Stockholms universitet

Performance of ARMA-GARCH models in Value at Risk estimation

Gabriella Sawma

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:2
Matematisk statistik
Januari 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2019:2**
http://www.math.su.se

Performance of ARMA-GARCH models in Value at Risk estimation

Gabriella Sawma*

January 2019

Abstract

Value at Risk (VaR) measures the maximal possible loss that may occur under normal market conditions and is the most widely used measure of risk in financial institutions and risk management practices today. In a statistical sense, VaR can be formulated as a quantile of the lower tail of the return distribution, i.e. the loss tail, given a certain confidence level and a time period. In this thesis we introduce how conditional volatility for a log return time series can be modelled by implementing the conditional heteroskedastic GARCH models. The aim of this thesis is to compare the performances of the GARCH models in estimating daily Value at Risk by making distributional assumptions regarding the residuals of the GARCH models. The performance of the models are evaluated using backtesting methods. We apply the ARMA(1,1)-GARCH(1,1) model on the OMXS30 index log return series with Normal and Student's t distributed error terms. In order to forecast the one step ahead VaR we use the ARMA(1,1)-GARCH(1,1) models in a rolling window estimation on an out-of-sample window of one thousand observations. The backtesting results reveal that the Student's t distributed model outperforms the Normal model in estimating daily VaR over the forecasting period. However, rejecting the Normal model is not justified since the evaluation tests disclose that both models are specified adequately enough to predict the volatility process of our forecast period, and do not underestimate the Value at Risk.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: gabriella_sawma@hotmail.com. Supervisor: Chun-Biu Li and Disa Hansson.

Preface and Acknowledgements

This is a Bachelor thesis covering 15 ECTS that will result in an Bachelor's Degree in Mathematical Statistics at Stockholm University, Department of Mathematics.

I wish to express my gratitude to my supervisors Chun-Biu Li and Disa Hansson for their encouraging advices in the process of writing this thesis. I would also like to sincerely thank my rock Jan Mikael Yousif for consistently supporting me.

Contents

1 Introduction

1.1 Background

In finance, volatility is related to the uncertainty of financial asset's returns and is used as a measure of risk. More specifically, the volatility is defined as the conditional standard deviation of the underlying asset's return (Tsay 2010, pp.109). To properly model the volatility, it is important to understand its characteristics and how they display in the financial asset's return over time.

A well-known property of asset returns are volatility clusterings that are presented as extended periods of high volatility followed by periods of low volatility (Mandelbrot 1963). The presence of volatility clusterings is a consequence of non-normality in the return set resulting from the non-constant variance of the error terms. The clusterings of fluctuations are an indication that observations in the return series that are close to each other tend to be correlated in time. This violates the assumption of constant error variance, also known as homoskedasticity. Since the series is correlated in time and has non-constant error variance, it suffers from conditional heteroskedasticity.

In order to capture the behavior of conditional volatility the Autoregressive conditional heteroskedastic model (ARCH) was introduced by Engle (1982) and was the first model to allow for such changes in the error variance . The ARCH model describes the conditional variance by regressing on its preceeding values of the error terms. In 1986 the ARCH model was further extended by Bollerslev (1986) who presented the Generalized ARCH model (GARCH) where he not only included the previous values of the error terms, but also regressed on the previous values of the conditional variance.

The models have been proven to successfully capture the characteristics of conditional volatility and are widely used in financial activities for risk analysis as well as being a guidance in financial decision makings. Typically in risk management, the volatility models are used for financial forecasting aspects to predict the severity of returns, the density or the quantiles of the returns (Engle et al. 2001).

A popular risk measure that is frequently applied in risk management practices is Value at Risk (VaR). Value at Risk is commonly formulated as the worst possible loss that will not be exceeded with a given high probability over a given time period (McNeil et al. 2005, pp.37-38). In a statistical sense, VaR can be described as a model that is built on the distribution of the profits and losses and is simply a quantile of the loss tail, i.e. the lower tail of the return series at a given confidence level and a time horizon (Jorion 2007).

The idea of VaR came up in the search for an improved risk measure after big financial crises lead to several banks going bankrupt, which raised discussion and skepticism concerning the existing market risk practices. The concept of Value at Risk was later developed by the researcher Till Guldimann from JP Morgan in the 1990s (Dowd 1998) and has since then increasingly gained recognition. In recent years, the enlarged price movements induced by globalization and emerging of financial markets has further expanded the popularity and need of VaR, to the point where it has become one of the most used measures of market risks in financial institutions today.

In VaR framework there are some simplified assumptions made considering the distribution of assets returns that are quite unrealistic (Christoffersen 2003, pp.50). For instance, it is often assumed that financial asset returns are normally distributed. Thus, ignoring the non-normal properties of returns series, as a consequence of the conditional volatility. Since the normal distribution fails to capture the extreme data points of a volatile series the assumption may in practice generate models that underestimate VaR. As a result this could bring unreliable or inaccurate VaR predictions.

Undoubtedly, the VaR models are only useful if they generate accurate predictions of future risks. In order to determine the reliability and validity of the VaR forecasts, the models should always be evaluated by backtesting methods. The backtesting procedure is a statistical framework designed to verify that the estimates of VaR are in line with the actual observed data of profits and losses (Jorion 2007).

1.2 Aim and Purpose

In this thesis the purpose is to study if distributional assumptions of the return series have a significant effect on the predictive accuracy of VaR estimations. The idea is to introduce how conditional volatility of a log return time series can be modelled for a financial stock index, by implementing methods inspired by Tsay (2010). Concisely, the ARMA(1,1)-GARCH(1,1) models are applied on an in-sample period for the OMXS30 index where the parameters of the models are estimated with maximum likelihood estimation. The residuals of the models are assumed to follow a Normal and respectively a Student's t distribution. The aim is to evaluate the performance of the obtained volatility models in predicting one step ahead VaR on an out-of-sample period of thousand observations with a 99% level of confidence. To determine the accuracy of the VaR estimates the useful methods of backtesting are conducted. The performance of the models are evaluated specifically by Christoffersen's and Kupiec's coverage tests. It is in our interest to find out how, or even if, the different distributions of the underlying models' innovations significantly effect the backtest results of the VaR models.

1.3 Previous Research

There are countless collections of studies and papers considering proper volatility modeling and forecasting of VaR. The initial sources of inspiration for this thesis are among others: Angelidis, Benos, Degiannakis (2004); Christoffersen (2003); Mcneil, Frey, Embrechts (2005); and Tsay (2002,2010).

Originally, Engle (1982) presented the ARCH model which was the first model to take into account the properties of volatility in financial asset returns. Bollerslev (1986) introduced the GARCH models which were proven to be superior of ARCH since it required fewer parameters to properly describe the conditional volatility (Tsay 2002). There are over one hundred variations of GARCH models presented in Bollerslev (2008), yet there is still is no specific answer to which GARCH family model that provides the best estimates of conditional volatility.

Mandelbrot (1963) released a paper where he demonstrated the departures of normal properties in conditional volatility and suggests that high peaked (leptokurtic) distributions would be a better fit. Angelidis et al. (2004) generates several combinations of GARCH models with different specifications for the conditional mean process AR(p) and compares the performance of the models with normal and non-normal distributions in VaR estimation. The authors find that the experimented mean modeling does not significantly affect the outcome of the VaR forecasts, and state that no superior model is found. However, they conclude that the non-normal GARCH models, especially the leptokurtic Student's t distributed models outperforms the normal models.

Some studies criticize the VaR framework for the lack of sub-additivity (Artzner et al. 1997, 1999), meaning that the VaR of two merged portfolios should not exceed the sum of the two separate VaRs (Tsay 2010). Despite this there still is no better quantifying risk measure (Orhan et al. 2012). Jorion (2007) states that VaR models that are rejected in the statistical tests of backtesting procedures should be reexamined for untrue assumptions and inaccurate modeling and refers to the backtesting methods as final "diagnostic check" (Christoffersen 2003), which is why we choose to evaluate our models by using backtesting methods.

2 Theoretical Framework

This section consists of the essential concepts and theoretical aspects that are needed to understand the analysis of this thesis. We introduce log returns and time series, discuss the main idea of volatility and present the concept of Value at Risk.

2.1 Log Returns

In financial studies that involve stock prices it is commonly seen that the observations that are analyzed are the logarithmic return series of the given asset, rather than the actual stock prices. This is because the log returns of an asset are more manageable and have better statistical behavior. For instance, when differencing the stock prices of an asset, (as is done to obtain log returns), the minor fluctuations and variations in the asset are reduced and become more consistent over time. The log return also brings a scale-free assessment and the multi-period log returns are evidently the sum of all the one-period log returns (Tsay 2010, pp.5).

We let P_t denote the price of an asset at time t. The simple one-period gross return is then defined as

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

The continuously compounded return, or the log return, r_t is simply the natural logarithm of the one-period gross return. The log return series r_t is expressed as

$$r_t = \log(1 + R_t) = \log(\frac{P_t}{P_{t-1}}) = \log(P_t) - \log(P_{t-1}) = p_t - p_{t-1}. \tag{2}$$

2.2 Time Series and Stationarity

Time series are formed when a variable is sequentially measured over time in a fixed interval. The measurements could be daily, monthly or quarterly observations over a certain time horizon, say one year. In this thesis our data sample consists of daily observations of the closing prices for OMXS30 index from Jan 1st 2005 to Dec 29 2017, forming a time series of 3286 observations in total. Calculating the log return of the given time series asset, it results in a sequence of random variables over time also measured in a fixed interval. Thus, the log returns $\{r_t\}_{t=1}^{T}$ form a time series where T is the number of daily observations.

For a linear log return series we express

$$r_t = \mu_t + a_t \tag{3}$$

where μ_t detones the mean of r_t, a_t denotes the error term a sequence of independent and identically distributed (i.i.d) random variables with mean zero and constant variance σ_a^2. That is, the residual series a_t is assumed to follow a white noise process. The error term a_t is often referred to as the mean-corrected return since $a_t = r_t - \mu_t$. Throughout this thesis, a_t will be referred to as the *innovations* or *shocks* of the log return series.

The conditional mean and variance of the log return series, given the information set F at time *t-1* where F_{t-1} consists of all linear functions of past returns (Tsay 2010, pp.111), are introduced as

$$\mu_t = E(r_t|F_{t-1}), \quad \sigma_t^2 = Var(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}]. \tag{4}$$

A lag operator l is a function that computes a lagged version of the time series by shifting the time base by a number of observations l. For example, the lag-1 function of r_t is r_{t-1}. The time series r_t is said to be weakly stationary if the mean is constant, that is, if $E(r_t) = \mu$ for all t and the covariance only depends on the lag l such that $Cov(r_t, r_{t-l}) = \gamma_l$. In other words, the mean does not change over time and the covariance between the return series r_t and its lagged values r_{t-l} are also time invariant (Tsay 2002, pp.23).

2.3 Heteroskedasticity

When graphically examining financial time series data it can be seen that the series fluctuates a lot over time. The series usually exhibits positive and negative shocks that could result from different external factors such as political, environmental or economically driven issues. This may trigger a non-constant variance in the error terms of the log return series, so that the variance of the residuals a_t in Equation 3 changes over time. When a model suffers from this condition of variety in the variance of the error terms it is assumed to be *heteroskedastic*, (Berry et al. 1985), and tends to manifest non-normality.

Heteroskedasticity can occur in two different forms: conditional or unconditional (Hayashi 2000). When time series exhibits unconditional heteroskedasticity the variations in the data can usually be tied to certain cycles or variables, such as seasonal variations or trends. In these types of time series, periods of low or high variability can be predicted and identified. Conditional heteroskedasticity on the other hand is not as easily predicted or recognized. In a conditional heteroskedastic time series the observations that are close to each other tend to be correlated. The changes and correlations in the variance of the error terms results in conditional volatility (Cowpertwait et al. 2009, pp.137). This means that the evolution of the conditional variance is directly tied to the volatility process of the given asset return (Tsay 2002, pp.17-18).

2.4 Volatility Clustering

When a market is volatile there are large price movements in both directions that are hard to predict. Since there is only one observation per trading day, volatility is not directly observable (Tsay 2010, pp.110-111). Despite this, one can properly understand and predict the nature of volatility by studying how it behaves inasset returns. When studying financial asset returns the special features of volatility commonly revealed. Generally, the three most important properties of volatility are that the volatile periods tend to cluster together, the volatility is mean reverting and persistent. We will briefly explain and discuss these properties in order.

When studying graphics of the underlying asset return over a given time period it is usually seen that the return series exhibits extended periods of low volatility and other periods with high volatility. The periods with large fluctuations tend to cluster together and vice versa. This phenomenon is known as volatility clustering and has been an important consideration in the development of volatility models (Tsay 2002, pp.80).

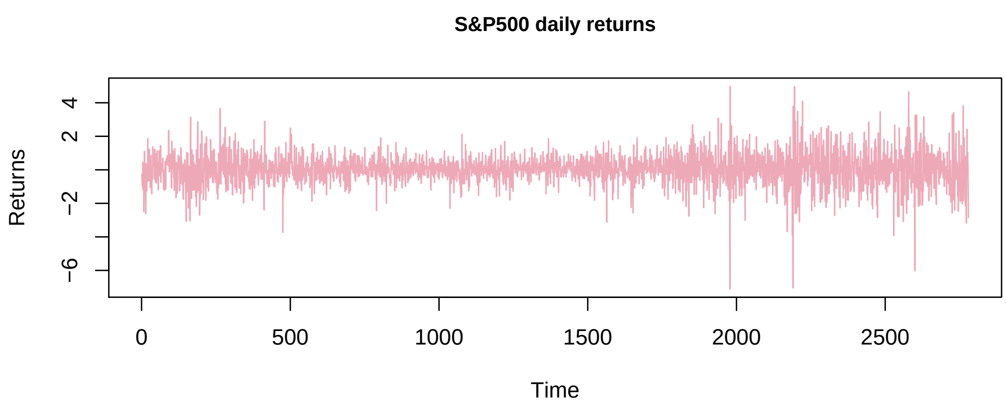


Figure 1: Daily returns of S&P500 from Jan 1st 1990 to Dec 27th 1999. Source: MASS package i R-studio.

We demonstrate the daily returns of The Standard And Poors 500 index in Figure 1 where the presence of volatility clusterings are clear. This property implies that large changes tend to be followed by large changes and small changes tend to be followed by small changes and informs the financially involved that when the market suffers a volatile shock, more volatility should be expected.

Mean reverting conditional volatility implies that in the long run, the conditional variance should revert to its average, which is the unconditional variance. To give an informal explanation of this property, the mean reversion of volatility lets us know that a highly volatile market will eventually become calm again. The persistence of volatility indicates that big price movements tend to endure for a while after that first shock. How the conditional heteroskedastic models replicate these properties is discussed later in section 3.2.2.

2.5 Value at Risk

Value at Risk is a measure of market risk that estimates the maximal loss that may occur under normal market conditions, given a corresponding confidence level and a time horizon (Jorion 2007). To exemplify, we consider an investor holding a stock where the daily VaR may be 100 SEK at 99% level of confidence. The VaR then indicates that there is a 1% chance that a loss greater than 100 SEK could occur the next day under normal market conditions, or equivalent, the worst expected loss will not exceed 100 SEK with 99% confidence over the next day. From a financial point of view, VaR is treated as an estimate of loss associated to rare events under normal market conditions, which is equivalent to VaR being defined as the minimal loss under rare market conditions. Although the definitions seem to differ, both interpretations will generate same measures of VaR, (Tsay 2010, pp.325-326).

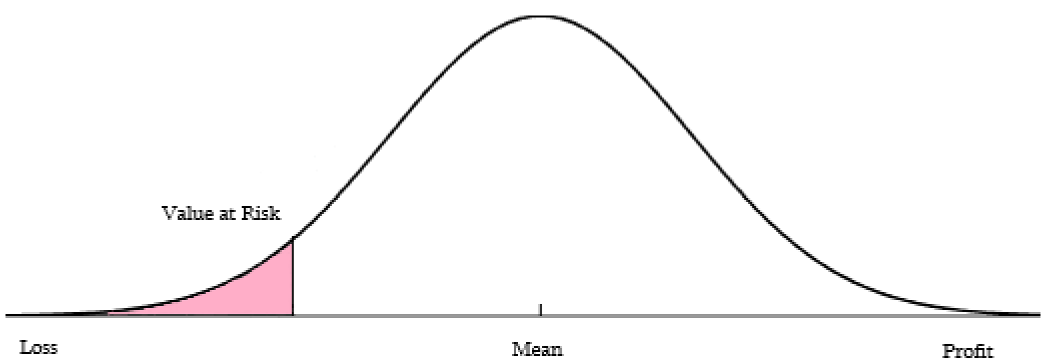


Figure 2: Illustrating Value at Risk as a quantile of a normal density function at a certain confidence level.

Statistically speaking, VaR can be described as the left tail or the “loss” quantile of the distribution function of the profits and losses in underlying asset return over a certain time period (Patton 2011, pp.17). We graphically illustrate the VaR as a quantile of a normal density function in Figure 2. Since the VaR estimates are projected from the distribution model of the log return series, the observed volatility in the series has an important impact on the outcome of the VaR estimates.

We define Value at Risk in terms of probabilities for the portfolio as in Christoffersen 2003, pp.48 where r is the observed return and $\alpha \in [0, 1]$ denotes the confidence level between zero and one. Then the probability that VaR exceeds the return is expressed as

$$\Pr(r < -VaR) = \alpha. \tag{5}$$

To calculate the VaR in Equation 5 one can make an assumption regarding the distribution of the underlying return set. Previous studies (Jorion 2007 and Mandelbrot 1963) suggest that fat-tailed distributions are a better fit than normal. In this thesis we use Normal and Student's t distributed innovations to estimate the one step ahead (daily) VaR on a 99% level of confidence for $\alpha = 0.01$.

We assume that the returns are normally distributed with mean μ and standard deviation σ where α the quantile of which VaR will be calculated. Then the VaR^{α} can be expressed by a function of the quantile distribution of profits and losses calculated as (Christoffersen 2003, pp.49)

$$VaR^{\alpha} = \mu - \sigma \Phi^{-1}(\alpha), \tag{6}$$

where Φ denotes the cumulative distribution function for a standard normal distribution and $\Phi^{-1}(\alpha)$ calculates quantile mass for α. This could of course be applied for other distributional assumptions. For the standardized Student's t distribution we express the VaR as the quantile function

$$VaR^{\alpha} = \mu - \sigma t_v^{-1}(\alpha), \tag{7}$$

where $\sigma^2 = \frac{v}{v-2}$ when $v > 2$, so that σ is the standard deviation and v detones the degrees of freedom, and $t_v^{-1}(\alpha)$ denotes the α quantile of the standard t distribution with v degrees of freedom (McNeil et al. 2005, pp.40).

2.5.1 Shortcomings of Value at Risk

Although VaR is the most popular and frequently used measure of risk the VaR framework has some shortcomings to it that has in fact been criticized, originally by Artzner et al. (1997, 1999). In this section we discuss some of these shortcomings.

As presented in previous section in Equation 6 and Equation 7, we see that the VaR is the loss tail quantile of the underlying return distribution function. However, the VaR does not fully describe the lower tail behavior of the function (Tsay 2010, pp.328). This means that VaR considers *if* an exceedance occurs yet fails to describe the severity of that loss. The consequence of this could in practice mean that two separate assets with the same VaR measure may experience different losses when an exceedance is encountered.

Furthermore, Artzner et al. (1997, 1999) showed that the VaR model is not subject to the sub-additivity property. The property states that the risk measure of two merged portfolios should be no greater than the sum of the individual risk measures of the two portfolios (Tsay 2010, pp.328), which reasonable risk measures are believed to have.

Since the VaR is calculated on the predictive distribution of future returns the VaR model is dependent on the distribution that it projects its estimations on (Tsay 2010, pp.328). Thus the VaR model also, to some extent, becomes subject to model risk and parameter uncertainty. This means that if the risk management models that are being used for estimating the VaR are miss specified or make assumptions about the distribution that are unrealistic, then the financial institutions will incur losses (McNeil et al. 2005, pp.40-41). In a properly specified model the predictive distribution takes into account the parameter uncertainty, however it is difficult to obtain the predictive distribution and most VaR methods ignore the effects of the uncertainty (Tsay 2010, pp.328).

A common assumption in VaR framework is that the distribution of the underlying asset return is normal. In reality, the conditional volatility in the return series is heavy tailed and displays non-normal behavior. This assumption might lead to a model that does not fully capture the extreme values so that the generated forecasts of the model might underestimate the Value at Risk (Tsay 2010, pp.332). We note that underestimating VaR is *not* the same as underestimating risk. A model that is underestimating VaR means that the model does not resemble the properties of the observations in the portfolio. This naturally generates a VaR that is too low, which in turn leads to more exceedances in the backtesting.

3 Time Series Models

This section provides the theoretical aspects and definitions regarding the time series models that are used in the analysis of this thesis.

3.1 Autoregressive Moving Average (ARMA)

The autoregressive model (AR) is a model in which future predictions are estimated by regressing on its past lagged values. The model explains the current value of a return series expressed as a function of its previous values. The AR model with p parameters, or AR(p) is given by

$$r_t = \phi_o + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t \tag{8}$$

where a_t is a white noise series with zero mean and variance σ_a^2, p is a non-negative integer determining the order, that is, the number of lags to include in the AR model, and r_{t-i} $(i = 1, ..., p)$ jointly determines the conditional expectation of r_t given the previous values, (Tsay 2010, pp.38).

The moving average (MA) model is also a model that regresses on its previous values but instead includes the lagged error terms a_{t-j} and current error term. The MA(q) model is described as

$$r_t = \mu_t + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t. \tag{9}$$

Similar to the AR model, a_{t-j} is a white noise series with zero mean and variance σ_a^2 and q is a non-negative integer determining the order of the MA model. Since the MA model is a linear combination of a white noise sequence, the model is always weakly stationary (Tsay 2002, pp.44).

Previous researchers has observed that it requires many parameters to adequately describe the structure of the data when modeling with the AR(p) and MA(q) models separatedly (Tsay 2002, pp.48). In order to surpass this problem and reduce the number of required parameters, the ARMA model was introduced by Box, Jenkins and Reinsel (1994). The ARMA model combines the AR(p) with MA(q), to handle the impact of dependencies in the series. The equation for the conditional mean in Equation 4 specified by the ARMA(p,q) model is expressed as

$$\mu_t = E(r_t|F_{t-1}) = \phi_o + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j}, \tag{10}$$

where it is seen that the first part of the equation represents the AR(p) model and the second represents the MA(q) model. In this thesis we later specify the ARMA(1,1) model for our conditional mean which is directly given by

$$\mu_t = E(r_t|F_{t-1}) = \phi_o + \phi_1 r_{t-1} - \theta_1 a_{t-1}. \tag{11}$$

We chose the simplest order of the ARMA(1,1) model based on (Angelidis et al. 2004) who suggested that the order of the mean model does not significantly affect the outcome of the VaR estimates.

3.2 Conditional Heteroskedastic models

3.2.1 Autoregressive Conditional Heteroskedasticity (ARCH)

The autoregressive conditional heteroskedastic model was introduced by Engle (1982). The ARCH model was the first model to provide systematic framework for modeling conditional volatility and has since then been widely used in financial institutions.

An informal explanation of what the ARCH model does is that it examines the data and predicts what risky outcome we might expect in the short-term future. The model does this by saying that when the market is volatile today we expect it to be volatile tomorrow, but that this eventually goes away. Respectively, if the market is calm today it is likely to be calm tomorrow, but it will eventually become more volatile again.

Te ARCH model lets the conditional variance be described by a quadratic function of its lagged error terms (Tsay 2010, pp.115-116). The concept of the model is that the innovations a_t of the log return series are serially uncorrelated but dependent.

The ARCH model of order p, or the ARCH(p) model assumes the following,

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2, \tag{12}$$

where ϵ_t are i.i.d random variables with zero mean and unit variance, $\alpha_0 > 0$, $\alpha_i \geqslant 0$ for $i > 0$. The structure of the model implies that if past values of the innovations $\{a_{t-i}^2\}$ are large then the conditional variance σ_t^2 may be large itself, hence, imitating the behavior of conditional volatility clusterings (McNeil et al. 2005, pp.139).

In this thesis we aim to introduce how conditional volatility can be modelled and used in estimating daily Value at Risk. However, we will not provide any empirical results of any ARCH model because we believe that the purpose can be carried out without the empirical results of ARCH model. The method of modeling an ARCH model is very similar to modeling the GARCH model and we provide all the necessary information regarding building and evaluating a volatility model in section 4 and 5.

3.2.1.1 Shortcomings of ARCH

Despite the fact that the ARCH model has many advantages, it is quite restrictive. First, as seen by the structure of the model in Equation 12 it requires many parameters (or high orders p) to appropriately describe the volatility process. Secondly, the predictions of the model are estimated by regressing on the *squared* previous innovations $\{a_{t-i}^2\}$. This means that the ARCH model suggests that positive and negative shocks have the same effect on the conditional volatility (Tsay 2010, pp.119). This is of course not true since stock prices should clearly react differently to positive and negative shocks. The ARCH model can decently describe how the conditional variance of an asset return operates, however, it does not explain the source or motive of such behavior to occur in the first place.

3.2.2 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

In 1986, Bollerslev presented an extension of the ARCH model, namely, the Generalized ARCH model. The GARCH model is widely used in financial institutions for estimating conditional volatility as well as being a guidance in financial decisions makings concerning risk analysis and portfolio selection (Engle 2001). The GARCH model not only depends on the past squared innovations of the log returns but also includes the preceding values of the conditional variance $\{\sigma_{t-j}^2\}$. We let $a_t = r_t - \mu_t$ denote the mean corrected innovation set at time t. Then a_t follows a GARCH(p,q) model if

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{13}$$

where ϵ_t again is i.i.d. with zero mean and unit variance, $\alpha_0 > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$ (Tsay 2010, pp.131-132). We note that the GARCH model has the ARCH(p) model in Equation 12 as a special case GARCH(0,p), when $q = 0$.

Similair to the ARCH model, the GARCH model fails to recognize the difference between positive and negative shocks. Despite this it has been proven to be capable of successfully describing the volatility process of a return series even in its simplest form (Engle 2001). It provides a better way of modeling the persistent volatility since it requires less parameters, where the ARCH model would need high order p to model persistent volatility. Thus, we chose to limit our analysis to implementing the most commonly applied form of GARCH, that is, the GARCH(1,1) model. The GARCH (1,1) model is defined by

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{14}$$

where $\alpha_1 + \beta_1 < 1$, which is the criteria that need to be fulfilled for the conditional volatility to be mean reverting (Engle 2001). Mean reverting volatility model, as mentioned in section 2.4, implies that the long-run forecasts of volatility should converge to its unconditional variance. This means that if we are using the GARCH model to forecast the h:th step ahead observation r_{t+h} where $h \to \infty$, then the current value r_t would not have a great impact on the forecast (Engle et al. 2001). The closer the coefficients $\alpha_1 + \beta_1$ are to adding up to one, the more persistent the volatility. That is, the closer $\alpha_1 + \beta_1$ is to one, the changes in the variance will have a more persistent effect to the model.

In this thesis we forecast one step ahead conditional volatility with the GARCH(1,1) model and we express daily forecasts by

$$\sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \tag{15}$$

where h is the forecast origin (Tsay 2010, pp.133).

4 Method

In the oncoming sections we provide essential information on *how* we carry out the analysis in this thesis. We explain the methods of in-sample and out-of-sample and give detailed information on how the forecasting and backtesting procedures are conducted. A brief explanation on how we build volatility models is also provided, inspired by Tsay (2002,2010).

4.1 Software

The statistical software used in this thesis for the graphical and empirical analysis is the open source program "R". Open source software means that anyone has free access to the program, as well as packages with built in codes that can be used for a specific type of statistical issue. In this thesis we use the package "tseries" (Trapletti 2018) for processing the data, and the "rugarch" package (Ghalanos 2014) for the model fitting and the backtesting forecast of Value at Risk. The official website for R is "https://www.r-project.org/", where R can be downloaded.

4.2 In-sample and Out-of-sample

Model selection should not only be based on the goodness-of-fit of the model to the data, but also on the objective of the analysis. When forecasting is of interest, the best fitted model does not necessarily generate more accurate forecasts (Tsay 2008). To overcome this problem, many to use the performance of the out-of-sample forecasts as a guidance in the selection of a model. By out-of-sample, we mean that the data that is used in model fitting is different from those that are used in the forecasting evaluation.

Typically, one may choose to divide the total sample period into two sub-periods. The first period is called the *in-sample* (or estimation sample) and the second period is called the *out-of-sample* (or forecasting sample). The observations in the in-sample period are used for estimating the parameters of the models and the obtained models will generate forecasts of observations of the out-of-sample period. The out-of-sample consists of observations that will be used for evaluating the predictive accuracy of the in-sample models forecasts (Tsay 2008). The purpose of this is generally to determine how well the model performs in a period where the observations were not used to estimate the models' parameters and is most commonly done in risk management practices. Consider for example the series $\{r_1, .., r_T\}$ where T is the total number of observations in the sample. We divide the total sample by $\{r_1, .., r_n\}$ for the in-sample and $\{r_{n+1}, .., r_T\}$ for the out-of-sample where n denotes the forecast origin (Tsay 2008).

Choosing the in-sample size for the estimation window is an important part of generating good forecasts. If the in-sample size is to small then the generated forecasts may not be accurate or reliable. On the other hand, if the sample size is too large then the forecasted volatility might be smoothed out since it is mean reverting, which would bring forecasts that do not represent the structure of the volatility process. However, we determined the sample sizes based on Tsay (2008) where he claims that a reasonable in-sample size n is $n = \frac{2T}{3}$.

In this thesis the gathered data consist of 3285 observations in total for the OMXS30 index from Jan 1st 2005 to Dec 29th 2017. The in-sample period that is used for estimating our models are on the first 2285 observations, whereas the remaining thousand observations will be the out-of-sample period used for evaluating the corresponding thousand forecasts of the models.

4.3 Forecasting method

To generate forecasts we use a rolling window forecast method where the estimation window moves forward one step at a time. First we let all observations between time $\{1, .., n\}$ produce a one step ahead forecast $\widehat{n+1}$, which is saved as the first observation in the forecast vector with length 1000. The sample window then jumps one step forward and now uses the observations $\{2, .., n+1\}$ for generating observation $\widehat{n+2}$, which is stored as the second element in the forecast vector. It is vital to note that the included observation $n+1$ used in the rolling window is *not* the prediction from the first step but represents the real observed data from the out-of-sample. The procedure continues to iterate until all thousand observations of the out-of-sample period are forecasted, and the estimation sample size stays constant at every iteration.

Some argue that the moving window forecast may remove important information once it rolls past data points. However, this rolling forecast method is often operated since it should capture the changes in the volatility more properly by including the preceding observation for every forecast.

4.4 Backtesting

A VaR model is only useful if it provides accurate predictions of future risks. Therefore, it is important to determine the reliability of the estimate which can conveniently be done using backtesting methods. The backtesting procedure is a statistical framework designed to verify if the VaR estimates are in line with the actual profits and losses (Jorion 2007, pp.139).

An exceedance means that the actual returns are below the estimated VaRs, or equivalent, the observed loss return is greater than the estimated VaR. In this thesis exceedance will also be referred to as an exception or failure.

Generally there are two important properties that should be satisfied for a proper VaR model. Those are that the total exceedances, or failures, match the expected exceedances given by the level of confidence, and that the exceedances occur independently.

To put this in picture, we consider our case where we estimate 1000 daily VaRs with a 99% level of confidence. The 99% confidence level implies that we expect a failure to occur once every 100 days. Since we estimate a thousand VaR forecasts with $\alpha = 0.01$, the total amount of expected exceedances is $1000 * 0.01 = 10$.

In the backtesting, the actual observed returns in the out-of-sample are compared with the predicted VaR estimates, and every exceedance is counted. The accumulated exceedances are then compared with the expected exceedances to make sure that the frequencies of the failures agree with the determined level of confidence. If the total failures are lesser than the expected, then the model is considered to overestimate the VaR, and if the actual failures are greater than the expected, the model underestimates VaR. These types of tests are straightforward and do not take into consideration for when the exceedances occurred and are known as tests of *unconditional coverage* (Jorion 2007).

Moreover, there are tests of *conditional coverage* that take the time variation of the exceedances into account. The conditional coverage tests examines if the exceedances are evenly spread or if they cluster together. If the exceedances are clustered, the VaR model has not been specified accurately enough to capture the changing volatility and correlations of the data (Jorion 2007).

In this thesis we conduct the backtesting with two of the most popular evaluation tests, namely, Kupiec's unconditional coverage test and Christoffersen's conditional coverage test.

4.4.1 Kupiec's Unconditional Coverage test

Let N denote the number of days where the loss is greater than the forecasted VaR over T days calculated by $N = \sum_{t=1}^{T} I_t$. The indicator variable I_t counts every exception by,

$$I_{t+1} = \begin{cases} 1, & if \quad r_{t+1} < -VaR_{t+1}^{\alpha}, \\ 0, & if \quad r_{t+1} \geqslant -VaR_{t+1}^{\alpha}, \end{cases} \tag{16}$$

where r_{t+1} represents the actual return at time $t+1$ and $-VaR_{t+1}^{\alpha}$ represents the one step ahead forecast of VaR at time $t+1$.

Kupiec (1995) argues that for a proper model the number of failures N should follow a Binomial distribution $N \sim Bin(T, \alpha)$, so that the expected failure frequency is $\frac{N}{T} = \alpha$. He introduces the likelihood ratio statistic LR_{UC} where the expected frequency of exceedances under the null hypothesis is,

$$LR_{UC} = 2\ln\left[(1-\frac{N}{T})^{T-N}(\frac{N}{T})^{N}\right] - 2\ln\left[(1-\alpha)^{T-N}\alpha^{N}\right] \sim \chi^2(1) \qquad (17)$$

In our case the null hypothesis is rejected if the test statistic LR_{UC} is greater than 6.635 on a 5% significance level. It is worth mentioning that the test does not recognize if the failures were less or greater than the expected, so the test does not give an answer to if the model overestimated or underestimated the VaR.

4.4.2 Christoffersen's Conditional Coverage test

Christoffersen (1998) developed an extension to Kupiec's unconditional coverage test and argues that the independence of the failures must be taken to account. If the exceedances are clustered together then the estimated model might not be specified well enough so that the conditional volatility is not fully captured. According to Christoffersen's test this would mean that even if the number of failures hold the expected, the model would still be rejected if the failures are not independent.

Let n_{ij} be the number of observations with value i followed by j, for $i, j = 0, 1$. Further, let $\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$ be the probabilities

$\pi_{01} = \mathrm{P}(I_{t+1} = 1 | I_t = 0) = \mathrm{P}(\text{Exceedance tomorrow} \mid \text{No exceedance today})$
$\pi_{11} = \mathrm{P}(I_{t+1} = 1 | I_t = 1) = \mathrm{P}(\text{Exceedance tomorrow} \mid \text{Exceedance today})$

If failures I_t are independent, then the probabilities should be equal to the confidence level, that is, $\pi_{01} = \pi_{11} = \alpha$. The likelihood ratio statistic LR_{CC} for Christoffersen's conditional coverage test is computed by,

$$LR_{CC} = 2\ln[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}] - 2\ln[(1-\alpha)^{T-N}\alpha^{N}] \sim \chi^2(2). \quad (18)$$

The null hypothesis of correct exceedances and independence of failures is rejected if LR_{CC} is above the critical value 9.21 in our backtest.

4.5 Building a volatility model

The concept of volatility modeling is that the log return series r_t is serially uncorrelated, but dependent (Tsay 2010, pp.111). To explain how this is possible, we remember that serial correlation measures the strength of the *linear* dependency between a return series and its lagged values (see Appendix A for ACF and more on serial correlation). However, the function does not capture the more complicated behavior of dependency such as changing volatility, since it is non-linear. Thereby, a serially uncorrelated but dependent return series means that the return series r_t and its lagged values r_{t-l} are indeed correlated and thus dependent, but the dependency is non-linear and cannot be described by a linear correlation function.

The first step of building model for the conditional volatility is to examine the serial dependence to determine if the log return series is serially uncorrelated but dependent. This will be done by graphical analysis of the ACF plots of our log return series and complemented by Ljung-Box tests (see Appendix B for Ljung-Box). For this property to be fulfilled the ACF plot of the log return should resemble a white noise series (e.g show no correlation), whereas the ACF plot of the squared log returns should show strong correlation. If the criteria is not fulfilled, we might need to specify a mean equation (ARMA model) to extract any linear dependencies in the log return series. Once the model for the conditional mean in 4 is fitted the linear dependencies should be removed from the residuals but left in the squared residuals. Checking for significant correlations in the squared residuals of the ARMA model is sometimes refered to as "testing for ARCH effects". The objective of building the volatility model is essentially to construct a variance measure σ_t^2 that has the property that the standardized squared returns $\frac{r_t^2}{\sigma_t^2}$ has no autocorrelation patterns. We express the standardized residuals as

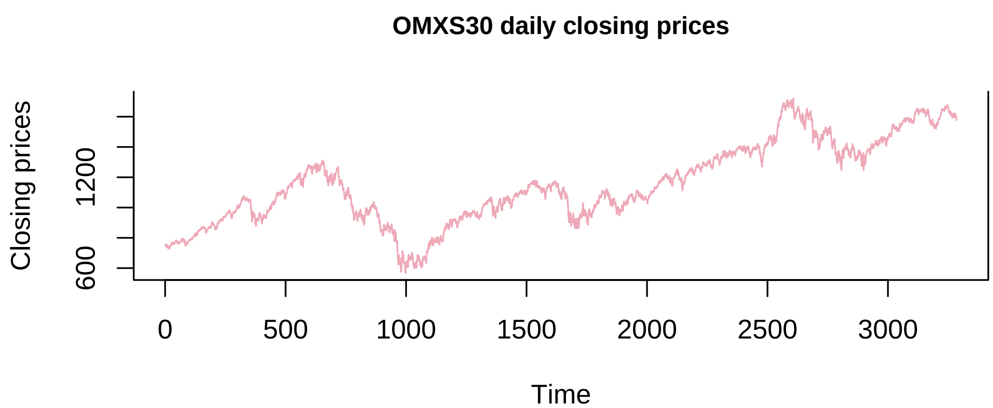
$$\tilde{a}_t = \frac{a_t}{\sigma_t}. \tag{19}$$

If there exists significant serial correlation in the ACF plot of the squared standardizes residuals of the ARMA model it is suitable to use conditional heteroskedastic models such as ARCH and GARCH. Once the GARCH models are fitted we test the GARCH models standardized residuals to make sure that all conditional volatility is captured. If the volatility is properly captured, the ACF plot of the innovations should follow a white noise series.These steps of building a model for conditional volatility of asset returns are inspired by Tsay (2010, pp.111-115).

5 Data

5.1 Descriptive statistics

The data used in this thesis consists of the daily closing prices from the OMXS30 index. Tsay (2010) suggests that volatility tends to be higher for a single stock since it is more sensitive to market news. We consciously choose the OMXS30 index so that external factors such as false rumors, product releases or other stock specific actions will not have a significant impact on the volatility process of our data. The data is gathered from finance.yahoo.com with a sample period from Jan 1st 2005 to Dec 29 2017 with a total of 3286 observations. For the continuously compounded return series as in Equation 2, the sample size is 3285 since it naturally loses one observation in the beginning. We illustrate the daily closing prices and the calculated log returns in Figure 3.



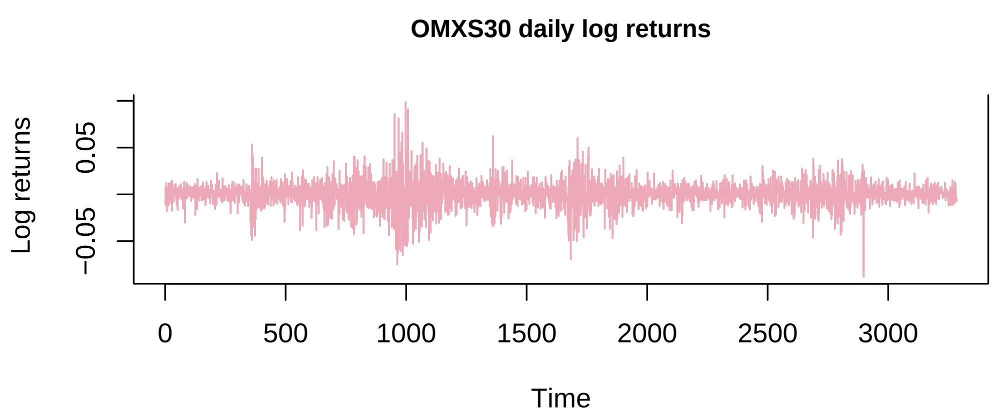


Figure 3: Time plot of OMXS30 daily closing prices (Upper), and the log return series OMXS30 (Lower) from Jan 1st 2005 to Dec 29th 2017.

In Figure 3 (Upper) it is seen that the stock market crash of Sep 29 2008, around the 1000^{th} observation, has had a major influence on the closing prices. The closing price decreases sharply but then continues to follow a positive trend. In Figure 3 (Lower) the log return series has its most volatile period during the financial crisis of 2008, in accordance with the upper plot in Figure 3. The log return plot also reveals volatility clusterings which is seen by periods of high volatility followed by periods of low volatility, as discussed in section 2.3.

For a normally distributed sample, the skewness is expected to be zero and the kurtosis 3, where the zero skewness indicates symmetric distribution (see Appendix C for more details about skewness and kurtosis). The Student's t distribution is symmetric just as the Normal distribution but with higher peak since it is heavy-tailed. We provide the descriptive statistics for the log return series of OMXS30 in Table 1. In Table 1 is can be seen that the skewness of the log return series is close to zero which indicates that the symmetry seems to be fulfilled. However, the high kurtosis in Table 1 reveals that the data is to heavy-tailed in order to fit the normal assumption. The kurtosis implies that a high-peakes, or *leptokurtic*, distribution would be a better fit than a normal distribution.

Mean	0.00027
Median	0.00060
Minimun	-0.07513
Maximum	0.09865
Std.Dev	0.01444
Skewness	0.04270
Kurtosis	7.82603

Table 1: Descriptive statistics of the OMXS30 log return series from Jan 1st 2005 to Dec 29th 2017.

To conlude this section we state that the volatility clusterings of the log return series in Figure 3 and the descriptive statistics in Table 1 clearly indicate that the log return series does not fit the normal assumption. Despite this we will build our volatility models assuming both Normal and Stundent's t distributed innovations.

5.2 Testing for ARCH-effect

To determine whether log return series in Equation 4 is serially uncorrelated but dependent, we check for significant correlation in the squared log return series and no correlation in the non-squared log return series. As explained in section 4.5, we test for ARCH-effect by examining the ACF plot of the log return series and the squared log return series, shown in Figure 4. The pink dashed lines in the ACF plot denotes the 95% confidence interval. If the number of significant lags that exceed the confidence interval are more than 5% of the total lags, the series is said to be correlated. Note that the zero lag is always strongly significant since it describes the correlation function between the series with no lag, that is $Cov(r_t, r_t)$, and is therefore equal to one.

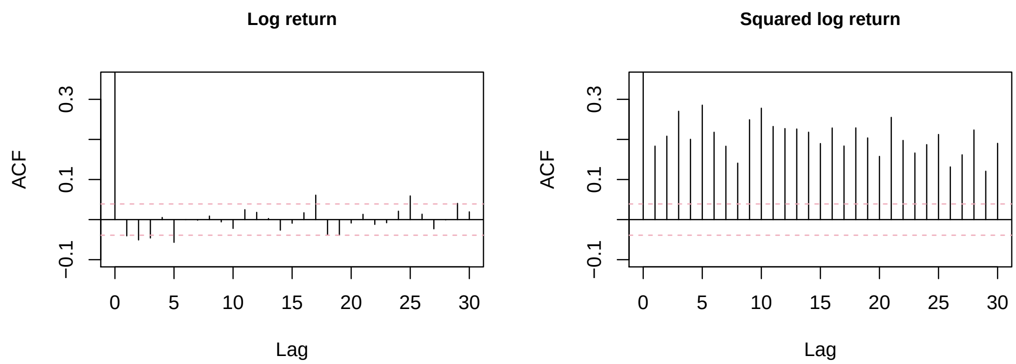


Figure 4: Autocorrelation function of the log returns (Right) and squared log returns (Left), with the pink lines denotes the 95% confidence interval.

In Figure 4 the strong correlation in the squared returns confirms the presence of conditional heteroskedasticity, i.e., the ARCH effect. Although the ACF of the non-squared series exhibits relatively weaker correlation compared to the squared returns, it is seen that the number of significant lags of the log returns seem to be greater than 5% of total lags. This could indicate that the serial correlation is non-zero and violate the assumption of serially uncorrelated return series.

The graphical analysis is complemented by performing a Ljung-Box test, where the null hypothesis assumes no serial correlation up to the k:th lag in the return series, whilst the alternative hypothesis assumes non-zero autocorrelation. The results are summarized in Table 2.

OMXS30 Series	**Chi.squared statistic**	**p-value**
Log returns	58.831	0.000151
Squared log returns	2969.94	<2.2e-16

Table 2: Statistics of the Ljung-Box test for the log returns of OMXS30 up to the 35:th lag.

The null hypothesis of the Ljung-Box test is rejected for the squared log return, indicating that the series is conditionally heteroskedastic which verifies the usage of an ARCH model. The test also stipulates that correlation in the non-squared series is present, which means that the series does not hold the assumptions of the serially uncorrelated log return series. To extract any serial correlation from the observed return series we specify a model for the conditional mean in Equation 4 by building an ARMA(1,1)-model, as in Equation 11.

As mentioned in section 4.5, we check the standardized residuals to ensure that the linear dependencies are removed from the standardized residuals of the ARMA(1,1) model. A Ljung-Box test is performed and the p-value for the test is 0.07881, so the null hypothesis of zero correlation holds. The ACF plot of the residuals of the ARMA(1,1) are provided in Appendix E. It is seen in Figure 8 in that the significant lags in the residuals are reduced, which means that we managed to extract the serial correlations. We can conclude that the fitted model has uncorrelated residuals and clear evidence of strong correlation in the squared residuals, in other words conditional volatility. Hence, the ARMA(1,1) model seems to have all desired properties, which brings us in position to fit the and GARCH models to capture the conditional heteroskedasticity of the error terms.

6 Empirical Results and Analysis

6.1 GARCH estimation

As mentioned before we fit two models: an ARMA(1,1)-GARCH(1,1) model with Normal innovations and one ARMA(1,1)-GARCH(1,1) model with Student's t distributed innovations. We refer to the models as the Normal model and the Student's t model. In this section we present the obtained models and check how well the criteria of the GARCH model hold. The main subjects that are analyzed are that the residuals of the fitted models should follow a white noise series, the empirical quantiles should match the theoretical quantiles of the assumed distribution and we also present the parameters of the models. Note that the models are estimated on the in-sample period for the OMX30 log return series (see section 4.2).

Fitting an ARMA(1,1)-GARCH(1,1) model we expect the residuals to resemble a white noise series for the model to be adequate since it should have captured the conditional volatility of the error terms. The Student's t model and the Normal model has very similar correlation plots which is why we choose to only present the ACF plot of the standardized residuals and the squared standardized residuals of the Normal model in Figure 5. In Figure 5 the standardized residual series seem to imitate a white noise series. The squared standardized residuals in the left plot in Figure 5 show one significant observation at lag 11, but this implication is not strong enough for us to reject that the series is i.i.d. We support this statement by using Figure 10 in Appendix E as a reference where we illustrate the outcome of significant lags for a simulated white noise series with 10'000 standard normal random variables.

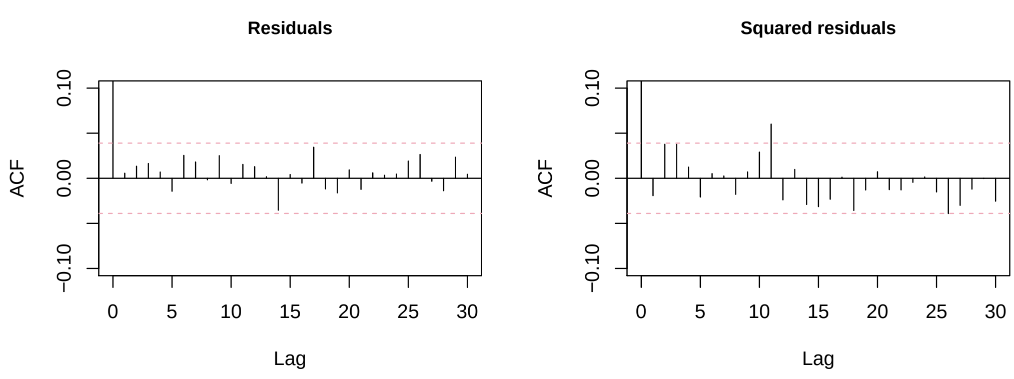


Figure 5: The ACF plot of the standardized residuals (Left) and the squared standardized residuals (Right) for the ARMA(1,1)-GARCH(1,1) model with Normal conditional distributed innovations. The pink dashed lines denote a 95% confidence interval.

We compare the serial correlations in Figure 5 with the ones in Figure 4 and we can state that the models have successfully filtered out the dependencies in the squared series. This is also confirmed by Ljung-Box test on the squared residuals that show a p-value of 0.08827 for the Normal squared innovations and 0.1488 for the Student's t squared innovations. Thus, the null hypothesis is not rejected for both models which means that there is no serial correlation in the variance of the error terms of the Student's t model and Normal model. This indicates that the standardized residuals of the models contain no conditional heteroskedasticity so that the volatility seems captured by the fitted models.

If the models are fitted accurately then the assumed distributions for the innovations of the models should be in line with the empirical distribution of the innovations. In Figure 6 we demonstrate the quantile-quantile plots for the standardized residuals of the models, where the left figure represents the Normal model's innovations and the right figure represents the Student's t model's innovations. The Student's t innovations are plotted against a quantile from a standardized Student's t distribution with seven degrees of freedom, and the Normal innovations are compared to a standard Normal distribution. The degrees of freedom were determined by testing different numbers and choosing the one where the line fitted most observations.

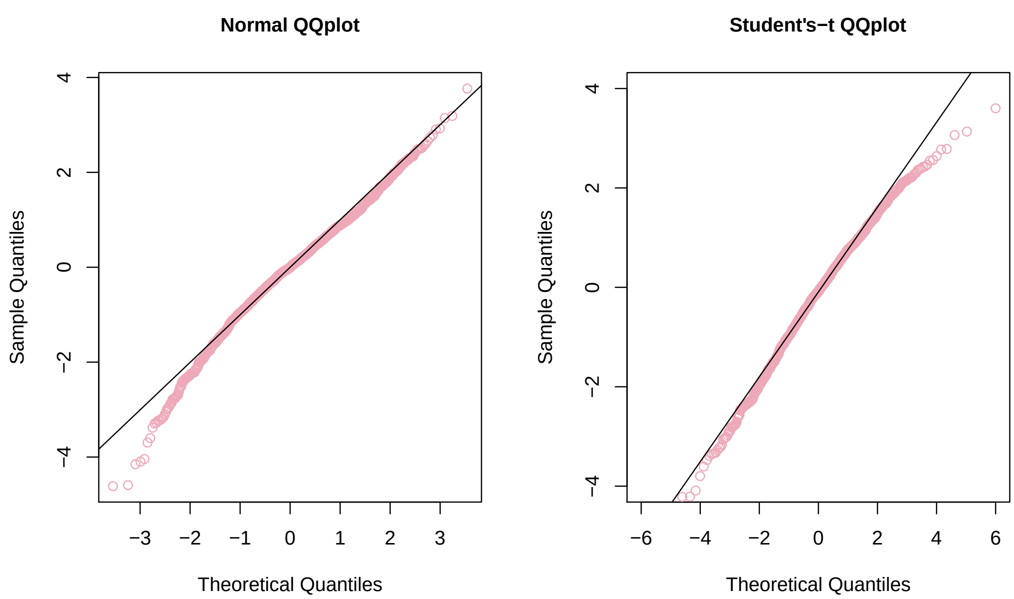


Figure 6: QQ-plots for the standardized model with Normal innovations compared to standard Normal distribution (Left), and the for the model with Student's t innovations with 7 degrees of freedom (Right).

It is seen i Figure 6 (Left) that the lower tails are much heavier than those of a standard Normal distribution. A closer look reveals the observations from the Normal quantile plot deviates from the line more than those for the Student's t distribution, yet the latter does not seem to capture the upper tails. Although none of the options are optimal since the quantiles are not shown as a straight line, we assume that the model with Student's t distribution is be a better fit than the Normal distribution.

The parameters of the two obtained ARMA(1,1)-GARCH(1,1) models are calculated with Maximum Likelihood Estimation (see Appendix D). The estimations are presented in Table 3 (Upper) for the Normal model and (Lower) for the Student's t model. In Table 3 it is seen that the estimated β_1 (beta1) parameter for both models is highly and has a higher value than other parameters. Since β_1 is the coefficient in front of σ_{t-1}, it indicates that including the past values of the conditional variance provides an accurate estimation of the current conditional variance. In other words, we interpret it as that yesterday's volatility indeed has great impact on today's volatility.

The ARCH parameter α_1 (alpha1) of the conditional squared residuals has its estimate close to zero, but the parameter is significant. Thus, the previous shock term a_{t-1}^2 does not affect the volatility as much as one might have expected, despite the clear evidence of correlation in the squared residuals of the ARMA(1,1) model. It is seen that the only parameter that is non-significant is the constant omega, which is common for financial time series.

Normal	**Estimates**	**Std.Error**	**t-value**	**p-value**
mu	0.000708	0.000152	4.6727	3e-06
ar1	0.819545	0.084528	9.6955	0
ma1	-0.866815	0.073667	-11.7667	0
omega	2e-06	1e-06	1.4449	0.148483
alpha1	0.083738	0.01513	5.5346	0
beta1	0.905592	0.016005	56.5804	0

Student's t	**Estimates**	**Std.Error**	**t-value**	**p-value**
mu	0.000866	0.000141	6.15832	0
ar1	0.80376	0.100704	7.98145	0
ma1	-0.854198	0.088632	-9.63753	0
omega	2e-06	2e-06	0.94859	0.34283
alpha1	0.083737	0.020396	4.10549	4e-05
beta1	0.910084	0.020702	43.9607	0

Table 3: Maximum likelihood estimations of in-sample parameters of the ARMA(1,1)-GARCH(1,1) model with Normal innovations (Upper) and Student's t innovations (Lower), for the OMX30 log return series.

From Table 3 we observe that the $\alpha_1 + \beta_1 < 1$ for both models, which means that both models hold the assumption of mean reversion. The $\alpha_1 + \beta_1$ are very close to adding up to one which implies that the volatility is persistent, specially for the Student's t model.

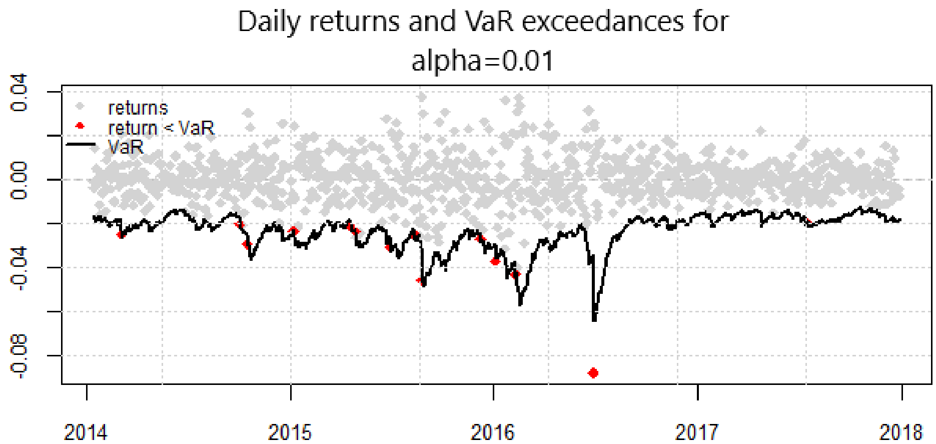
The log likelihood value is 7660 for the Student's t model and 7626 for the Normal model. Hence, it suggests that the assumption of Student's t distribution of the ARMA(1,1)-GARCH(1,1) model is the better fit to the data.

So far, the Normal model and Student's t model seem to be a proper fit for the data, but according to Angelidis et al. (2004) and Tsay (2008), good or bad results of the in-samples does not indicate good or bad out-of-sample results. Thus, before making any statements about the models forecasting abilities, we evaluate the performance of the models in estimating Value at Risk by conducting backtests.

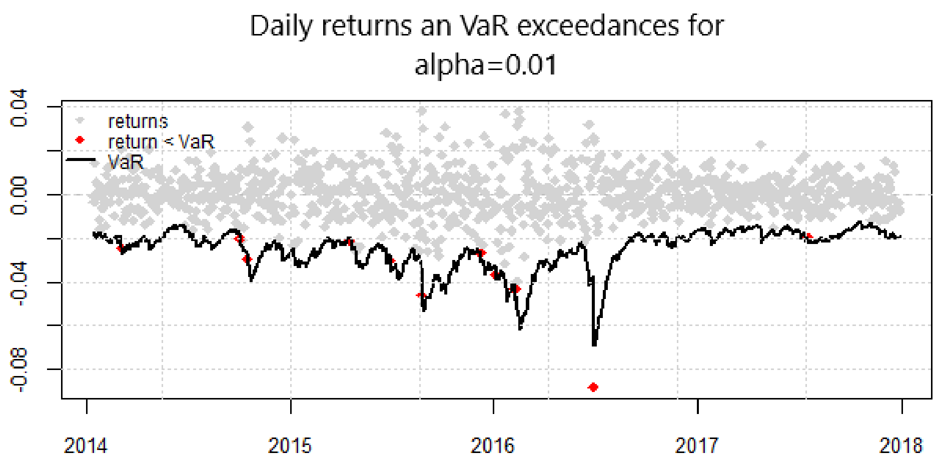
6.2 Value at Risk Backtest

We use the ARMA(1,1)-GARCH(1,1) models in a rolling window approach (see section 4.3) to forecast one thousand daily VaR estimates, with the level of confidence α set for 99%. The outcome of the backtest forecasting are illustrated in Figure 7. The actual observations of the log returns are plotted in gray, the black line represt the Value at risk level and the red marked observations are the numbers of exceedances. The exceedances are the observations where the returns are lower than the Value at Risk.

In Figure 7 it can be seen that the Normal and Student's t VaR model have similar alterations of the line and seem to follow the same trend. However, it is observed that the position of the Normal VaR model lies on a higher level than the Student's t model, which naturally generates more exceedances. The expected exceedances of the VaR models at a 99% level of confidence is 10, and the Normal VaR models accumulates in total 14 exceedances, whereas the Student's t model generated 11 ones. Thus, we have reasons to believe that the non-normal model outperforms the normal and that the Normal VaR model underestimates the VaR. Underestimating VaR, as mentioned earlier, means that the model estimates a VaR that is too low, which generates more exceedances in the backtests. We also observed that some of the exceedances in Figure 7 seem to be somewhat clustered, specifically those around the time 2016. This could imply that the models do not fully capture the conditional volatility of the data and that the exceedances therefore still might reveal some correlations.



(a) Normal VaR model



(b) Student's t VaR model

Figure 7: Backtests of the ARMA(1,1)-GARCH(1,1) models' rolling forecasts of 99% Value at Risk for the (a) Normal VaR model and (b) Student's t VaR model.

We present the summarized results of the backtest coverage of the ARMA(1,1)-GARCH(1,1) models estimations of daily VaR in Table 4. It reveals that the likelihood ratio statistics LR_{uc} and LR_{uc} of the coverage tests are below the critical value. Even though the exceedances are more than the expected the p-values also prove that the null hypothesis is not rejected on the 5% level for the tests. These results indicate that both VaR models hold the expected number of exceedances, and that the exceedances occur independently. Even though we could observe that some failures were clustered together, the Christoffersen's test of independence holds:

we cannot reject the hypothesis of independent failures.

VaR Backtest	**Normal**	**Student's t**
Backtest Length:	1000	1000
Alpha:	0.01	0.01
Expected Exceed:	10	10
Actual VaR Exceed:	14	11
Unconditional Coverage (Kupiec)		
Null-Hypothesis:	Correct Exceedances	
LR.uc Statistic:	1.437	0.098
LR.uc Critical:	6.635	6.635
LR.uc p-value:	0.231	0.754
Reject Null:	NO	NO
Conditional Coverage (Christoffersen)		
Null-Hypothesis:	Correct Exceedances	& Independence of Failures
LR.cc Statistic:	1.835	0.343
LR.cc Critical:	9.21	9.21
LR.cc p-value:	0.399	0.842
Reject Null:	NO	NO

Table 4: Backtest statistics of Normal and Student's t VaR model evaluated by Kupiec's and Christoffersen's coverage tests.

It is clear that the Student's t VaR model outperforms the Normal VaR model. However, we cannot reject the Normal VaR model according to the backtests, since it does not significantly underestimate VaR as we thought it would. Although the Normal model is not rejected, we still state that the leptokurtic Student's t model is preferred for VaR framework. The results of the backtests however indicate that both models generate accurate forecasts of VaR and are specified well enough to cope with the changing volatility in our log return series. Thus, the fitted models of the estimation sample are valid to use for predicting daily VaR in the out-of-sample period.

7 Discussion

This thesis analyzes the performance of GARCH(1,1) models in estimating a thousand daily Value at Risk predictions where the confidence level is set for 99%. The ARMA(1,1)-GARCH(1,1) model is fitted on an in-sample period of 2285 observations for the OMXS30 index. We make a distributional assumption of regarding the innovations of the models where we assume normal distribution and student's t distribution. We use the two ARMA(1,1)-GARCH(1,1) models to generate VaR predictions with a rolling window forecast over the out-of-sample period. The performances of the models are evaluated using backtesting methods, specifically, Christoffersen's conditional coverage test and Kupiec's unconditional coverage test. The tests reveal that the null hypothesis holds for both coverage tests. This implies that the exceedances in the estimates occur independently, hence are not clustered and the accumulated exceedances matches the expected exceedances given by the level of confidence $\alpha = 0.01$. According to the backtest results, the models are specified well enough to adequately capture the volatility process of the forecast period and generate accurate estimates of VaR.

There are many subjects in the construction of the backtesting procedure that may have had a great impact on the obtained results. To begin with, the chosen length of the in-sample period for the estimated models, as well as the length of the out-of-sample period for the forecasts, may have generated different results if the samples had different sizes.

Another important factor that significantly effects the outcome of the VaR estimates is the actual *type* of observed data in the in-sample period. For example, if the observed data is not so volatile then the estimated model may not be capable of generating forecast of a highly volatile forecast period, and vice versa. In this thesis, the in-sample is highly volatile because of the financial market crash of Sep 28th 2008, and the out-of-sample period is rather calm but also exhibits a relatively weaker volatility around observations $2600 - 2800$, Figure 3. The backtesting results implied well performance of the models' estimations of VaR. However, considering the highly volatile data used for fitting the models, and the relatively calm volatility in the forecast period, we interpret the backtest results with caution. As a complement to overcome this problem one might use the estimated models for "stress testing", proposed by Christoffersen (2003). Stress testing the model means that we expose the estimate model to extreme volatile data, to test the model's performance under extreme market conditions.

Type I error means rejecting a true null hypothesis, whereas type II means failing to reject a false null hypothesis. According to Jorion (2007), the 1% Value at Risk level might sometimes cause low power to the backtest since it does not generate enough exceedances, which initially could induce type I and II errors in the coverage tests for borderline values. This may be the explanation for why the Normal VaR model was not rejected in the coverage tests of the backtesting. This could of course be judged by comparing different levels of VaR.

It is worth mentioning that Christoffersen's conditional coverage test only considers dependencies between two following exceedances, that is, today given yesterday. Put in other words, the test only measures independencies in one step, but fails to recognize longer period clustering, for instance, today given a week ago. Thus, a complement to the applied coverage tests in this thesis would be the conditional duration test suggested by Christoffersen and Pelletier (2003), that do take longer step dependencies into account.

As mentioned in section 3.2.2 the GARCH model, as well as the ARCH model, fails to distinguish between negative or positive shocks since the model includes the squared values of past innovations and variance. Perhaps a better fit would have been obtained if we extended the GARCH model to an EGARCH model by including a third parameter that allows the volatility to react differently to positive and negative shocks.

It is worth noting that the results from the conditional volatility modeling are in line with Tsay (2002, 2010). The results of the backtesting seems promising. Although the GARCH model with the Student's t distributed innovations outperforms the Normal model, we cannot state that the Normal model is not specified well enough to be used for the estimation of VaR. However, we can suggest that according to our results, it is more suitable to use a non-normal leptokurtic distribution for the innovations that captures the high peak of extreme data points that asset returns have.

Moreover, the distributional assumption appears to be a matter of trial and error. In the analysis of the quantile plots (Figure 6) it was seen that the empirical quantiles did not match neither of the assumed distributions for the innovations. In Angelidis et al. (2004), who examined the Normal, Student's t and GED distributions concluded that no model was clearly superior but that the leptokurtic Student's t model generated the most adequate VaR predictions. In the backtesting in this thesis, the Student's t VaR model performed better than the Normal VaR model, yet we are unable to reject the Normal VaR model. Hence, our results are in line with Angelidis et al. (2004).

Having this said the conclusions we draw based on our results are: we cannot state that the distributional assumptions have a significant effects to the predictive accuracy of the VaR models predictions, and the backtesting method is rather fragile in the sense that it is very dependent of the type of data in the in-sample and out-of-sample, as well as the sizes of the sub-samples.

8 Conclusion

The backtests of the models suggests that both models are able to properly imitate and predict the volatility process of the OMXS30 index log return series. Despite the results seen that the Student's t VaR model outperforms the Normal VaR model, however there are not enough evidence to prove that normal distributed innovations are not legitimate enough to use in VaR framework. Therefore, we cannot state that that there is a significant difference in the outcomes of the VaR estimate based on the distributional assumptions. Nonetheless, we conclude that Student's t distribution is preferred when estimating daily VaR, since the high-peak allows for more extreme values that log return series has, as a result of conditional volatility.

We find the method of backtesting rather fragile since the generated forecasts are considerably affected by the volatility levels in the in-sample observations, the estimation and forecast sample sizes, and the chosen evaluation tests. To overcome these problems we suggest: stress testing the estimation sample, testing the VaR estimates on several confidence levels, and using more evaluation tests in the backtesting to give a proper perspective of the predictive accuracy of the models. If researches are aware of all flaws associated with the evaluation of VaR estimates, then the VaR can be very useful in risk managements.

9 References

Angelidis T., Benos A., Degiannakis S. (2004), 'The use of GARCH models in VaR estimation', Statistical Methodology 1, 105–128.

Berry W.D., Feldman S. (1985), 'Multiple Regression in Practice', Sage University, Paper series on Quantitative Applications in the Social Sciences.

Bollerslev T. (1986), 'Generalized autoregressive conditional heteroskedasticity.', Journal of Econometrics, 307–327.

Bollerslev T. (2008), 'Glossary to ARCH (GARCH)', CREATES Research Paper 2008-49, School of Economics and Management University of Aarhus.

Christoffersen P.F., Pelletier D. (2003) 'Backtesting value-at-risk: a duration based approach', Journal of Financial Econometrics, 84-108.

Christoffersen P. F. (2003), 'Elements of financial risk management', Academic Press.

Cowpertwait S.P , Metcalfe V.A (2009) , 'Introductory Time Series with R', Springer.

Cryer J.D., Chan K.S (2008), 'Time Series Analysis: With applications in R', Second edition, Springer.

Dowd K. (1998), 'Beyond Value at Risk: The new Science of Risk Management', John Wiley.

Engle R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom ination', Econometrica, 987-1007.

Engle R. (2001) 'GARCH 101: The use of ARCH/GARCH models in applied econometrics. Journal of Economic Perspectives, 157-168.

Engle R.F, Patton A.J (2001) 'What good is a volatility model?', Quantitative Finance, 234-245.

Esch L., Kieffer R, Lopez T (2005), 'Asset and Risk Management: Risk Oriented Finance', John Wiley.

Hayashi F. (2000), 'Econometrics', Princeton University Press.

Jorion P. (1997), 'Value at Risk: The New Benchmark for Controlling Market Risk', The McGraw-Hill Company.

Jorion P. (2007), 'Value at Risk: the new benchmark for managing financial risk' , Third edition, The McGraw-Hill Company.

Mandelbrot B. 1963, 'The Variation of Certain Speculative Prices', The Journal of Business, 394-419.

McNeil A.J., Frey, R., Embrechts P (2005), 'Quantitative Risk Management: Concepts, Techniques and Tools', Princeton University Press.

Orhan M., Köksal B. (2012), 'A comparison of GARCH models for VaR estimation.', Expert Systems with Applications, 3582–3592.

Patton A.J. (2011), 'Volatility forecast comparison using imperfect volatility proxies.', Journal of Econometrics, 246-256.

Paul Wilmott (2007), 'Paul Wilmott on Quantitative Finance', Second edition, John Wiley & Sons, Chapter 9 and 22.

Poon S., Granger C.W.J. (2003), "Forecasting volatility in financial markets: a review", Journal of Economic Literature, 478-539.

Robert F. Engle and Tim Bollerslev (1986), 'Modeling the persistence of conditional variances', Econometric Reviews.

Tsay R. (2002), 'Analysis of financial time series'. John Wiley & Sons.

Tsay R. (2010), 'Analysis of financial time series', Third edition, John Wiley & Sons.

Tsay R. (2011) 'Conditional Heteroscedastic Models', Lecture Note of Bus 41202 Univariate Volatility Models.

Tsay R. (2008) 'Out-of-Sample Forecasts', Lecture Note of Bus 41402 Univariate Volatility Models.

Appendix

A Autocorrelation function (ACF)

The correlation coefficient measures the strength of linear dependence between two random variables X and Y defined by,

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2 E(Y-\mu_y)^2}} \tag{20}$$

where μ_x and μ_y is the mean of X and Y, and the variances of the variables are assumed to exist where the variables are uncorrelated if $\rho_{x,y} = 0$ (Tsay 2010 pp.30). In time series studies the theory of correlation is generalized to autocorrelation. The autocorrelation function is a measure of the correlation between observations in a time series r_t that are separated by l time units r_t and r_{t-l}. The lag operator l shifts the series r_t such that the lagged values r_{t-l} are aligned with the return series r_t itself. Just as the correlation coefficient describes the similarity between two random variables X and Y, the autocorrelation function determines the serial correlation between the time series and its own lagged values. When observations in a time series are serially correlated, it means that the series are linearly dependent and that future values are affected by past values. We define the autocorrelation function (henceforth ACF) by

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{Cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}. \tag{21}$$

Note that this equality holds because the series are assumed to be weakly stationary. For a white noise process the autocorrelation is zero at all lags except for lag zero where the autocorrelation is one.

B Ljung-Box

The Ljung-Box test is a tool much used in time series analysis to jointly test autocorrelations of a log return time series r_t. The null hypothesis and the alternative hypothesis is formulated by

$H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$
$H_a : \rho_i \neq 0, \text{for some } i \in (1, ..., m).$

The Ljung-Box test statistic is defined as

$Q(m) = N(N+2)\sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{N-l}$

where N is the size of the series, ρ_l is the autocorrelation at lag l and m is the number of lags which we are testing in our model. The null hypothesis is rejected if $Q(m) > \chi_\alpha^2$ where χ_α^2 denotes the (1-α)th quantile of a chi-squared distribution with m degrees of freedom (Tsay 2010, pp.32-33).

C Skewness and Kurtosis

Consider a random variable X. The skewness $S(x)$ and kurtosis $K(x)$ of X are given by the normalized third and fourth central moments defines as (Tsay 2010, pp.9),

$S(x) = E\left[\frac{X-\mu_x^3}{\sigma_x^3}\right], \qquad K(x) = E\left[\frac{X-\mu_x^4}{\sigma_x^4}\right]$

The skewness measures the symmetry of X with respect to its mean which tells the amount of departure from horizontal symmetry. The kurtosis measures the tail thickness and describes how tall and sharp the peak is. For a Normally distributed sample x_i ($i = 1, \ldots, N$), the skewness is 0 and the kurtosis is 3. The quantity $K(x) - 3$ is called the *excess kurtosis*. If the kurtosis is larger than three the set is leptokurtic, for which it is better to assume a Student's t distribution.

D Maximum Likelihood Estimation (MLE)

In order to estimate the parameters of the obtained models we use Maximum Likelihood Estimation (MLE). In this time series analysis the observations of the log return series $\{r_t\}_{t=1}^{T}$ are not independent. By considering the observations $\{r_1, r_2 \ldots, r_T\}$ fixed, the joint probability density function can be expressed by conditioning:

$f(r_t, r_{t-1}, \ldots, r_1; \theta) = f(r_t|r_{t-1}, \ldots, r_1; \theta) \times \ldots \times f(r_2|r_1; \theta) \times f(r_1; \theta).$

Here θ detones vector that consists of the unknown parameters vector which in our case is $\theta = (\alpha_0, \alpha_1, \beta_1)$. The general likelihood function $L(\theta; r_t)$ can then be described by:

$L(\theta; r_t) = \prod_{t=1}^{T} f(r_t|r_{t-1}, \ldots, r_1; \theta).$

Despite the formal definition of the likelihood function it is more common to use the logarithmic likelihood function in practice, also known as the log-likelihood function:

$l(\theta; r_t) = \log\left(\prod_{t=1}^{T} f(r_t|r_{t-1}, \ldots, r_1; \theta)\right) = \sum_{t=1}^{T} \log f(r_t|r_{t-1}, \ldots, r_1; \theta).$

Since we are interested in the innovations series a_t rather than the actual log return series r_t we recapitulate that $r_t = \mu_t + a_t = \mu_t + \sigma_t \epsilon_t$ since $a_t = \sigma_t \epsilon_t$ (Equation 3 and Equation 13). Under the distributional assumption of standard normal ϵ_t in the conditional log-likelihood function for a GARCH(p,q) model with normal innovations, $a_t = \sigma_t \epsilon_t \sim N(0, \sigma_t^2)$ can be approached. Since then, given the history up to time $t-1$, $\epsilon_t \sim N(0,1)$ we can express the density of ϵ_t by

$$f(\epsilon_t | \epsilon_{t-1}, \ldots, \epsilon_0) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\epsilon_t^2}{2}\right).$$

and the density of a_t by

$$f(a_t | a_{t-1}, \ldots, a_0) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right).$$

For the GARCH(p,q) model where the paramater vector θ is $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q)$, we calculate the conditional log-likelihood function by

$$l(\theta; a_{t-1}, \ldots, a_0) = \sum_{t=q+1}^{T} \log f(a_t | a_{t-1}, \ldots, a_0)$$
$$= \sum_{t=q+1}^{T} \log \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right)$$
$$= \sum_{t=q+1}^{T} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{a_t^2}{2\sigma_t^2}\right]$$

The MLE are obtained by deriving the conditional log-likelihood function with respect to the parameters in θ and set the derivation functions equal to zero. Thereby maximizing the log-likelihood functions to obtain the estimates of $\hat{\theta}$. For more details and information of obtaining MLE of GARCH(p,q) models under the assumption of Student's t distrion we refer to Tsay (2010, pp.120-121).

E Additional figures

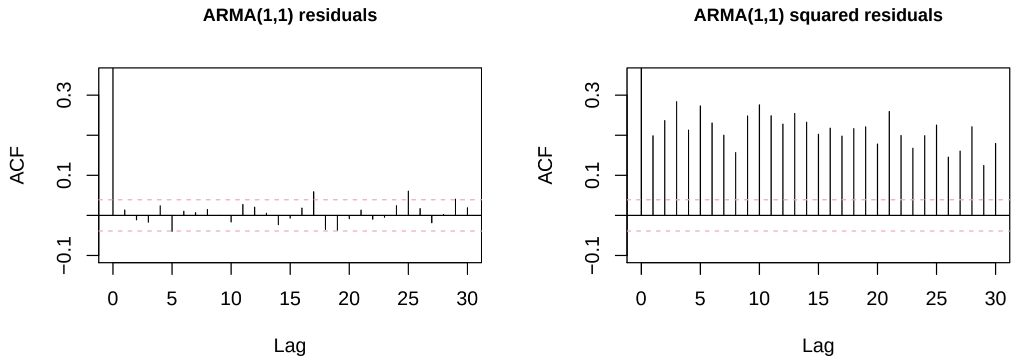


Figure 8: ACF plot of standardized residuals (Left) and the squared standardized residuals (Right) of the ARMA(1,1)-model. The pink dashed lines denote a 95% confidence interval.

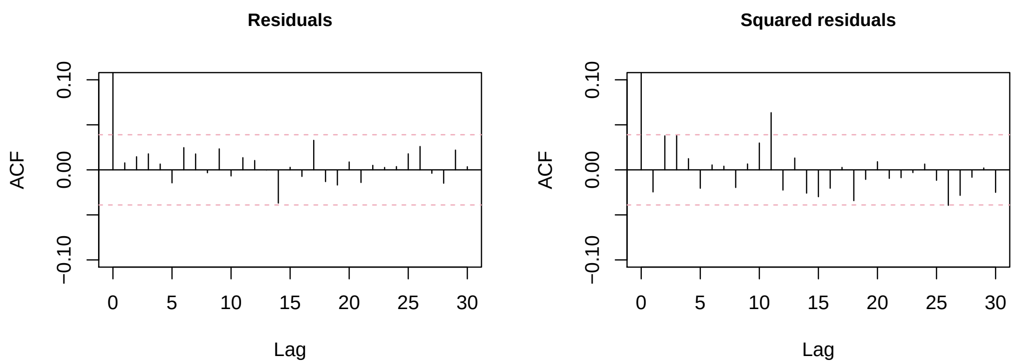


Figure 9: The ACF plot of the standardized residuals (Left) and the squared standardized residuals (Right) for the ARMA(1,1)-GARCH(1,1) model with Student's t conditional distributed innovations. The pink dashed lines denote a 95% confidence interval.

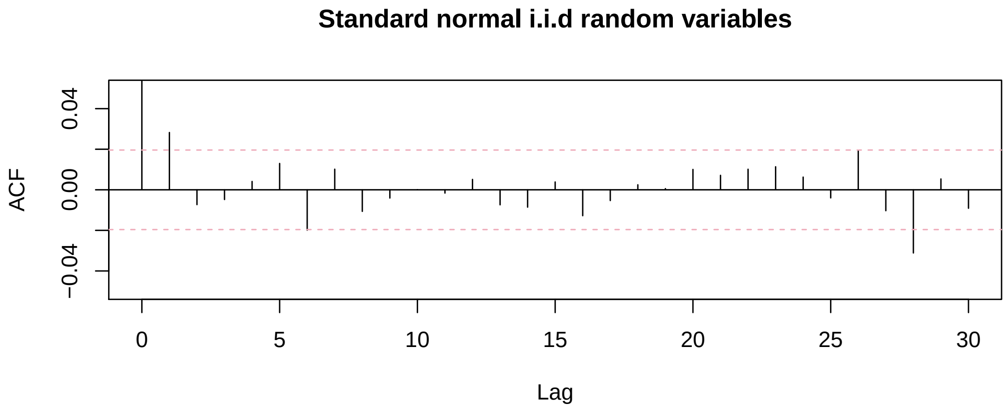


Figure 10: Illustrating the outcome of significant lags in a simulated white noise series of 10'000 standard normal variables. The pink dashed lines denote a 95% confidence interval.