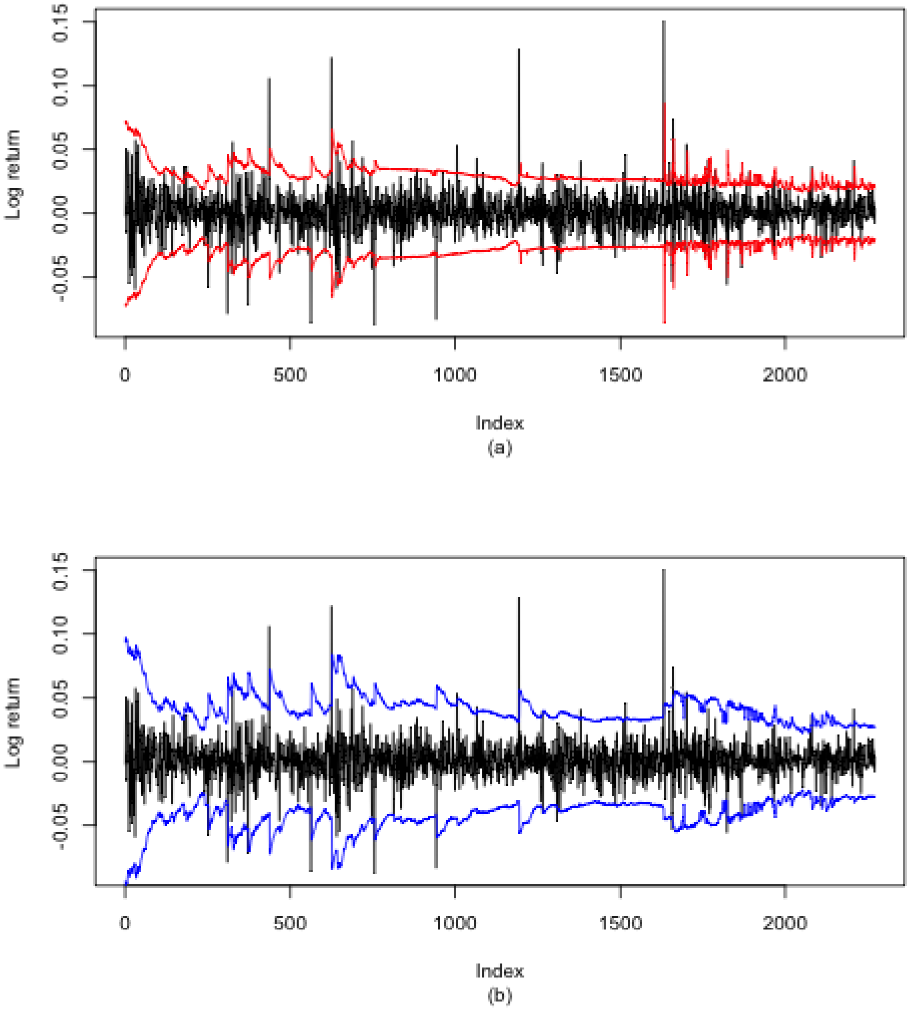



**Figure 10:** Log-returns with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 750 days,
(b) GARCH(1,1) student-t distribution, rolling windows length 750 days

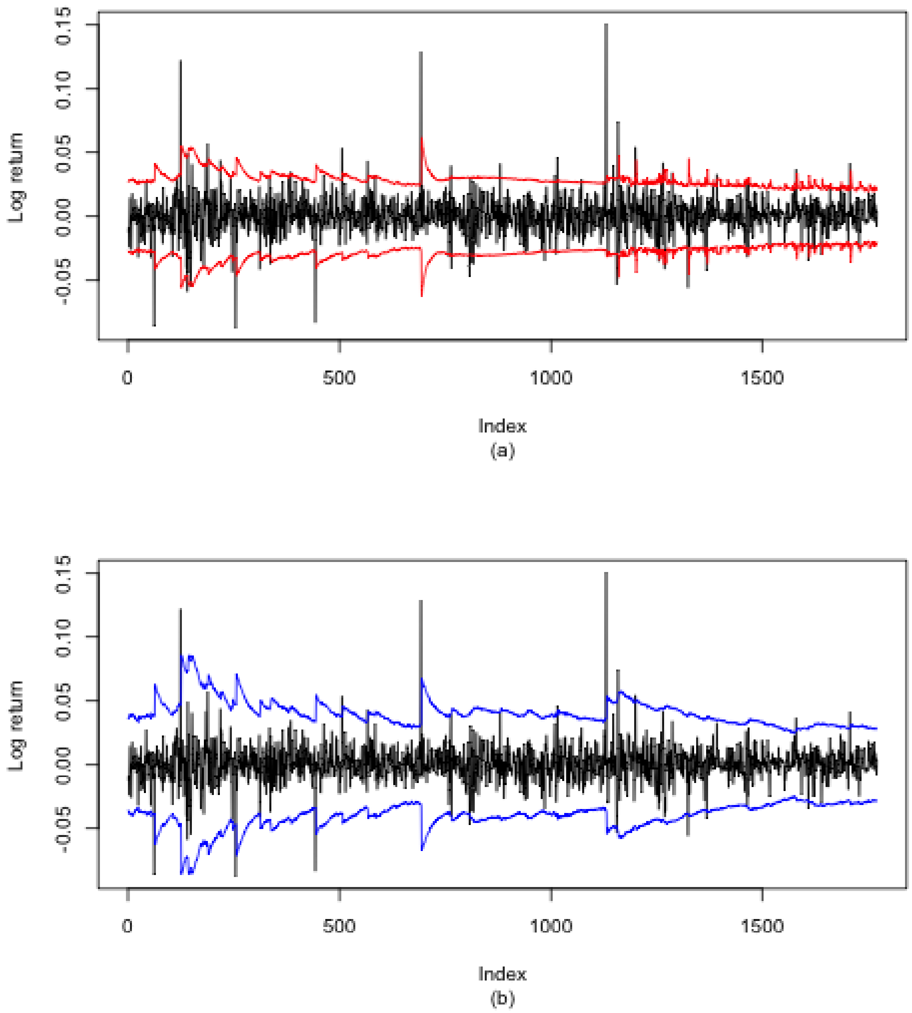



**Figure 11:** Log-returns with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 1250 days,
(b) GARCH(1,1) student-t distribution, rolling windows length 1250 days

### 5.1.1 Unconditional coverage

From Figure 9 we know that the prediction intervals assuming the student-t distribution is much wider than the ones assuming a normal distribution. Hence, it is not a surprise that the estimated coverage $1-\hat{\pi}_l-\hat{\pi}_u$ (see section 3.2.11) in Table 3, is higher for the student-$t$ distribution compared to the normal distribution, regardless of the window length. It seems as when assuming a student-$t$ distribution the coverage is constantly over estimated, that is, all of them are greater than 95%. However, when assuming a normal distribution the coverage is under estimated using windows of length 250 days, almost 95% for window length 750 days and over estimated when using window length 1250 days.

| distribution | window length | $\hat{\tau}_l$ | $1 - \hat{\tau}_l - \hat{\tau}_u$ | $\hat{\tau}_u$ |
|---|---|---|---|---|
| normal | 250 | 0.03211837 | 0.9400938 | 0.0277878 |
| student-t | 250 | 0.0111873 | 0.9790689 | 0.009743775 |
| normal | 750 | 0.02421841 | 0.9559665 | 0.01981506 |
| student-t | 750 | 0.006164685 | 0.9876706 | 0.006164685 |
| normal | 1250 | 0.01976285 | 0.9621683 | 0.01806889 |
| student-t | 1250 | 0.005646527 | 0.9875776 | 0.006775833 |

**Table 3:** Estimated probabilities, $\tau_l$ is the lower tail probability and $\tau_l$ is the upper tail probability

The coverage of the intervals in the section 6.1 is expected to be 95%, due to the fact that the predicted intervals are calculated with a lower quantile equal to 2.5% and upper quantile equal to 97.5% . To test this at the 5%-level the LR-test of unconditional coverage from section 3.2.11 can be used. From Table 4 that shows the results from the unconditional coverage test it can be seen that the null hypothesis of correct unconditional coverage is only accepted at the 5% significance level for window length 750 days assuming a normal distribution. Otherwise, the null hypothesis of correct unconditional coverage is rejected.

| distribution | window length | test | test-statistic | $\chi^2_{0.05}$ |
|---|---|---|---|---|
| normal | 250 | unconditional coverage | 6.266627 | 5.991465 |
| student-t | 250 | unconditional coverage | 62.80462 | 5.991465 |
| normal | 750 | unconditional coverage | 2.771981 | 5.991465 |
| student-t | 750 | unconditional coverage | 96.0462 | 5.991465 |
| normal | 1250 | unconditional coverage | 6.140014 | 5.991465 |
| student-t | 1250 | unconditional coverage | 74.60935 | 5.991465 |

**Table 4:** Test result for unconditional coverage of the log returns

### 5.1.2 Independence

The unconditional coverage test does not take into account that the observations, $S = \{1, 2, 3\}$ might come clustered together. However, the independence test stated in section 3.2.12 takes this into account, that is, it will test if the probability of being in the observed state at time period $t$ is dependent on the state observed at time period $t - 1$. In Table 5 that shows the result of the independence test, it can be seen that the null hypothesis of independence can only be rejected at the 5% level, for the student-$t$ distribution, with window length 250 days. However, this does not necessarily mean that there exist some dependence with the observation made in time period $t-1$. Rather, it is most likely due to too few observations.

| distribution | window length | test | test-statistic | $\chi^2_{0.05}$ |
|---|---|---|---|---|
| normal | 250 | independence | 7.796351 | 9.487729 |
| student-t | 250 | independence | 11.14131 | 9.487729 |
| normal | 750 | independence | 2.455526 | 9.487729 |
| student-t | 750 | independence | 3.719931 | 9.487729 |
| normal | 1250 | independence | 2.533295 | 9.487729 |
| student-t | 1250 | independence | 4.100778 | 9.487729 |

**Table 5:** Test result for independence of the log returns

### 5.1.3 Conditional coverage

The next step is to jointly test for independence and correct coverage with the conditional coverage test stated in section 3.2.12. The null hypothesis of the unconditional coverage test is tested against the alternative of the independence test. Due to the fact that regardless of the window length, the student-t distribution always overestimates the unconditional coverage, it is not a surprise that they do not have a correct conditional coverage, as can be seen in Table 6. Regarding the normal distribution the test of complete coverage can only be rejected at 5% significance level, when using window length 250 days. For longer windows assuming a normal distribution, the null hypothesis of conditional coverage can not be rejected.

| distribution | window length | test | test-statistic | $\chi^2_{0.05}$ |
|---|---|---|---|---|
| normal | 250 | conditional coverage | 14.06298 | 12.59159 |
| student-t | 250 | conditional coverage | 73.94593 | 12.59159 |
| normal | 750 | conditional coverage | 5.227506 | 12.59159 |
| student-t | 750 | conditional coverage | 99.76613 | 12.59159 |
| normal | 1250 | conditional coverage | 8.673309 | 12.59159 |
| student-t | 1250 | conditional coverage | 78.71013 | 12.59159 |

**Table 6:** Test result for conditional coverage of the log returns

### 5.1.4 Density forecast

To test how well the GARCH(1, 1) model predicts over all possible percentiles the density forecast method explained in section 2.2.13 is used. If the model predictions are correct then the cumulative distribution function of the one step ahead standardized residuals should be $U(0, 1)$ distributed. When assuming a normal distribution with window length 250 days, the histogram of the standardized residuals in Figure 12 does not have a rectangular form as it should have. Instead it has to many observation in the middle, indicating that the normality assumption does not hold. The same results holds for longer windows, as can be seen from the histograms in Appendix D, see Figure 23 and 24. To formally test these results the one sample Kolmogorv Smirnov test from section 3.2.14 is applied. The result from this test states that the cumulative distribution function of the one step ahead standardized residuals are not $U(0,1)$ distributed. These results also holds for the simulated data as can be seen from the histograms in Appendix C, Figure 17, 19 and 20.

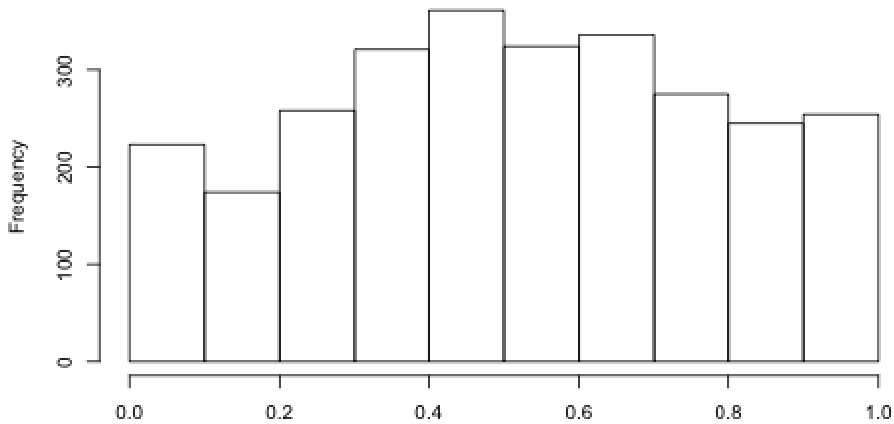



**Figure 12:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 250 days. Data used is the Google stock daily log returns.

Moreover, when assuming a student-t distribution with window length 250 days, the histogram of the standardized residuals seems to be quite rectangular, except for the high bar to the right, as can be seen in Figure 13. When using windows longer then 250 days the bars seems to be higher near the value 1, as can be seen from the histograms in Appendix D, see Figure 25 and 26. When applying the Kolmogorov Smirnov test, the null hypothesis that the standardized residuals are $U(0,1)$ distributed, is only accepted for window length 250 days. For longer windows the null hypothesis is rejected. However, when using the simulated data the null hypothesis of the uniform distribution is always accepted, regardless of the window length. The interested reader can see these histograms in Appendix C. This result is expected since we know that the underlying distribution is the student-t distribution, since the data is simulated from a student-t distributed GARCH$(1,1)$ process.

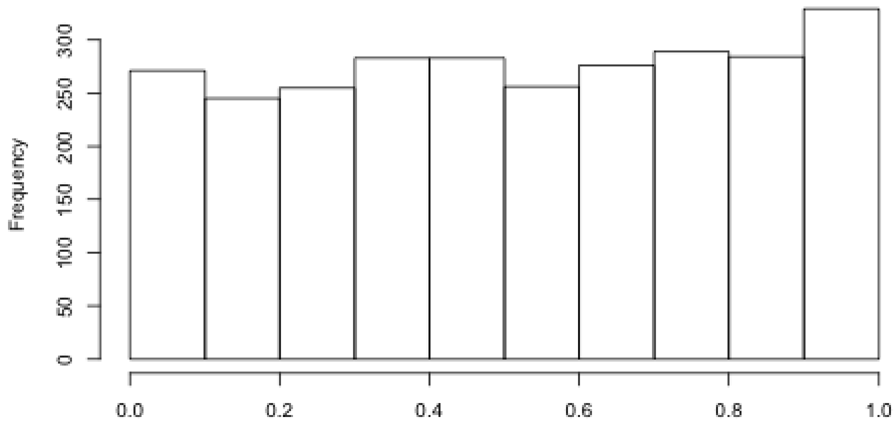



**Figure 13:** Histogram of the standardized residuals, for GARCH(1,1) student-t distribution using rolling window length of 250 days. Data used is the Google stock daily log returns.

# 6 Discussion and conclusion

In section 5.1 it is concluded that the Google stock daily log returns seems to follow a student-$t$ distribution, see Figure 4, and that the financial time series data of the log returns are containing ARCH effects, as can be seen in Figure 5(d). These results where expected and are agreeing with the theory from Tsay(2010). When fitting a GARCH(1,1) model to the log returns series it is revealed that the squared volatility seems to depend heavily on the past squared volatility observed at time $t-1$ but not so much on the previous shock observed at time $t-1$, as can be seen in Table 2. It is also concluded from Figure 6 that the standardized residuals from the GARCH$(1,1)$ model seems to follow a student-$t$ distribution. However, it should be noted that the right tail deviates from the straight line, indicating that the distribution of the standardized residuals might be skewed. Hence, a skewed student-$t$ distribution might have a better fit.

Moreover, the interval forecasts in section 6.1 reveals that the intervals assuming a student-$t$ distribution is much wider than the ones assuming a normal distribution, which is not a surprise since the student-$t$ distribution has heavier tails compared to the normal distribution. It can also be concluded from Figure 10 and Figure 11, that regardless of the distribution assumption, the GARCH$(1,1)$ model has trouble to capture the sudden changes of the log returns, when using a window length greater than 250 days. This indicates that it might be better to use more recent data when trying to forecast the one step ahead volatility. The unconditional coverage test in section 6.1.1 states that the null hypothesis of unconditional coverage, $p = 0.95$, can be rejected regardless of the window length when using a student-$t$ distribution. But when assuming a normal distribution the null hypothesis of unconditional coverage, $p = 0.95$, is rejected for window length of 250 days respectively 1250 days. But when using windows length 750 days the null cannot be rejected.

Regarding the independence test explained in section 2.2.12 and applied in section 5.1.2, it should be noted that it only test for dependency in one step. Hence, this test does not consider whether the observations outside of the interval comes in clusters for a longer period. Another problem is that there might be too few observations outside the interval. To get around this problem, Christoffersen(2012, page 306) reefers to the Monte Carlo method, which can be used to simulate p-values. Our result from the independence test, see Table 5, states that the null hypothesis of one step independence can only be rejected for the student-$t$ distribution using window length 250 days. This result does not necessarily mean that there exist a dependence with the observation made the day before. Rather, it could be due to the fact that we have too few observations outside the interval. Thus, for further research it is recommended to simulate p-values.

The results from the conditional coverage test in section 5.1.3 states that the null hypothesis of correct conditional coverage, $p = 0.95$, is rejected regardless of the window length, when assuming a student-$t$ distribution, see Table 6. This indicates that the intervals assuming a student-$t$ distribution are too cautious. These results are in line with the ones Christoffersen(1998) gets when the true coverage probability is, $p = 0.95$. When we assume a normal distribution, the null hypothesis is only rejected for window length 250 days, otherwise the null hypothesis of correct conditional coverage can't be rejected, see Table 6. These results implies that assuming a normal distribution with window length greater than 250 days, yields a correct conditional coverage for $p = 0.95$. Which indicates that intervals that are less sensitive to changes of the log returns, yields a more precise conditional coverage for $p = 0.95$. However, it should be noted that this does not mean that assuming a normal distribution using windows length 750 days respectively 1250 days, will always yield good interval forecasts. For instance, we have only applied intervals with coverage probability, $p = 0.95$. Thus, we could get different results using different values of $p$. Christoffersen(1998) who were the one first suggesting the conditional coverage test, let the actual coverage probability, p, vary between 0.50 and 0.95. He concluded that interval forecasts from the student-$t$ distributed GARCH$(1, 1)$ model where to cautious for $p = 0.95$ but when the actual coverage probability, $p$, where less than 0.95 the GARCH model provided good interval forecasts.

Moreover, to test how well the GARCH$(1, 1)$ model predicts over all possible percentiles, we use the density forecast method. The one sample Kolmogorov Smirnov test in section 5.1.4, states that when assuming a normal distribution the one step ahead standardized residuals are not $U(0, 1)$ distributed, regardless of the window length. And the histogram in Figure 12 shows that there is too many observations in the middle. When assuming a student-$t$ distribution the one sample Kolmogorov Smirnov test states that the null hypothesis of uniformly distributed one step ahead standardized residuals is only accepted using rolling windows of length 250 days, otherwise it is rejected. When using windows longer than 250 days the histograms in Figure 25 and 26 shows that there seems to be more observations near the value one, indicating that the underlying distribution might be skewed. This result in combination with the outliers in the right tail of the QQ-plot, see Figure 6, implies that it might be better to assume a skewed student-$t$ distribution. It should be noted that these plots indicates that there seems to exist outliers of positive returns.

Since the results from the interval forecasts respectively density forecasts, contradicts each other. For further research it is suggested to calculate interval forecast, letting the actual coverage probability, p, for instance vary between 0.8 and 0.95. It is also suggested that in addition to the normal and student-$t$ distribution also include the skewed student-$t$ distribution.

# 7 References


Bagdonavičius, V., Julius, K. and Nikulin, M. S., (2011). *Goodness-of-fit Tests Based on Empirical Processes, in Non-parametric Tests for Complete Data.* Hoboken, NJ, USA: John Wiley Sons, Inc.

Christoffersen, P.F., (1998). Evaluating Interval Forecasts. *International Economic Review*, 841–862.

Christoffersen, P.F., (2012). *Elements of financial risk management. 2 edn.* 225 Wyman Street, Waltham, MA 02451, USA: Elsevier.

Gut, L. (2009). *An Intermediate Course in Probability.* Springer-Verlag New York

Held, L., & Sabanés Bové, D. (2014). *Applied statistical inference.* Berlin Heidelberg: Springer.

Li, C. (2007). *On estimation of GARCH models with an application to Nordea stock prices*, Uppsala University.

Tay, A. S., Wallis, K. F. (2000). Density forecasting: a survey. *Journal of Forecasting*, 19(4), 235.

Tsay, R.S., (2010). *Analysis of Financial Time Series. 3 edn.* Hoboken, NJ, USA: John Wiley Sons, Inc.

# A Appendix

## A.1 Normal distribution

The theory in this section is from Gut(2009, page 292)

Let X be a normally distributed random variable, with mean $\mu$ and variance $\sigma^2$. Then the density function of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma_t^2}\right\}, \quad x \in (-\infty, \infty).$$

## A.2 Student-$t$ distribution

The theory in this section is from Tsay(2010, page 121)

Let X be a student-$t$ distributed random variable with $\nu$ degrees of freedom

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

## A.3 Chi-square distribution

The theory in this section is from Gut(2009, page 291)

Let X be a chi-square distributed random variable, with degrees of freedom $\nu$, mean $\nu$ and variance $2\nu$. Then the density function of X is

$$f(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{1}{2}\nu-1} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} e^{-\frac{x}{2}}, \quad x > 0,\, \nu = 1, 2, 3, \ldots$$

# B Appendix

## B.1 Maximum Likelihood (ML)

The theory in this section is from Held and Sabane's Bove'(2014, section 2.1.1, section 5.1 and appendix C.1.3) and Li(2007, section 3.2)

Let $X = (x_1, \dots, x_n)$ be a vector, containing random variables, assumed to be independently and identically distributed, with probability mass or density function $f(x_1, \dots, x_n; \theta)$. Where $\theta$ is an unknown parameter-vector. The maximum likelihood estimate $\hat{\theta}_{ML}$ is the most plausible estimate of $\theta$

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta).$$

Where $L(\theta)$ is the likelihood function

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

due to computational convenience the log-likelihood is frequently used

$$\log L(\theta) = l(\theta) = \sum_{i=1}^{n} f(x_i; \theta).$$

The Score vector is the gradient of the log-likelihood

$$S(\theta) = \nabla l(\theta) = \frac{\partial l_t(\theta)}{\partial \theta}$$

and the expected Fisher information matrix

$$J(\theta) = E\left[-\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T}\right].$$

The Newton-Raphson (NR) Method is then applied to numerically maximize the log-likelihood function. This method is frequently utilized to find the root/roots of a equation. In every iteration t of the NR method, the derivative of $\theta$ is approximated using a Taylor expansion around the current approximation. Let $\theta^{(t)}$ denote the parameter-vector after the t:th iteration. The next iteration is then given by

$$\theta^{(t+1)} = \theta^{(t)} + J(\theta)^{-1} S(\theta) \tag{28}$$

### B.1.1 GARCH$(m, s)$ assuming normal distribution

The theory in this section is from Li(2007, section 3.2) and Tsay(2010, section 3.4.3)

Let $a_t = \epsilon_t \sigma_t$ and assuming that the log returns are conditionally normally distributed, we have that $\epsilon_t \sim N(0,1)$. Hence, $a_t/\sigma_t \sim N(0,1)$ and $(a_t/\sigma_t)^2 \sim \chi^2(1)$. Then the likelihood function is given by

$$L(\theta) = \prod_{t=m+1}^{T} f(a_t|\mathcal{F}_{t-1},\theta) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{a_t^2}{2\sigma_t^2}\right\}$$

and the log-likelihood

$$\begin{aligned} l_t(\theta) &= \sum_{t=m+1}^{T} \log f(a_t|\mathcal{F}_{t-1},\theta) \\ &= \sum_{t=m+1}^{T} \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{a_t^2}{2\sigma_t^2}\right\}\right) \\ &= \sum_{t=m+1}^{T} \left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_t^2 - \frac{a_t^2}{2\sigma_t^2}\right). \end{aligned}$$

Where $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_s)$. Thus,

$$\frac{\partial l_t(\theta)}{\partial\theta} = \left(-\frac{1}{2\sigma_t^2} + \frac{a_t^2}{2(\sigma_t^2)^2}\right)\frac{\partial\sigma_t^2}{\partial\theta} = \left(\frac{a_t^2}{2(\sigma_t^2)^2} - \frac{1}{2\sigma_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta}$$

and

$$\frac{\partial^2 l_t(\theta)}{\partial\theta} = \left(\frac{a_t^2}{2(\sigma_t^2)^2} - \frac{1}{2\sigma_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T} + \left(\frac{1}{2(\sigma_t^2)^2} - \frac{a_t^2}{(\sigma_t^2)^3}\right)\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T},$$

where $\frac{\partial\sigma_t^2}{\partial\theta} = (1, a_{t-1}^2, \ldots, a_{t-m}^2, \sigma_{t-1}^2, \ldots, \sigma_{t-s}^2)$. Hence, the Score vector

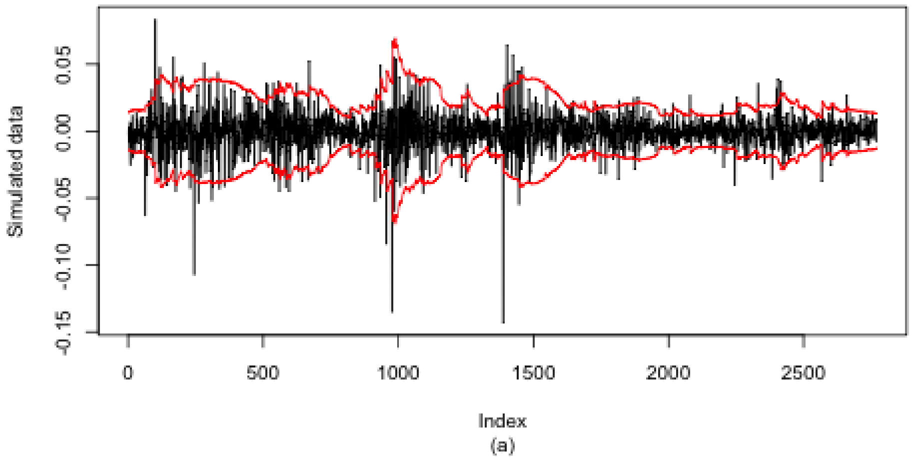
$$S(\theta) = \frac{1}{2}\sum_{t=m+1}^{T}\left(\frac{a_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta}$$

and the Fisher information matrix

$$
\begin{aligned}
J(\theta) &= E\left[-\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T}\right] \\
&= \sum_{t=m+1}^{T} E\left[-\left(\frac{a_t^2}{2(\sigma_t^2)^2} - \frac{1}{2\sigma_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T} - \left(\frac{1}{2(\sigma_t^2)^2} - \frac{a_t^2}{(\sigma_t^2)^3}\right)\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T}\right] \\
&= \frac{1}{2}\sum_{t=m+1}^{T} E\left[\left(\frac{1}{\sigma_t^2} - \frac{1}{(\sigma_t^2)^2} + \frac{2a_t^2}{(\sigma_t^2)^3} - \frac{a_t^2}{(\sigma_t^2)^2}\right)\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T}\right] \\
&= \frac{1}{2}\sum_{t=m+1}^{T} E\left[\left(\frac{\sigma_t^2}{(\sigma_t^2)^2} - \frac{1}{(\sigma_t^2)^2} + \frac{2a_t^2}{\sigma_t^2(\sigma_t^2)^2} - \frac{a_t^2\sigma_t^2}{\sigma_t^2(\sigma_t^2)^2}\right)\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T}\right] \\
&= \left\{\text{utilizes that } \frac{a_t^2}{\sigma_t^2} \sim \chi^2(1)\text{, with mean } 1\right\} \\
&= \frac{1}{2}\sum_{t=m+1}^{T} E\left[\left(-\frac{1}{(\sigma_t^2)^2} + \frac{2}{(\sigma_t^2)^2}\right)\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T}\right] \\
&= \frac{1}{2}\sum_{t=m+1}^{T} E\left[\frac{1}{(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta\partial\theta^T}\right]
\end{aligned}
$$

The next step is to apply the Newton-Raphson method (see equation 28) to numerically decide the maximum likelihood estimates.

# C Appendix



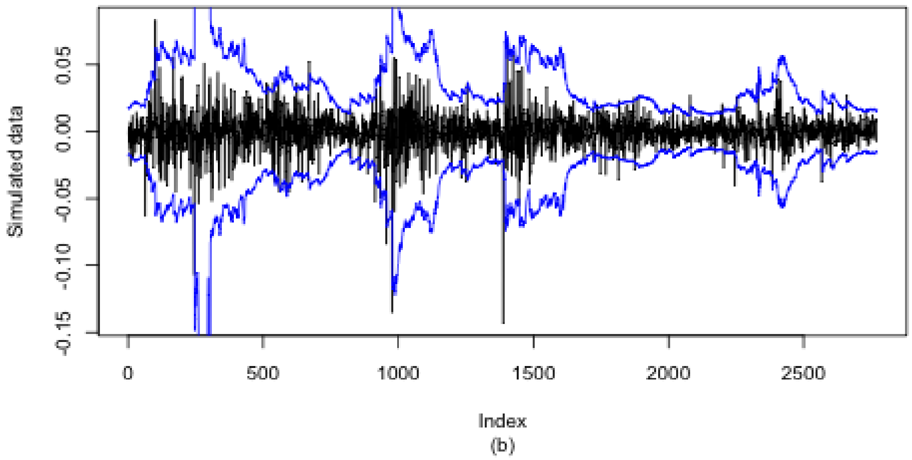




**Figure 14:** Simulated data with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 250 days,
(b) GARCH(1,1) student-$t$ distribution, rolling windows length 250 days

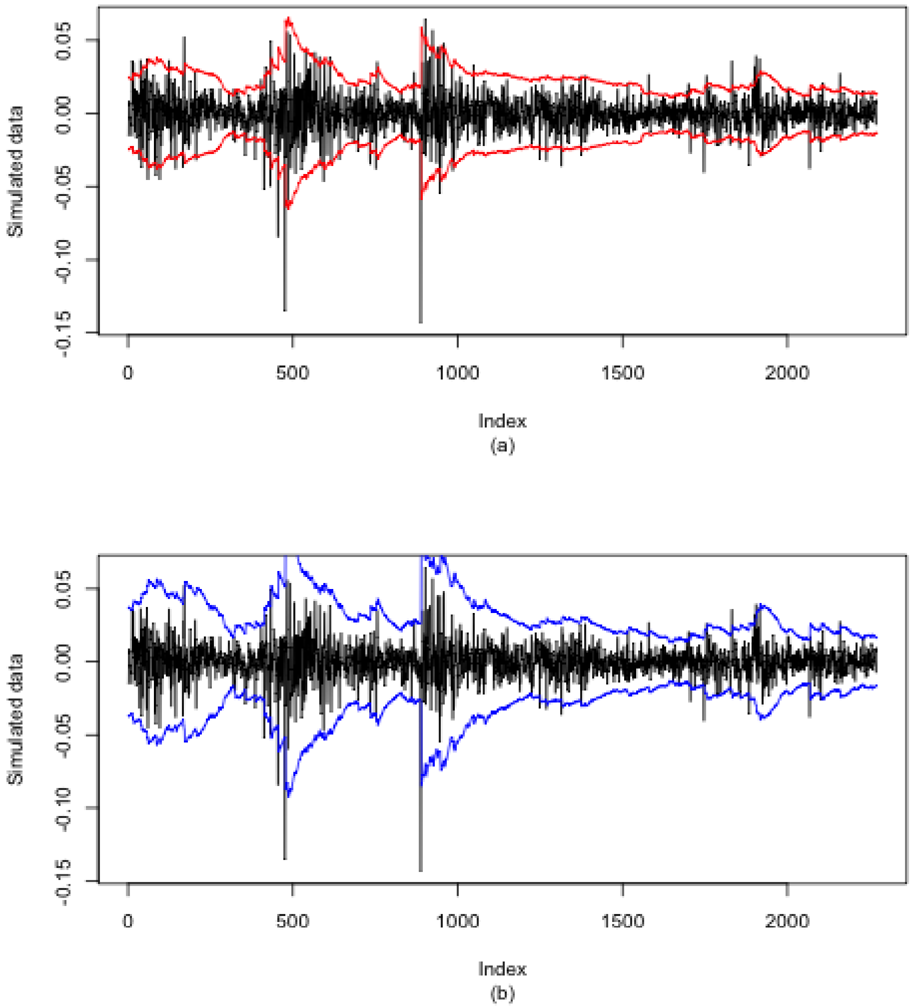



**Figure 15:** Simulated data with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 750 days,
(b) GARCH(1,1) student-*t* distribution, rolling windows length 750 days

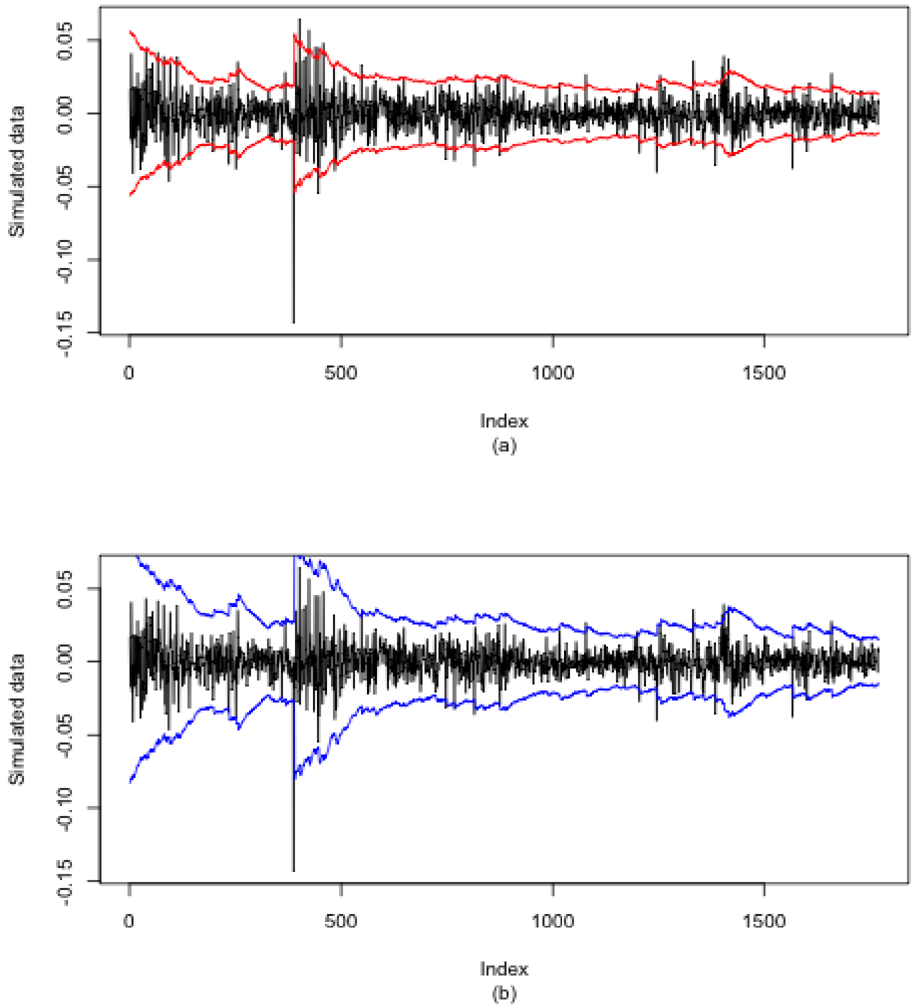



**Figure 16:** Simulated data with 95% prediction intervals,
(a) GARCH(1,1) normal distribution, rolling windows length 1250 days,
(b) GARCH(1,1) student-$t$ distribution, rolling windows length 1250 days

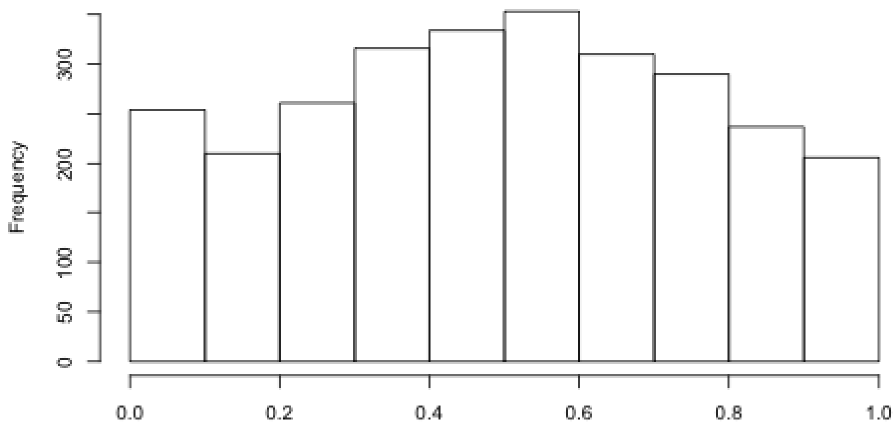



**Figure 17:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 250 days with simulated data

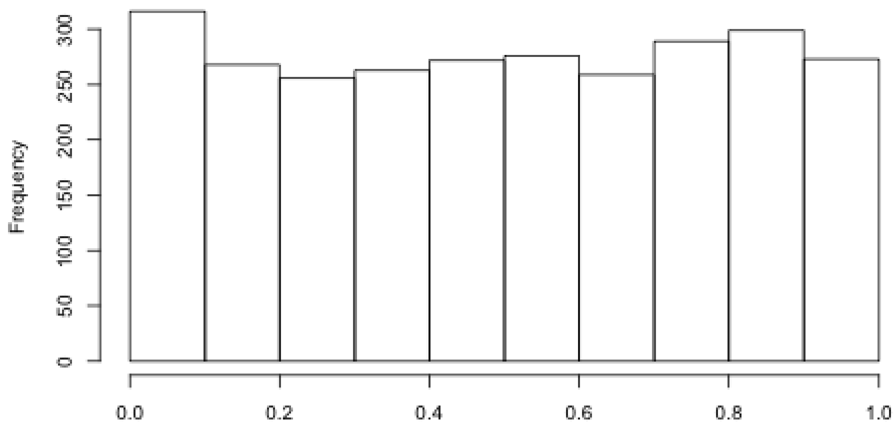



**Figure 18:** Histogram of the standardized residuals, for GARCH(1,1) student-$t$ distribution using rolling window length of 250 days with simulated data

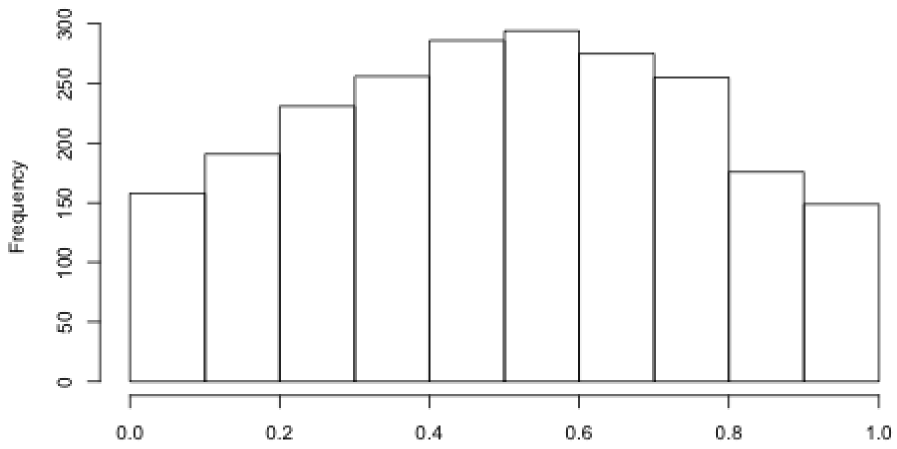



**Figure 19:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 750 days with simulated data

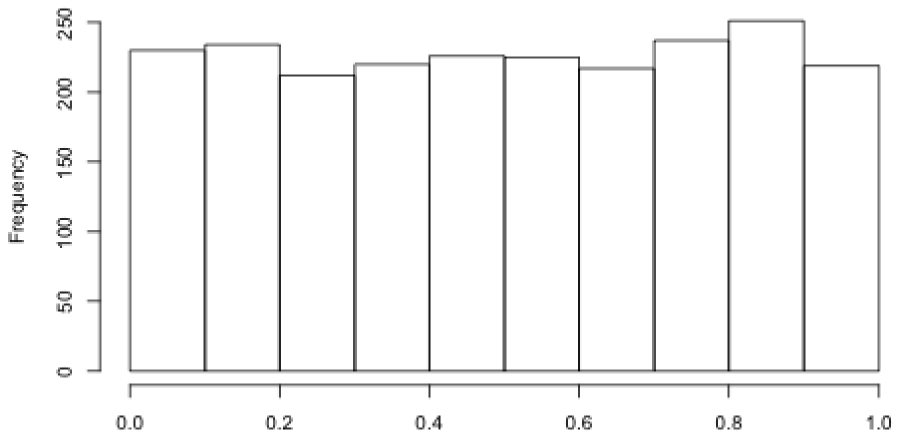



**Figure 20:** Histogram of the standardized residuals, for GARCH(1,1) student-$t$ distribution using rolling window length of 750 days with simulated data

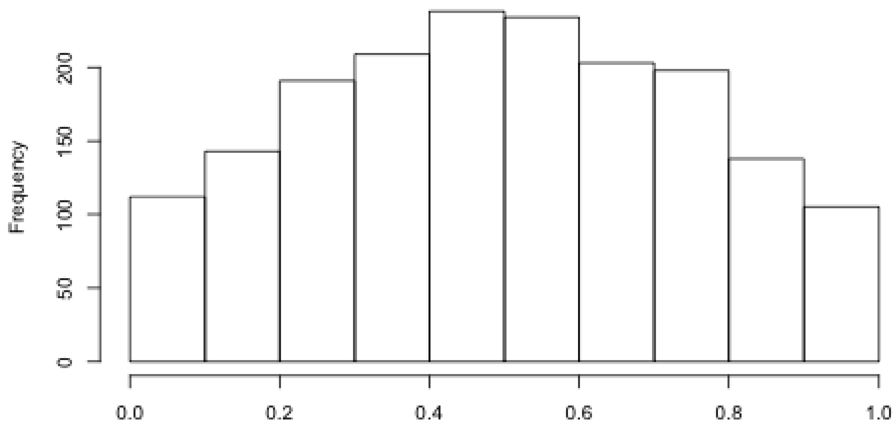



**Figure 21:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 1250 days with simulated data

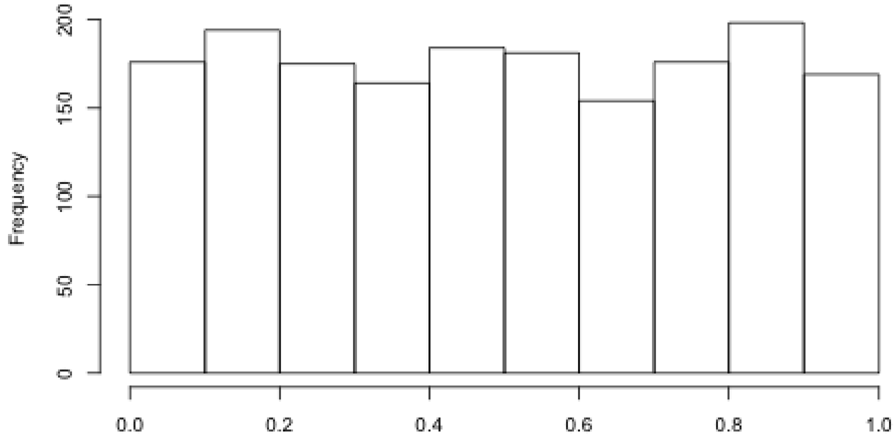



**Figure 22:** Histogram of the standardized residuals, for GARCH(1,1) student-$t$ distribution using rolling window length of 1250 days with simulated data

# D Appendix

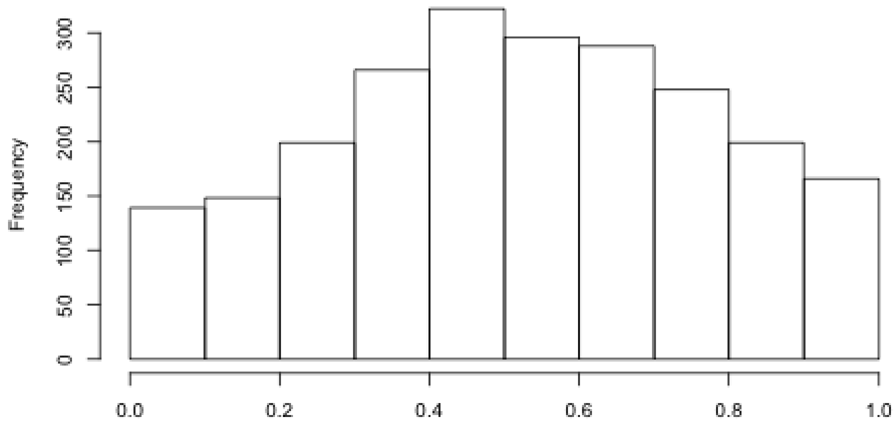



**Figure 23:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 750 days, Data used is the Google stock daily log returns.

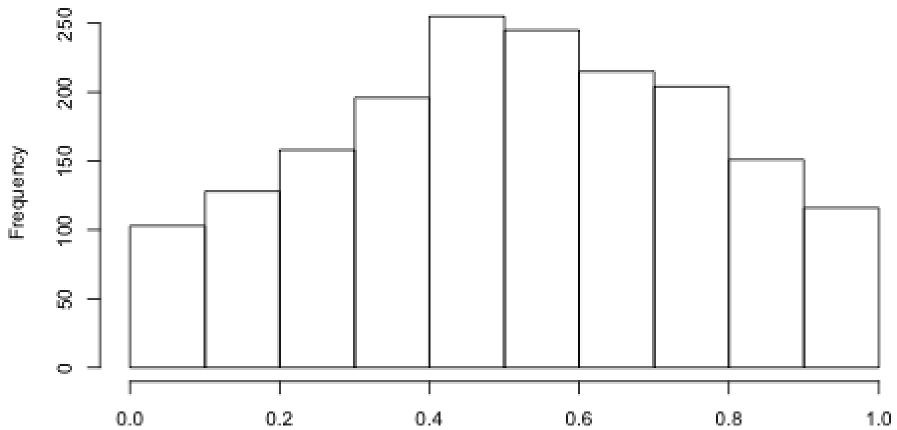



**Figure 24:** Histogram of the standardized residuals, for GARCH(1,1) normal distribution using rolling window length of 1250 days. Data used is the Google stock daily log returns.

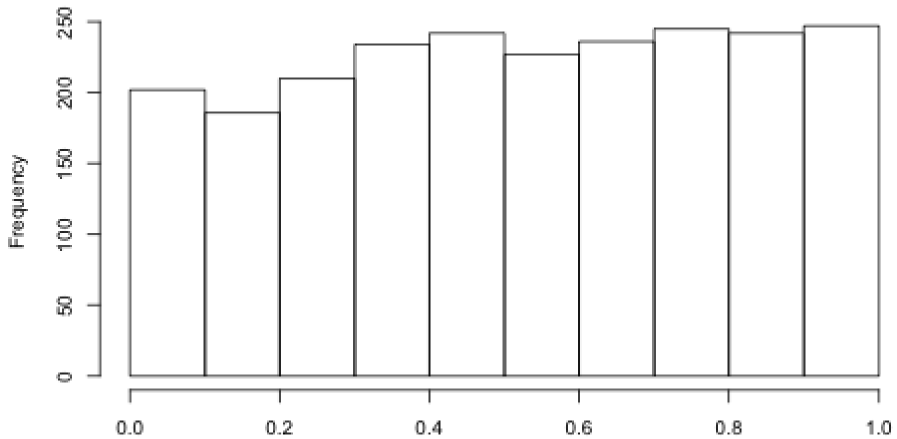



**Figure 25:** Histogram of the standardized residuals, for GARCH(1,1) student-$t$ distribution using rolling window length of 750 days. Data used is the Google stock daily log returns.

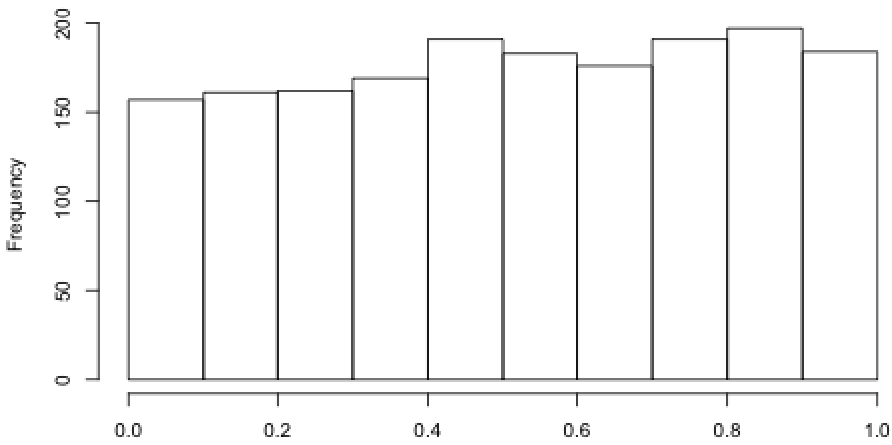



**Figure 26:** Histogram of the standardized residuals, for GARCH(1,1) student-$t$ distribution using rolling window length of 1250 days. Data used is the Google stock daily log returns.