Stockholms universitet

Evaluating the performance of the GARCH(1,1) model. A real life simulated saga.

Sebastian Sjögren

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2020:8
Matematisk statistik
Juni 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2020:8**
http://www.math.su.se

Evaluating the performance of the GARCH(1,1) model. A real life simulated saga.

Sebastian Sjögren*

June 2020

Abstract

Financial data is known to have non-constant variance, also known as volatility clustering, or heteroscedasticity. In this thesis we are concerned with the VaR prediction of the GARCH(1,1) model on both real life data (Nasdaq compsite index), and for a simulated data set. We will assume both a standard normal and a t-distributed error term for the innovation, and the model performance will be evaluated by measuring the fraction of violations, and their independence. For real life data, both model versions suffer from underestimation of the 95% VaR, in particular when considering time periods including financial crises. Both models performs satisfactory on the simulated data, indicating that the modelling and backtesting procedures are working properly. In general the results seem to favour the GARCH(1,1) model with conditionally normal error terms over the t-distribution when predicting VaR.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: sebastian_sjogren@hotmail.com. Supervisor: Mathias Millberg Lindholm, Kristofer Lindensjö, Felix Wahl.

Acknowledgement

I would like to thank my supervisors Mathias Millberg Lindholm, Kristofer Lindensjö and Felix Wahl for their guidance and weekly discussions, which were instrumental in completing my thesis. I would also like to thank my father who took the time to proofread my writing, despite perhaps not understanding the content fully. And finally I would like to thank my fellow students and friends for their support and kind words, for which I am very grateful.

Contents

1 Introduction

Financial data, and in particular exchange rates or the return of assets tend to be serially uncorrelated over time, but have historically been characterized by tranquil and volatile periods. As noted by B.Mandelbrot (1963), it appears that large price changes in the market are usually followed by large changes, and equivalently in periods of tranquility it can be expected that a calm day is followed by another calm day. This behaviour of non constant volatility is normally referred to as *volatility clustering*, or heteroscedasticity.

The phenomenon called volatility clustering may be well known for us today, but traditionally the econometric models assumed a constant one-day forecast variance, which when trying to predict the volatility of tomorrow may cause the models to either over- or underestimate it. R.Engle (1982) further recognizes the heteroscedasticity of financial data, and the issues of using unconditional variance in current models, quoting: "This standard solution to the problem seems unsatisfactory, as it requires a specification of the causes of the changing variance, rather than recognizing that both conditional means and variances may jointly evolve over time". Later in the paper, the model we know today as the ARCH model is suggested, which allow the innovation to depend on its lagged values, and thus evolve over time as a function of past errors.

The ARCH model was later Generalized by T.Bollerslev (1986), as some empirical applications of the ARCH-model turned out to require a large amount parameters in the variance equation. Following this, the extension to the ARCH model was proposed in order to allow for a longer memory and a more flexible lag structure, which lead to what we today call the *GARCH* model. What differentiate the GARCH model from it's precursor is the inclusion of the lagged conditional variance term itself. Making the model more persistent to fluctuations of the innovation, while simultaneously allowing for a more parsimonious model.

We want to evaluate the proposed GARCH model and see how well they can fit the selected data by implementing it on an existing data set. For this thesis we have selected the Nasdaq composite index, which could be considered as a portfolio of stocks as it is an index following the development of multiple separate stocks. The model will be fitted to the data, and evaluated through a series of model diagnostics such as the standardized residuals and coverage tests.

2 Theoretical Framework

This section will cover the necessary theory required for the analysis which has been gathered from Tsay - Analysis of Financial time series 2010 in Chapter 1 and 2 unless otherwise stated.

In most cases when performing financial studies they will involve returns rather than prices of assets. There are mainly two reasons for this, being that for the general investor returns are a complete and scale free summary of the investment, and secondly return series are easier to handle due to more attractive statistical properties.

2.1 Asset returns

Let P_t be the price of an asset at time index t. Holding any given asset from time $t-1$ to t results in a *simple gross return* defined as

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

And corresponding *simple net return*, or only *simple return*

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

Further we consider that $P_t > 0$ and $P_{t-1} > 0$ for all t, which means that $\frac{P_t}{P_{t-1}} > 0$. This means that the gross return has a lower bound $1 + R_t > 0$. If we take the logarithm of the simple gross return we receive what is called the continuously compounded return, or *log return* defined as

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

The log return enjoy some advantages over the simple return, e.g. if we consider a multiperiod return, the log return is simply the sum of the one period log returns involved. Also the log return has more tractable statistical properties. Thus from here on we will be using the log returns solely in our analysis.

2.1.1 Skewness

The third central moment measures the symmetry of the distribution which is called *Skewness*, and in application is used to determine the characteristics of the distribution. The Skewness is defined as

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

and if we let $\{x_1, x_2, ..., x_T\}$ be a random sample of X with T observations, the sample Skewness can be estimated by

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

2.1.2 Kurtosis

The fourth central moment measures the tail behaviour of the distribution, and this measurement is called *Kurtosis*. It is defined as

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

and the quantity $K(x) - 3$ is called *excess kurtosis* because $K(x) = 3$ for a normal distribution. If a distribution has positive excess kurtosis it is said to have fat tails, implying that the distribution tend to feature more extreme values than a normal distribution, and thus, having fatter tails. A distribution featuring these characteristics is said to be *leptokurtic*.

In application this quantity can also be estimated by the sample, and again if we let $\{x_1, x_2, ..., x_T\}$ be a random sample of X with T observations, the sample Kurtosis is

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

2.2 Linear time series

A time series r_t is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty}\psi_i a_{t-i}, \tag{1}$$

where μ is the mean of r_t, ψ_i the weights of r_t and $\{a_t\}$ is a sequence of i.i.d. random variables with zero mean and a well defined distribution. a_t is often denoted the *innovation* or *shock* at time t.

2.2.1 Stationarity

A time series $\{r_t\}$ is said to be *strictly stationary* if the joint distribution $(r_{t_1}, ..., r_{t_k})$ is identical to that of $(r_{t_1+t}, ..., r_{t_k+t})$ for all t where k is an arbitrary positive integer. Meaning that the joint distribution is required to be invariant under time shifts, which can be hard to verify empirically.

Often in application one is more concerned with *weak stationarity*, and a time series $\{r_t\}$ is said to be weakly stationary if both the mean of r_t and the covariance between r_t and $r_{t-\ell}$, $Cov(r_t, r_{t-\ell}) = \gamma_\ell$ are invariant under time shifts where ℓ is an arbitrary integer.

2.2.2 White noise

A time series r_t can be referred to as a *White noise* series if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance. An example of a white noise series is the sequence $\{r_t\}$ where $r_t \sim N(0, \sigma^2)$, which is called a Gaussian white noise series. What can be said for all white noise series is that the ACF's are zero, or in practice, not significantly different from zero, meaning there exist no correlation between the series and its previous values.

2.2.3 Autocorrelation function (ACF)

If we consider a weakly stationary time series r_t where the linear dependence between r_t and its past values are of interest we use the concept of correlation. In this form it is generalized to autocorrelation, and the correlation between r_t and $r_{t-\ell}$ is called the lag-ℓ autocorrelation of r_t. It is usually denoted ρ_ℓ, which is defined as

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

where the property of r_t being invariant under time shift is used, such that $\mathrm{Var(r_t)} = \mathrm{Var}(r_{t-\ell})$. In addition a weakly stationary series is not correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.

For a given realized sample of returns r_t with sample mean $\bar{r} = \sum_{t=1}^{T} r_t/T$, the lag-$\ell$ sample autocorrelation is defined as

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^{T}(r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}.$$

If r_t is a weakly stationary time series that satisfy $r_t = \mu + \sum_{i=1}^{q} \psi_i a_{t-i}$, where $\psi_0 = 1$ and $\{a_j\}$ is a sequence of i.i.d. random variables with zero mean, then $\hat{\rho}_\ell$ is asymptotically normal with mean zero and variance $1/T(1 + 2\sum_i^q \rho_i^2)$ for $\ell > q$ as per Bartlett's approximation, from literature; Box, Jenkins, and Reinsel (2008).

2.2.4 Testing ACF

Previous results from section 2.2.3 can be used in order to test the null hypothesis $H_0 : \rho_\ell = 0$ against the alternative $H_a : \rho_\ell \neq 0$ for any given positive integer $\ell > q$. This is done by using the test statistic

$$z = \frac{\rho_\ell}{\sqrt{(1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_i^2)/T}}.$$

The test statistic z is asymptotically distributed as a standard normal random variable. The decision rule is to reject H_0 if $|z| > Z_{\alpha/2}$ where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard normal distribution.

2.2.5 Ljung-Box test

Ljung and Box proposed another statistic $Q(m)$ in order to jointly test that several autocorrelations of r_t are zero. The null hypothesis $H_0 : \rho_1 = .. = \rho_m = 0$ is tested against $H_a : \rho_i \neq 0$ for some $i \in (1, ..., m)$. The statistic is defined as

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

which under the assumption that $\{r_t\}$ is an i.i.d. sequence, this test statistic is asymptotically χ^2-distributed with m degrees of freedom. The decision rule is to reject H_0 if $Q(m) > \chi_\alpha^2$ where χ_α^2 is the $100(1 - \alpha)$th percentile of the chi-square distribution with m degrees of freedom. It has been suggested through simulation studies that the choice of m should be proportional to the sample size T through $m = ln(T)$ as per R.Tsay (2010 p.33) but during analysis multiple choices of m could be used.

2.2.6 AR Models

Consider a time series data r_t with a significant lag-1 autocorrelation. If we would be interested in making use of the lagged return r_{t-1} in order to predict r_t a simple model that could be used is the Autoregressive (AR)-model. As we in this case only will be considering the lag-1 autocorrelation this is referred to as an AR(1)-model which can be written

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \tag{2}$$

where a_t is assumed to be a white noise series with mean zero and variance σ^2 and one explanatory variable r_{t-1}. What can be noted from equation (1) is that conditional on the previous return r_{t-1} it follows that

$$E(r_t|r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \quad Var(r_t|r_{t-1}) = Var(a_t) = \sigma^2.$$

What this tells us is that given the past return r_{t-1} the mean is centered around $\phi_0 + \phi_1 r_{t-1}$ with a standard deviation of σ. As the AR(1)-model only has one parameter and is solely depending on r_{t-1} this is not always

sufficient to determine the conditional expectation of r_t. Thus this model can be generalized to an AR(p) model with $p > 0$ parameters according to

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t. \tag{3}$$

2.2.7 MA Models

In order to introduce a MA model we will consider an infinite order AR-model with some parameter constraints. In theory this infinte-order model we would entertain may be defined as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + a_t.$$

In practice however, this is not realistic due to it containing infinite parameters, and one way to get around this is by introducing some constraints to the ϕ_i's so they are defined by a finite number of parameters. One idea of this approach is

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t \tag{4}$$

where the coefficients depend on a single parameter θ_1 via $\phi_i = -\theta_1^i$ for $i \geq 1$. For the model in equation. (4) to be stationary we put the constraint that θ_1 must lower than 1 in absolute value, as otherwise the series would explode in value when i goes towards infinity. We can rewrite the model in (4) to compact form as

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \cdots = \phi_0 + a_t. \tag{5}$$

Following this we can also write the model for r_{t-1} as

$$r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \cdots = \phi_0 + a_{t-1} \tag{6}$$

and multiplying equation (6) by θ_1 and subtracting the result from equation (5) we then end up with

$$r_t = \phi_0 (1 - \theta_1) + a_t - \theta_1 a_{t-1} \tag{7}$$

which tells us that apart from the constant term $\phi_0(1 - \theta_1)$, r_t is a weighted average of the shocks a_t and a_{t-1}, hence this model is what we would call a MA(1) model as it only depend on the shocks of today and yesterday.

This model can extended to include higher order lags if necessary, which is referred to a MA(q) model defined by

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}.$$

2.3 Conditional heteroscedasticity models

This section will cover the conditional heteroscedasticity models that will be considered in this thesis. One keyword for these models are volatility, and specifically volatility for an asset return. We will use these models as an approach to estimate Value at Risk (VaR) which is a concept that will be introduced in section 2.5.

Even though volatility is hard to observe, there are some characteristics that are quite common. The first is that volatility tend to come in clusters, meaning that a high volatile trading day is usually followed by another. Equivalently, it also means that a calm day is usually followed by another calm day. Another characteristic is that it normally evolves over time, and big jumps are rather uncommon. Finally it also

appears that volatility does not diverge to infinity, and instead varies within a finite range. Which statistically speaking means that volatility is often stationary.

It also appears that when it comes to financial data, the volatility seem to react differently to price increases and drops, which is called the *leverage* effect. There are some models that attempt to correct for that (for these leverage effects), such as the *EGARCH* model, but it will not be considered in this thesis.

2.3.1 Structure of the models

If we let r_t be the log returns of an asset at time t we can use the basic idea behind volatility studies that the time series $\{r_t\}$ is serially uncorrelated, or with some minor correlation but dependent. To introduce volatility models it is informative to begin by considering a return series r_t and it's conditional mean and variance

$$\mu_t = E\left(r_t|F_{t-1}\right), \quad \sigma_t^2 = \mathrm{Var}\left(r_t|F_{t-1}\right) = E\left[\left(r_t - \mu_t\right)^2|F_{t-1}\right] \tag{8}$$

where F_{t-1} denotes the information set available at time $t-1$. As stated in Tsay (2010, pp.111) after some empirical research the serial correlation is typically weak if it exists at all for an asset return series. If we assume that r_t follows a simple time series model it can be written as

$$r_t = \mu_t + a_t \tag{9}$$

in compact form, where the mean equation μ_t should be simple, such as an AR(p) or MA(q) model. If we combine the equation (8) and (9) we have

$$\sigma_t^2 = \mathrm{Var}\left(r_t|F_{t-1}\right) = \mathrm{Var}\left(a_t|F_{t-1}\right). \tag{10}$$

As the conditional heteroscedastic models are concerned with the evolution of σ_t^2, the goal is to find a dynamic equation that can describe the evolution of the conditional variance over time.

2.3.2 ARCH Model

It was the first model to provide a systematic framework of modelling volatility and it's called a Autoregressive conditional heteroscedastic (ARCH) model. The basic idea of the model is that the innovation a_t is serially uncorrelated but dependent. This dependency can be described by a quadratic equation of its lagged values. Specifically we can write an ARCH(m) model as

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2 \tag{11}$$

where ϵ_t is a sequence of independent and identically distributed random variables with mean zero and variance one. And normally in practice it is assumed to follow a standard normal or Student's t-distribution.

What can be observed from equation (11) is that large past innovations $\{a_{t-i}\}_{i=1}^m$ will result in a large conditional variance of a_t which is in line with our previous statement of volatility clustering. However it should be noted that this does not imply that a large innovation is always followed by another large innovation but rather that the probability for this to occur is higher.

2.3.3 GARCH Models

The ARCH model may be simple but at times can require many parameters in order to describe the volatility process adequately. This led T.Bollerslev to propose an extension of the model known as the Generalized ARCH (GARCH) model. If we let $a_t = r_t - \mu_t$ denote the innovation at time t then it follows a $GARCH(m, s)$ model if

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{12}$$

where ϵ_t is a sequence of independent and identically distributed random variables with zero mean and variance 1 as for the ARCH model. Some further constraints are $\alpha_0 > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0$ and $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$, where we underline that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. These constraints assures that the unconditional variance of a_t is finite while still allowing σ_t^2 to evolve over time. Again as for the ARCH model ϵ_t is often assumed to follow a standard normal or standardized students-t distribution (which are the assumptions that will be used in this thesis).

The simplest form of a GARCH-model is the GARCH(1,1) where the conditional variance of a_t will have the form

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{13}$$

2.4 Model selection

Once a certain type of model has been selected we need to determine what order the model should have. This section will introduce the Partial Autocorrelation function as well as Akaike Information Criteria and Bayesian Information Criteria which can aid in this selection.

2.4.1 Partial Autocorrelation function (PACF)

The PACF of a stationary time series can be useful when selecting the order of a linear time series model. The PACF provides the added contribution of the next lag-term in a model. One intuitive way of introducing this could be considering a set of consecutive AR-models

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

where the terms $\phi_{0,p}$ are the constant terms, $\phi_{i,p}$ the coefficient of r_{t-i} and $e_{p,t}$ the error term of an AR(p)-model. We note that all the models are in form of a multiple linear regression and can be estimated through the least-square method. The PACF for a given model is the highest order $\hat{\phi}_{p,p}$ which holds the added contribution of r_{t-p} to r_t and the selection of model is made with a cut off where the lag-p+1 PACF is no longer significant. Let $\ell > p$, then we note that the asymptotic variance of $\hat{\phi}_{\ell,\ell}$ is $1/T$ where T denotes the sample size (as per R.Tsay pp.47) and also that $\hat{\phi}_{\ell,\ell}$ converges to zero. So when a model is to be selected the lag-p PACF is computed for a set of AR-models and the cut off is made where the sample PACF is no longer significant.

In practice, there can be multiple cut offs making the selection not as straightforward, but one could also use some criteria like AIC or BIC, presented in following section. As well as considering the noise additional parameters may cause in the model.

2.4.2 AIC and BIC

The theory for this subsection has been gathered from Held & Sabanés Bové (2014, Chapter 7).

Both Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are likelihood based methods of evaluating the quality of a model. It is based on the estimated maximum likelihood function

$\hat{\theta}_{ML}$, where some measure of model complexity is integrated in order to allow for comparison of non-nested models. For both AIC and BIC this measure will depend on the dimension p of the parameter vector θ. We define AIC and BIC as

$$\text{AIC} = -2l(\hat{\theta}_{ML}) + 2p \tag{14}$$

$$\text{BIC} = -2l(\hat{\theta}_{ML}) + p\,\log(T) \tag{15}$$

where in both (14) and (15) p denotes the dimension of the parameter vector, and T in (15) denotes the sample size. Both criterion are negatively oriented and we select the model which has the lowest value. We note that both criterion penalizes the number of parameters but BIC also penalizes the number of parameters based on the sample size T as $\log(T) \geq 2, \quad T \geq 8$. What this means in practice is that if we consider a situation where we have multiple models with different number of parameters that are close to the truth, we would like to select the model that is the most parsimonious, i.e. the model with fewer parameters.

An increasing sample size from the unknown population distribution aids in selecting a model that is closer to the truth, but when using AIC this may lead to overfitting due to the penalty term's independence of the sample size while for BIC the complexity penalty increases as the sample size does.

However in practice it is common to use both criterion when selecting a model.

2.5 Value at Risk

This subsection is based on Christoffersen (2012 - chapter 1), and Value at Risk is a measurement that aid in answering a common question within the business industry: Given a portfolio, what is the loss that will only be exceeded $p \cdot 100\%$ of the time where p denotes the probability that the loss is greater than the Value at Risk defined as $P(\text{Loss} > \text{VaR}) = p$.

As we are concerned with log returns in this thesis and will focus on one-day-ahead forecasts, we can define a more specific VaR. First let VaR^p_{t+1} denote the $p \cdot 100\%$ one day ahead Value at Risk, and consider that our portfolio consists of only one asset. If we assume that the logreturns are normally distributed with zero mean and σ_t^2 variance, then

$$\begin{aligned} \mathrm{P}(r_{t+1} < -\,\mathrm{VaR}^p_{t+1}) = p \Leftrightarrow \\ \mathrm{P}(r_{t+1}/\sigma_t < -\,\mathrm{VaR}^p_{t+1}/\sigma_t) = p \Leftrightarrow \\ \mathrm{P}(z_t < -\,\mathrm{VaR}^p_{t+1}/\sigma_t) = p \Leftrightarrow \\ \Phi(-\,\mathrm{VaR}^p_{t+1}/\sigma_t) = p \end{aligned} \tag{16}$$

where $\Phi(\cdot)$, in this particular case, denotes the cumulative density function (cdf) of the standard normal distribution. If we then take the inverse of the cdf $\Phi^{-1}(\cdot)$ of the preceding equation we get

$$\begin{aligned} -\,\mathrm{VaR}^p_{t+1}/\sigma_t = \Phi^{-1}(p) \Leftrightarrow \\ \mathrm{VaR}^p_{t+1} = -\sigma_t \Phi_p^{-1} \end{aligned} \tag{17}$$

which yields the one day ahead VaR given a predetermined probability p and the standard deviation of the asset. It is also possible to compute the VaR for other distributions as well, and where $\mathrm{E}[r_t] \neq 0$ which we show later in equation (22).

2.6 Forecasting method

To generate a forecast we want to use a rolling window approach where we pre specify the rolling window length and then perform out of sample predictions based on in sample estimates of our parameters. In our forecast we will be calculating the one day ahead VaR at each time point t based on the conditional variance σ_t and the distribution assumption for our model. As we will be considering two models where the error term is assumed to follow a standard normal and students-t distribution respectively the conditional standard deviation will be multiplied with the inverse of the cumulative density function to provide us with the VaR.

To provide an explanatory example of how the rolling window forecast will be used, we consider a data set $R_t = \{r_1, r_2, ..., r_T\}, \quad T = 1000$, containing daily log returns from an arbitrary stock. We assume the logreturns are normally distributed with mean zero and unconditional variance σ^2.

Then we have to decide how many observations we want to forecast, the length of our rolling window, how often we want to refit the parameters and for what confidence level p we want to calculate the VaR.

Let us determine that we want to forecast 400 observations, use a rolling window length of 300 and to re estimate the parameters after every prediction. This means that initially, we are going to estimate the parameter for our model based on $\{r_{301}, r_{302}, ..., r_{600}\}$. Given the estimated parameters, and a confidence level, lets say $p = 0.05$, we can then predict the one day ahead Value at Risk for observation r_{600+1} as per

$$\text{VaR}^p_{600+1} = -\sigma_{601}\Phi^{-1}_{0.05}.$$

Once the prediction has been made, the data sample used to estimate the parameters will be moved one step, and the parameters will be re estimated, now using observations $\{r_{302}, r_{303}, ..., r_{601}\}$. This procedure carries on each day until we reach the end of our observations where we now will have a set of predicted VaR values: $\{\text{VaR}^p_{600+1}, \text{VaR}^p_{600+2}, ..., \text{VaR}^p_{1000}\}$.

These predicted VaR's can then be used to evaluate our model performance, and how this can be done will be presented in following section.

2.7 Forecasting evaluation

The theory for this subsection has mainly been taken from Christoffersen (2012 - chapter 13). If any other sources has been used they will be referred to individually.

Once a model has been selected and fitted to our data it is of interest to make sure this model is performing well. To make sure this is the case one can use backtesting which can be seen as a final diagnostic check of the selected model. As mentioned in subsection 2.5, VaR is a measurement which promises that at a pre decided probability p the actual return r_{t+1} will only be worse than the VaR^p_{t+1} forecast $p \cdot 100\%$ of the times. Given our time series we define a "hit sequence" of these violations with an indicator function as

$$I_{t+1} = \begin{cases} 1, \text{ if } r_{t+1} < -VaR^p_{t+1} \\ 0, \text{ if } r_{t+1} \geq -\text{VaR}^p_{t+1} \end{cases}$$

The hit sequence return a 1 on days where the predicted VaR was exceeded, and 0 if it was not violated. If we are using a perfect model we should not be able to predict in advance if the VaR will be violated, meaning that these violations should simply occur with probability p every day. Thus this hit sequence should be completely unpredictable and could be described as a sequence of independent Bernoulli variables that takes the value 1 with probability p. Meaning we can define our null hypothesis as

$$H_0 : I_{t+1} \ \sim \ \text{i.i.d. Bernoulli}(p).$$

which will later be used to evaluate the forecast.

2.7.1 Unconditional Coverage testing

We want to test if the fraction of violations obtained from our risk model, call it π is significantly different than the pre determined fraction p. To do this we write the likelihood of our assumed i.i.d. Bernoulli sequence

$$L(\pi) = \prod_{t=1}^{T}(1-\pi)^{1-I_{t+1}}\pi^{I_{t+1}} = (1-\pi)^{T_0}\pi^{T_1}$$

where T_1, T_0 represent the number of 1's and 0's from the hit-sequence. π is estimated from $\hat{\pi} = T_1/T$, that is the number of violations divided by the number of observations, also the observed fraction of violations from the sequence. We can then plug the maximum likelihood estimate back into the likelihood function to retain the optimized likelihood as

$$L(\hat{\pi}) = (1 - T_1/T)^{T_0}(T_1/T)^{T_1}.$$

And under the unconditional coverage null hypothesis we assume that $\pi = p$, where p is the determined VaR coverage rate, we have the likelihood function

$$L(p) = \prod_{t=1}^{T}(1-p)^{1-I_{t+1}}p^{I_{t+1}} = (1-p)^{T_0}p^{T_1}. \tag{18}$$

With these likelihood functions we can then test the null hypothesis with a standard Likelihood ratio test

$$LR_{uc} = -2\ln[L(p)/L(\hat{\pi})]$$

which we know is asymptotically χ^2-distributed with one degree of freedom, as the number of observations goes towards infinity. We added the indexing uc to differentiate the Likelihood ratio tests we will be using.

2.7.2 Independence testing

As mentioned in the previous section, the likelihood ratio assumes that all violations occur independently, and that each day the probability for the return to be higher than the estimated VaR should be p. The unconditional coverage testing does not take this assumption of independence into account and only assures the amount of violations are not significantly different from the expected number.

Instead, if we consider a poor financial model that does not capture the conditional heteroscedasticity to full extent, we could assume most violations would come in a cluster, as the model would not be able to capture the increase in volatility. This is clearly unsatisfactory, as the hit-sequence should consist of independent Bernoulli(p) trials. And if we think in financial terms for a company, it would most likely be worse to have all violations come at once, rather than scattered over the full period.

So in order to test if these violations occur independently we first consider a situation where the hit-sequence is dependent. This can be described by a first-order Markov sequence with transition probabilities

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{19}$$

where $\pi_{01} = P(I_{t+1} = 1|I_t = 0)$ denotes the probability of tomorrow being a violation given that today was not, and $\pi_{11} = P(I_{t+1} = 1|I_t = 1)$ in a similar way denotes the probability of tomorrow being a violation given that today was also a violation. The first-order markov property states that only the outcome of yesterday matters for the outcome of today. Which means that the matrix in equation (19) covers all possible transitions.

If we again consider a hit-sequence with T observations, we can write the likelihood function of the first-order markov process as

$$L(\Pi_1) = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}, \tag{20}$$

where T_{ij}, $i,j = 0,1$ is the number of observations that given the starting state i, is followed by state j. Taking the first derivative with respect to π_{01} and π_{11}, and putting these equal to zero we can solve for the maximum likelihood estimates

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}$$
$$\hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}.$$

Recall that by allowing a dependence in the hit-sequence, it corresponds to allowing π_{01} to be different from π_{11}. However, under the null hypothesis the violations should occur independently, meaning that $\pi_{01} = \pi_{11} = \pi$, giving us the transition matrix

$$\hat{\Pi} = \begin{bmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{bmatrix} \tag{21}$$

.

Once this has been established we can test the independence hypothesis by again using a likelihood ratio test

$$LR_{ind} = -2\ln[L(\hat{\Pi})/L(\hat{\Pi}_1)] \sim \chi^2(1).$$

2.7.3 Conditional Coverage testing

Finally what we want to accomplish is a simultaneous test that can assess whether the VaR violations are independent and the average number of violations correct for the selected VaR-level. This can be done by jointly testing for independence and correct coverage by using a conditional coverage test

$$LR_{cc} = -2\ln\left[L(p)/L\left(\hat{\Pi}_1\right)\right] \sim \chi^2(2).$$

We note that this test combines the null hypothesis from the Unconditional Coverage test with the alternative hypothesis from the Independence test by using the likelihood in equation (18) with the likelihood established in equation (20).

2.7.4 Testing for higher order dependencies

The backtesting methods we have defined only bothers to validate the independence for violations in $[VaR_t^p, VaR_{t+1}^p]$, $\quad t = (0,1,2...T-1)$, meaning we can only make sure the validations are not correlated day by day. However as per Christoffersen (2012 - page 306) we can use the ACF defined in section 2.2.2 to assess whether there exists dependencies in the hit-sequence or not by plotting it as any lag-ℓ dependency will then show a rise in ACF plot.

It is also possible to use the Ljung-Box test statistic defined in section 2.2.5 to formally test if the autocorrelation between lag $1, 2..., m$ are jointly zero.

3 Methodology

This section will cover where data was gathered, how models were created and which software was used. The layout of this thesis, in regards to sections and subsections was inspired by Robert Cedergren - Forecast evaluation of 1-step-ahead predictions using GARCH(1,1) on the Euro/US Dollar FX Spot Rate (2018).

3.1 Data

The data was downloaded from finance.yahoo.com on $2019-03-05$ where the Nasdaq Composite Index was selected providing daily data during the time period 1989-2019. Available for use in the data were multiple variables, such as *Date*, *Opening Price*, *Closing Price*, *Highest Price*, *Lowest Price* and *Traded Volume*.

The data set contains 7558 observations, and thus 7557 daily log returns as they are computed through $ln\left(\frac{P_t}{P_{t-1}}\right)$ where P_t for our analysis has been selected as the Closing price of the day.

3.2 Model building

When building a volatility model there are a few general steps to follow as per Tsay (2010 - page 113).

1. First we want to specify a mean equation by testing for serial dependence in the data with ACF's and, if necessary fit a suitable econometric model to remove any linear dependence. Normally for log return series there are no or minor serial correlation and the mean equation may often be reduced to the sample mean but there are examples where it's necessary to fit another model.
2. Once a mean equation has been specified we can use the residuals of the mean equation, i.e $a_t = r_t - \mu_t$ to test for ARCH effects.
3. If ARCH effects are statistically significant we may specify a volatility model and perform joint estimations of the mean and volatility equations.
4. Once a model has been fitted check and if necessary adjust it.

3.3 Tools used

This thesis and it's content such as figures, tables and models has been written and created with the open end software R and R studio. In particular the packages rugarch, fGarch and forecast has been used for statistical modelling while some of the figures has been created with the use of ggplot. Coverage testing functions described in section 2.7 have been created by the author by implementing suggested methods from Kupiec and Christoffersen.

3.4 Reasoning behind methods chosen

The reasoning in this section was inspired by Chapter 1 in Christoffersen - Elements of Risk Management (2012).

What drives the corporate firm to interest in risk management when, if we for example look at classic portfolio theory this tells us that investors can eliminate asset-specific risk by diversification. Which by putting everything else aside mean the investor does not care about firm-specific risk . Thus not giving any incentive for a firm to invest resources into risk management. This goes in line with the famous Modigliani-Miller theorem which says the value of a firm is independent of its financial structure; and should maximize its expected profits regardless of the risks, since the investor can reduce risk by portfolio diversification.

However in practice, the strict conditions required to uphold the Modigliani-Miller theorem are frequently violated as it is, as far as I know, not possible to avoid taxes which is one rule the theorem is founded on.

Other more realistic descriptions to why a firm should bother with risk management are

- *The cost of bankruptcy.* There are both direct costs, and indirect costs and they are large. The real cost being if an actual bankruptcy occur, and the indirect cost amounting to the fact if an investor consider

there is a nontrivial possibility of bankruptcy, the potential cost will reduce the current valuation of the company. Thus, meaning risk management can increase the value of the company by lowering the odds of not going bankrupt.

- *Compensation packages.* Due to the nature of humans being to some various degree, risk averse, there is an implied cost for a risky company to keep human capital and other key employees. Meaning that the riskier the company, the more compensation would be required in return to stay. Leading us to the conclusion again that by reducing risk, the cost may be reduced and thus the value of the company.

There seem to be consensus that risk management improve firm performance as researchers have found that reducing cash flow also leads to reduced costs of capital, allowing for more investment. There seem to be some evidence from the corporate sector where the benefits of risk management can be clearly seen. While if looking at some recent studies of the risk management at some of the largest US banks, it appears to be not as good. On average the risk forecast seem to be overly cautious, while the realized losses far exceed the risk forecasts. It is also stressed that not only does the losses far exceed the risk forecast, they also occurred consecutively. This serial dependency is a motivation for development and implementation of tools that can manage this, and among those tools are the GARCH model, which aim to capture this serial dependency by conditioning on past volatility.

4 Analysis

4.1 Data exploration

This section will provide an exploratory introduction to to familiarize the reader with the data and its properties. The data set consists of the closing price of Nasdaq composite index, from which we have derived the log returns, in this thesis known as r_t which indicates the daily percentage change of the index. Just to get a grasp of how the index has moved over the times we will display three plots where the closing price, the log return as well as the squared log return are plotted against time.

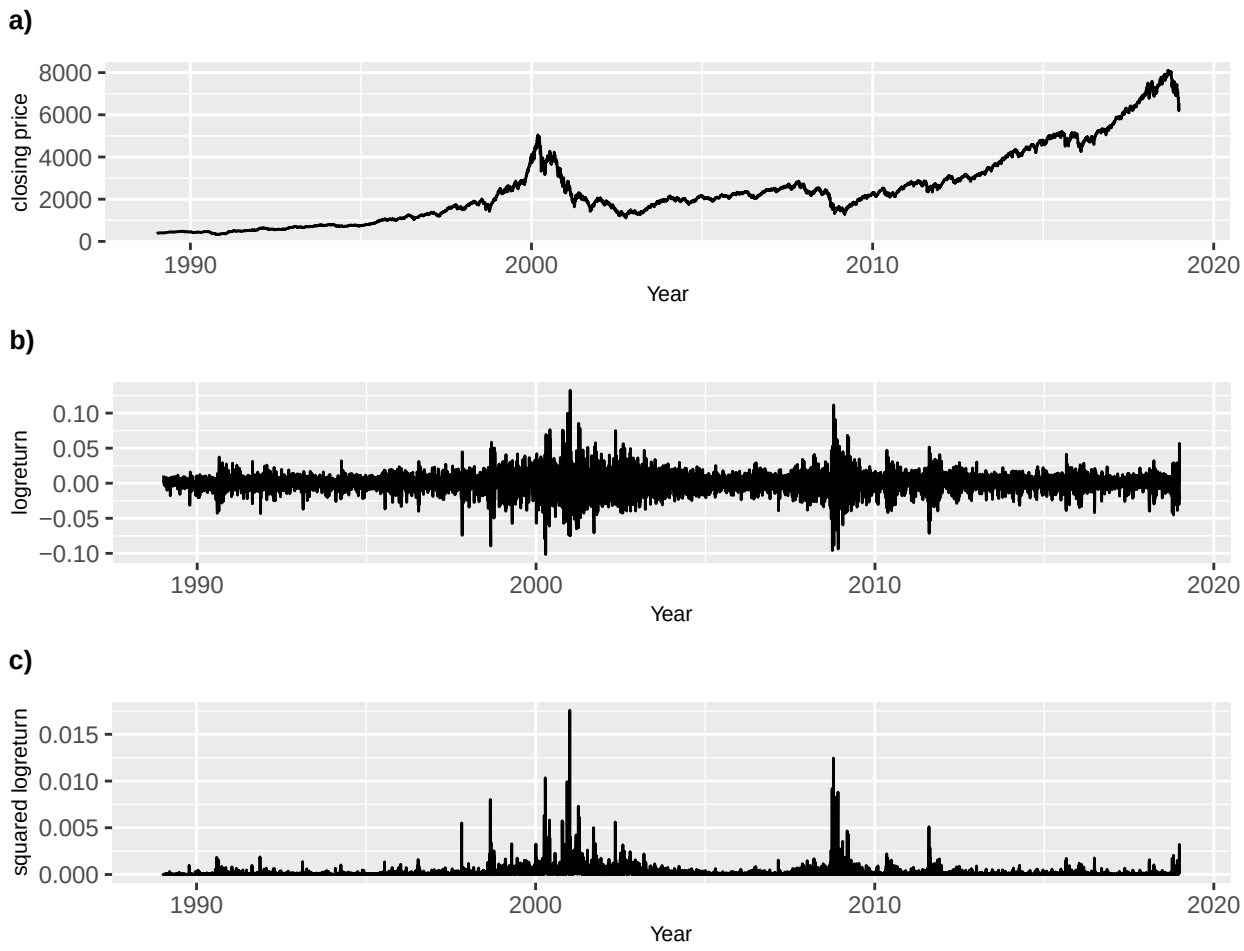

Figure 1: Time plots of a) Nasdaq closing price over time, b) Nasdaq logreturns over time and c) Nasdaq squared logreturns over time during the period 1989-2018

Figure 1 provides a visual illustration of our data set, showing how the Nasdaq index has moved over time, daily logreturns and the squared log returns. As many may be familiar with there were some financial difficulties during 2000-2002 due to the dot-com bubble, as well as the global financial crisis during 2008. This is well represented in the plots where in plot a) we can see a significant price drop in the stock while plot b) and c) shows spikes in the logreturn in comparison to the calmer period in between.

As we in this thesis are concerned with modelling the volatility process through GARCH models, these spikes are of particular interest as they indicate the data contains ARCH-effects, which is the foundation of why these models are being used. The ARCH effects will be confirmed later through ACF plots and the Ljung-Box

test, but for the sake of knowledge some further descriptive statistics will be presented.

Mean	Variance	Standard deviation	Skewness	Kurtosis
0.00038	2e-04	0.0142	-0.11368	9.65656

Table 1: Statistical properties for the logreturns of Nasdaq index.

Table 1 presents the statistical properties of the log return series $\{r_t\}$, where we can see that the mean is approximately zero and slightly negatively skewed. The kurtosis is 9.65656 which means we have an excess kurtosis of 6.65656 confirming our data being leptokurtic and more fat-tailed than the normal distribution, thus featuring more extreme values. This does not say anything regarding the data and whether it is serially correlated or dependent at all, which we need to deal with prior to modelling. If we recall equation (9) and assume our data can be written as $r_t = \mu_t + a_t$ we can re write this expression as $a_t = r_t - \mu_t$, where μ_t denotes the mean equation. As an initial step, in accordance to section 3.1, we want to test if our data is serially correlated which will be done by ACF plots and solidified by Ljung Box test.

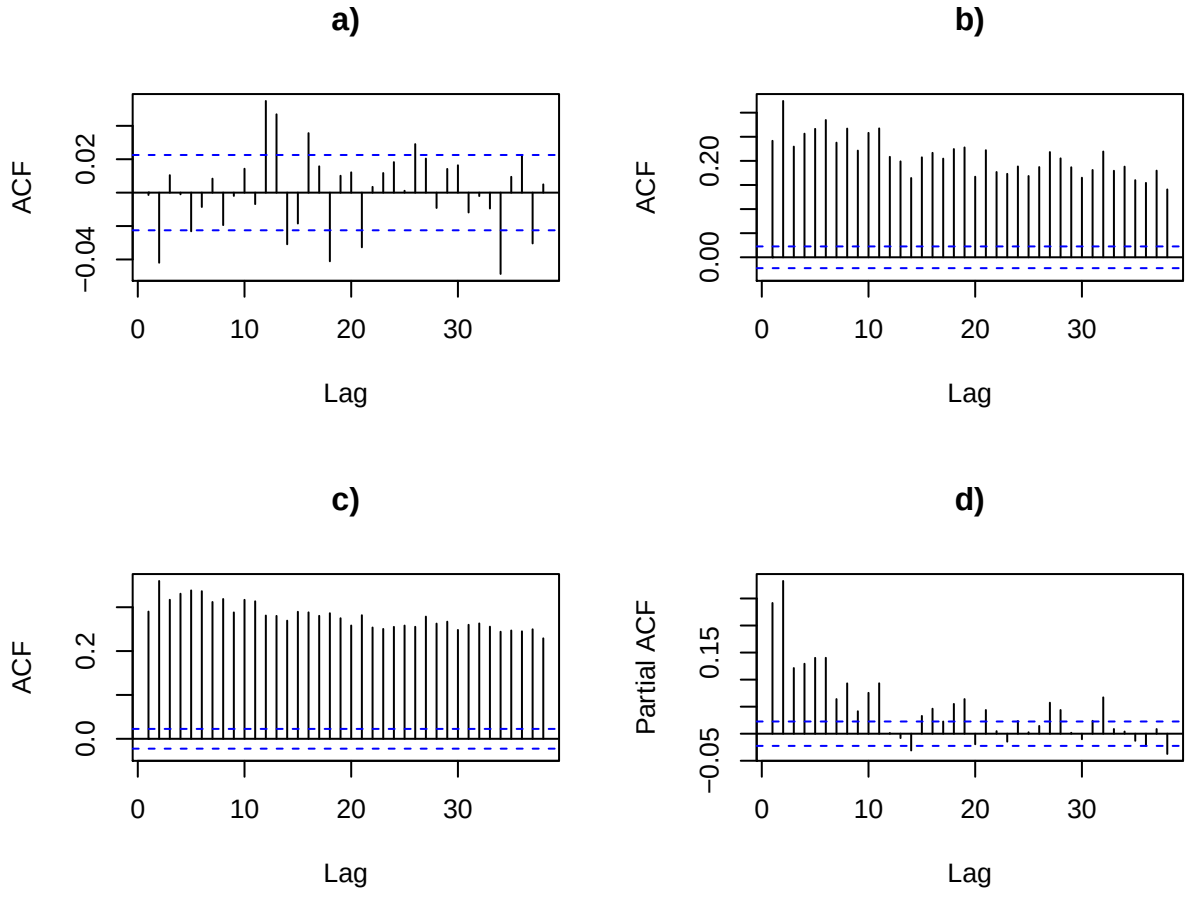

Figure 2: ACF and PACF plots for the log returns of the Nasdaq composite index during the time period 1989-2018 of a)ACF plot of the log return serie, b)ACF plot of the squared log return serie, c)ACF plot of the absolute valued log return serie and d)PACF of the squared log return serie

Figure 2 shows the auto-correlation functions for the log return series and in order for us to make the

assessment that our data series is serially uncorrelated it should contain the majority of auto-correlation within the dashed blue lines. Figure 2 *a*) is however displaying a number of auto-correlations exceeding this which would violate the assumption that our series is serially uncorrelated. Figure 2 *b*) which contains the ACF's for the squared log return series is indicating a strong existence of an ARCH effect. To further solidify this statistically we will use the Ljung-Box statistic Q(m) to test the null hypothesis that the first m lags are serially uncorrelated for both the return series and the squared return series. The selection of how many lags m to be checked is made through the method mentioned in section 2.2.6, which is $m = \ln(T)$ where T is the number of observations we have.

Series	Q(m)	p-value
Log return	22.132	0.008
Squared log return	4249.609	0.000

Table 2: The test statistic Q(m) and p-value for respective series.

Table 2 contains the Ljung-Box statistic and corresponding p-value. Since the p-value < 0.05 we reject the null hypothesis that the return series $\{r_t\}$ contains no serial correlation. The squared returns show even stronger correlation with the p-value close to zero which means the series is exhibiting strong ARCH effects. So, in order to move on with the process of modelling the volatility, we first need to address the fact that our data is serially correlated. As mentioned in R.Tsay (p.114), the serial correlation for most asset return series are weak, if any, and it is suggested that building a mean equation would amount to simply remove the sample mean from the series, if the sample mean would show to be significantly different from zero. To check if this is the case or not we will use a one sample, two sided t-test under the null hypothesis that the mean equals zero,

$$H_0 : \mu = 0 \quad \text{against} \quad H_a : \mu \neq 0.$$

The t-ratio statistic is calculated through

$$\text{t-ratio} = \left| \frac{\hat{\mu} - \mu}{s/\sqrt{T}} \right| = 2.314 > t_{0.975}(T-1) = 1.96$$

where $\hat{\mu},\ s,\ T$ are the estimated sample mean, standard deviation and population size respectively. The null hypothesis states that the sample mean equals zero and is rejected at the 95% level but not at the 99% level. Not disclosed here, but in the Appendix an attempt to remove the serial correlation in the data by simply subtracting the sample mean from the return series was made. Through the ACF plots in Figure 10, and the Ljung Box test in Table 18 it was concluded that it was not enough, indicating the mean of the series is dependent of previous values and thus we will attempt to specify an econometric model that can aid in removing the serial correlation.

The selection of the econometric model to function as our mean equation has been made through the open end R software, and in particular the package forecast and method *auto.arima*. The function uses an AIC based selection where some initial models are considered and then a step-wise algorithm is used to find the model with the lowest AIC. After specifying a MA(2) model we will again plot the ACF's to determine if this is enough to remove the serial correlation our data previously exhibited. We remind ourselves that we are examining the residuals of the mean equation, i.e. $a_t = r_t - \mu_t$ where $\mu_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ is the representation of the MA(2) model.

Mean	Variance	Standard deviation	Skewness	Kurtosis
-3e-08	2e-04	0.01418	-0.16838	9.55258

Table 3: Statistical properties for the residuals of the fitted MA(2)-model to the logreturns from Nasdaq composite index.

A summary of the descriptive statistics for the residual series is presented in Table 3 and we note that the mean of the residual series is smaller and very close to zero now. Performing a t-test of this series under the null hypothesis that the sample mean equals zero provides a p-value of 0.999 which is satisfactory. Apart from that, the variance and standard deviation have not changed much, the data is slightly more negatively skewed but nothing major and the series still exhibits excessive kurtosis.

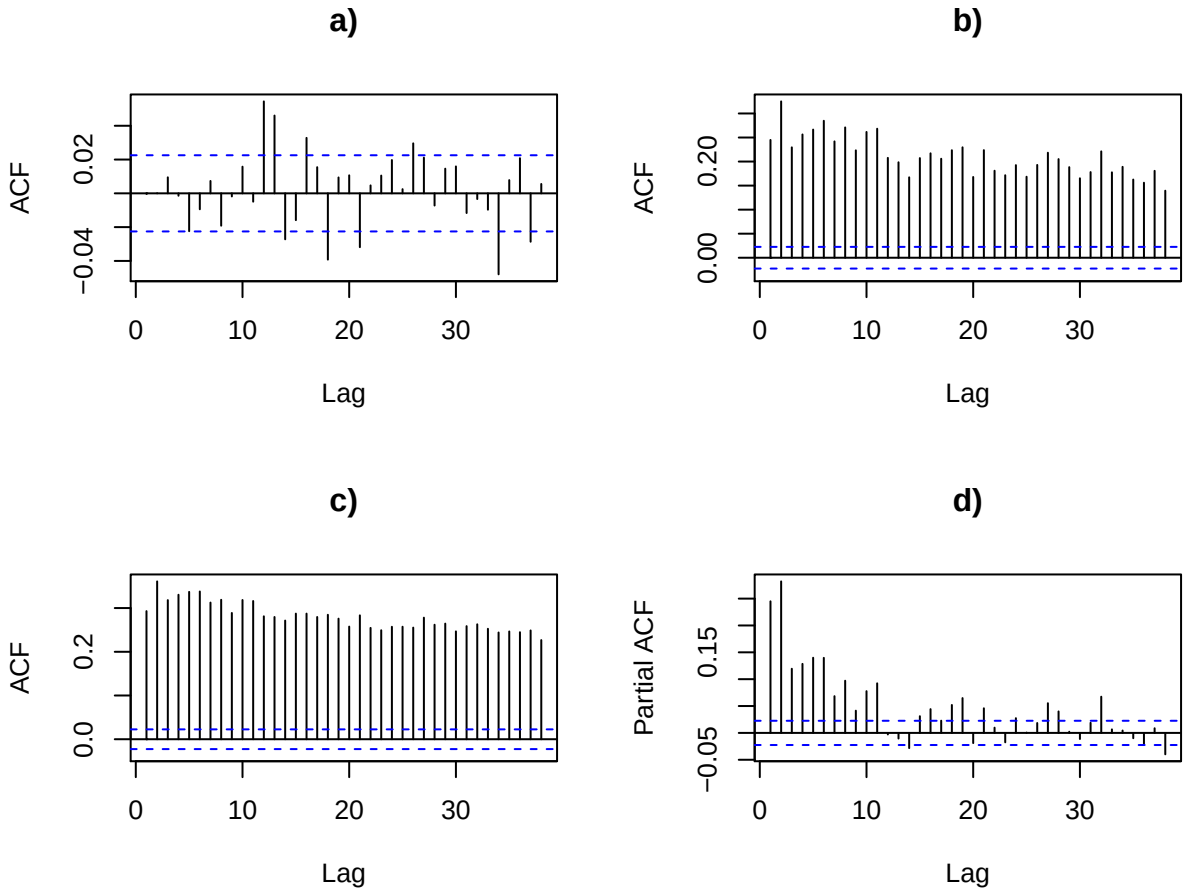

Figure 3: Autocorrelation plots containing the following series: a)residuals of the MA(2) model, b)squared residuals of the MA(2) model, c)absolute value of the residuals from the MA(2) model, d)PACF of the squared residuals

Figure 3 holds the ACF plots for the residuals of the mean equation where, if we take a closer look at plot a), it appears the implementation of the MA(2) model successfully has reduced the serial correlation. The Ljung-Box test statistic in Table 4 further solidifies the removal of serial correlation as the p-value for the residual series $\{r_t - \mu_t\}$ is now close to 0.5 and thus, we cannot reject the null hypothesis that data is not serially correlated at the 95% level anymore. However, as was quite clearly visible in subplot b) and c) from Figure 3, the Ljung-Box test verifies the presence of ARCH effects from the squared residuals $\{a_t^2\}$, which indicates that fitting a heteroscedastic model to the data should be a suitable approach.

Series	Q(m)	p-value
Residuals of Mean Equation	8.349	0.492
Squared Residuals of Mean Equation	4298.271	0.000

Table 4: The test statistic Q(m) and p-value for respective series.

4.2 GARCH modelling

As the preparation of the data has been taken care of by implementing a MA(2) model we are now ready to fit the GARCH(1,1) model. We ended the previous section by establishing it would be a good idea to fit a heteroscedastic model to the residuals of the mean equation as this series contains the presence of ARCH-effects. We remind ourselves from equation (13) of the form of the GARCH(1,1) model

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

where we will be implementing two separate models where the ϵ_t's are assumed to follow a standard normal distribution as well as a standard students-t distribution. To motivate the selection of two models with separate underlying distributions for the error term we can look at Calzolari. G et al. (2014). In the introduction of the article it is mentioned that the distribution of an asset return series may remain leptokurtic, even after controlling for volatility clustering. What this means is that even after applying for example a GARCH model to some arbitrary asset return series, in some cases it still exhibit volatility clustering. This indicates that the distribution for the conditional variance provided by the model is more light tailed than the distribution for the innovation. Two methods are mentioned to handle this where the second one consists of specifying a leptokurtic distribution for the error term, such as the students-t distribution. It is later concluded that the model with underlying t-distribution provided a better fit than the Gaussian one, and in addition it also provided more reliable quantiles for VaR calculations.

The degrees of freedom for the students-t distributed error term will not be pre-specified and will be jointly estimated together with the model. As we have modelled a mean equation for our log returns with the MA(2) model we can now consider the innovation to be the residual of the mean equation $a_t = r_t - \mu_t$ where $\mu_t = c_0 - \theta_1 a_{t-1} - \theta_2 a_{t-2} + a_t$ takes the form of the MA(2) model.

Table 5 contains the estimated coefficients, the standard error and their p-values, where we can observe that all parameters are significant at the 95% level for both models apart for θ_2 and α_0. For where we have assumed students-t distributed error terms the parameter θ_2 is almost significant, but one could consider dropping these parameters and evaluate if this would result in an improvement of the model.

	MA(2)-GARCH(1,1)-N			MA(2)-GARCH(1,1)- t		
Parameter	Estimate	Std Error	p-value	Estimate	Std Error	
c_0	8e-04	1e-04	0	0.0010	0.0001	0.0000
θ_1	0.0462	0.0125	2e-04	0.0475	0.0118	0.0001
θ_2	-0.0125	0.0129	0.3318	0.0220	0.0118	0.0615
α_0	0	0	0.6603	0	0	0.5739
α_1	0.098	0.0456	0.0317	0.0909	0.0385	0.0183
β_1	0.8903	0.0532	0	0.9077	0.0355	0.0000
ν	-	-	-	7.1689	1.0119	0.0000

Table 5: Summary of parameter estimations, standard error and p-values of GARCH(1,1) model with normal and t-distributed error terms of log returns from Nasdaq composite index 1989-2019.

For our two models $\{\epsilon_t\}$ is assumed to follow a standard normal respectively students-t distribution with mean zero and variance 1 and these error terms should behave like a white noise series. If we again remind ourselves of the representation of the innovations $a_t = r_t - \mu_t$, and also that $a_t = \sigma_t \epsilon_t$ in accordance to the GARCH framework it follows

$$
\begin{aligned}
a_t = r_t - \mu_t, \quad a_t = \sigma_t \epsilon_t \\
\Rightarrow r_t - \mu_t = \sigma_t \epsilon_t \\
\Rightarrow \epsilon_t = \frac{r_t - \mu_t}{\sigma_t} = \frac{a_t}{\sigma_t}
\end{aligned}
$$

where r_t is the observed log return, μ_t a stochastic value estimated by our MA(2)-model and a_t is the innovation consisting of the conditional standard deviation σ_t and an error term ϵ_t. What this means is that the ϵ_t's can be estimated by $\hat{\epsilon}_t = \frac{a_t}{\sigma_t}$ which is commonly referred to as the standardized innovations and denoted $\tilde{a}$. These innovations will be used for further validating our model and since they should behave like a white noise series. This can be done by again utilizing ACF plots as well as the Ljung-Box test to verify there still does not exist serial correlation from standardized innovation series $\{\tilde{a}_t\}$ as well as verifying that we successfully have removed it from the squared series $\{\tilde{a}_t^2\}$. In the theory section 2.3.3 for the GARCH model we saw that ϵ_t is assumed to have zero mean and variance one. Presented in Table 6 are some descriptive statistics for the standardized innovations for both models and we note the mean is close to zero and the standard deviation is also approximately 1 as was suggested in the framework for a GARCH-model.

	MA(2)-GARCH(1,1)-N	MA(2)-GARCH(1,1)-t
Mean	-0.0331412	-0.0576483
Standard deviation	0.9986451	0.99715
Skewness	-0.4947936	-0.5513839
Kurtosis	4.5547595	4.8692942
Excess Kurtosis	1.5547595	1.8692942

Table 6: Descriptive statistics for the standardized innovations from both MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t model

After conducting a two-sided t-test where we tested the null hypothesis that the sample mean equals zero

$$
H_0 : \mu = 0, \quad H_a : \mu \neq 0
$$

the statistics -2.885 and -5.026 are derived for the standardized normal and standardized t-distributed innovations respectively. They are to be compared with the critical value 1.96 when tested at the 95% level, from which it follows we reject the null hypothesis, even at the 99% level. It appears both models are showing some negative skewness as well as some excessive kurtosis, which indicates heavy tails, and in particular the lower tail and will later be verified in Figure 5 with a QQ-plot.

Figure 4 displays the ACF's for the standardized innovations $\{\tilde{a}_t\}$ and the squared standardized innovations $\{\tilde{a}_t^2\}$ for both models. We see there are some minor serial correlation in the squared series at lag 2 and 10 but it's not possible to draw any conclusions from the plots only. The Ljung-Box statistic is computed and presented together with the p-value in Table 7 which confirms these series does not contain any serial correlation.

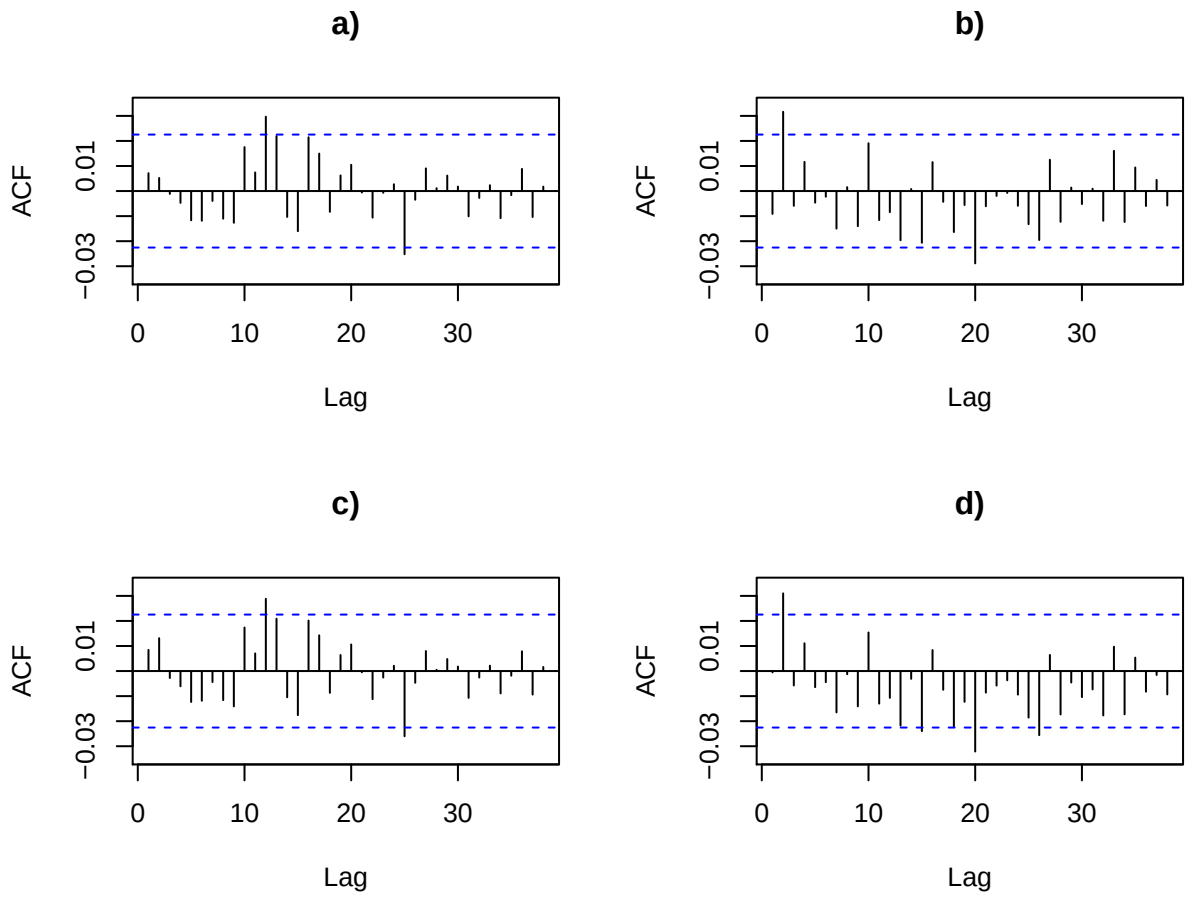

Figure 4: ACF-plots of the Standardized innovations for daily log returns of Nasdaq composite index 1989-2019. Plot a) and b) contains the ACF of the standardized innovation series as well as the squared standardized innovations of the normal distributed series. Plot c) and d) contains the standardized and squared standardized innovations for the t-distributed series.

Series	Q(m)	p-value
MA(2)-GARCH(1,1)-N Standardized residuals	3.906	0.915
MA(2)-GARCH(1,1)-N Standardized squared residuals	11.410	0.244
MA(2)-GARCH(1,1)-t Standardized residuals	5.563	0.777
MA(2)-GARCH(1,1)-t Standardized squared residuals	10.982	0.271

Table 7: Ljung-Box test statistic and corresponding p-value for the standardized residuals and the squared standardized residuals for both MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t models.

As a final examination of our model we will plot the sample quantiles against the theoretical quantiles for both series. The MA(2)-GARCH(1,1)-N model will be plotted against the theoretical quantiles of the N(0,1) distribution and MA(2)-GARCH(1,1)-t against the theoretical quantiles for a Students-t distribution with 7.2 degrees of freedom as was estimated for the model. Figure 6 displays the QQ-plots for the standardized innovations and while we can see that the standardized innovations from both models appear to have heavy tails in the lower quantile, the MA(2)-GARCH(1,1)-t model seem to be light tailed in the upper quantile.

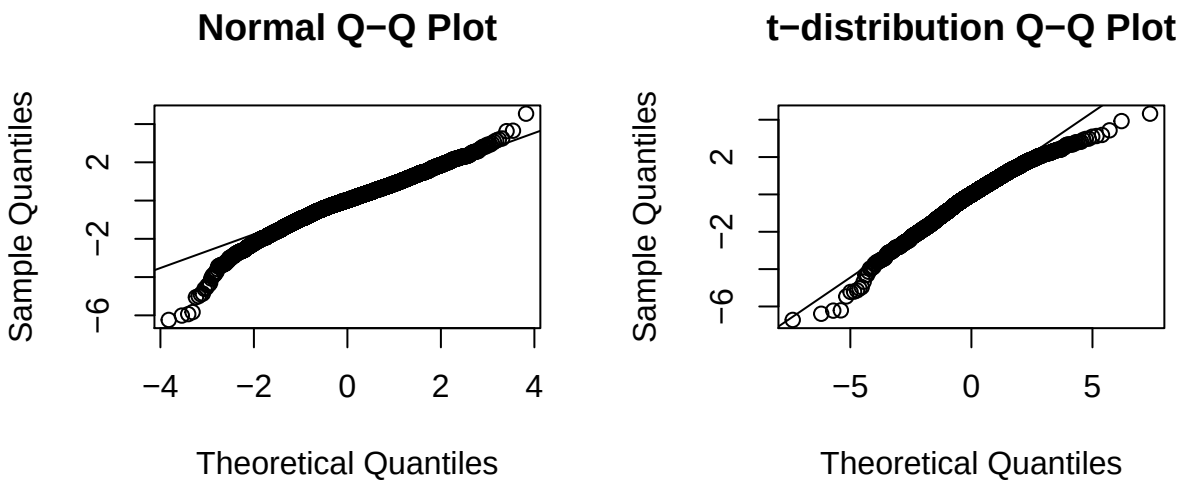

Figure 5: QQ-plots for the standardized residuals of the MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t model respectively.

4.3 Forecasting

This subsection will cover the one-day ahead rolling window forecast. It will be performed as described in section 2.6 where we initially perform an in sample estimation of our model parameters and then use the rolling window to predict the one-day-ahead VaR. We will be using the entire data sample of 7557 daily log returns and perform three separate rolling window forecasts with different length for out of sample predictions. The length of our out of sample intervals will be 1800, 3000 and 6000 trading days in which we will obtain the VaR. All sample intervals have the same end date but naturally span over different time periods, which are:

- Sample interval with 1800 trading days: 2011-11-02 - 2018-12-28
- Sample interval with 3000 trading days: 2007-01-31 - 2018-12-28
- Sample interval with 6000 trading days: 1995-03-02 - 2018-12-28.

If we recall Figure 1 from section 4.1 we know that during the years 2000-2002 and 2008-2009 we suffered a global financial crisis which led to increased volatility in the stock market. The shortest interval does not contain any particular period of extreme volatility, while the medium interval contains one and the longest two. We remember this for future references as the inclusion of a financial crisis may have some impact on our models prediction capabilities.

It would be possible to calculate a symmetric prediction interval in order to cover both the maximum and minimum logreturns for each trading day. But we will be focusing on the lower part of the interval which probably would be considered more valuable to correctly model in finance. We want to calculate the VaR's which in theory should allow us to with a 95% security say that an eventual loss tomorrow will not be greater than our estimated VaR.

The window length has been selected quite arbitrarily to include 500 observations for each parameter estimation, and the parameters will be refitted each day prior to calculating the VaR. Figure 6 contains the plots where in black we can see the realized values of the daily log returns while the plotted line in red corresponds to the model with a standard normal distributed error term and the line in blue to the models with standardized t-distributed error term.

If we remember the motivation for including two models with separate underlying distributions, it was

because in some cases using just a standard gaussian distribution was not enough to account for the volatility clustering exhibited by the data. If we combine this knowledge with our general knowledge of the students-t distribution being more fat-tailed than the standard gaussian one, we would expect the predicted VaR's from the MA(2)-GARCH(1,1)-t model to be smaller and thus putting the blue line below the red. However this does not seem to be the case from Figure 6 and as this was not entirely expected and the predicted VaR's were gathered from the R-package 'rugarch' and package function 'ugarchroll' the VaR's where manually calculated as well. To manually compute the VaR we followed the steps from equation (16) and (17) where we swapped the cumulative density function for the students-t instead of the standard normal giving us

$$\text{VaR}_{t+1}^{0.05} = \mu_t - \sigma_t \Psi_{0.05}^{-1} \sqrt{\frac{\nu}{\nu - 2}} \tag{22}$$

where μ_t and σ_t denotes the expected value and conditional standard deviation at time t, $\Psi_{0.05}^{-1}$ the cumulative density function for the students-t distribution at the 5% quantile and $\sqrt{\nu/(\nu-2)}$ is the standard deviation for a t-distribution with ν degrees of freedom. This has to be multiplied with the quantile of the cdf in order to standardize the VaR as we want t-distribution to have variance 1.

After calculating the VaR manually we verified that the values provided by the package are indeed correct. By selecting a more restrictive confidence level for the VaR, such as 99% we find that the the VaR for the student-t model then exceeds the standard normal model.

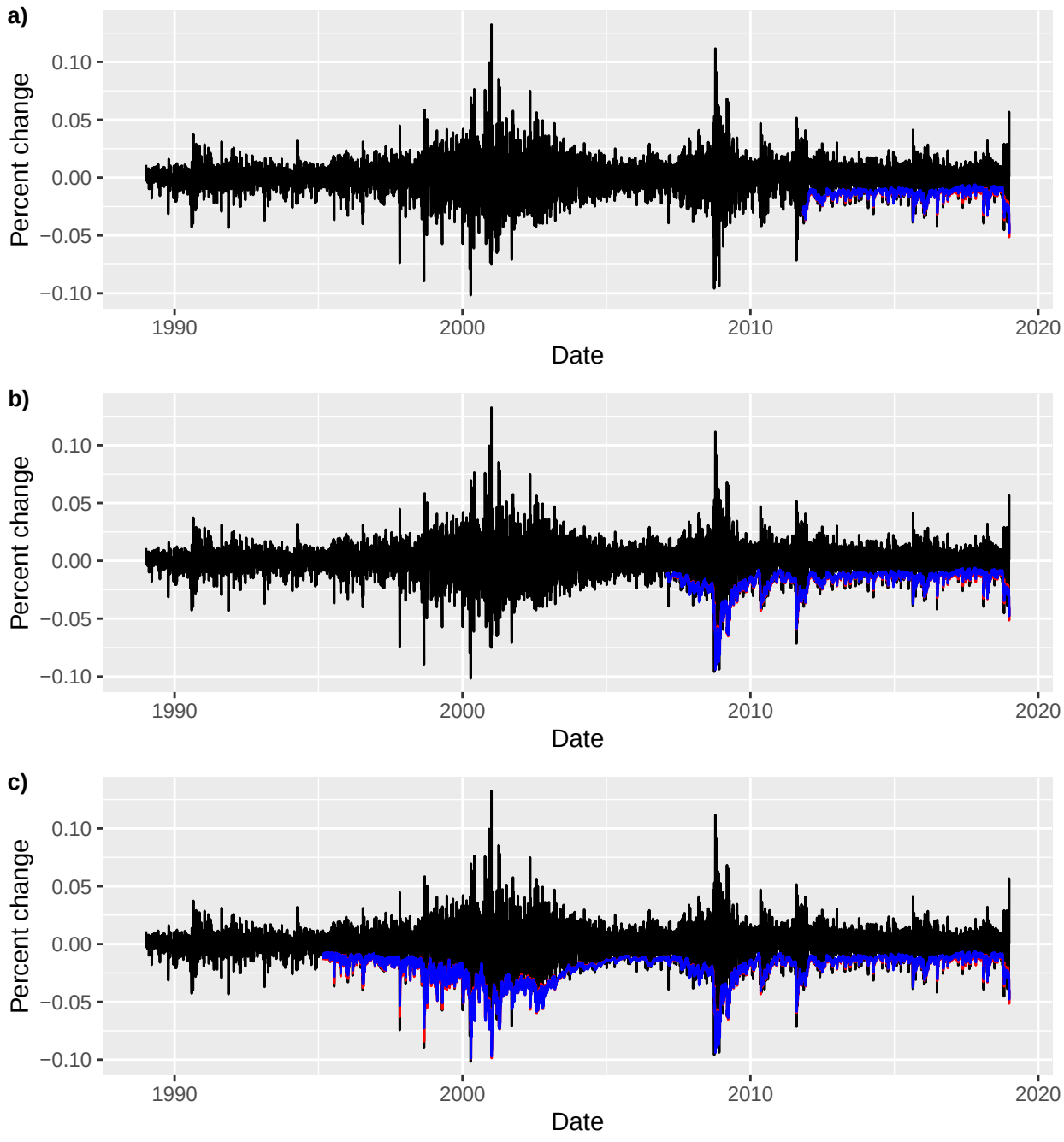

Figure 6: Plots for predicted VaR represented by red and blue line for MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t model respectively where the black line equals the realized log returns. Forecasted intervals for subplots are a) 2011-11-02 - 2018-12-28, b)2007-01-31 - 2018-12-28 and c)1995-03-02 - 2018-12-28.

5 Results

In this section the results of the forecasts, including a number of backtests will be presented.

5.1 Backtesting

Backtesting is a method used to measure the prediction capabilites and to evaluate the performance of a model. The methods used for this section are the ones presented in the theoretical framework in section 2.7, which we remind ourselves are

- Unconditional Coverage test
- Independence test
- Conditional coverage test
- Testing for higher order dependencies.

In following subsections the results of all tests are presented, however no comments regarding the results will be presented as we will discuss the models in section 5.2

5.1.1 Unconditional Coverage test

The *Unconditional Coverage test* measures the fraction of violations. Given the previously defined hit-sequence I_t, it is derived from the assumption that this hit-sequence is i.i.d. Bernoulli(p) random variables where $p = 0.05$ is the probability used to determine the VaR's. Under this assumption we can form the null hypothesis that the fraction of realized values lower than the estimated $\text{VaR}_t^{0.05}$ should be equal to 5%, i.e.

$$H_0 : I_t \sim \text{Be}(p), \quad p = 0.05.$$

In Table 8 we can see the summarized statistics and critical values from the unconditional test and for both models we note a trend that the longer the forecasted interval, the worse the model seem to perform. None of the models, for neither of the forecast intervals pass the test and the null hypothesis of correct fraction of violations is rejected at the 95% level for all forecasts. If we would chose a more restrictive confidence, such as the 99% level, the MA(2)-GARCH(1,1)-N model forecast of the shortest interval is not rejected, but only by a very small margin.

Model	Forecast length	χ^2 Critical value	LR_{uc} Statistic	p-value
MA(2)-GARCH(1,1)-N	1800	3.84	6.2351	0.0125
MA(2)-GARCH(1,1)-N	3000	3.84	21.8801	2.9×10^{-6}
MA(2)-GARCH(1,1)-N	6000	3.84	31.9104	1.6×10^{-8}
MA(2)-GARCH(1,1)-t	1800	3.84	10.1928	0.0014
MA(2)-GARCH(1,1)-t	3000	3.84	30.2501	3.8×10^{-8}
MA(2)-GARCH(1,1)-t	6000	3.84	40.9805	1.5×10^{-10}

Table 8: Unconditional coverage testing with the Likelihood ratio statistics for our MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t together with the chisquare critical value and corresponding p-value for each model and different forecast lenghts.

5.1.2 Independence test

Further more we also want to test whether all violations occur independently of each other or not. This is done through the *Independence test* where we evaluate if the violations, where the realized value exceeds the Value at Risk, are really independent. The reason to why this is important is that the aim of our model is to capture the volatility by conditioning on the known information set at each given time point. And if the violations do not occur independently of each other, and come in cluster, then that would imply once we

observe a violation of the estimated VaR, the following day will not be another violation with the probability p, but rather some other probability q where $q > p, \quad q, p \in [0, 1]$. This indicates that our model is not optimal and there **should** exist another model which could better capture the volatility and thus improving our prediction capabilities We remind ourselves of the matrices defined in equation (19) and (21) where the latter contains the null hypothesis defined as

$$H_0 : \pi_{01} = \pi_{11} = \pi.$$

Table 9 contains the likelihood ratio statistics, critical χ^2 value and corresponding p-value from the Independence test for our models and the different forecast intervals. We note that at the standard 95% confidence level, none of the models has the null hypothesis of independent violations rejected. The p-values are higher than 0.05 for all forecast intervals, even if some are so by just a small margin. This test only checks for dependency between adjacent days, but we will perform another test in section 5.1.4 to evaluate the dependency between multiple lags as well.

Model	Forecast length	χ^2 Critical value	LR_{ind} Statistic	p-value
MA(2)-GARCH(1,1)-N	1800	3.84	0.2231	0.6367
MA(2)-GARCH(1,1)-N	3000	3.84	2.937	0.0866
MA(2)-GARCH(1,1)-N	6000	3.84	1.6173	0.2035
MA(2)-GARCH(1,1)-t	1800	3.84	0.241	0.6235
MA(2)-GARCH(1,1)-t	3000	3.84	2.1208	0.1453
MA(2)-GARCH(1,1)-t	6000	3.84	0.6775	0.4104

Table 9: Independence testing with the Likelihood ratio statistics for our MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t together with the chisquare critical value and corresponding p-value for each model and different forecast lenghts.

5.1.3 Conditional Coverage test

Here, we combine previous results in order to simultaneously test if the fraction of violations are correct, and, in addition, that they occur independently. Presented in Table 10 we see that none of the models have a p-value larger than 0.05. This means that the null hypothesis of jointly correct fraction of violations, and their indepedence are rejected for all model variations. Considering that this test evaluates the combined performance of the two previous tests, it is not surprising that we reject this hypothesis. This is due to the fact that neither of the models did pass the conditional coverage test for any forecast interval, and they all had p-values very close to zero.

Model	Forecast length	χ^2 Critical value	LR_{ind} Statistic	p-value
MA(2)-GARCH(1,1)-N	1800	3.84	6.4581722	0.0395937
MA(2)-GARCH(1,1)-N	3000	3.84	24.8171	4.0835249×10^{-6}
MA(2)-GARCH(1,1)-N	6000	3.84	33.5277	5.2426691×10^{-8}
MA(2)-GARCH(1,1)-t	1800	3.84	10.4338219	0.0054241
MA(2)-GARCH(1,1)-t	3000	3.84	32.5139	8.7035479×10^{-8}
MA(2)-GARCH(1,1)-t	6000	3.84	41.658	$8.9966378 \times 10^{-10}$

Table 10: Conditional coverage test with the Likelihood ratio statistics for our MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t together with the chisquare critical value and corresponding p-value for each model and different forecast lenghts.

5.1.4 Testing for higher order dependencies

As a final diagnostic we will use the Ljung-Box statistic to test the correlation of violations in the hit sequence between multiple lags instead of only adjacent days $[t, t + 1]$. The max lag length considered will again be the logarithm of the sample size as we established this to be a valid candidate in the theory section 2.2.5.

Table 11 contains the test statistics together with their critical values and p-values. What can be seen here is that for both model variations, and all sub intervals except one, the null hypothesis of no serial correlation cannot be rejected. It is the MA(2)-GARCH(1,1)-N model with the 3000 days forecast interval for which we cannot reject the null hypothesis of no correlation.

Model	Forecast length	χ^2 Critical value	Q(m)	p-value
MA(2)-GARCH(1,1)-N	1800	14.78	4.0345	0.8177
MA(2)-GARCH(1,1)-N	3000	15.52	17.9653	0.0216
MA(2)-GARCH(1,1)-N	6000	16.5	12.2515	0.1806
MA(2)-GARCH(1,1)-t	1800	14.78	4.6678	0.7488
MA(2)-GARCH(1,1)-t	3000	15.52	15.0834	0.0577
MA(2)-GARCH(1,1)-t	6000	16.5	9.7512	0.3439

Table 11: Ljung-Box test statistics for our MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t together with the chisquare critical value and corresponding p-value for each model and different forecast lenghts.

5.2 Model discussion

The results from our final models are somewhat inconsistent and not as satisfactory as we would have hoped. We have quite poor prediction capabilities for both models as noted in Table 8, section 5.1.1 where not a single forecast for either model have the correct number of violations. It is the MA(2)-GARCH(1,1)-N model with the shortest forecast interval that holds the best results, but for all of our forecasts, the number of times the realized value exceed our VaR are far too many, as can be seen in Table 12. Where in contrast to expectations, the model using an underlying t-distribution underestimated the VaR more frequently.

Model	Forecast interval	Expected violations	Actual violations	Fraction
MA(2)-GARCH(1,1)-N	1800	90	114	6.3%
MA(2)-GARCH(1,1)-N	3000	150	209	6.9%
MA(2)-GARCH(1,1)-N	6000	300	400	6.6%
MA(2)-GARCH(1,1)-t	1800	90	121	6.7%
MA(2)-GARCH(1,1)-t	3000	150	220	7.3%
MA(2)-GARCH(1,1)-t	6000	300	414	6.9%

Table 12: Expected number of violations, actual violatons and the corresponding fraction for both MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t models.

Some considerations can be done regarding the performance of our models. One attempt to explain the performance of our models could be if we recall section 4.3, and in particular the comment "We remember this for future references as the inclusion of the financial crises may have some impact on our models prediction capabilities". This could be one explanation for the poor prediction capability of our models since throughout the tests, it appears the models performs best for the short interval which does not contain any financial crisis.

If we compare the results from the Independence test with the test for higher order dependency we find that the p-value has increased for the forecasts made on the short interval, while the p-values have been lowered for all forecasts made on the longer intervals. It has even been lowered to the point where we no longer can reject the null hypothesis of no serial correlation for the MA(2)-GARCH(1,1)-N model forecast on the forecast interval 3000. What this indicates is that once a violation has occurred, we actually have some added information that within the following 9 days (the number of lags tested), we can expect another violation with a higher probability than the desired 5%.

This could perhaps be explained by the *memory* of the model, i.e the persistence of a previous spike in volatility not being long enough. This would lead the model to properly estimate the VaR for a day following a spike, but this conditional volatility would then fall off too fast causing the model to underestimate the

VaR. However while the persistence of a previous spike in volatility maybe could be good during a financial crisis following previous argument, this could probably lead to grossly overestimating the VaR during calm periods. Considering if a spike of high volatility remains significant for the conditional variance for a longer period of time, the model would be very sensitive for every spike.

To summarize, it seems the suggested models manages to quite adequately capture the heteroscedasticity of our data, as the null hypothesis is not rejected for any forecast in the Independence test. When adding the higher order dependency test to check for serial correlation between multiple lags, all forecasts but one passes the test which strengthen the assumption that the models are able to capture the volatility clustering. No forecast is however able to pass the conditional coverage test, indicating the models consistently underestimate the size of the volatility, even if they are able to recognize and account for the clustering.

Since our aim is to evaluate the GARCH(1,1) model we are also going to consider a simulation study where we can govern the parameters used to create the data set and assure there are, in some sense, no random events other than the general randomness of leptokurtic data.

6 Simulation study

As discussed in the previous section, the results from our models are not completely satisfactory. So it would probably be preferable to make use of another model during periods of extreme volatility such as a financial crisis, since from our study it appears the standard GARCH(1,1) model does not suffice.

In an attempt to establish some positive results for our model we are going to perform a simulation study. We will simulate observations from the very same MA(2)-GARCH(1,1) models we estimated for the Nasdaq log returns and then refit it again using the same techniques as previously. Considering we will apply the same techniques they will not be presented with the same detail and we will focus on the results instead. As we will have full knowledge of the true parameter values it will be easier to compare our new estimates with the true values and draw inference from this.

In order to not confuse our simulated values from our original return series $\{r_t\}$, from here on we will refer to the simulated returns as $\{\tilde{r}_t\}$. We are going to simulate two separate return series, one from the MA(2)-GARCH(1,1)-N model, and one from the MA(2)-GARCH(1,1)-t model and we will differentiate these as $\{\tilde{r}_t^T\}$ and $\{\tilde{r}_t^N\}$ where the subscript t denotes the time point and the superscript T, N if the simulated value was simulated from a process with underlying normal or students-t distribution.

To each return series we are going to fit two GARCH(1,1) models with normal- and students-t distributed underlying error terms as we did previously. These models will be evaluated in order to assure their goodness of fit, by both comparing the parameter values to those used for simulating the data, and by looking at the standardized residuals. The models will also be used to perform out of sample forecasts with the rolling window method, where the window size will remain at 500 observations, and the parameters will be re estimated each day.

The same backtesting methods will be applied to the forecasted result

6.1 Simulated data and descriptive statistics

The data is simulated from the exact same parameters that were presented in Table 5 which were fitted from the entire Nasdaq logreturn set. The simulation was done using the rugarch package 'ugarchsim' where the fitted model is supplied, and a burn in sample length is specified in order for the simulated data to settle and properly match the model. We have chosen to simulate 7000 observations with a burn-in period of 2000, meaning the first 2000 simulated values will be discarded.

Figure 7 contains the simulated logreturn series $\{\tilde{r}_t^T\}$ and $\{\tilde{r}_t^N\}$ where both series appears to have some sort of volatility clustering as expected.

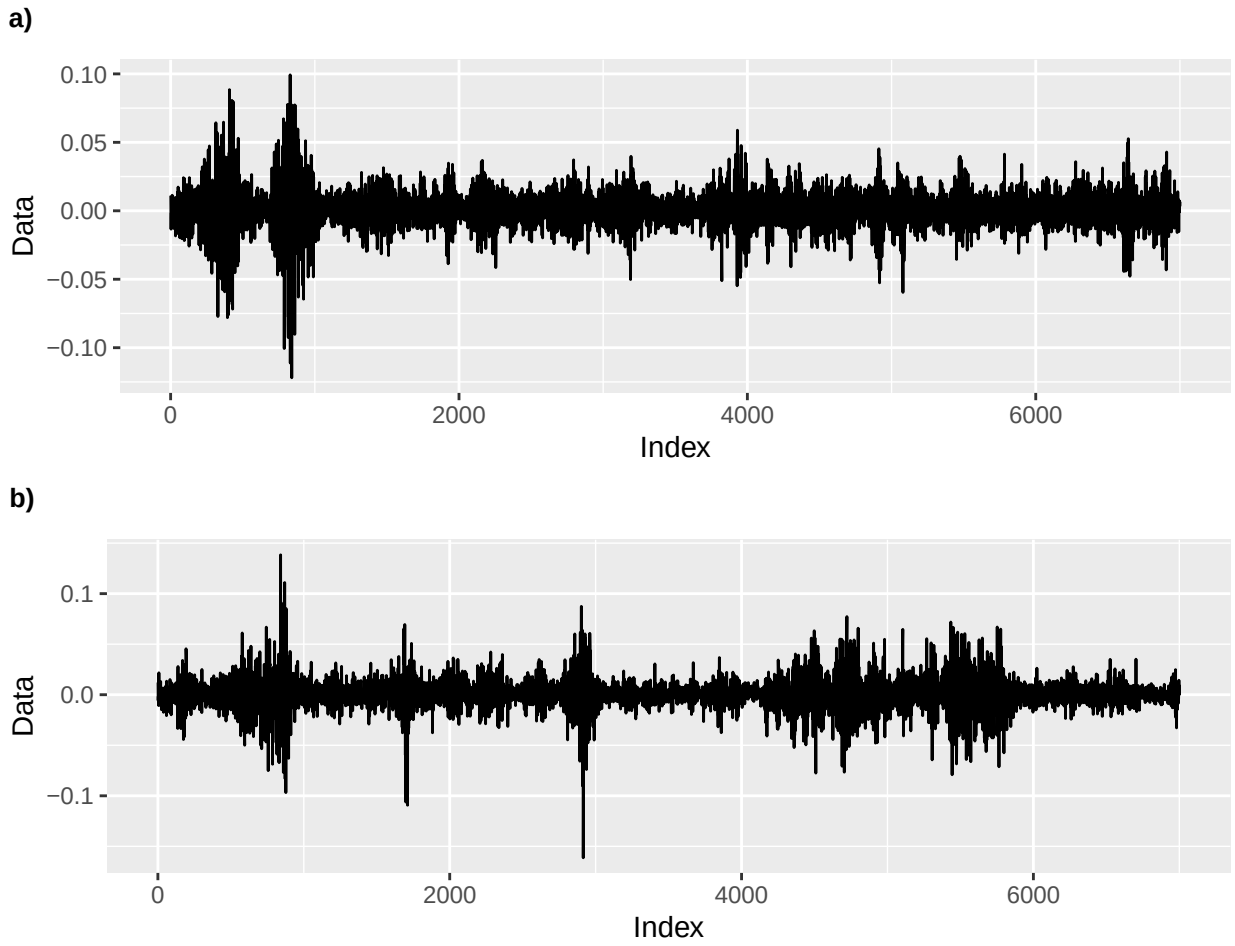

Figure 7: Time plots of simulated data from a)MA(2)-GARCH(1,1)-N model and b)MA(2)-GARCH(1,1)-t model. 6000

We compute the statistical properties for both series which is presented in Table 13. The mean, variance and standard deviations are very similar for both series, with marginally higher values for $\{\tilde{r}_t^T\}$ and both series are slightly negatively skewed and hold excessive kurtosis. We can compare this with the real data from our initial study and note that they are very similar, exhibiting approximately the same variance, skewness and kurtosis. By conducting a t-test we can confirm the mean for both series to be significantly different from 0, and through ACF plots and the Ljung-Box test presented in Appendix A, it is verified both series exhibit ARCH effects. The null hypothesis of no serial correlation is rejected for both regular series $\{\tilde{r}_t^T\}$ and $\{\tilde{r}_t^N\}$, and their squared values as well.

Series	Mean	Variance	Standard deviation	Skewness	Kurtosis
$\{\tilde{r}_t^N\}$	7.3e-4	2.1e-4	0.01388	-0.24317	9.69625
$\{\tilde{r}_t^T\}$	9.9e-4	2.7e-4	0.01499	-0.19791	11.77205

Table 13: Statistical properties for the simulated return series.

To adjust for the serial correlation in the logreturn series we again apply the R-function auto.arima. Given the data, the auto.arima method suggests a MA(2) model to best fit $\{\tilde{r}_t^N\}$ while the model suggested for $\{\tilde{r}_t^T\}$ is an ARMA(3,2) model. A combined AR(p)-MA(q) model has not been presented for this thesis and

considering that our main interest lies in evaluating the performance of the volatility process, in addition to the knowledge it was actually derived from a MA(2) model, the MA(2) specification is what we will used for both series.

Once the serial correlation has been adjusted for, we can now consider the residuals of the mean equation to be the innovation at time t, $a_t = r_t - \mu_t$. The last step prior to fitting the GARCH(1,1) model is verifying that the residuals of the mean equation is no longer correlated, whereas the squared residuals still exhibits ARCH effects. ACF plots for the residuals can be found in Appendix A, Figure 10, while the Ljung Box statistic and corresponding p-value is presented in Table 14. As the p-values for both regular series are higher than 0.05, we cannot reject the null hypothesis of no serial correlation, while the squared residuals are still showing strong ARCH effects, indicating that a GARCH-model could be suitable.

Series	Q(m)	p-value
$\{\tilde{r}_t^N - \mu_t\}$	6.653	0.66
$\{(\tilde{r}_t^N - \mu_t)^2\}$	5606.433	2.2e-16
$\{\tilde{r}_t^T - \mu_t\}$	6.699	0.655
$\{(\tilde{r}_t^T - \mu_t)^2\}$	2181.23	2.2e-16

Table 14: The Ljung-Box test statistic Q(m) and p-value for respective series.

6.2 GARCH modelling and forecasting for simulated data

We fit the two separate models with the specifications presented in the previous section, i.e. the MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-N model. These specifications will be fitted to each of the simulated series $\{\tilde{r}_t^N\}$ and $\{\tilde{r}_t^T\}$, entailing four models in total. Once the models have been fitted to the simulated data we want to look at the estimated parameters which we can compare with the parameters from the original model. In a perfect world we want these parameters to conform considering the simulated data was based on them. We are also going to look at the standardized residuals to verify the goodness of fit for our models to the data.

The full Tables containing all estimated parameters, standard errors and p-values for our models, based on the series $\{\tilde{r}_t^N\}$ and $\{\tilde{r}_t^T\}$, can be found in table 20 and 21 in Appendix B. Whereas for clarity, presented in Table 15, we have singled out the GARCH parameters α_0, α_1 and β_1 from the fitted models, based on the simulated data, and put them next to the parameter values from where they were simulated. We see that the estimated parameter values come very close to the real value, in particular for the MA(2)-GARCH(1,1)-t simulated data.

MA(2)-GARCH(1,1)-N simulated data			
Parameter	Baseline value	GARCH(1,1)-N	GARCH(1,1)-t
α_0	2.1×10^{-6}	2.1×10^{-6}	2.2×10^{-6}
α_1	0.098	0.1084	0.1124
β_1	0.8903	0.8798	0.881

MA(2)-GARCH(1,1)-t simulated data			
Parameter	Baseline value	GARCH(1,1)-N	GARCH(1,1)-t
α_0	1.1×10^{-6}	9.5×10^{-7}	9.7×10^{-7}
α_1	0.0909	0.0912	0.0973
β_1	0.9077	0.9074	0.9028

Table 15: Estimated parameter values for GARCH(1,1)-N and GARCH(1,1)-t models compared to respective Baseline value originating from the value used to simulate data.

As we now are familiar with, the standardized residuals $\tilde{a}_t = \frac{a_t}{\sigma_t}$ should behave like a white noise series and no serial correlation should remain. To validate that this is the case, the Ljung-Box test statistics for both the

standardized residuals and the squared standardized residuals are computed for the fitted models. Presented in Table 16 we find test statistics and corresponding p-value which confirms the standardized residuals does not contain any significant correlation, nor does the squared residuals.

Presented in Table 16 we find the test statistics and corresponding p-value. All p-values are higher than 0.05, indicating that neither of the standardized residuals or their squared values contain any serial correlation as the null hypothesis is rejected for all series.

MA(2)-GARCH(1,1)-N simulated data		
Series	Q(m)	p-value
MA(2)-GARCH(1,1)-N Standardized residuals	10.427	0.305
MA(2)-GARCH(1,1)-N Standardized squared residuals	8.262	0.4935
MA(2)-GARCH(1,1)-t Standardized residuals	10.434	0.304
MA(2)-GARCH(1,1)-t Standardized squared residuals	8.3735	0.4826

MA(2)-GARCH(1,1)-t simulated data		
Series	Q(m)	p-value
MA(2)-GARCH(1,1)-N Standardized residuals	5.882	0.740
MA(2)-GARCH(1,1)-N Standardized squared residuals	6.688	0.656
MA(2)-GARCH(1,1)-t Standardized residuals	6.117	0.715
MA(2)-GARCH(1,1)-t Standardized squared residuals	8.327	0.487

Table 16: Ljung-Box test statistic Q(m) and corresponding p-value for the standardized residuals and the squared standardized residuals

As it appears the models adequately describe the volatility process, it is now time to perform the rolling window forecast. Figure 8 contains the results of the forecasts with our models fitted to the MA(2)-GARCH(1,1)-N simulated data and Figure 9 the forecasts fitted to MA(2)-GARCH(1,1)-t simulated data. The red lines denote the forecasted VaR from our models with an underlying normal distribution, and the blue to the VaR from the models with a students-t distribution. If we take a closer look at Figure 8 it appears the MA(2)-GARCH(1,1)-t model shows a more conservative prediction in VaR. The blue line corresponding to the MA(2)-GARCH(1,1)-t model lies below the red line with a few exceptions. The same trend is not clear in Figure 9 containing the predicted VaR from the other subset of simulated data, where the models in periods seem to provide the more conservative VaR.

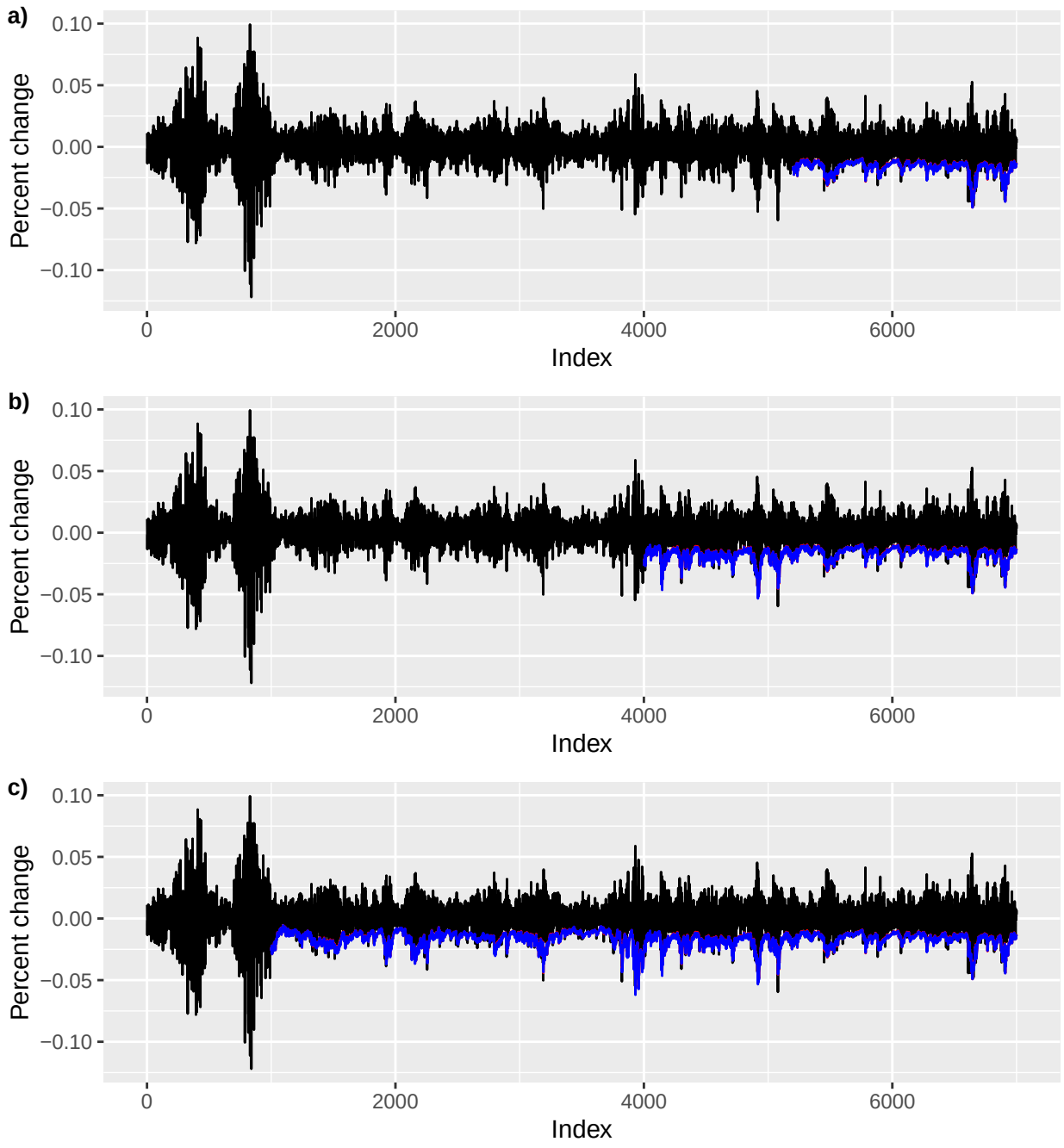

Figure 8: Plots for predicted VaR represented by red and blue line for MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t model respectively where the black line equals the simulated value from a MA(2)-GARCH(1,1)-N process. Forecasted interval length for subplots are a)1800 b)3000 and c)6000.

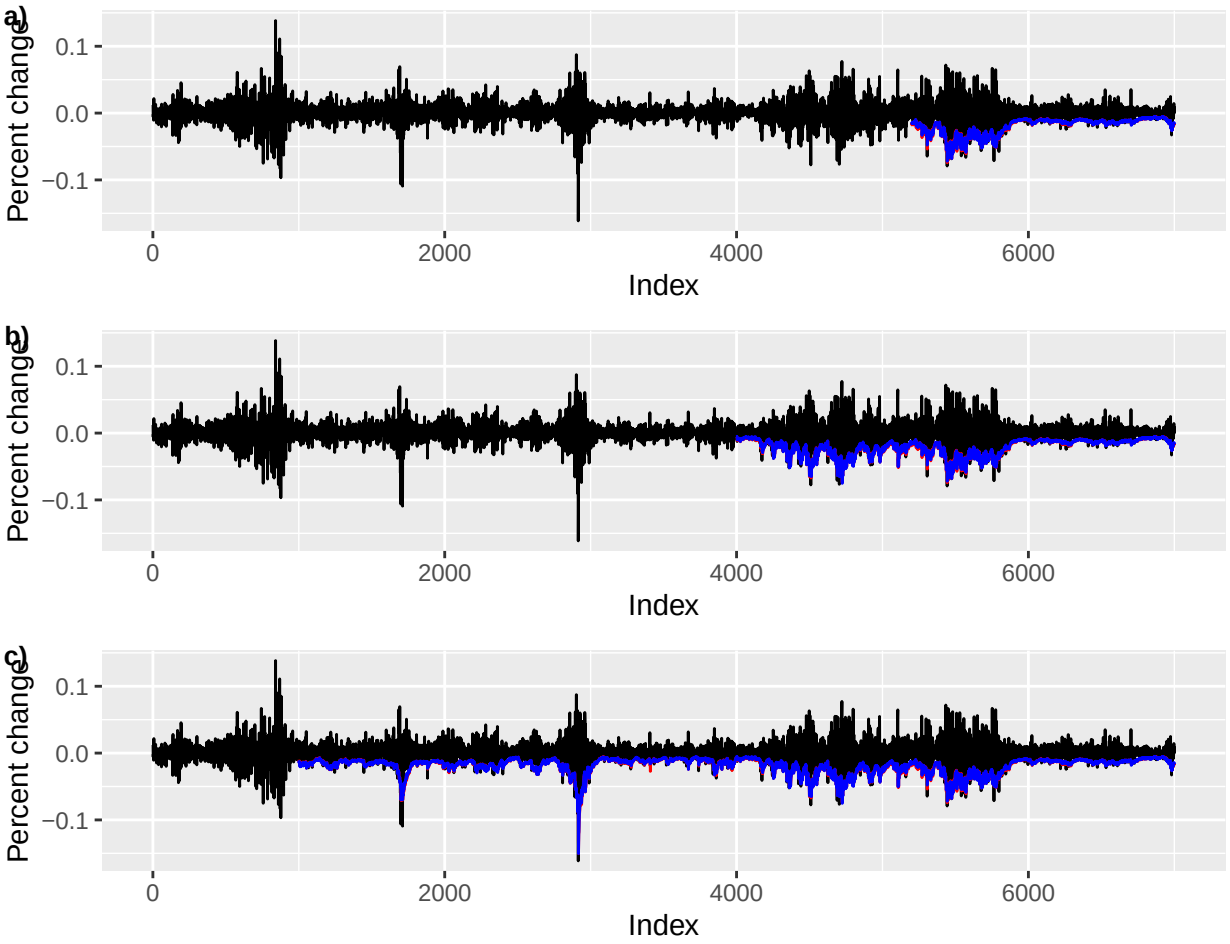

Figure 9: Plots for predicted VaR represented by red and blue line for MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t model respectively where the black line equals the simulated value from a MA(2)-GARCH(1,1)-N process. Forecasted interval length for subplots are a)1800 b)3000 and c)6000.

6.3 Backtesting simulated data

Due to the extensive amount of Tables needed to show all results from the backtests, only the Conditional Coverage test is presented immediately in Table 17. The full result from the conditional coverage test, the unconditional coverage and independence test for all models can be found in Appendix B, Table 21,22 and 23.

The names of the models were abbreviated to save space and so the "MA(2)" part of the model names were omitted. The name "GARCH(1,1)-N" still implies the same model as specified previously, i.e. "MA(2)-GARCH(1,1)-N".

	GARCH(1,1)-N simulated data		GARCH(1,1)-t simulated data	
Forecast length	Model	LR_{cc} p-value	Model	LR_{cc} p-value
1800	GARCH(1,1)-N	0.4532	GARCH(1,1)-N	0.4875
3000	GARCH(1,1)-N	0.7054	GARCH(1,1)-N	0.203
6000	GARCH(1,1)-N	0.471	GARCH(1,1)-N	0.8226
1800	GARCH(1,1)-t	0.1586	GARCH(1,1)-t	0.3151
3000	GARCH(1,1)-t	0.2205	GARCH(1,1)-t	0.0724
6000	GARCH(1,1)-t	0.2284	GARCH(1,1)-t	0.6185

Table 17: Independence testing with the Likelihood ratio statistics for our MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t together with the chisquare critical value and corresponding p-value for each model and different forecast lenghts.

If we recall the results from our original models, there were not a single one out of the six forecasts for which the conditional coverage test could not reject the null hypothesis of independent and correct number of violations. Now, however, we note that all models are accepted and the null hypothesis is not rejected for any model. The MA(2)-GARCH(1,1)-t model fitted to the GARCH(1,1)-t simulated data at the forecast interval 3000 is almost rejected, but at the 95% level it barely passes the test.

To comment on Tables 22,23 and 24 in Appendix B, the unconditional coverage test in Table 22 verify that the number of violations for all submodels are within the boundaries of the test, not rejecting any of the submodels. The Independence test in Table 23 verifies the violations occurs independently of each other for all models but one, which of course is the same GARCH(1,1)-t model that nearly did not pass the Conditional Coverage test just mentioned. This is to be expected, as if we recall section 2.7.3 in the theoretical framework we noted the Conditional Coverage test combines the likelihood of the null hypothesis from the Unconditional Coverage test with the alternative hypothesis from the Independence test.

By comparing the results from backtesting the forecasts on simulated data, to the ones obtained when forecasting real life data, it is clear the former present a more satisfactory result. Even if no forecast made on real life data contained a correct fraction of violations, we did note the performance seemed better for the short, 1800 observation forecast. The same trend continued when looking at the dependency of violations where the shorter interval displayed the highest p-values by quite a large margin.

If we take a closer look at the real data and the forecasts in Figure 6, there are some differences in the separate intervals. If we compare the short interval to the longer, we note that it acts more homogeneous, i.e. it does not contain as much variability as the data does during the longer intervals, which includes two financial crises. Seeing this, and taking into consideration that the models with the same specifications performed better on simulated data we could probably assume that the MA(2)-GARCH(1,1) model works better during calmer and more homogeneous periods. Even if the simulated data contains heteroscedasticity, we know for a fact the simulated data is stationary, which is harder to verify for real data. So perhaps it could be that a more complex model would be required in order to predict real life financial data.

7 Conclusion

We have been looking at how to fit a heteroscedastic model, in particular the MA(2)-GARCH(1,1) model and how it can be used in order to make one-day-ahead predictions. In the model we assume the residuals to follow either a standard normal distribution, or a students-t distribution, thus creating two model versions for each data set. We use well established tools to evaluate the goodness of fit of our models prior to using the rolling window approach to calculate predicted Value at Risks. Our predictions are then backtested as a final diagnostic to evaluate and solidify the suggested models.

These methods are initially applied to a real data set, consisting of the logreturns for the Nasdaq composite index. We find that the model seem to account for the heteroscedasticity in the series, but it underestimate the VaR frequently, in particular for time periods including financial crises.

To examine whether the issue lies in the data or the model, two separate data sets are simulated using a MA(2)-GARCH(1,1) process with the exact same parameter values that was estimated for the model fitted to the real life data. We fit two MA(2)-GARCH(1,1) models to each simulated data sets, one assuming residuals to be normally distributed and one assuming residuals to be t-distributed. We find the models to adequately predict the VaR, as the number of violations are correct and they appear independent of each other. This indicates that there is nothing wrong with our modelling and backtesting procedures.

So it appears the MA(2)-GARCH(1,1) model, in periods of crisis, do not provide a perfect fit when it comes to predict the one-day-ahead VaR for real life financial data, and it could be preferable to consider a more complex model.

In the comparison of assuming the model to follow a standard normal-, or a students-t distribution, it appears the GARCH(1,1)-N specification is slightly favoured, as in general it seem to provide a more correct fraction of violations.

It should be stressed though, that this thesis only covers the evaluation of the Nasdaq composite index, and the results do not apply to all financial data.

8 Further research

This section will cover a selection of topics that could have been investigated further in order to deepen the analysis and assist in evaluating the performance of the suggested models.

A first thought would have been to evaluate the model at a more restrictive confidence level, as in practice, the Basel accord stipulate a 99% VaR to measure the market risk. However in my opinion it could have been a good inclusion, rather than the sole subject of investigation. As by considering a large variety of confidence levels, it would better capture the differences in the assumed distribution for the models.

I also realized a bit too late through the analysis, that the approach of evaluating the models could have been better. It would probably have been more interesting and rewarding in terms of performance, to evaluate different window lengths rather than forecast lengths Because when we only increase the forecast interval, we will always have the same parameter estimations when predicting VaR for data that is contained within all intervals.

Some further considerations that could be made is the comparison of models, as the ARCH and GARCH family offers a wide variety of submodels, such as the *IGARCH*, *EGARCH* and *GARCH-M* models to provide some examples.

The backtesting methods used in this thesis in order to evaluate the performance of our models amount to looking at the predicted VaR, which is defined as the conditional lower quantile of the return distribution. If we recall section 2.7 from the theoretical framework, we defined a hit-sequence

$$I_{t+1} = \begin{cases} 1, \text{ if } r_{t+1} < -VaR^p_{t+1} \\ 0, \text{ if } r_{t+1} \geq -\text{VaR}^p_{t+1} \end{cases}$$

which returns a 1 if a violation occurs, and 0 else. This sequence should be i.i.d. Bernoulli(p) and we defined three tests to evaluate this assumption. A key problem with these tests, as mentioned by Berkowitz. J (2001) is that the hit sequence I_t only takes on two values, 0 or 1. Also if the models are evaluated at the more restrictive level, for instance the 99% level stipulated by the Basel accord, it also takes on the value 1 very rarely. He suggests that it could be preferable to use density evaluating methods instead, as quoted: "Density evaluation methods make use of the full distribution of outcomes and thus extract a greater amount of information from the available data." (Berkowitz, 2001, p. 466). An intuitive way of seeing this is that when using the hit sequence, the only information we deduct from a violation is the fact that it happened, but we do not consider the largeness of it, thus overlooking a lot of information.

The approach for density evaluating methods would amount to using an integral transformation dating back to at least Rosenblatt (1952), which is then extended by Berkowitz. J (2001). As evaluation method for the density forecast, Diebold et al. (1998) suggests a less formal, but more revealing approach, which would be to visually inspect histograms to validate the unconditional uniformity of the density in addition to more formal tests. It is stressed that more formal tests such as Kolmogorov-Smirnov or Cramer-von Mises alone, do not hold much value as they do not provide any guidance to *why* the rejection occurred. However, if we consider a model that fails to capture fat tails properly, the histogram will exhibit peaks near 0 and 1.

Another topic that could be of interest to research further, and that is often mentioned when discussing backtesting methods of VaR, is the fact that violations happen very rarely, in particularly when adopting a more restrictive confidence level. Under the circumstance of limited data, one may expect a very low count of violations which may cause the finite samples distribution intractable. Dufour. J-M (2005) presents a Monte-Carlo method, which could aid in controlling the size of test.

9 References

Berkowitz. J (2001) - *Testing Density Forecasts, with Applications to Risk Management.* Journal of Business & Economic Statistics, Vol. 19, Issue 4, p. 465–474.

Bollerslev. T (1986) - *Generalized autoregressive conditional heteroscedasticity.* Journal of Econometrics, Vol. 31, Issue 3, p. 307-327.

Box. G & Jenkins. G & Reinsel. G (2008) - *Time series analysis, Forecasting and Control.* 4th ed. John Wiley & Sons, Inc. Hoboken, New Jersey.

Calzolari, G., Sentana, E., & Fiorentini, G. (2003). - *Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models with Student t Innovations.* Journal of Business & Economic Statistics, Vol. 21, no. 4. p. 532–546.

Cedergren. R (2018) - *Forecast evaluation of 1-step-ahead predictions using GARCH(1,1) on the Euro/US Dollar FX Spot Rate*, Bachelor Thesis in Mathematical statistics, Stockholm University.

Christoffersen. P.F. (2011) - *Elements of Financial Risk Management (2nd ed).* Academic Press.

Christoffersen. P.F & Pelletier. D. (2004) - *Backtesting Value-at-Risk: A Duration-Based Approach.* Journal of Financial Econometrics, no. 1, p. 84.

Diebold, Francis X., Todd A. Gunther, & Anthony S. Tay. (1998) *Evaluating Density Forecasts with Applications to Financial Risk Management.* International Economic Review, Vol. 39, no. 4, p. 863–883.

Dufour. J-M (2006) - *Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics.* Journal of Econometrics, Vol. 133, Issue 2, p. 443-477.

Engle. R (1982) - *Autoregressive conditional heteroscedasticity with estimates of the variance of u.k. inflation.* Econometrica. Vol. 50, Issue 4, p. 987-1007.

Held. L & Sabanés Bové. D (2014) - *Applied Statistical Inference.* Springer-Verlag Berlin Heidelberg.

Tsay. R.S. (2010) - *Analysis of Financial Time series.* 3d ed. John Wiley & Sons, Inc. Hoboken, New Jersey.

Appendix A

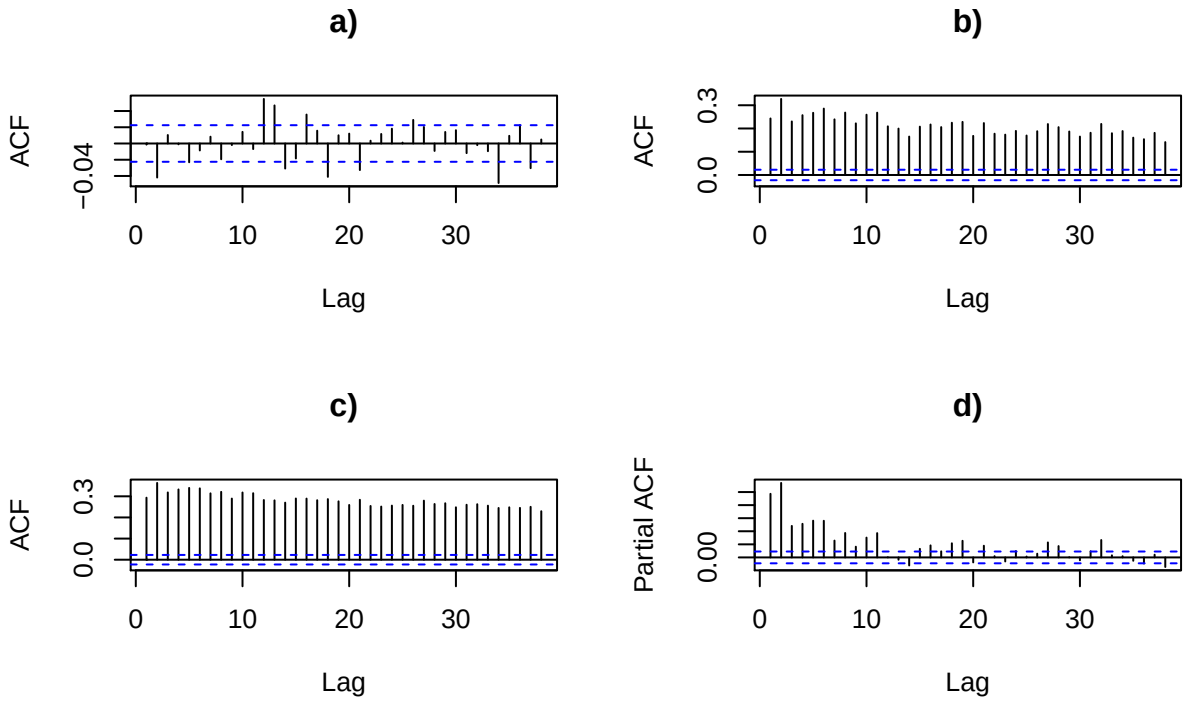

Figure 10: Autocorrelation plots containing the a)mean corrected series, b)squared mean corrected series, c)absolute value of the mean correted series, and d)PACF of the squared mean corrected series

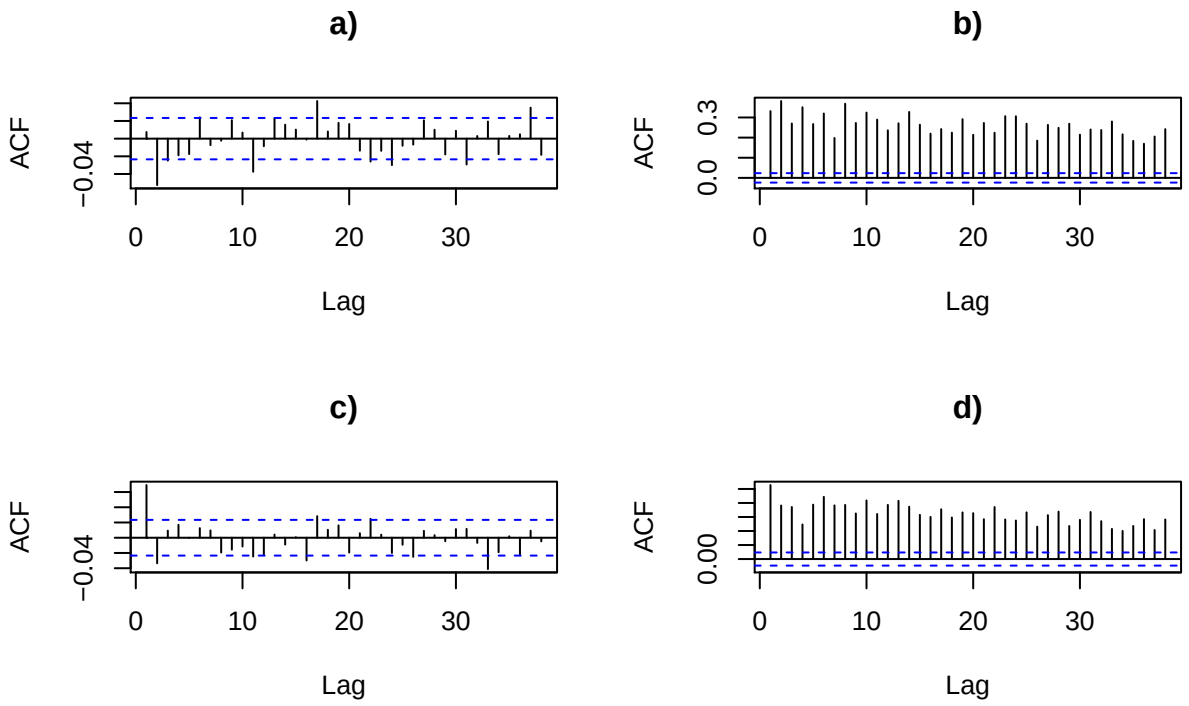

Figure 11: ACF plots of the simulated time series, plot a) and b) contains the series simulated from a MA(2)-GARCH(1,1,)-N process and plot c) and d) was simulated from a MA(2)-GARCH(1,1,)-t process.

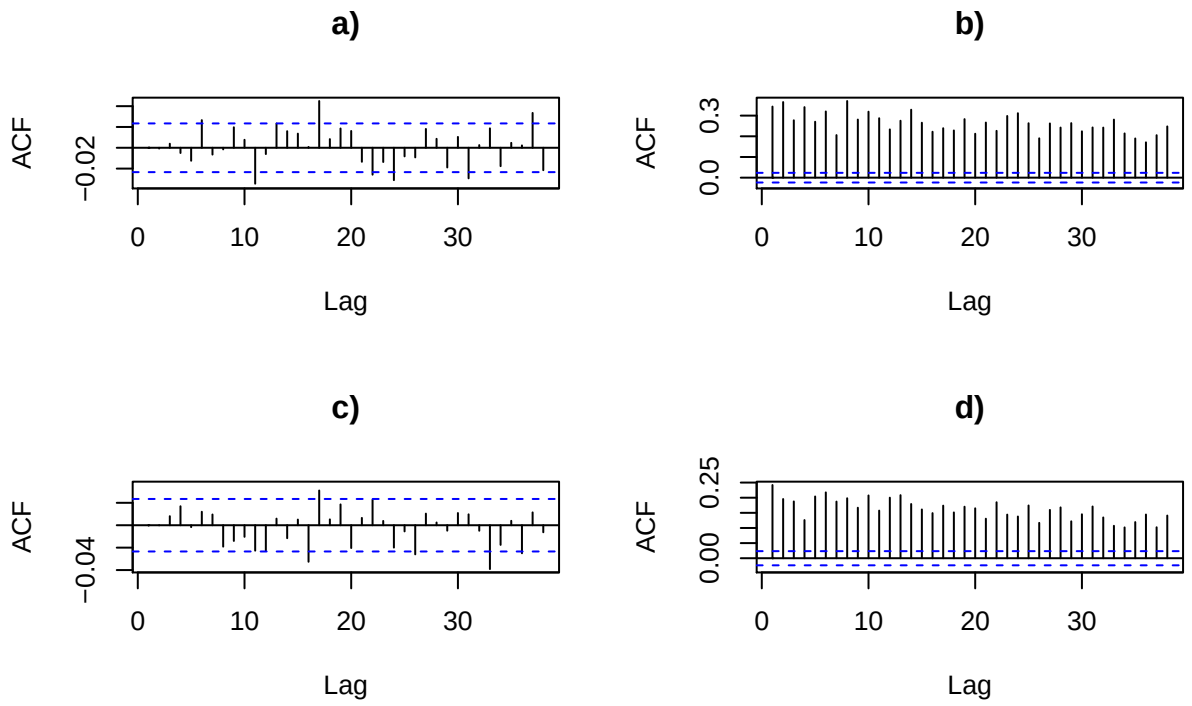

Figure 12: Plot a) and b) originates from the normally distributed series containing the ACF from the residuals of the mean equation as well as the squared residuals . Plot c) and d) originates from the t-distributed series containing the ACF of the residual and squared residuals.

Appendix B

Series	Q(m)	p-value
$\{ r_t - \mu\}$	22.1316	0.0081
$\{(r_t - \mu)^2\}$	4294.9887	2.2e-16
$\{\|r_t - \mu\|\}$	6513.7305	2.2e-16

Table 18: The Ljung-Box test statistic Q(m) and p-value for the uncorrected simulated series $\tilde{r}_t^N$ and $\tilde{r}_t^T$.

Series	Q(m)	p-value
$\{\tilde{r}_t^N\}$	33.216	1.1×10^{-4}
$\{\tilde{r}_t^N\}^2$	5607.122	2.2e-16
$\{\tilde{r}_t^T\}$	48.441	0
$\{\tilde{r}_t^T\}^2$	2218.44	2.2e-16

Table 19: The Ljung-Box test statistic Q(m) and p-value for the uncorrected simulated series $\tilde{r}_t^N$ and $\tilde{r}_t^T$.

MA(2)-GARCH(1,1)-N simulated data						
	MA(2)-GARCH(1,1)-N			MA(2)-GARCH(1,1)- t		
Parameter	Estimate	Std Error	p-value	Estimate	Std Error	p-value
c_0	8.4×10^{-4}	1.13×10^{-4}	1.3×10^{-13}	8.1×10^{-4}	1.1×10^{-4}	8.5×10^{-13}
θ_1	0.0257	0.0124	0.0392	0.0239	0.0125	0.05625
θ_2	-0.0263	0.0123	0.0327221	-0.0298	0.0124	0.01622
α_0	2.1×10^{-6}	3.4×10^{-7}	8.4×10^{-10}	2.2×10^{-6}	4.1×10^{-7}	9.4×10^{-8}
α_1	0.1084	0.0074	0	0.1124	0.0088	0.0000
β_1	0.8798	0.0077	0	0.881	0.0088	0.0000
ν	-	-	-	10	0.7453	0.0000

Table 20: Summary of parameter estimations, standard error and p-values of GARCH(1,1) model with normal and t-distributed error terms of data simulated from a MA(2)-GARCH(1,1)-N process.

MA(2)-GARCH(1,1)-t simulated data						
	MA(2)-GARCH(1,1)-N			MA(2)-GARCH(1,1)- t		
Parameter	Estimate	Std Error	p-value	Estimate	Std Error	p-value
c_0	8.6×10^{-4}	1.1×10^{-4}	0	8.8×10^{-4}	1.03×10^{-4}	0
θ_1	0.045	0.0128	4.17×10^{-4}	0.041	0.0123	8.5×10^{-4}
θ_2	-0.0307	0.0125	0.0141278	-0.0314	0.0121	0.0093
α_0	9.5×10^{-7}	1.6×10^{-7}	6.7×10^{-9}	9.71×10^{-7}	2.03×10^{-7}	1.67×10^{-6}
α_1	0.0912	0.0058	0	0.0973	0.0073	0.0000
β_1	0.9074	0.0052	0	0.9028	0.0065	0.0000
ν	-	-	-	7.0623	0.5777	0.0000

Table 21: Summary of parameter estimations, standard error and p-values of GARCH(1,1) model with normal and t-distributed error terms of data simulated from a MA(2)-GARCH(1,1)-t process.

	GARCH(1,1)-N simulated data			GARCH(1,1)-t simulated data		
Forecast length	Model	Statistic	LR_{uc} p-value	Model	Statistic	LR_{uc} p-value
1800	GARCH(1,1)-N	0.43	0.512	GARCH(1,1)-N	0.287	0.5921
3000	GARCH(1,1)-N	0.58	0.4463	GARCH(1,1)-N	0.986	0.3207
6000	GARCH(1,1)-N	0.0564	0.8123	GARCH(1,1)-N	0.1707	0.6795
1800	GARCH(1,1)-t	1.76	0.1846	GARCH(1,1)-t	0.046	0.8302
3000	GARCH(1,1)-t	2.642	0.1041	GARCH(1,1)-t	1.3367	0.2476
6000	GARCH(1,1)-t	1.1591	0.2817	GARCH(1,1)-t	0.499	0.4799

Table 22: Unconditional coverage test Likelihood ratio statistics for the MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t models fitted to both simulated data sets. Together with corresponding p-value for each model and different forecast lenghts.

	GARCH(1,1)-N simulated data			GARCH(1,1)-t simulated data		
Forecast length	Model	Statistic	LR_{indep} p-value	Model	Statistic	LR_{indep} p-value
1800	GARCH(1,1)-N	1.153	0.2829	GARCH(1,1)-N	1.149	0.2838
3000	GARCH(1,1)-N	0.119	0.7301	GARCH(1,1)-N	2.203	0.1377
6000	GARCH(1,1)-N	1.4496	0.2286	GARCH(1,1)-N	0.2198	0.6392
1800	GARCH(1,1)-t	1.922	0.1656	GARCH(1,1)-t	2.263	0.1325
3000	GARCH(1,1)-t	0.382	0.5365	GARCH(1,1)-t	3.9141	0.0479
6000	GARCH(1,1)-t	1.7939	0.1805	GARCH(1,1)-t	0.4617	0.4968

Table 23: Independence testing Likelihood ratio statistics for the MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t models fitted to both simulated data sets. Together with corresponding p-value for each model and different forecast lenghts.

	GARCH(1,1)-N simulated data			GARCH(1,1)-t simulated data		
Forecast length	Model	Statistic	LR_{cc} p-value	Model	Statistic	LR_{cc} p-value
1800	GARCH(1,1)-N	1.583	0.4532	GARCH(1,1)-N	1.437	0.4875
3000	GARCH(1,1)-N	0.698	0.7054	GARCH(1,1)-N	3.189	0.203
6000	GARCH(1,1)-N	1.5059	0.471	GARCH(1,1)-N	0.3905	0.8226
1800	GARCH(1,1)-t	3.683	0.1586	GARCH(1,1)-t	2.31	0.3151
3000	GARCH(1,1)-t	3.024	0.2205	GARCH(1,1)-t	5.2508	0.0724
6000	GARCH(1,1)-t	2.9529	0.2284	GARCH(1,1)-t	0.9608	0.6185

Table 24: Conditional coverage test Likelihood ratio statistics for the MA(2)-GARCH(1,1)-N and MA(2)-GARCH(1,1)-t models fitted to both simulated data sets. Together with corresponding p-value for each model and different forecast lenghts.