Stockholms universitet

Evaluation of volatility performance of GARCH models on Carnegie Strategyfund

Aleksandra Vojcic

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2017:11
Matematisk statistik
Juni 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2017:11**
http://www.math.su.se

Evaluation of volatility performance of GARCH models on Carnegie Strategyfund

Aleksandra Vojcic*

June 2017

Abstract

This thesis examines three commonly used forecasting models, the GARCH model, the EGARCH model and the TGARCH model. Three main themes will be covered throughout this paper. The evaluation of time horizons that creates the best conditions for future forecasts, determining which distribution suits the error term and the evaluation of which GARCH model provides the best sample-fit in terms of AIC and BIC. The results indicated that both EGARCH and TGARCH that are more complex models outperformed the symmetric GARCH. When it comes to the distribution term it was quite evident that the Student-t distribution provided better sample-fit compared to the Gaussian distribution. This was quite expected given the fact that negative shocks tend to have larger impact on the volatility market compared to positive shocks, resulting in heavier tails in the distribution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: aleksandra_vojcic@hotmail.com. Supervisor: Mathias Millberg Lindholm, Filip Lindskog.

Contents

1 Introduction

The study of financial markets often exhibits volatility clustering with periods of high and low volatility. This is often interpreted in financial time series models that captures the nonconstant volatility. GARCH models are commonly used time series models especially when the volatility is randomly varying in the stock market. It has been shown that negative shocks have a larger effect on stock pricing than positive shocks of the same magnitude. The stock market have shown indications of a longer recovery from the negative skocks, resulting in a long lasting impact. This also indicates that a symmetric distribution is not always a realistic assumption. Tsay[2] argues that the extended GARCH models captures the asymmetric effects between negative and positive asset returns. In this paper three different GARCH models will be evaluated. First, the symmertric GARCH model and then the asymmetric EGARCH and TGARCH that account for the asymmetric effects in the financial market.

In this paper there will be three main themes being studied. First, the evaluation of which time interval creates the best conditions for future forecast in terms of AIC and BIC. The second theme will be to determine which distribution for the error term give the best sample-fit and the third theme will be to evaluate and find a model from the GARCH family models that have the best sample-fit in terms of AIC and BIC.

The first part of this paper will cover the theoretical framework which explaines the background of the analysis. Terms and definitions will be explained in order to increase the understanding behind the results and the analysis. The second part of this paper will cover the methodology framework. Carnegie Strategyfund is the sample data used in this paper and will be used as the basis for the analysis throughout this paper. The methodology framework will also contain different figures and plots that describe different characteristics of the sample data. Afterwards the results will be concluded where the different time horizons will be evaluated together with the evaluation of the GARCH models chosen in this paper. This paper will end with a discussion part where all the results are processed and concluded with the most important highlights from the results.

1.1 Aim

The aim of this thesis is to evaluate three different GARCH models in order to find the best fit for sample data. The time horizons will also be evaluated to determine which interval creates the best conditions for volatility forcasting performance, but also to find which distribution creates better sample-fit.

2 Theoretical framework

In this section the theoretical framework will be presented. Terms, definitions and statistical tests that are used for the analysis will be explaind in order to understand the background of the analysis and the methodology.

2.1 Return and Volatility

Let the asset return at time t and a price index P_t be defined as:

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

where R_t corresponds to a one-period simple return, see Tsay[2]. Taking the logarithm the equation follows:

$$r_t = \log \frac{P_t}{P_{t-1}} = \mu_t + a_t$$

where $a_t = \sigma_t \epsilon_t$, μ is the mean value of returns, σ is the standard deviation and ϵ is the error term. Return of an asset is the most common measure for investors when analyzing financial stocks. The return measures the gain or loss of an asset given a particular time period. A common rule is that the more risk an investor takes, the greater the potential for higher returns and losses. The greater the risk of an asset there is, the higher the volatility clustering becomes. Standard deviation is the measure for stock volatility, defined as:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(r_t - \mu)^2}$$

where μ is the mean value of returns and t = 1,2,3,...,n with n observations.

2.2 Normal distribution and Student-t distribution

When estimating different GARCH models it is important to determine which distribution of the error term ϵ that fits data. An inappropriate distribution for the model may lead to either overestimation or underestimation of future risk. Different distributions of GARCH may also lead to different results regarding option pricing. It is therefor important to find the best fit of the distributions term for the different GARCH models, see Zhou[5]. In this paper two of the most commonly used distributions will be presented and examined. The first one is the Gaussian (or Normal) distribution, defined as follows:

$$f(x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_t-\mu)^2}{2\sigma^2}}$$

where μ is the mean value and σ^2 is the variance.

The second one is the Student-t distribution which captures the heavy-tails in financial time series, given by:

$$f(x_t) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{(v-2)\pi}} (1 + \frac{x_t^2}{v-2})^{-\frac{v+1}{2}}$$

where Γ is the usual Gamma function and $v > 2$ is the number of degrees of freedom, see Tsay[2].

2.3 GARCH

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) is the extension of the ARCH process allowing for a more flexible lag structure. The ARCH process was first introduced in Engle (1982) and later extended in Bollerslev[1] (1986) with the GARCH process which recognizes randomly varying volatility that occurrs in the financial markets. The GARCH process often provides a more real-world context compared to the ARCH process when dealing with prediction of prices of financial instruments. Hence, if $a_t = \sigma_t \epsilon_t$, the GARCH(p,q) follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{max(q,p)} (\alpha_i + \beta_i) < 1$, $p \geq 0$, $q > 0$ and ϵ_t is a sequence of iid random variables. α_i is referred as ARCH parameters while β_j is referred as GARCH parameters.

In this paper the GARCH(1,1) process will be implemented. It is the simplest process but often very useful. GARCH(1,1) is given by:

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

GARCH(1,1) represents how many ARCH terms appears in the model which is represented by the first notation in GARCH(1,1), the second notation indicates how many GARCH terms are included in the model, also known as moving average lags.

GARCH(1,1) is set up to forecast for one period ahead which in turn can forecast a two-period forecast. By using this technique, it has been shown that long-horizon forecasts can be constructed when based on the one-period forecast.

2.4 EGARCH

Tsay[2] argues that EGARCH (Exponantial GARCH) became an improvement of the GARCH model proposed in Nelson(1991). In order to overcome some weaknesses of the symmetric GARCH model Nelson[3] has presented EGARCH that allows asymmetric effects between negative and positive asset returns. EGARCH can in fact account for leverage effect when handling volatility models with financial time series. Nelson[3] defined EGARCH as follows:

$$\log \sigma_t^2 = a + \sum_{i=1}^{p} \alpha_i g(Z_{t-i}) + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2$$

where $g(Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|))$, σ_t^2 is the conditional variance, a, α, β, θ, λ are parameters. There are no further restrictions due to the fact that $\log \sigma_t^2$ may be negative. $g(Z_t)$ allows the conditional variance to respond asymmetrically to negative and positive shocks in stock price.

Defining $Z_t = \frac{\epsilon_t}{\sigma_t}$ and using the fact that $g(Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|))$, the function can thus be expressed as follows:

$$\log \sigma_t^2 = a + \sum_{i=1}^{p} \alpha_i \lambda \left(|\frac{\epsilon_{t-1}}{\sigma_{t-1}}| - E(|\frac{\epsilon_{t-1}}{\sigma_{t-1}}|) \right) + \sum_{i=1}^{p} \theta_i \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2$$

Depending on the distribution for ϵ_t, Z_t varies according to Tsay[2]:

$$E(|Z_t|) = \sqrt{\frac{2}{\pi}}$$

for the standard Gaussian dirtribution and,

$$E(|Z_t|) = \frac{2\sqrt{v-2}\Gamma[(v+1)/2]}{(v-1)\Gamma(v/2)\sqrt{\pi}}$$

for Student-t distribution. The use of Z_t enables the model to respond asymmetrically to negative and positive values.

In this paper EGARCH(1,1) is used which can be expressed as followes:

$$\log \sigma_t^2 = a + \alpha_1 \lambda \left(|\frac{\epsilon_{t-1}}{\sigma_{t-1}}| - E(|\frac{\epsilon_{t-1}}{\sigma_{t-1}}|) \right) + \theta_1 \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \log \sigma_{t-1}^2$$

which is again set up to forecast for one period ahead.

2.5 TGARCH

Another commonly used model is the TGARCH (threshold GARCH) that accounts for handling leverage effects. TGARCH also captures the asymmetry as earlier mentioned with the EGARCH model. TGRACH is defined as followed, see Glosten, Jagannathan and Runkle[4]:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i})\, a_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2,$$

where N_{t-i} is an indicator for negative a_{t-i}, defined as:

$$N_{t-i} = \begin{cases} 1, & a_{t-i} < 0 \\ 0, & a_{t-i} \geq 0 \end{cases}$$

where α_i,γ_i and β_j are nonnegative parameters. TGARCH uses zero as its threshold to separate the impacts of past negative and positive shocks. a_{t-i}^2 can have different effects on the conditional variance σ_t^2 depending on whether a_{t-i} is above or below the threshold. When the value a_{t-i} is positive, it contributes with $\alpha_i a_{t-i}^2$ to σ_t^2 compared to when the value of a_{t-i} is negative it has a larger impact with $(\alpha_i + \gamma_i)a_{t-i}^2$ when $\gamma_i > 0$, see Tsay[2].

TGARCH(1,1) which forecasts one period ahead is going to be implemented throughout this paper, defined as follows:

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

2.6 Parameter estimation

As mentioned earlier two different distributions will be used throughout this paper, the Normal distribution and the Student's t distribution. The results will differ when estimating parameters for the different GARCH models when using different distributions. Depending on the distributions the likelihood function will differ when obtaining the maximum likelihood estimators.

2.6.1 Maximum likelihood

Maximum likelihood is the most commonly used method when estimating parameters for time series models. $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}$ is the estimate of the estimator $\boldsymbol{\theta}$ that maximizes the likelihood funcion, see Held and Bové[6]:

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} l(\theta)$$

where $l(\theta) = \ln(L)$ is the log-likelihood function.

When assuming the normal distribution, the likelihood function is defined as follows:

$$L = f(a_{n+1}, ..., a_K | \alpha, a_1, ..., a_n) = \prod_{t=n+1}^{K} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{a_t^2}{2\sigma_t^2}}$$

where $\alpha = (\alpha_0, \alpha_1, ..., \alpha_n)$ are the parameters to be estimated and $a_1, a_2, ..., a_K$ is a set of K independent and identically distributed random variables. Recall that $a_t = r_t - \mu_t \approx r_t$ when $\mu_t \approx 0$. The log-likelihood function is therefore as follows:

$$l(\theta) = \sum_{t=n+1}^{K} \ln(\frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{a_t^2}{2\sigma_t^2}}) =$$

$$-\frac{(K-(n+1))}{2} ln(2\pi) - (K-(n+1)) ln(\sigma_t) - \frac{1}{2\sigma_t^2} \sum_{t=n+1}^{K} a_t^2$$

When Student's t distribution is assumed the likelihood function is defined as:

$$L = f(a_{n+1}, ..., a_K | \alpha, a_1, ..., a_n) = \prod_{t=n+1}^{K} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{(v-2)\pi}} \frac{1}{\sigma_t} (1+\frac{a_t^2}{(v-2)\sigma_t^2})^{-\frac{v+1}{2}}$$

with the log-likelihood function:

$$l(\theta) = \sum_{t=n+1}^{K} ln(\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{(v-2)\pi}} \frac{1}{\sigma_t} (1+\frac{a_t^2}{(v-2)\sigma_t^2})^{-\frac{v+1}{2}}) =$$

$$(K-(n+1)) ln(\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{(v-2)\pi}}) - \frac{1}{2} \sum_{t=n+1}^{K} ln(\sigma_t^2) - \frac{v+1}{2} \sum_{t=n+1}^{K} ln(1+\frac{a_t^2}{(v-2)\sigma_t^2})$$

2.7 Statistical tests and Model selection

2.7.1 AIC

Akaike Information Criterion (AIC) measures the goodness of fit for different statistical models and compares different models, see Tsay[2]. AIC makes adjustmenst to the likelihood function to account for the number of parameters. If the number of parameters in the model is K, the AIC is given by:

$$AIC = 2K - 2l$$

where l is the maximized value of the loglikelihood function. The best fit will have the lowest AIC value.

2.7.2 BIC

Bayesian Information Criterion (BIC) is a criterion for model selection. BIC measures the goodness of fit and is based on the likelihood function, see Tsay[2]. BIC is closely related to the AIC with the same notation but with one additional parameter. BIC is given by:

$$BIC = Kln(N) - 2l$$

where N denotes the number of observations in sample size, K is the number of parameters in the model and l is the maximized value of the loglikelihood function. BIC resolves the problem of overfitting a model that have many parameters by introducing a penalty term for the number of parameters in the model. BIC generally penalizes more strongly than AIC. A smaller value of BIC provides better sample fit and is therefore prefered.

2.7.3 T-test

The t-test is a statistical hypothesis test to check whether the mean of the test statistic is equal or different from zero. The distribution of the sample is normally assumed to follow a Gaussian distribution. The null hypothesis is defined as follows, see Britton & Alm[9]:

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis:

$$H_A : \mu \neq \mu_0$$

where μ is the mean from a sample $X = (x_1, x_2, ..., x_n)$. The test statistic can be defined as follows:

$$T = \frac{\bar{X} - \mu_0}{s(X)/\sqrt{n}} \sim t(n-1)$$

where the test statistic is Student-t distributed with $(n-1)$ degrees of freedom and s(X) is the standrad deviation. The null hypothesis is rejected and the mean is different from zero when $|T| > t_{\alpha/2}(n-1)$.

2.7.4 Autocorrelation function and White noise

Autocorrelation measures if there is linear dependence between two observations. It is of interest to compare the dependency between the return series R_t and its past values R_{t-i}, see Tsay[2]. A linear time series model is defined as follows:

$$R_t = \mu_t + a_t$$

where μ is the mean of the return series and a_t is a sequence of iid random variables with mean zero and variance σ^2.

For a given return series R_t, the autocorrelation is defined as follows, see Tsay[2]:

$$\hat{\rho}_l = \frac{\sum_{t=2}^{T}(R_t - \bar{R})(R_{t-l} - \bar{R})}{\sum_{t=1}^{T}(R_t - \bar{R})^2}$$

where $\bar{R}$ is the sample mean, l is the number of lags and $0 \leq l < T-1$. A stationary series is not serially correlated if $\rho_l = 0$ for $l > 0$.

In order to determine if serial correlations occur it is convenient to set up statistical tests that test for zero correlations. As earlier mentiond, Ljung-Box test caculates p-values in order to find out whether or not to reject the null hypothesis.

If all autocorrelation functions are close to zero, then the series is said to be a white noise series. A time series R_t is white noise if if R_t is a sequence of iid (identical and independently distributed) random variables with finite mean and variance σ^2.

2.7.5 Ljung-Box

Ljung-Box test tests if the autocorrelations of a time series are different from zero. The hypothesis H_0 is defined as follows:

$$H_0 : \rho_1 = \rho_2 = \rho_3 = ... = \rho_m = 0$$

with the alternative hypothesis that autocorrelations are different from zero. If that is the case, then they exhibit serial correlation.

According to Ljung and Box[7], the test statistic can be defined as follows:

$$Q(\rho) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{\rho}_k^2 \sim \chi_m^2$$

where $\hat{\rho_k}$ is the estimated autocorrelation, k number of lags and n observations in the time series. The Ljung-Box test is χ^2 distributed with m degrees of freedom. If $Q(\rho) > \chi^2$ the decision is to reject H_0. Based on the number of lags, Ljung-Box test tests the overall randomness in the time series.

2.7.6 Heteroscedasticity

In a classical linear regression model the variance of each disturbance term ϵ_j, conditional on explanatory variables is constant and equal to σ^2 is said to be homoscedastic. On the other hand when the variance of the disturbance term is not constant it is a case of heteroscedasticity, see Tyrcha[8]. Heteroscedasticity is defined as the conditional variance that changes over time. In other words the variance is nonconstant over time.

2.7.7 Skewness and Kurtosis

Skewness is the third central moment to measure the symmetry of a data set. A distribution is called symmetric if it looks the same to the left and right of the center point. Negative values indicates left skewness and positive values indicates right skewness. A value of zero indicates that sample data comes from a normal distribution. Skewness is defined as follows, see Tsay[2]:

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

where $x_1, x_2, ..., x_T$ is a random sample of X with T observations, μ_x is the mean and σ_x is the standard deviation.

Data sets with high kurtosis tend to have heavy tails. Similarly, data sets with low kurtosis tend to have light tails. Kurtosis is the fourth moment and is an easy way of measuring the size of distribution tails. A value of 3 indicates that sample data comes from a normal distribution. With the same assumptions as above, Tsay[2] defines kurtosis as follows:

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

2.7.8 Jarque-Bera Test

Jarque-Bera test is a normality test based on skewness and kurtosis. The test statistic is defined as follows, see Tsay[2]:

$$JB = \frac{\hat{S}^2(r)}{6/T} + \frac{[\hat{K}(r) - 3]^2}{24/T}$$

which is chi-squared distributed with 2 degrees of freedom. The value of 3 in the test statistic represents a normal distribution as mentioned above. The assumption of normality under H_0 is rejected when the p-value of JB test is less than the significance level.

3 Methodology framework

In this section the methodology framework will be presented. The sample data is downloaded from Carnegie Fund's database and represents daily closing prices for Carnegie Strategyfund.

3.1 Data analysis

In this part of the paper the sample analysis will be evaluated. As earlier mentioned, Carnegie Strategyfund is the sample data used for analysis with 2517 observed closing prices, covering the time 2007/01/26 - 2017/01/25. The sample data will be observed from two point of views. The first one where all the 2517 observations are covered. The second point of view is were the sample is divided in two subintervals, before and during the financial crisis, and the second subperiod that covers the time after the financial crisis. The reason for this divition is to find out whether the same GARCH model fits all time intervals, but also to find out which interval is most benficial to use for future forecast.

3.2 Carnegie Strategyfund

Carnegie Strategyfund is a mixed fund with 60 percent shares and 40 percent interest rates. As mentioned above, the return series will be presented from two points of views where the first analysis covers the whole 10 year period, afterwards the sample data will be divided in two subperiods in order to find the interval with best conditions for future forecast. The sample data will be analyzed with plots that brings out different characteristics from the sample.

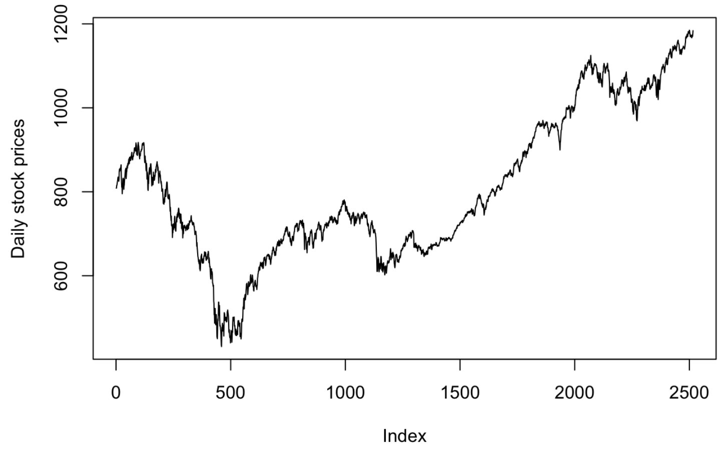

Figure 1: Daily closing prices

Figure 1 demonstrates how the price for Carnegie Strategyfund have fluctuated during a 10 year period between 2007/01/26 - 2017/01/25. Before the crisis the stock price was on a downfall and reached the lowest point in 2008. After the crisis it was evident that the price rebounded and started to rise. The price process will be one main variable throughout this paper.

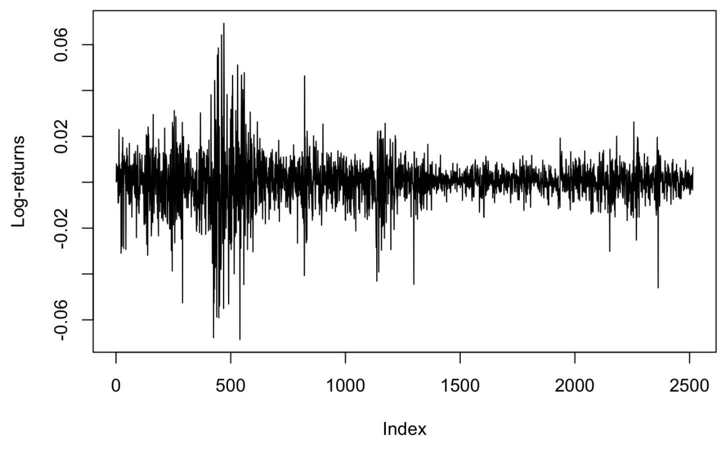

Figure 2: Daily log-returns

In figure 2 the daily log-returns are presented. The figure illustrates how the returns have fluctuated over time, with especially high fluctuations around 2008 when the financial crisis occured, followed by calmer periods efter the crisis. As mentioned before, the return is defined by $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ and taking the logarithm gives the following relation: $R_t \approx log(\frac{P_t}{P_{t-1}})$. By the judgement of the shape, it becomes eveident that the return series indicate traits of heavier tails. The deviation around index 500 is much larger than the rest of the series which indicates heavier tails in sample data.

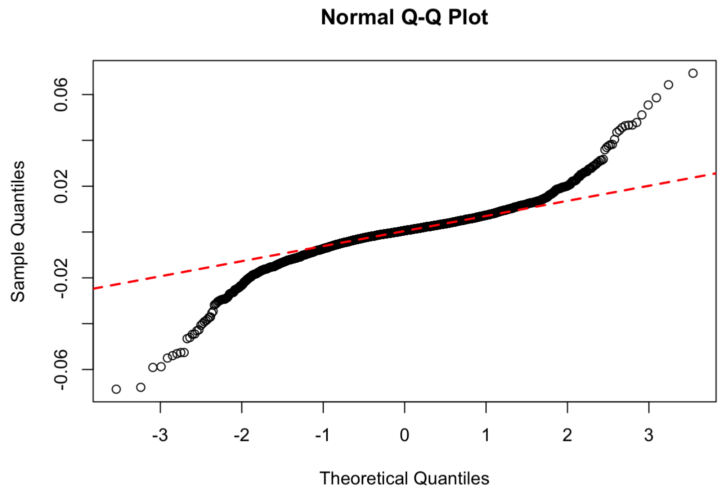

Figure 3: QQ-plot of daily log-returns

To find out whether the error term from the sample data comes from a Gaussian or Student-t distribution it can be convenient to observe the following plot in figure 3. By plotting the Quantile-Quantile plot (QQ-plot) it becomes clear if the data follows a Gaussian distribution or if the sample data indicates heavier tails. In this case it is demonstrated by figure 3 that data is not normaly ditributed because of the tails that deviates from the staight red line. The Gaussian distribution would therefore not provide a particularly good fit. The QQ-plot exhibits significantly heavier tails and the behavior indicates that the Student-t distribution could be a better fit.

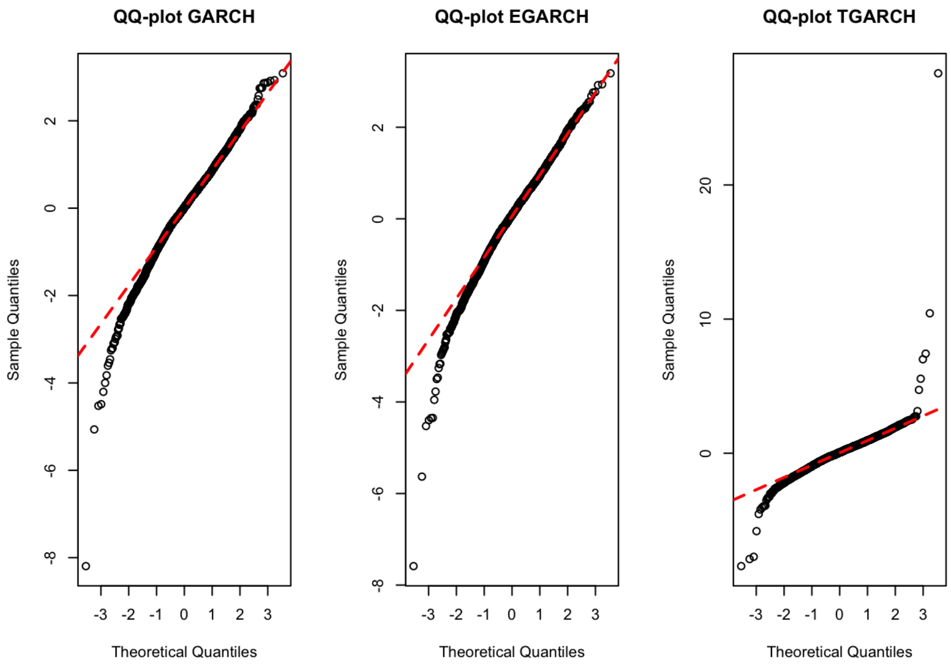

Figure 4: QQ-plot for GARCH, EGARCH & TGARCH with Gaussian distribution used on the whole time interval

It is also demonstrated by the QQ-plots in Figure 4 that the residuals exhibit heavier tails for all the three GARCH models. Once again it could be said that all the three models does not follow a Gaussian distribution because of the deviation from the straight line. The QQ-plots for the three GARCH models with Student-t distribution are presented in appendix.

Figure 5 is observed in order to find out if the financial crisis had an impact on the QQ-plots or not. When looking at Figue 5 the QQ-plots only indicate a significant small improvement. Although it is still the same situation as in Figure 4 where both GARCH and EGARCH looks asymmetric and the TGARCH model looks slightly more symmetric but with deviated observations on both sides. It is hard to make any assuptions on which model has the best distribution fit, the QQ-plots are very much alike but with the exception that TGARCH have some larger deviations than GARCH and EGARCH. QQ-plots with Student-t distribution are presented in appendix.

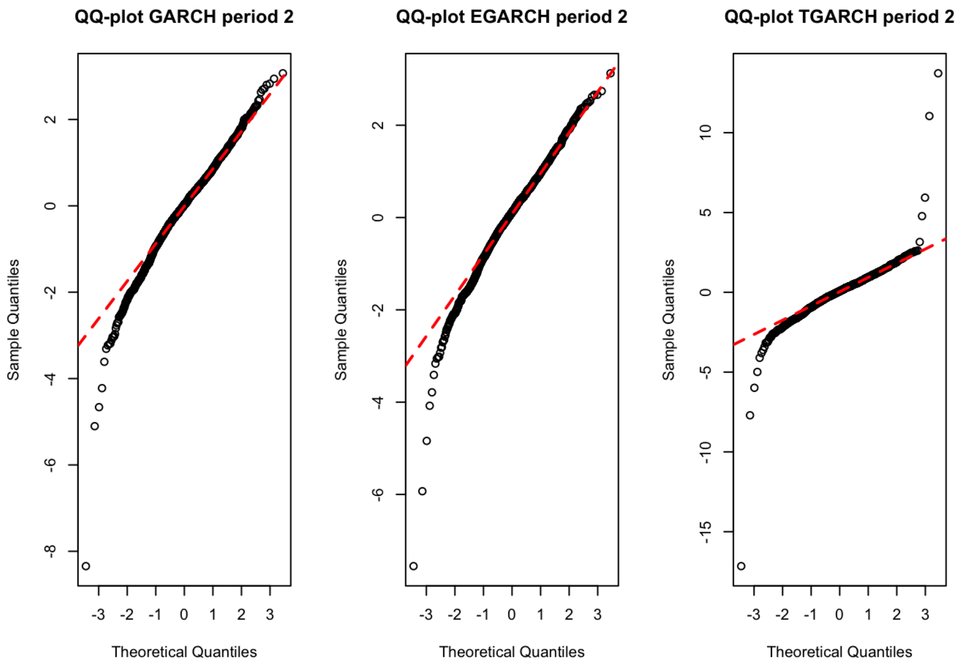

Figure 5: QQ-plot for GARCH, EGARCH & TGARCH with Gaussian distribution used on period 2 without the financial crisis

3.2.1 Characteristics of sample data

Sample Data		
Skewness	Kurtosis	Jarque-Bera
-0.3540795	10.92358	p <2.2 E-16

Table 1: Sample tests

A characterization of the sample data includes skewness and kurtosis in order to detemine the location and the variability of the data set. As earlier mentioned, skewness is a measure of symmetry and it discovers whether a distribution is symmetric and look the same to the left and right of the center point. Data follows a normal distribution if skewness is near zero. In table 1 the sample characteristics are summerized with a skewness of -0.3540795 which indicates that the sample is negatively skewed meaning that sample data does not come from a normal distribution.

Kurtosis measures whether the sample is heavy tailed or light tailed relative to a normal distribution. A value of 3 indicates that sample comes from a Gaussian distribution. The sample data presented in table 1 have a kurtosis of 10.92358 which indicates a heavy tailed distribution.

The Jarque-Bera test is used in order to find out whether the error terms are normally distributed based on skewness and kurtosis. A p-value of less than 2.2e-16 indicates a significant result where the assumption of normality under the null hypothesis is rejected.

The statistically significant result recieved from the Jarque-Bera test and a negative skewness together with a high value of kurtosis indicates heavier tails to the left, and it becomes evident that the sample does not follow a Gaussian distribution. The sample data has a better fit with the Student-t distribution.

3.2.2 Testing for Autocorrelation

In order to find out if any autocorrelation occurs in data it is convinient to compute autocorrelation plots. It is of interest to find out whether there are any linear dependencies between the return series R_t and the past values of R_{t-i}. In figure 6 the autocorrelation function is observed and indicates a weakly stationary series with a weak correlation, if any. The series is therefore a white noise series. In order to use GARCH models and apply them to data it is important that the return series is uncorrelated but still dependent, see Tsay[2]. The squared autocorrelation function, which is the second moment of the returns, is also presented in figure 6 which demonstrates and confirmes that there is a dependency, which is recuired in order to predict the future.

In order to be certain that the series truly is uncorrelated it is necessary to test for zero correlation. This is easily done by using Ljung-Box test which is χ^2-distributed, and test the null hypothesis $H_0 : \rho = 0$ versus the alternative hypothesis $H_A : \rho \neq 0$. The null hypothesis is rejected if the p-value is approximately smaller than 0.05 on a 95% level. According to Tsay[2], the number of lags are calculated by taking the logarithm of the number of observations from the data, giving the following: $\ln(2517) \approx 7.8$. Computations of Ljung-Box test gives a p-value equal to 0.06436 which is non-significant and the null hypothesis is not rejected on a 95% level. The return series is therefore uncorrelated.

It is also necessary to test if the series mean is equal to zero. This is provided with a t-test with the null hypothesis $H_0 : \mu = 0$ versus the alternative hypothesis $H_A : \mu \neq 0$. As earlier mentioned, the null hypothesis is rejected if the p-value is less than 0.05. The t-test gives a p-value of 0.4654 with the 95% confidence interval [-0.0002554677, 0.0005585943]. This means that the output is non-significant and H_0 can not be rejected. The return process

$r_t = \mu_t + a_t$ can now be simplified as $r_t = a_t$

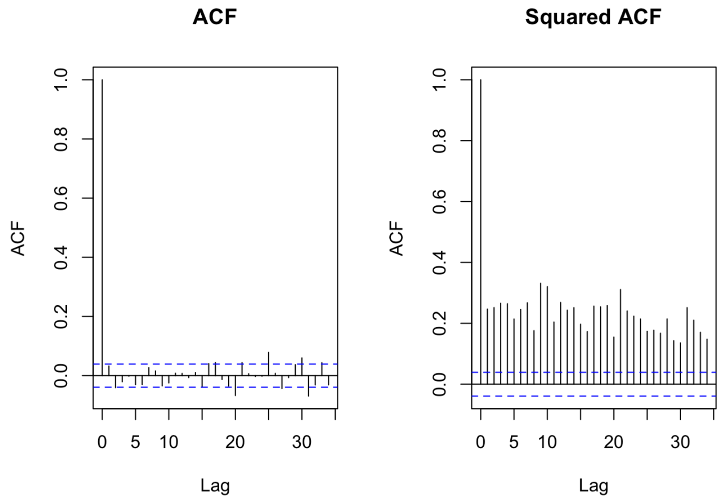

Figure 6: Autocorrelation functions of daily log-returns

Continuing on with the observation of Figure 7 it becomes evident that the series is not autocorrelated. The plots illustrate autocorrelation functions of the GARCH model. All the spikes lies within the confidence interval which indicates that there are no linear dependencies within the time series, the series is therefore a white noise series. The same can be concluded when applying ACF on EGARCH and TGARCH. The results are found in appendix.

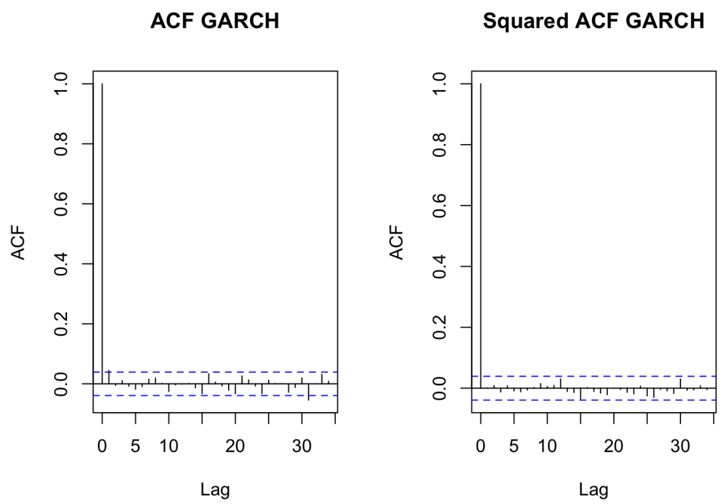

Figure 7: Autocorrelation functions of the standardized residuals for GARCH(1,1)

3.2.3 Testing for Heteroscedasticity

To determine whether or not heteroscedasticity occurs in sample data it is essential to set up a hypothesis that test if heteroscedasticity is present. This could be done with for example plots of residuals that show in what way the variance is behaving; if it is constant, increasing, decreasing etc. Heteroscedasticity can also be tested with Whites test to determine whether the variance of the error term is constant or varies over time. If heteroscedasticity is present it is said that the disturbance term σ^2 do not have constant variance, see Tyrcha[8].

Another way to test for heteroscedasticity is to use the Ljung-Box test as mentioned above. If autocorrelations are present it indicates that heteroscedasticity can be assumed. Otherwise, if the disturbances would be autocorrelated then the assumption would be violated. The result from the Ljung-box test and the QQ-plots above proved that the return series is uncorrelated and therefore heteroscedasticity can be assumed in data. When heteroscedasticity is present in the residuals it also means that there are ARCH-effects in the data. It is therefore convinient and approved to use GARCH models for the existing sample data.

4 Results

In the following section the results will be evaluated. Two different distributions of the error term are used when evaluating the models. Data will be analyzed separately with the GARCH models chosen in this paper; GARCH(1,1), EGARCH(1,1), TGARCH(1,1).

There are different ways to evaluate models and to compare the goodness of fit for different GARCH models. Different methods can be used depending on if they are nested or not. In other words how complex a model is and if it is possible to transform the model into a simpler model by setting contraints on some parameters. In this case the likelihood ratio would be evaluated when determining whether some parameters are necessary or not. This is not the case in this paper and will therefore not be used. Other evaluation methods are AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) which are the most commonly used information criterion tests. These tests can be used when comparing any model's goodness of fit against each other. AIC and BIC will be used in this paper for evaluation and finding the most appropriate model for future forcasts. When using AIC and BIC a smaller value is preffered. A small value indicates that the model have a better fit for data. The GARCH family models will also be evaluated by comparing the statistical significance of the parameters estimated by the maximum likelihood.

After having presented the results for the whole 10 year period the data will be divided in two periods. In this way it can be determined if the same model fits the separate subperiods or if different models are more appropriate when divinding the periods. The data is divided as follows; Period 1 with 736 observations between 2007/01/26 - 2009/12/30 which represents the time before and during the financial crisis and Period 2 with 1781 observations between 2010/01/04 - 2017/01/25 which represents the time after the financial crisis.

4.1 GARCH(1,1)

GARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	4.079 E-07**	3.943 E-04**
α_1	9.373 E-02***	9.427 E-02***
β	9.057 E-01***	9.060 E-01***
AIC	-6.916841	-6.875109
BIC	-6.905251	-6.865837
*** = Statistically significant on 0.1% level ** = Statistically significant on 1% level		

Table 2: Parameter estimation

When looking at the results from Table 2 all the coefficients show that they are statistically significant, both for the Student-t and the Gaussian distribution. One difference is that the Gaussian and Student-t value for α_0 have a slightly lower significance level than the rest of the parameters. α_0 is significant on a 1% level when in the meantime the other parameters are significant on a 0.1% level which is preferable. When looking at Table 2 the results show the value obtained from AIC and BIC. The Student-t distribution has a slightly smaller AIC compared to the Gaussian distribution. The Student-t distribution also have a slightly smaller BIC value compared to the Gaussian. This indicates that the Student-t distribution both have the lowest AIC and BIC value. On the other hand the results are so close that it is hard to make any conclusions that Student-t is definitely a better choice.

4.2 EGARCH(1,1)

EGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	-0.13150***	-0.116270***
α_1	-0.10408***	-0.092278***
β	0.98632***	0.987482***
θ	0.12706***	0.130911***
AIC	-6.9388	-6.9058
BIC	-6.9226	-6.8919
*** = Statistically significant on 0.1% level		

Table 3: Parameter estimation

Furthermore when analysig the EGARCH model in Table 3 all the parameters show a statistical significance on a 0.1% level. The Student-t distribution has slightly lower AIC and BIC values just as the results from GARCH. It is however difficult to draw any conclusions or assumptions when the results only slightly varies.

4.3 TGARCH(1,1)

TGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	7.975 E-04***	8.430 E-04***
α_1	6.550 E-02***	6.825 E-02***
β	9.384 E-01***	9.370 E-01***
γ	8.451 E-01***	7.389 E-01***
AIC	-6.886106	-6.456248
BIC	-6.872197	-6.444657
*** = Statistically significant on 0.1% level		

Table 4: Parameter estimation

When looking at Table 4 the parameters are statistically significant as before. The statistical tests are presented and when comparing them it is as before stated that the Student-t distribution have lower values. The AIC and BIC value for Student-t are smaller and could therefore be prefered. The values differ slightly more than the results from GARCH and EGARCH and it could be said that in this case Student-t distribution could be a better choice when determining the distribution. Although this does not come as suprise that the Student-t distribution would have slightly better results when only looking at the values from AIC and BIC. The sample data indicated heavier tails that were present in the QQ-plots.

4.4 Dividing data in two periods

The question whether or not the same model fits data after dividing it in two periods will be presented below. The time before and after the crisis is analyzed in order to determine if the same model fits the subperiods versus the whole period. But also to find out which interval creates the best conditions for future forecast.

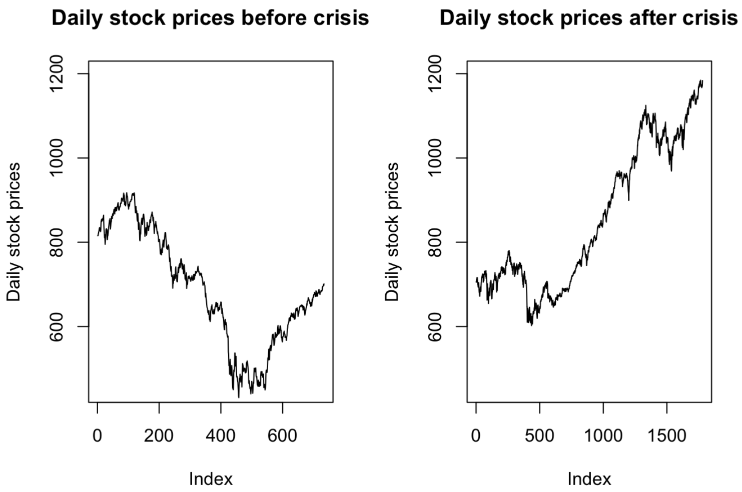

Figure 8: Closing prices divided in two periods

Figure 8 demonstrates the divition of the whole time period. The first period represents the time bofore and during the financial crisis when the stock market had a huge breakdown. The prices declined and reached rock bottom. Afterwards came a better period when the crisis was over, the stock market began to rise and have recovered since the breakdown.

4.4.1 Period 1

As before mentioned Period 1 represents 736 observations between 2007/01/26 - 2009/12/30, the time before and during the financial crisis. This period is characterized by a downfall when the prices reached bottom in the stock market. The dataset will be analyzed in order to find the best fit and determine which model suits Period 1 the best.

The results of the parameter estimation and the statistical tests of the GARCH models are presented in appendix. In Table 5 GARCH(1,1) is evaluated and the results indicates a statistical significance of α_1 and β on a 0.1% level. The intecept α_0 is on the other hand not significant for both the Student-t distribution and the Gaussian distribution. The AIC and BIC value is slightly lower for Student-t than the values for the Gaussian distribution.

Continuing on with EGARCH(1,1) in Table 6, all the parameters show a statistical significance as in the results for the whole period presented above. The Student-t distribution could as before be a better choice when only looking at the values calculated from AIC and BIC. In Table 7 TGARCH(1,1) indicate a statistical significance for all the parameters besides the intercept that has no statistical significance. With the same motive as before the Student-t distribution would be chosen, due to the lower AIC and BIC values.

4.4.2 Period 2

Period 2 represents the time 2010/01/04 - 2017/01/25 with 1781 observations after the financial crisis when the stock market started going upwards. The price of the fund began to rise and have not declined by that much ever since. This interval has no influences from the crisis, it will therefore be analyzed if the sample fit improves. The results are presented in appendix.

When applying the same analysis as before it is evident that EGARCH(1,1) with Student-t distribution has the lowest AIC and BIC values. All the parameters are statistically significant on 0.1% level compared to GARCH(1,1) that have a lower significance level on the intercept. The difference is however slightly distinct between the models and does not vary by that much, just as the results presented within the other time intervals.

When comparing the whole timeperiod with the single subperiods it becomes evident that period 2 without the financial crisis would be chosen when applying a GARCH model for forecasts. The time interval 2010/01/04 - 2017/01/25 indicates a upgoing trend and is not affected by the breakdown from period 1. This interval presents slightly stronger results compared to the whole time interval in terms of AIC and BIC but also stronger significance when comparing the parameters. Period 1 is on the contrary worse which also was expected by the judgement of fewer observations, but also a hectic period were a downfall in the stock market was observed. This of course had negative impact on the results and the ability to be used for further forecasts.

4.5 Sensitiveness of model selection

One important perspective of this paper is to analyze and address the question how sensitive the result is to the correctness of an assumed model. How big is the uncertainty within model selection and how strong conclusions can be made when the results between different AIC values only slightly varies. According to Pawitan[10], the distance between a model and the true underlying distribution that generates data is essential to consider. It is important to recognise what we get if we assume a wrong model, since the real data analysis will always be wrong regardless of what is assumed. There are for example no real data that are exactly Normal or Student-t distributed data.

In order to find the best model with a distribution closest to the true distribution it is essential to maximize the likelihood. In principal, the model with the closest distance to the true distribution should be chosen. The AIC is used when comparing models and deciding which is the best and most supported by the data. It is especially simple if the models are nested to compare the Loglikelihood ratio, but that is not the case in this paper. One of the themes of this paper has been to test whether H_0 : normal or student-t modeles are equally good for data, versus H_A : one model is better than the other. Choosing between two different distribution models is more diffcult than comparing the Loglikelihood ratio between two nested models and deciding wheter or not one extra parameter enhaced the goodness of fit.

When comparing the AIC values between non-nested models it is important to consider some criterias that affects the result. Firstly, when a model is tested the AIC value always improves when there are a lot data observations available. Secondly, when the AIC values only slightly varies between the models it could be difficult to be absolutely sure which model is the best one. Recall that $AIC = 2K - 2l$ where l is the maximized loglikelihood and K is the number of parameters in the model. In this case where there are a lot of data observations available, the variable K becomes negligible and the maximized loglikelihood that yields a larger value becomes a more important variable.

5 Discussion and conclusion

In this paper there were three main themes being studied, first the evaluation of which time interval created the best conditions for future prediction. Then the second theme to determine which distribution for the error term gave the best sample-fit and the third theme was to evaluate and find a model from the GARCH family that had the best sample-fit in terms of AIC and BIC.

One main theme was to find out whether the whole timeperiod or a subperiod created the best conditions for volatility forecasting. It was evident that period 2 without the financial crisis was preferred because of the best sample-fit and the best values from the information criterion tests. This could again be assumed and a possible reason might be that the dynamics of the volatility changes over time, especially during the financial crisis where the volatility and the dynamics of the market shifted. It was evident that EGARCH(1,1) were slightly better than GARCH(1,1) and TGARCH(1,1) in terms of AIC and BIC. The conslusion here by the judgement of the resluts is that more complex models provide better sample-fit than simpler models with fewer characteristics. If that would not be the case then EGARCH(1,1) and TGARCH(1,1) would be pointless and they would have been reduced to the simpler GARCH(1,1). But, in this case the GARCH model is nested in both EGARCH and TGARCH, the results and the conslusion was therefore quite expected. Further evaluation indicates that EGARCH(1,1) had slightly better values than TGARCH(1,1). Both AIC and BIC resulted in smaller values and all the parameters were statistically significant. But on the other hand, the results between the models were only slightly different and it is difficult to be absolutely sure which model will perform the best in reality. When comparing two non-nested models with sligtly different AIC values it becomes more difficult to decide which one is the best.

The impact of the distribution error term was quite evident when looking at the sample fit. It was clear that the Student-t distribution provided a slightly better sample fit than the Gaussian distribution whithin all time intervals in terms of AIC and BIC. Based on the QQ-plots that indicated heavier tails it was quite expected that the Student-t distribution would have better sample fit. When comparing the characteristics of sample data the results indicated a negative skewness and a high value of kurtosis, it was further quite evident that sample data was heavy-tailed. The return series also indicated significantly heavier tails when looking at the shape of the plot. Theoreticaly, a symmetric distribution as the Gaussian is not always a realistic assumption because of the leverage effect. The effect considers the larger impacts from negative shocks that occurs in the financial market and therefore the Student-t indicates a distribution closest to the true distribution.

When closely looking at the results it is obvious that the parameters and the statistical test does not vary by that much. The GARCH models belong to the same GARCH family and only have some different characteristics between each other. For instance, EGARCH and TGARCH are improved models that overcome the weakness of the symmetric GARCH and allows for asymmetric effects between positive and negative shocks in stock price, which is coherent to the reality. In theory it does not come as a suprise that complex models outperformes simpler models, it is also explicable that the results only slightly varies because of the closeness within the GARCH family.

5.1 Further research

For further investigation and to increase the understanding of volatility models it could be convinient to find out whether a higher order of GARCH models would perform a better fit. The next natural step would also be to perform a forecast for the three different GARCH models in order to evaluate the performance and compare the models against each other. Another thing to research would be to reduce the time intervals. One reason is that the volatility might have changed during this long time horizon. The volatility over this sample time is not stationary and the dynamics of the volatility might have shifted. One major impact was the financial crisis that especially changed the dynamics of the volatility. Further investigation with back-testing could be done in order to estimate the performance for the models, this would enhance the evaluation of the performance for the different GARCH models in this paper.

Acknowledgement

I would like to express my sincere gratitude to my supervisors Mathias Millberg Lindholm and Filip Lindskog for their guidance and support throughout the work of this thesis. I would also like to thank my family and friends for their help and support.

6 Appendix

6.1 ACF

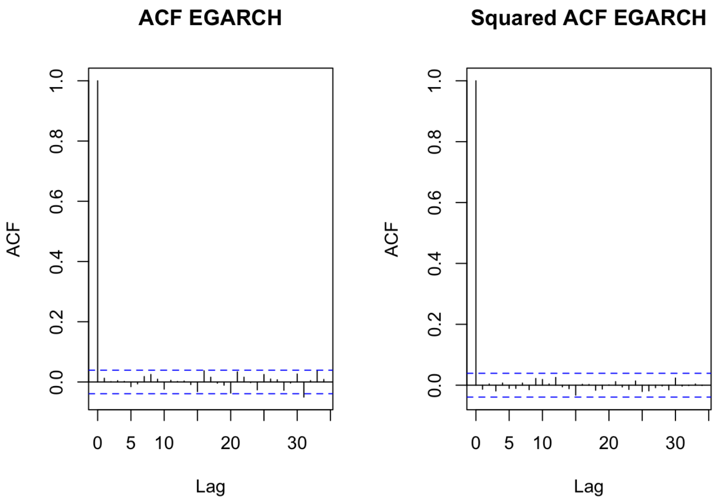

Figure 9: Autocorrelation functions of the standardized residuals for EGARCH(1,1)

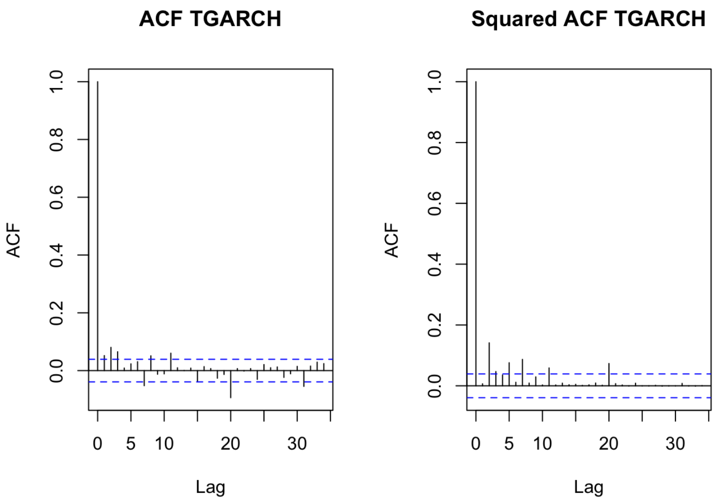

Figure 10: Autocorrelation functions of the standardized residuals for TGARCH(1,1)

6.2 QQ-plot

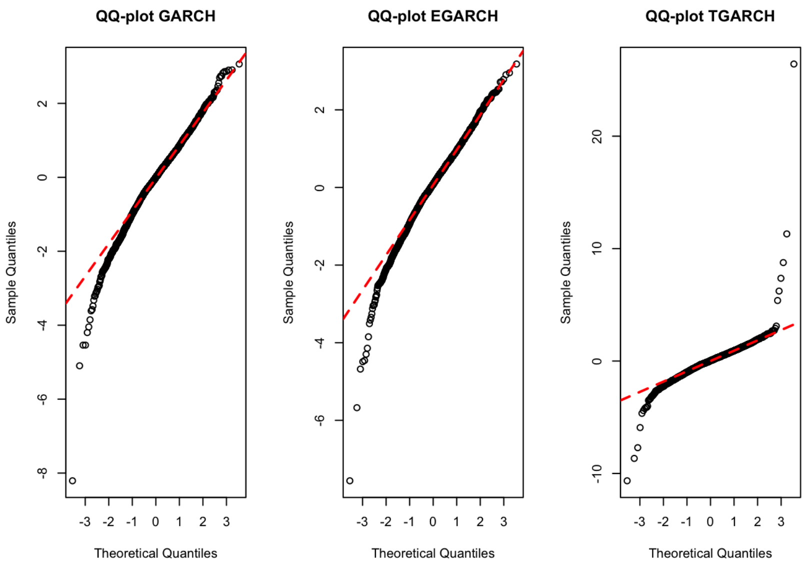

Figure 11: QQ-plot for GARCH, EGARCH & TGARCH with Student-t distribution used on the whole time interval

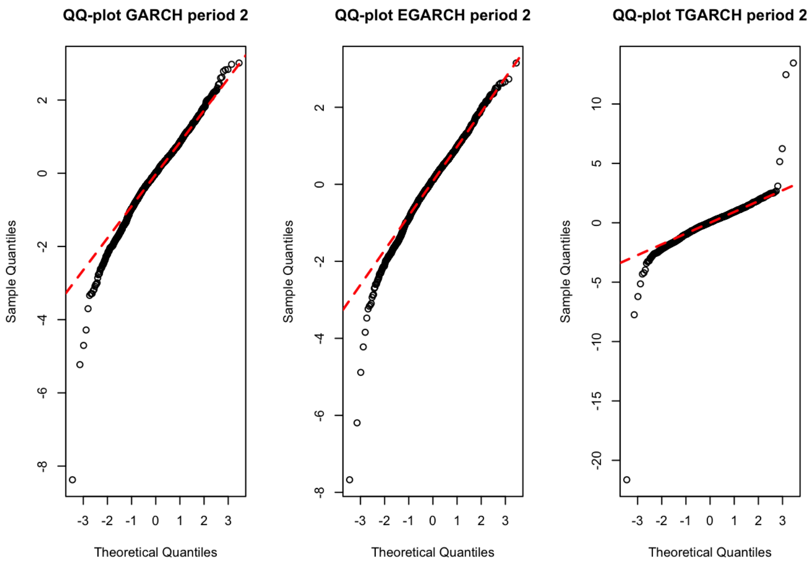

Figure 12: QQ-plot for GARCH, EGARCH & TGARCH with Student-t distribution used on period 2 without the financial crisis

6.3 Period 1

GARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	2.277 E-06	1.742 E-06
α_1	9.798 E-02***	9.147 E-02***
β	8.945 E-01***	9.029 E-01***
AIC	-5.799185	-5.784202
BIC	-5.767826	-5.759115
*** = Statistically significant on 0.1% level		

Table 5: Parameter estimation and statistical test for Period 1

EGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	-0.067647***	-0.048526***
α_1	-0.108492***	-0.094206***
β	0.992841***	0.994789***
θ	0.106202***	0.091662***
AIC	-5.8256	-5.8167
BIC	-5.7817	-5.7791
*** = Statistically significant on 0.1% level		

Table 6: Parameter estimation and statistical test for Period 1

TGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	1.131 E-04	9.082 E-05
α_1	5.853 E-02***	5.059 E-02***
β	9.454 E-01***	9.537 E-01***
γ	9.745 E-01***	1.000 E-01***
AIC	-5.621123	-4.692115
BIC	-5.583493	-4.660756
*** = Statistically significant on 0.1% level		

Table 7: Parameter estimation and statistical test for Period 1

6.4 Period 2

GARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	5.918 E-07*	5.435 E-04**
α_1	9.158 E-02***	9.878 E-02***
β	8.986 E-01***	8.954 E-01***
AIC	-7.374472	-7.322697
BIC	-7.359059	-7.310367
*** = Statistically significant on 0.1% level ** = Statistically significant on 1% level * = Statistically significant on 5% level		

Table 8: Parameter estimation and statistical test for Period 2

EGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	-0.27059***	-0.26323***
α_1	-0.12039***	-0.11749***
β	0.97302***	0.97330***
θ	0.11187***	0.12796***
AIC	-7.4022	-7.3628
BIC	-7.3807	-7.3443
*** = Statistically significant on 0.1% level		

Table 9: Parameter estimation and statistical test for Period 2

TGARCH(1,1)		
Parameters	Student-t	Gaussian
α_0	1.349 E-04***	1.428 E-04***
α_1	6.455 E-02***	7.292 E-02***
β	9.278 E-01***	9.223 E-01***
γ	1.000 E+00***	8.507 E-01***
AIC	-7.342345	-7.002533
BIC	-7.323850	-6.987121
*** = Statistically significant on 0.1% level		

Table 10: Parameter estimation and statistical test for Period 2

7 References

[1] Tim Bollerslev(1986), Journal of Econometrics 31: Generalized Autoregressive Conditional Heteroskedasticity, pp. 307-327

[2] Ruey S. Tsay(2010), Analysis of Financial Time Series, 3rd ed. New Jersey: John Wiley and Sons, Inc

[3] Daniel B. Nelson(1991), Econometrica: Conditional Heteroskedasticity in Asset Returns: A New Approach, Vol 59, pp. 347-370

[4] Glosten, Jagannathan, Runkle(1993), The Journal of Finance: On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, Vol 48, pp. 1779-1801

[5] Chen Zhou(2013), Diagnosing the Distribution of GARCH Innovations

[6] Leonhard Held, Daniel S. Bové(2014), Applied Statistical Inference: Likelihood and Bayes

[7] Ljung, Box(1978), Biometrica: On a measure of lack of fit in time series models, Vol 65, pp.297-303

[8] Tyrcha, Andersson(2012), Notes in Econometrics, Department of Mathematics, Stockholm University

[9] Britton, Alm(2008), Stokastik: Sannolikhetsteori och Statistikteori med Tillämpningar, Liber

[10] Yudi Pawitan(2013), In All Likelihood: Statistical Modelling and Inference Using Likelihood, Oxford