Stockholms universitet

Evaluation of Value at Risk estimates using Extreme Value Theory

Andreas Rosén Björkhage

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2017:15
Matematisk statistik
Juni 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2017:15**
http://www.math.su.se

Evaluation of Value at Risk estimates using Extreme Value Theory

Andreas Rosén Björkhage*

June 2017

Abstract

Value at Risk (VaR) is a risk measure that quantifies the maximal loss we may incur under normal market conditions, given a confidence level and a fixed time horizon. Due to this intuitive interpretation and its applicability VaR has become the most widely used risk measure today. In this thesis we will compare one day ahead VaR forecasts from an AR(1)-GARCH(1,1) time series model with either normal or t distributed innovations with the corresponding models where the Peak-Over-Threshold (POT) method has been used in order to model the tails of the innovations. For further comparisons we also include an unconditional model where the VaR estimates are the quantile estimates based on the General Pareto Distribution. By using several backtest procedures on historical daily log-returns for five stock indices we find that the models made using the POT method outperforms the other included models at the higher confidence levels.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: anro9094@student.su.se. Supervisor: Filip Lindskog & Mathias Lindholm.

Contents

Preface

This is a Bachelor thesis to an extent of 15 ECTS. Its completion will result in a Bachelor's Degree in Mathematical Statistics at the Department of Mathematics at Stockholm University.

Acknowledgements

I sincerely would like to thank my supervisors Filip Lindskog and especially Mathias Lindholm for their guidance, encouraging help and suggestions throughout the work of this thesis. I would also like to express my gratitude to Benjamin Allévius for his motivation and suggestions in the model making. Last but definitely not least I also wish to thank my friends and family for your unceasing support, without you this thesis would not have been completed.

1 Introduction

1.1 Background

The globalization of financial markets has naturally induced an increased volatility in the prices of the traded assets. All participants in the financial and commodity markets share the necessity to accept the risk of losing parts of or all of their investment, and are now more than ever dependent upon global speculation. This development emphasizes the need for risk management in order to control for e.g. market risk (the risk that a market participant incurs losses due to unfavorable market movements) where financial crises such as the subprime mortgage crisis in 2007 serves as a school book example of the potential consequences of deficient risk management.

Value at Risk (VaR) is the most widely used market risk measure today which in a simple way may be described as the worst potential loss over a given time period that will not be exceeded with a given level of confidence. This intuitive interpretation along with the fact that it may be used in order to encompass other sources of risks such as commodities, foreign currencies, and equities (Jorison 2007, pp. 16) has paved the way for its popularity.

From a mathematical point of view VaR is a quantile of the distribution of the returns series of a portfolio or any other asset over a prescribed time period and in this thesis we will be concerned with tail estimation for several negative log return series from a number of selected stock shares retrieved from Nasdaq OMX.

When estimating VaR you may either choose a non-parametric or parametric approach where historical simulation and the implementation of different generalized autoregressive conditional heteroscedastic (GARCH) models are the most frequent methods in each approach. Historical simulation is basically the empirical distribution of past gains and losses of the asset and the major advantage of this approach is that it is easily implemented and require few assumptions. The major drawback lies in the fact that extrapolation beyond past observations is impossible, meaning that historical simulation won't be able to forecast events or volatile periods which have not been previously observed (McNeil et al. 2005, pp. 50-51).

Previous parametric approaches using GARCH modeling for estimating VaR has been proven to successfully capture the volatility clustering observed in financial markets, however they often fail to capture the heavy-tailedness of the asset returns by assuming that they are normally distributed, thereby underestimating the occurrence of extreme outcomes or events.

When estimating VaR there is no optimal model or general feasible approach available and in this thesis we will compare some of the previous used models (where the residuals or innovations are either standard normal or t distributed) with a relatively new approach based on Extreme Value Theory (EVT).

1.2 Aim and purpose of the thesis.

In this thesis we aim to model conditional volatility for the negative log return series for a number of selected stock shares retrieved from Nasdaq OMX inspired by the approach first introduced by McNeil & Frey in their paper from 2000. In short we will use GARCH models in which the parameters are obtained through maximum likelihood estimation. The innovations or the distribution of the residuals will be modeled using Extreme Value Theory and more specifically by applying the Peaks-Over-Threshold method (POT method). From this distribution we will be able to obtain a quantile of chosen significance level which along with the estimates of the conditional mean and volatility allows us to calculate the one day ahead VaR forecasts. The evaluation methods used are the back testing procedures described in the oncoming section 5.5.

1.3 Previous research

In this subsection we intend to present some previous research on how to calculate Value at Risk by the use of Extreme Value Theory. However this is a relatively difficult task because of the large amount of previous studies done on the subject. Therefore we chose to present a selection of papers we find particularly interesting and at the same time demonstrating the broad scope of applicability of EVT. The readers who wish to learn more about time series analysis and EVT are referred to Tsay 2010, McNeil et al. 2005, and Embrechts et al 1997 which we refer to several times in the sections covering EVT and the time series models used in this thesis.

Since McNeil & Frey published their paper in 2000 promoting a conditional approach when calculating VaR many other papers followed (e.g. Gençay et al. 2003 and Kuester et al. 2005), however the use of EVT has been found useful in several studies promoting an unconditional approach when calculating VaR. In fact Danielsson & de Vires 2000 propose the use of a semi-parametric method for estimation of tail probabilities after showing that conditional parametric methods, such as GARCH with normal innovations under predict the VaR for stock returns at the 1% risk level (or below). They do not favor an unconditional approach above a conditional one but argues that the choice of methodology should depend on the situation and question at hand (Danielson & de Vires 2000, pp. 241).

Bali argue that methods based on parametric distributions assume that interest change is normally distributed, however there is now strong evidence that they are not (Bali 2003, pp. 83). In his paper he uses Extreme Value Theory and a stochastic differential equation in order to calculate VaR for U.S. Treasury yields. Using this unconditional approach he finds that the statistical theory of extremes provides a more accurate approach to risk management and VaR calculations in relation to its unconditional normal counterpart (Bali 2003, pp. 106).

2 Theoretical framework

In the following section we will present the theoretical tools and concepts needed in order to understand the analysis related to the thesis.

2.1 Log returns

We start by defining simple return before moving on to the log returns.
Let P_t be the daily closing price of an asset at time point t. Then holding the asset for one time period from say $t-1$ to t would result in a simple gross return R_t we choose to express as

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

This is directly related to the continuously compounded return or log return r_t which is the natural logarithm of the simple gross return.

$$r_t = \log(1 + R_t) = \log\Big(\frac{P_t}{P_{t-1}}\Big) = \log\big(P_t\big) - \log\big(P_{t-1}\big) \tag{1}$$

This compounding method makes it easy to calculate the multi period return by simply summing over every one-period log return involved (Tsay 2010, pp. 5).

Usually our main concern when working with Value at Risk is the expected loss of our investment, therefore we will exclusively use the *loss series* when carrying out the numerical analysis. Note that the losses are simply the *negative* log returns. This is done because of the convenience that follows from estimating positive VaR limits.

Further it is important to keep in mind that Value at Risk does not contain any information regarding the size of the potential loss $-r_t$. This will be further addressed in the upcoming section where we formally define Value at Risk and account for a number of shortcomings.

2.2 Value at Risk

Given a fixed time horizon and confidence level, the Value at Risk (VaR) is an estimate of the maximal loss we may incur under normal market conditions. In order to exemplify, VaR can help us to calculate what the worst expected loss would be 99 days out of 100. This is equivalent that VaR gives us an estimate of the minimal loss associated with extraordinary market conditions. However both definitions imply the same method in order to quantify financial risk (Tsay 2010, pp. 325-326).

Mathematically VaR can be defined through the distribution of the asset return in the following way:

Definition (Value at Risk): Given a confidence level $\alpha \in (0,1)$ and a time horizon of k days, the k-day Value at Risk, expressed as a percentage of the value of the asset at time t, is given the smallest number x_k such that the probability that the loss X_t exceeds x_k is no larger than $(1-\alpha)$. Formally,

$$VaR_{\alpha,k} = \inf\Big\{x_\alpha \in \mathbb{R} : P\Big(X_{t;k} > x_\alpha\Big) \leq 1-\alpha\Big\} = \inf\Big\{x_\alpha \in \mathbb{R} : F_{x;k}(x_\alpha) \geq \alpha\Big\}. \tag{2}$$

Where $F_{X;k}$ is the distribution function of the losses over a k-day period (McNeil et al. 2005, pp. 38).

It is common to express VaR as a numerical value rather than a percentage which is easily calculated by multiplying the percentage VaR presented in definition (2) by the value of the portfolio or asset.

2.2.1 Alternatives to VaR

There are no substantial reasons to doubt that VaR is the most popular risk measure in use, but that does not mean that it alone guarantees a complete analysis that answers all the interesting lines of inquiry. In fact the usefulness of VaR has been questioned and complementing risk measures accounting for some shortcomings of VaR has been proposed (see e.g. Acerbi & Tasche 2002).

The popularity of VaR is usually explained by its ability to capture an essential aspect of risk, namely how bad things can get with a certain probability. The fact that it is probabilistic separates VaR from previous risk measures such as Risk-Adjusted Return On Capital (RAROC) or Sharpe's single factor beta model (Jorison 2007, pp. 16) which makes it easily communicated and somewhat intuitively understood.

The deficiencies commonly referred to when discussing VaR is its lack of accounting for parameter uncertainty. VaR is a forecast concerning possible loss of an asset given a certain confidence level and time horizon. Therefore it should be computed using the predictive distribution of future returns of the financial position. From a statistical viewpoint the predictive distribution takes into account the parameter uncertainty in a properly specified model. However, the predictive distribution is usually difficult to obtain, and most of the available methods for VaR calculation ignore the effects of parameter uncertainty (Tsay 2010 pp. 328).

From the definition presented earlier VaR can be viewed as a quantile of the negative log returns. It is important to note that it does not fully describe the upper tail behavior of the distribution for the negative log returns. This means that even if two different assets have the same VaR, they may still experience different losses when the VaR is exceeded. Put in other words, VaR accounts only for *if* an exceedance takes place, but not the magnitude of the exceedance.

In addition to what has been mentioned above, the lack of sub-additivity of VaR is recurrent in contemporary literature on the subject. If a risk measure is sub-additive the merging of two assets should not be greater than the sum of their individual risk measures before they were merged, which calls for extra care when using VaR in order to quantify risk for several assets (Tsay 2010, pp. 328)

One risk measure that accounts for the characteristics of the tails and often works as a complement to VaR is the Expected shortfall which can be described as the expected value of the loss *given* that the VaR quantile is exceeded. However we have due to limitations when writing this thesis chosen to only consider VaR in the rest of the analysis. For the interested reader, we refer to McNeil & Frey 2000 who complement their analysis of the VaR estimates by also forecasting expected shortfall.

3 Extreme Value Theory

Extreme value theory (EVT) has been proved to be a useful analytical tool in a wide spectrum of different fields. It can be used in order to analyze e.g. natural disasters, life spans and loss returns for a financial time series (de Haan & Ferreira 2006, pp. 7). All mentioned examples share the feature of in some way analyzing or forecasting the probability for rare and extreme events. In other words EVT concerns the tail behavior of random variables, which is useful when estimating VaR. Up until today there are two available approaches to the analysis of extremes that is offered by EVT.

The first is based on the Fisher-Tippett-Gnedenko theorem that mainly concerns the limit distribution of centered and normalized sample maxima. In short the theorem states that the maximum of a sample of i.i.d. random variables after proper renormalization can only converge in distribution to one of three possible distributions, the Gumbel distribution, the Fréchet distribution, or the Weibull distribution (Embrechts et al 1997, pp. 121).

The second approach is focusing on the excesses of random variables over a certain threshold, and this approach will be the one used in this thesis. Therefore we start by introducing the Generalized Pareto Distrubution (GPD), which is used to model this phenomena. In the following we assume that the distribution of the excesses X over some threshold u follows a GPD, and where $y = x - u$.
The cumulative distribution of the GPD is given by:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - (1 - \xi y/\beta) & \text{if } \xi \neq 0 \\ 1 - \exp(-y/\beta) & \text{if } \xi = 0 \end{cases}$$

Where the scale parameter $\beta > 0$, the support is $y \geq 0$ and $0 \leq y \leq \beta/\xi$ when $\xi < 0$.
We also introduce the excess distribution function:

$$F_u(y) = P\big(X - u \leq y | X > u\big) = \frac{F(y+u) - F(u)}{1 - F(u)} \tag{3}$$

According to the Pickands-Balkema-de Haan theorem y is well approximated by the GPD (Embrechts et al 1997, pp. 354), and

$$F_u(y) \approx G_{\xi,\beta}(y) \quad \text{as } u \to \infty$$

for a large class of distribution functions, including the Pareto distribution and log-normal distribution. This result directly relates to our implementation when calculating VaR by using the Peak over threshold model.

3.1 Implementation of EVT

We now move on to what is commonly referred to as the Peak-Over-Threshold method (POT method). Remember that in this framework $X_1, X_2, ..., X_n$ are assumed to be independent and identically distributed random variables. Again we chose a threshold u and define every exceedance Y_i as $Y_i = X_i - u$ for $i \in (1, 2, ..., N_u)$ where N_u represents the total number of exceedances.

Here we are particularly interested in the excess distribution function of X that was introduced in the previous section in (3). If we are to consider the complementary event

$$\overline{F}_u(y) = P\big(X - u > y | X > u\big) = \frac{\overline{F}(y+u)}{\overline{F}(u)} \Leftrightarrow \overline{F}(y+u) = \overline{F}_u(y)\overline{F}(u).$$

Then according to Embrechts et al. 1997 (pp. 354) a natural estimator for $\overline{F}(u)$ is given by the empirical distribution function

$$\hat{\overline{F}}(u) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i > u\}} = \frac{N_u}{n}.$$

The Pickands-Balkema-de Haan theorem motivates an estimator of $\widehat{F_u(y)}$ on the form

$$\widehat{\overline{F}}_u(y) = \overline{G}_{\hat{\xi},\hat{\beta}}(y)$$

for an appropriate $\hat{\xi} = \hat{\xi}_{N_u}$ and $\hat{\beta} = \hat{\beta}_{N_u}$. What is meant by appropriate will be further discussed in section 5.4.

By merging these results, we may obtain an estimator of the tail $\overline{F}(y+u)$ for $y > 0$ on the form

$$\widehat{\overline{F}}(y+u) = \frac{N_u}{n}\big(1 + \hat{\xi}\frac{y}{\hat{\beta}}\big)^{-\frac{1}{\hat{\xi}}}.$$

However, an estimator for the quantile x_q is obtained by inverting the formula. This inversion is of particular interest since it provides us a way of calculating the Value at Risk for a chosen quantile q denoted VaR_q

$$\mathrm{VaR}_q = x_q = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-q)\right)^{-\hat{\xi}} - 1\right).$$

In the implementation below, we will follow McNeil & Frey and apply the POT model to the standardized residuals of an AR(1)-GARCH(1,1) model, that will be applied and presented in the oncoming section.

4 Time Series Models

In this section, we present all of the time series models used in the analysis with a brief explanation in how they provide forecast estimates.

4.1 Autoregressive models

A model describing the linear dependence of X_t on previous observations is the autoregressive model of order p, or the AR(p)-model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t \tag{4}$$

Where ϵ_t is a white noise series with mean zero and variance σ_ϵ^2. Note further that in (4) $p \in \mathbb{N}$, demonstrating that the previous p variables jointly determine the conditional expectation of X_t given previous data (Tsay 2010, pp. 38).

The AR(p)-model makes it possible to calculate the ℓ-step ahead forecast based on the minimum squared error loss function (for details see Tsay 2010 pp. 54), which can be expressed as:

$$\hat{X}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{X}_h(\ell - i).$$

Where h is the current time index and ℓ the forecast horizon.

In this thesis we will use an AR(1)-model in order to model the dynamics and calculate the one day ahead forecast of the conditional mean. Knowing the general form of an AR(p) model the AR(1) model is simply expressed as:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t. \tag{5}$$

Where X_t represents the negative log return at time point t.

4.2 Conditional Heteroscedastic Models

4.2.1 Autoregressive conditional heteroscedastic model

The main idea behind the autoregressive conditional heteroscedastic model (ARCH) first introduced by Engel 1982 is that the shocks ϵ_t of an asset return are indeed dependent but serially uncorrelated and their dependence can be described by a function of its lagged values (Tsay 2010, pp. 116-117).

In order to be more accurate we may define an ARCH(p) process in the following way:
If we let Z_t be a white noise process with zero mean and unit variance then the mean adjusted process $X_t - \mu = \epsilon_t$, is an ARCH(p) if it is strictly stationary and satisfies for all $t \in \mathbb{Z}$ and for strictly positive valued process σ_t the equations:

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \end{aligned}$$

and $\omega > 0$, $\alpha_i \geq 0$ for $i \in (1, 2, ..., p)$ (McNeil et al. 2005, pp. 139).
The general ARCH model above illustrates that if any or more of $|\epsilon_{t-1}|, |\epsilon_{t-2}|, ..., |\epsilon_{t-p}|$ are particularly large, then ϵ_t will be drawn from a distribution represented by large variance, and may itself be large. This is how the ARCH model imitates volatility clustering (McNeil et al. 2005, pp. 139).

4.2.2 Generalized autoregressive conditional heteroscedastic model

Although the ARCH model is simple and widely implemented, it often requires many estimated parameters to adequately describe the volatility process of an asset return (Tsay 2010, pp. 131). In order to account for this Bollerslev (1986) introduced an alternative model called the Generalized ARCH model (GARCH). The GARCH model is a generalized form of ARCH in the sense that the variance σ^2 depends on previous values of the variance, as well as the previous values of the process.

If we again define Z_t as a white noise series with zero mean and unit variance, and ϵ_t as the mean adjusted series $\epsilon_t = X_t - \mu$.
Then the GARCH(p, q) model may be defined as:

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 . \end{aligned}$$

Where $\omega > 0$, $\alpha_i \geq 0$ for $i \in (1, 2, ..., p)$ and $\beta_j \geq 0$, $j \in (1, 2, ..., p)$.

In this thesis we will limit our focus to the simplest of all GARCH models, namely the GARCH(1,1) model, which conditional variance can be expressed by:

$$X_t = \mu_t + \epsilon_t \tag{6}$$
$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{7}$$

Where $\alpha_1 + \beta_1 < 1$ has to be fulfilled in order for ϵ_t to be strictly stationary.

Later on we will use the GARCH(1,1) model in order to calculate the one day ahead volatility forecast through the following:

$$\sigma_{h+1}^2 = \omega + \alpha_1 \epsilon_h^2 + \beta_1 \sigma_h^2.$$

Where h represents the forecast origin and where we assume that ϵ_h and σ_h^2 are assumed to be known at time h. For further details see Tsay (Tsay 2010, pp. 133).

5 Methodology

We start this section by presenting the software used to obtain the results and proceed with a brief overview of our included data sets. In the remaining we will carry out a fundamental data analysis in order to motivate the choice of time series models and the implementation of Extreme Value Theory. Furthermore we will introduce all the methods used when carrying out the analysis as well as briefly explain how they are implemented.

5.1 Software

The code used in the graphical and numerical analysis was written in the open source program **R**, and can easily be downloaded through the following link: https://www.r-project.org/.

The packages included in the making of this thesis are listed in table 1, along with their authors, recent publication year and in what purpose they were used.

Package	Authors	Year	Usage
evir	Pfaff, McNeil & Stephenson.	2012	Implementation of EVT.
fExtremes	Wuertz.	2013	Implementation of EVT.
fGarch	Wuertz & Chalabi.	2008	Fitting time series models.
ggplot2	Wickham & Chan.	2016	Graphical illustration of data.
gridExtra	Auguie & Antonov.	2016	Graphical illustration of data.
timeSeries	Wuertz & Chalabi.	2010	Process data.

Table 1: List of included R packages.

5.2 Initial data analysis

All of the included data sets can be accessed online via the link leading to the homepage of Nasdaq OMX Nordic http://www.nasdaqomxnordic.com/ and are listed in table 2 below.

Since we have chosen to include several stocks in our analysis we limit the initial data analysis to thoroughly present the data analysis to the reader in the case of the stock for Danske Bank, and we begin by plotting the negative log return series for the stock, which is calculated as presented in (1) and multiplied by (-1).

Stock	Acronym	Sector	Nr. of obs.	From	To
Astra Zeneca	AZN	Health Care	4525	1999-04-06	2017-03-31
Autoliv SDB	ALIV SDB	Automobiles & Parts	5007	1997-05-26	2017-04-26
Danske Bank	DANSKE	Financials	4125	2000-10-16	2017-04-05
Hennes & Mauritz B	HM B	Consumer Services	6762	1990-01-02	2017-03-29
Scandinavian airlines	SAS	Airports & Air Services	3963	2001-07-06	2017-04-07

Table 2: List of included data sets.

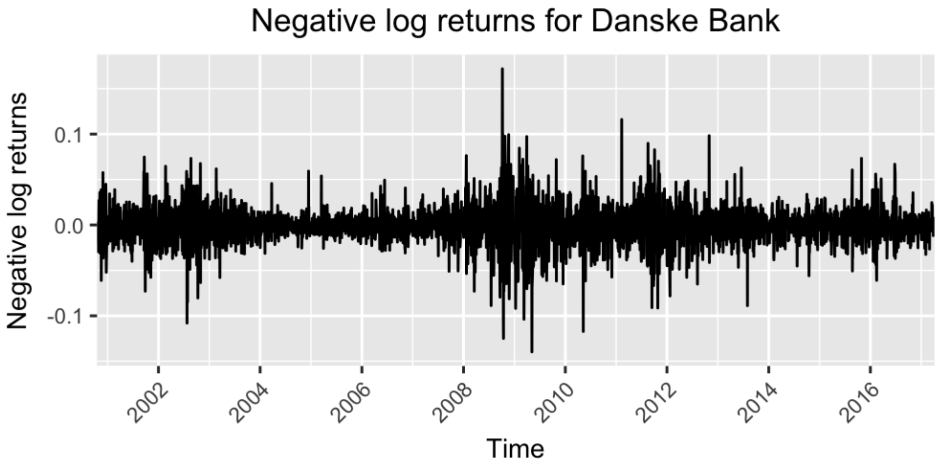

Figure 1: Negative log returns for Danske Bank

As seen in figure 1 the presence of volatility clustering is clear where we can observe several periods characterized by increased volatility, where the most volatile period seems to be located between the years of 2008-2010. This may not be surprising if we remember that this is the period following from the global financial crisis starting in 2007 in the subprime mortgage market in the USA.

Before we consider to fit a time series model to our data, it may be appropriate to inspect the data sets to guarantee that they fulfill the requirements that our models rely on them to do. Therefore, we must first ensure the presence of autocorrelation and fat-tailedness in our stocks. This will be done through graphical analysis, or more specific with the use of QQ plots and by plotting the autocorrelations functions.

ARCH models have been proven useful when analyzing financial time series

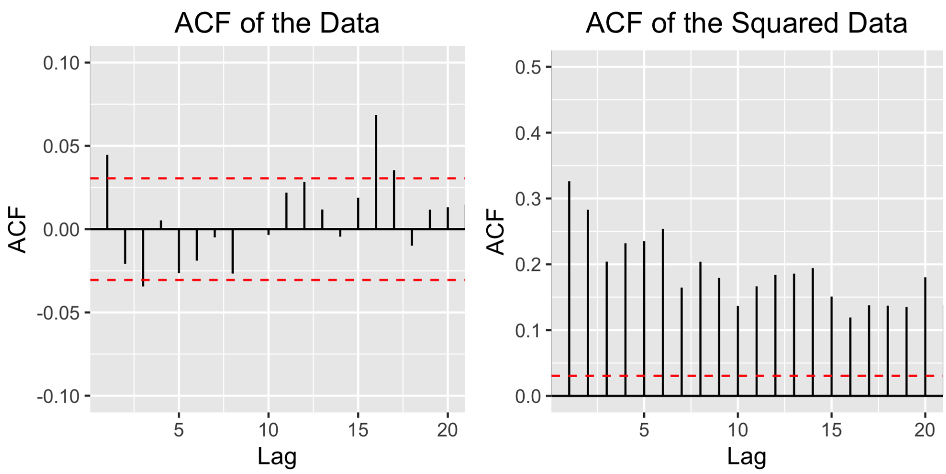

Figure 2: Autocorrelation function for the negative log returns (Left) and the squared negative log returns (Right) for Danske Bank. The dashed red lines denotes a 95% confidence bound.

characterized by little or no serial correlation in the daily returns but strong serial correlation in the squared returns (see e.g. Engel 2001). In figure 2 we can see how the autocorrelation in the squared data exhibit signs of relatively strong autocorrelation for the last 20 days, in relation to the autocorrelation in the non-squared data. The characteristics of the autocorrelation found in the negative log returns motivates the use of an ARCH model when estimating the volatility of the data series. We supplement the graphical analysis with the Ljung-Box test.

Series: Danske bank	χ^2 statistic	Lag	D.f.	p-value
Negative log returns	22.372	log(4124)	log(4124)	0.005225
Squared negative log returns	1942	log(4124)	log(4124)	$<$ 2.2e-16

Table 3: Summary of the Ljung-Box test for the negative log returns for Danske bank.

The null hypothesis of the Ljung-Box test stipulates that the autocorrelation up to the p:th lag is zero whilst the alternative hypothesis assumes that the autocorrelation is nonzero. As seen in table 3 the null hypothesis is rejected at all conventional significance levels. This means that even if the autocorrelation function is relatively weaker in the non-squared series (which can be seen by looking at the y-axis) the negative log returns exhibit significant serial correlation in the squared as well as in the non-squared series. When fitting an AR(1)-GARCH(1,1) model to the data series, we expect that the model extracts the serial dependence found in the data, which means that the model residuals should not be autocorrelated. This will be examined after we have inspected the fat-tailness of the data.

We have already mentioned that our aim is to use EVT in order to estimate the one day ahead Value at Risk for the negative log returns of our included stocks. It is therefore motivated to determine whether the tails of the distribution of the negative log returns are heavier than those of the standard normal distribution. This will be done using a quantile quantile plot (QQ-plot). If the observed empirical distribution correspond with the theoretical distribution, the observations should lie on a straight line.

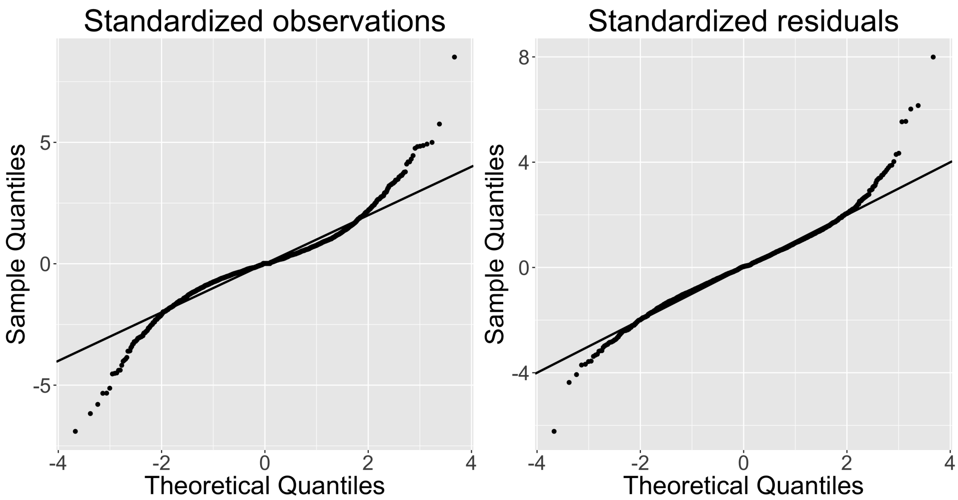

Figure 3: QQ-Plots of the standardized negative log returns for Danske Bank (Left) and standardized residuals resulting from an AR(1)-GARCH(1,1) model with normal innovations (Right), both with the quantiles of the standard normal distribution as reference.

To the left in figure 3 we are able to see that the observations are located below the left tail and above the right tail of the standard normal distribution, indicating that the tails of the distribution of the loss series are heavier than those for the standard normal, which motivates the use of EVT when estimating the one day ahead VaR.

As mentioned earlier it is appropriate to check the adequacy of the AR(1)-GARCH(1,1) model by ensuring that the model residuals are i.i.d. and thereby not subjected to serial correlation. Therefore we fit our time series model to the data set and check if the autocorrelation we earlier observed has been filtered. If our model is appropriate we expect the results from the ACF to be similar to white noise.

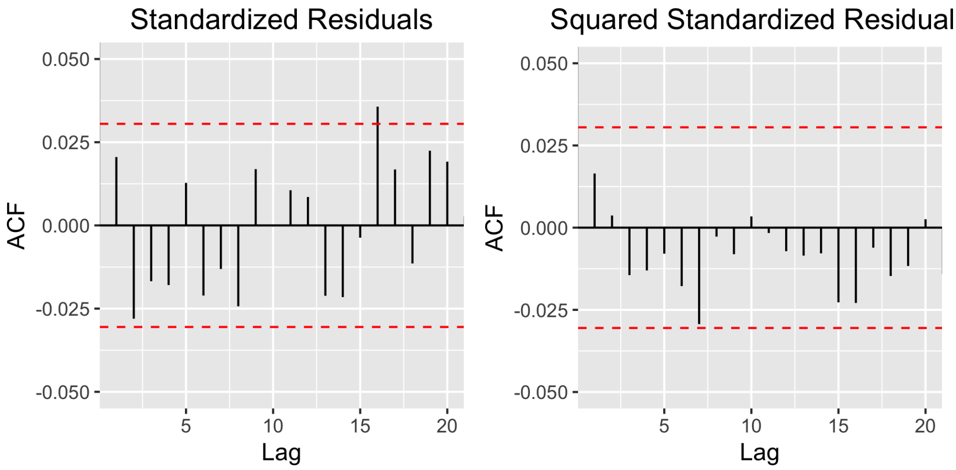

Figure 4: Autocorrelation function for the standardized residuals (Left) and the squared standardized residuals (Right) from the fitted AR(1)-GARCH(1,1) model. The dashed red lines denotes a 95% confidence bond.

In figure 4 we see that the majority of the standardized and the squared standardized residuals lies within the 95% confidence bound. The only observation located outside of the confidence bonds is the 18:th lag in the non-squared series. Having said that, this breach of the confidence bound is not to be considered major and is not a strong enough indication for us to reject the idea that the residuals are i.i.d. In figure 5 we complement the ACF of the standardized and squared standardized residuals by displaying the ACF of 10'00 i.i.d. standard normally distributed random variables. This is meant to illustrate the occurrence of significant test results, based on nothing but pure chance.

If we recapitulate the ACF of the loss series illustrated in figure 2, then we remember that the squared loss series showed definite signs of autocorrelation while the indication of presence of autocorrelation for the non squared series was not as clear. Therefore we conclude this section by illustrating the amount of significant parameters in the AR(1)-GARCH(1,1) model with normal innovations.

As seen in figure 6 below, a majority of the GARCH(1,1) parameters used in order to model the conditional volatility are significant while approximately

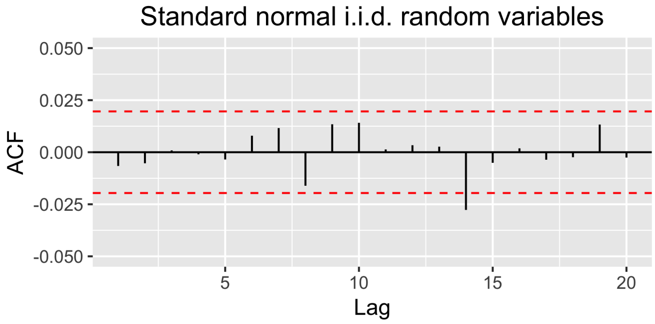

Figure 5: Illustration of the occurrence of a significant outcome for the ACF of 10'000 i.i.d. standard normal random variables.

25% of the intercept used in the modeling of the conditional mean are significant and roughly 50% of the estimated autoregressive parameters are significant. However, since the main purpose of our analysis is to evaluate the ability to forecast VaR for our included time series models, the significance for the parameters in the AR(1)-GARCH(1,1) models are of minor interest. However it is still mentioned for the sake of completeness. In figure 9 found in the appendix we also illustrate the estimated AR1, ALPHA1 and BETA1 parameters from our fitted AR(1)-GARCH(1,1) model with normal innovations using a window of 500 and 1000 observations.

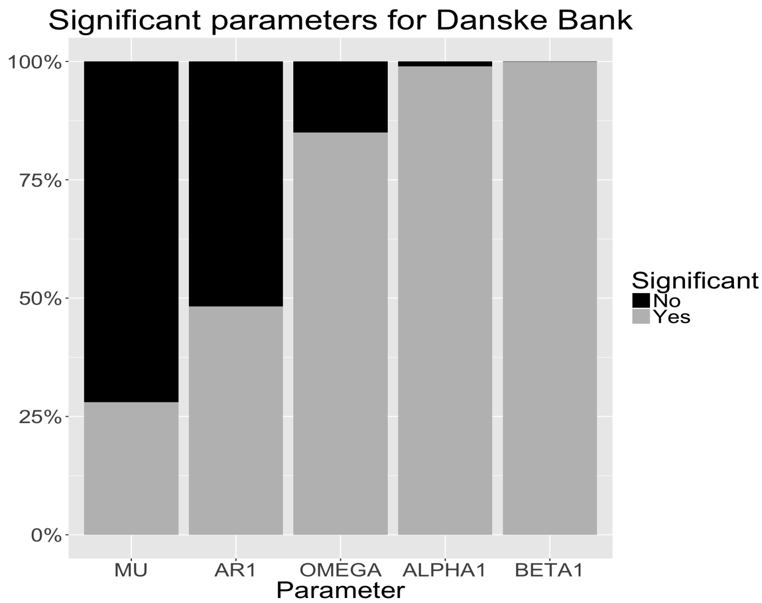

Figure 6: Illustration of the percentage of significant estimated parameters in the AR(1)-GARCH(1,1) model with normal conditional distribution.

5.3 Estimation of VaR

5.3.1 Forecasting using the AR(1)-GARCH(1,1) model

As we have mentioned earlier our aim is to use our fitted AR(1)-GARCH(1,1) model in order to estimate the conditional mean $\hat{\mu}_{t+1}$ and the conditional volatility $\hat{\sigma}^2_{t+1}$ using the previous 1000 observations in the negative log return series. We recapitulate the fact that the negative log returns X_t are assumed to be an strictly stationary time series on the form:

$$X_t = \mu_t + \sigma_t Z_t.$$

We have so far not made any assumption regarding the marginal distribution of the innovation Z_t besides that they are assumed to be i.i.d. with zero mean and unit variance. If we let $\mathbb{I}_t$ denote the known information up to and including time point t we may define the predictive distribution for $t+1$ as:

$$F_{X_{t+1}|\mathbb{I}_t}(x) = P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x|\mathbb{I}_t) = F_Z\Big(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\Big)$$

(McNeil & Frey 2000, pp. 277).

From the definition of Value at Risk presented in the previous section we are able to obtain the estimated Value at Risk quantile for the next day through

$$\begin{aligned} x_q &= \inf\left\{x : F_{X_{t+1}|\mathbb{I}_t}(x) \geq q\right\} && (8) \\ &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} z_q && (9) \end{aligned}$$

which highlights the importance of the choice of distribution for the innovations. If we don't choose a distribution for the innovation term a priori we will not be able to calculate the VaR quantile.

In this thesis we will assign two distributions for the innovation term. First the standard normal distribution, and secondly a standardized t-distribution with 4 degrees of freedom. From equation (9) one may realize that an estimate for the one day ahead VaR given at time t with innovations following a standard normal distribution can the be expressed as:

$$\hat{x}_q = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\Phi^{-1}(q).$$

Where Φ is the cumulative distribution function of a standard normal variable for the quantile q. For further comparisons we will also model the innovations for our time series models using a student t distribution with 4 degrees of freedom standardized in order to have variance 1. These will later be compared to the corresponding estimates calculated by fitting a GPD distribution to the residuals exceeding a chosen threshold. How this is implemented is presented in more detail in the following two sections.

5.3.2 Estimation using conditional EVT

When calculating VaR the choice of distribution for the innovations is decisive. In this section we will describe how to apply the Peak over Threshold model first mentioned in section 3.1 to the residuals resulting from the AR(1)-GARCH(1,1) model.

We start by fitting an AR(1)-GARCH(1,1) model to 1000 observations in our historical series of negative log returns using maximum likelihood estimation. By using the parameters from the fitted model we calculate the standardized residuals

$$(z_{t-n+1}, z_{t-n+2}, \cdots, z_t) = \left(\frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \frac{x_{t-n+2} - \hat{\mu}_{t-n+2}}{\hat{\sigma}_{t-n+2}}, \cdots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right)$$

where $n = 1000$. Given these residuals and a threshold u we estimate a quantile for the innovations in accordance to equation (10)

$$\hat{z}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u}(1-q) \right)^{-\hat{\xi}} - 1 \right). \tag{10}$$

Once this is done, we may proceed by calculating an estimate for VaR by using $\hat{z}_q$ and the parameter estimates from the fitted model through

$$\widehat{VaR}_q = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{z}_q.$$

In order for us to be able to compare VaR estimates from unconditional and conditional approaches we also include an unconditional model (later referred to as Unconditional EVT) where the VaR estimates is the quantile estimation of the negative log returns based on GPD.

5.4 Choice of Threshold

When applying the Peaks over Threshold model on the ordered residuals from our respektive AR(1)-GARCH(1,1) models the choice of threshold u is important since it effects the stability of the shape and scale parameters ξ and β. McNeil & Frey describes that the best GPD estimates of the excess distribution are obtained by trading bias off against variance (McNeil & Frey 2000, pp. 7). This means that we will chose a threshold u high enough to reduce the chance of bias while at the same time keeping N large in order to control the variance for the estimated parameters. In our analysis, we will follow the approach of McNeil & Frey by using the 101:st ordered residual as the threshold, thereby making it random whilst fixating the tails to always contain 10 % of the observations resulting from the residuals of the moving window. We will not analyze the choice of threshold further in this thesis, however, Kjellson & Koskinen Rosemarin showed that this choice of u resulted in stable parameter estimates for the GPD by using Monte Carlo simulation and quantifying the bias using the Root Mean square Error (Kjellson & Koskinen Rosemarin 2012, pp. 23.).

5.5 Backtesting

As in every case when working with statistical models, they are only helpful if they manage to accomplish what we expect of them. In this case if our fitted time series model favorably manage to forecast the financial risk.

In order to verify this, we will be using a method called backtesting. Backtesting may be described as a formal statistical framework that consists of verifying that actual losses are in line with projected losses (Jorison 2007, pp. 139). In fact, various backtesting methods exist. The tests used in this thesis are referred to as the exact binomial test, the test of independence, the unconditional- and the conditional coverage test. All tests will be presented and implemented in the oncoming sections.

5.5.1 The Exact Binomial Test

When implementing the exact binomial test we denote the number of exceedances as n_1, the total amount of VaR estimates as N and the ratio $\frac{n_1}{N}$ as the exceedance rate. In an ideal setup the exceedance rate should be an unbiased estimator of p in VaR_{1-p} (where $(1-p)$ denotes the coverage rate), and should converge to p as the sample size increased.

We are interested in testing whether n_1 is deviating to much in order for the null hypothesis to be accepted in a total sample size of N. This test makes no assumption regarding the distribution of the returns, making this a non-parametric approach (Jorison 2007, pp. 143). The data we are to test is basically a sequence of exceedances and non-exceedances (here coded as zeros and ones), which is often referred to as *bernoulli trials* in statistical literature. Under the null hypothesis these exceedances are independent making the total numer of exceedances n_1 to follow a binomial probability distribution on the following form:

$$f(n_1) = \binom{N}{n_1} p^{n_1} (1-p)^{N-n_1}.$$

Note that p denotes the probability of exceeding our estimated VaR limits which in our case will either be 0.05, 0.01 or 0.005.

5.5.2 The Conditional Coverage Test

The Conditional Coverage test first introduced by Christoffersen 1998 offers an extension to the unconditional coverage test. Briefly explained the unconditional coverage test (just as the name indicates) ignores conditioning and thereby time variation in the data. Hypothetically this means that we could have a situation where our estimated VaR limits hold the expected amount of exceedances for the whole tested period. However, if the exceedances that takes place are clustered in a short period of time our model would not be satisfactory. The conditional coverage test combines the unconditional coverage test and independence testing in order to account for previous events and can be expressed as:

$$LR_{CC} = LR_{UC} + LR_{IND} \sim \chi^2(2). \tag{11}$$

For the complete proof regarding the construction of the LR_{CC} test statistic, see the appendix of Christoffersens paper (Christoffersen 1998, pp. 861). The Unconditional Coverage test and Independence testing is explained in more detail in the two oncoming sections.

5.5.3 The Unconditional Coverage Test

The null hypothesis for the unconditional coverage test stipulates that every exceedance of the given VaR level is to be treated as a i.i.d Bernoulli-distributed random variable. This is intuitive since if our model is correctly specified the hit sequence of exceedances should be unpredictable and thus distributed over time as an independent Bernoulli variable.

We move on by testing whether the proportions of exceedances obtained from our fitted model π is significantly different from the stipulated proportion p, leading us to formulate the null and alternative hypothesis as

$$H_0 : \pi = p$$
$$H_A : \pi \neq p$$

We test the hypothesis first by defining an indicator variable that notifies if the negative log return in $t+1$ exceeds the one day ahead VaR forecast in time point t with a coverage rate of q

$$I_t = \mathbb{1}\{X_{t+1} > \widehat{VaR}_q^t\}$$

and by formulating the likelihood of an i.i.d. Bernoulli exceedance sequence as:

$$L(\pi) = \prod_{i=1}^{N} (1-\pi)^{1-I_t} \pi^{I_t} = (1-\pi)^{n_0} \pi^{n_1}.$$

Where n_0 and n_1 represents the amount of non-exceedances, and exceedances of the VaR quantile. The total amount of estimated VaR limits can then be expressed as $N = n_0 + n_1$. The maximum likelihood estimator of π can then be obtained using the conventional method as $\hat{\pi} = \frac{n_1}{N}$, which is simply the observed proportion of exeedances in our sample (Christoffersen 2012, pp. 302).

If we plug this estimator into the likelihood function we can create a likelihood ratio statistic on the form:

$$\begin{aligned} LR_{UC} &= -2\ln\Big[\frac{L(p)}{L(\hat{\pi})}\Big] \\ &= -2\ln\Big[(1-p)^{n_0} p^{n_1}\Big] + 2\ln\Big[(1-\frac{n_1}{N})^{n_0}(\frac{n_1}{N})^{n_1}\Big] \sim \chi^2(1). \end{aligned}$$

The test statistic LR_{UC} is asymptotically χ^2 distributed with one degree of freedom when the number of observations N goes to infinity (Christoffersen 2012, pp. 303). Our decision criterion will be based on the p-value and the conventional significance level of 0.05 will be the chosen level for all our tests.

5.5.4 Independence Testing

The main purpose of the independence test is to reject VaR models whose exceedances occur during the same time or appears as clusters. In order to establish a test that will account for this phenomena we assume that the exceendance frequency is dependent over time and can be described as a first order Markov sequence with transition probability matrix

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

Where π_{01} denotes the probability of an exceedence *tomorrow* conditioned on that for today we have no exceedence, or more explicitly

$$\pi_{01} = \Pr\big(\text{Exceedance tomorrow} \mid \text{No exceedance today}\big)$$
$$\pi_{11} = \Pr\big(\text{Exceedance tomorrow} \mid \text{Exceedance today}\big)$$

Remember that since a Markov Chain is 'memoryless' it means that only todays outcome effects the outcome of tomorrow. As a result of this the probabilities π_{01} and π_{11} describe the entire process, and we may write the likelihood function of the first order Markov chain as:

$$L\big(\Pi_1\big) = (1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}} \cdot (1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}.$$

Here, $n_{ij},\ i,j \in \{0,1\}$ denotes the amount of observations in the loss series with a j follwing an i. If we are to solve for the maximum likelihood estimates and utilize the fact that the probabilities in the rows of the matrix have to sum up to 1 we have that

$$\begin{aligned}\hat{\Pi}_1 &= \begin{bmatrix} \hat{\pi}_{00} & \hat{\pi}_{01} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} 1-\hat{\pi}_{01} & \hat{\pi}_{01} \\ 1-\hat{\pi}_{11} & \hat{\pi}_{11} \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{n_{00}}{n_{00}+n_{01}} & \dfrac{n_{01}}{n_{00}+n_{01}} \\ \dfrac{n_{10}}{n_{10}+n_{11}} & \dfrac{n_{11}}{n_{10}+n_{11}} \end{bmatrix}.\end{aligned}$$

If our observed exceendances are time dependent, it would mean that π_{01} and π_{11} would not be equal and we may test the independence hypothesis $\pi_{01} = \pi_{11}$ by the following likelihood ratio statistic.

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

Where $L(\hat{\Pi})$ is the likelihood under the alternative hypothesis of the Unconditional coverage test.

6 Results

We start this section by illustrating the one day ahead forecast VaR at the confidence level 0.995 from the unconditional EVT model and the forecasts obtained from from the AR(1)-GARCH(1,1) model with normal conditional distribution (used as a reference) and the corresponding model based on EVT.

Note that the length of the predicted period differs from that of the original data series. This length reduction in the series of predictions is partly due to the transformation from closing price to log returns, but mainly because of the rolling window of 1000 observations used in the model. What this means is that VaR will not be estimated for the 1000 first days since these observations are used in order to calculate the first prediction, which will be the VaR for the 1001:st day in each data series.

As seen in figure 7 the unconditional EVT model (see dashed blue line) appears to adopt slower to periods of increased volatility resulting in several violations of the estimated VaR limit. We can also observe that the VaR estimates from the conditional model with standard normal innovations (see black bars) seems to quickly respond to periods of increased volatility. Overall the two conditional models produce similar estimates where the VaR estimates from the conditional EVT model is slightly higher than those from the conditional normal model. This is to be expected since the main reason we applied the POT method to model the innovations was due to the fact that it has heavier tails in relation to the normal distribution, which is most clearly observed at the higher levels of significance (see figure 10 in appendix). We are also able to see that the conditional EVT model is more sensitive to volatility spikes since the red line in figure 7 is clearly above the black lines in the more volatile periods. Furthermore, we can in the period around 2013 observe how the VaR estimates from the conditional EVT model recede in a slightly slower phase after a volatility spike in relation to its normal counterpart.

Next, we summarize the numerical results from the exact binomial test and the conditional coverage test in the 2 tables that follows. The backtesting results from the two other backtests will also be mentioned, but in order to facilitate the reading we have chosen to include their numerical results in table 6 and 7 in the appendix. When deciding if we are to reject a model, we use a significance level of 0.05 but the p-values up to three digits is presented for all test statistics. To give the reader an easier overview of the results we will be using bold p-values if the tested model is rejected.

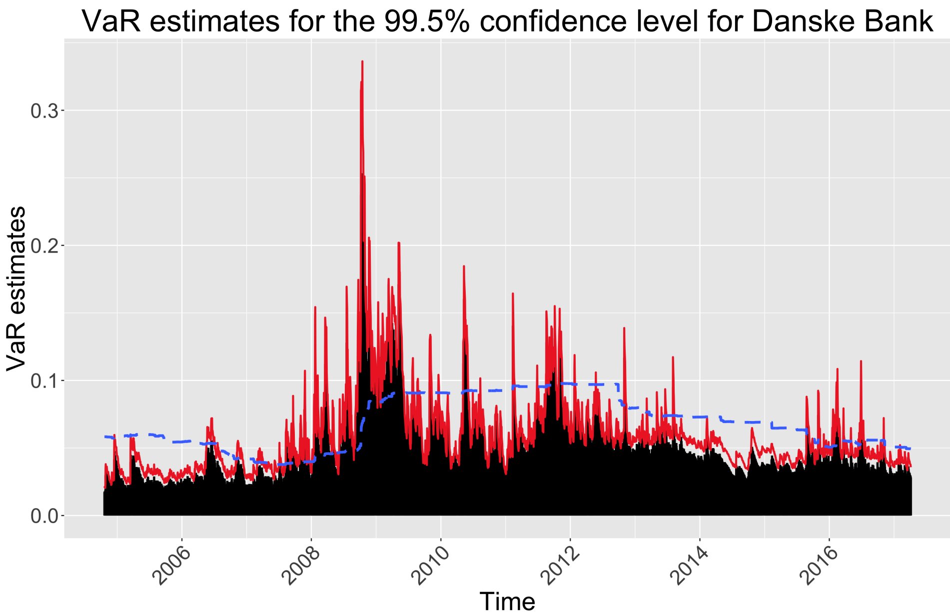

Figure 7: VaR estimates for the 0.995 confidence level for the negative log returns of Danske Bank. The black bars illustrates the estimates obtained form the fitted AR(1)-GARCH(1,1) model with normal innovations, the red line shows the estimates from corresponding model with innovations modeled by the POT method. The dashed blue line denotes the estimates from the unconditional EVT model.

The results from the unconditional coverage test should be in line with those of the exact binomial test since they are both unconditional tests based on the observed proportions of exceedances. As seen in table 4 and 6 the same models are rejected in both tests except for the conditional normal model at confidence level 0.95 for the Astra Zeneca data set, making them overlap to a large extent. The results indicate that the unconditional EVT model is one of the two worst performing models which in particular applies at the higher significance levels. In table 4 we see that at the 0.995 level the unconditional EVT model is rejected for three out of five data sets which indicates that the model fails to estimate VaR limits which encloses the stipulated proportion of observations.

The worst performing model is the conditional normal model which is rejected for all data sets at the higher significance levels in the case for the unconditional coverage and the exact binomial test. This clearly indicates that the tails for the standard normal distribution is not heavy enough to model the distribution for the negative log returns for our data series.

The conditional t model is rejected once at the 95% significance level and never at the higher significance levels, which means that it is clearly outperforming its standard normal counterpart. It has in many previous studies been shown that the t-distribution is preferred when estimating VaR for log returns (see e.g. Angelidis et al. 2003, pp. 21). That being said, we are still surprised by these great differences in performance between the conditional normal and t-distribution models.

The EVT model with normal and t conditional distribution is not rejected at any significance level and are to be considered the best performing models. The results show that both conditional EVT models produces VaR estimates enclosing a proportion of the observations which does not differ significantly from the stipulated proportion. These results are in line with those obtained by McNeil & Frey.

The results for the test of independence are summarized in table 6 where we can see that at the highest significance level, all models perform relatively well. In the two previous tests the unconditional EVT and conditional normal model were rejected for several (or all) data sets at this level of confidence while in this test both mentioned models are rejected only once each. For the lowest significance level 0.95 all models are rejected for Danske Bank and H&M indicating that all models are experiencing clustered violations of the estimated VaR limits. Worth noting is that the EVT model with innovations modeled by a t-distribution is just barely rejected at the 0.99. The rest of the results of the independence test are in line with those previously observed.

	Astra Zeneca	Autoliv SDB	Danske Bank	H&M	SAS
Total amount of predictions	3519	4006	3125	5761	2962
0.95 Quantile					
Expected amount of exceedances	176	200	156	288	148
Conditional Normal	151 (0.053)	185 (0.310)	146 (0.435)	221 **(0.000)**	115 **(0.005)**
Conditional t	195 (0.153)	221 (0.137)	188 **(0.011)**	280 (0.650)	128 (0.092)
Conditional EVT Normal	182 (0.643)	214 (0.328)	168 (0.325)	304 (0.333)	148 (1.000)
Conditional EVT t	186 (0.439)	212 (0.384)	176 (0.109)	309 (0.204)	153 (0.673)
Unconditional EVT	176 (1.000)	212 (0.384)	173 (0.175)	308 (0.227)	145 (0.833)
0.99 Quantile					
Expected amount of exceedances	35	40	31	58	30
Conditional Normal	54 **(0.003)**	60 **(0.003)**	51 **(0.001)**	85 **(0.001)**	41 **(0.042)**
Conditional t	44 (0.149)	38 (0.812)	33 (0.719)	61 (0.643)	23 (0.267)
Conditional EVT Normal	37 (0.735)	46 (0.340)	34 (0.589)	65 (0.320)	28 (0.853)
Conditional EVT t	38 (0.611)	42 (0.750)	36 (0.368)	61 (0.643)	32 (0.644)
Unconditional EVT	36 (0.865)	41 (0.874)	50 **(0.002)**	67 (0.208)	44 **(0.012)**
0.995 Quantile					
Expected amount of exceedances	18	20	16	29	15
Conditional Normal	44 **(0.000)**	47 **(0.000)**	35 **(0.000)**	59 **(0.000)**	27 **(0.004)**
Conditional t	25 (0.092)	19 (0.911)	15 (1.00)	33 (0.401)	14 (1.000)
Conditional EVT Normal	19 (0.719)	22 (0.653)	17 (0.702)	38 (0.092)	19 (0.294)
Conditional EVT t	21 (0.402)	21 (0.822)	16 (0.899)	39 (0.061)	19 (0.294)
Unconditional EVT	19 (0.719)	30 **(0.032)**	24 **(0.041)**	33 (0.401)	27 **(0.004)**

* The table includes the p-values from the two-sided binomial test, the observed and the expected number of exceedances. Bold numbers indicate rejection of the null hypothesis.

Table 4: Results from the Exact binomial test.

	Astra Zeneca	Autoliv SDB	Danske Bank	H&M	SAS
Total amount of predictions	3524	4006	3124	5761	2962
0.95 Quantile					
Conditional Normal	4.949 (0.084)	1.325 (0.515)	10.895 **(0.001)**	23.627 **(0.000)**	9.751 **(0.002)**
Conditional t	3.03 (0.22)	2.861 (0.239)	20.922 **(0.000)**	6.280 **(0.043)**	3.398 (0.065)
Conditional EVT Normal	2.404 (0.301)	0.984 (0.611)	10.539 **(0.001)**	5.757 (0.056)	0.024 (0.877)
Conditional EVT t	2.339 (0.310)	0.711 (0.701)	11.718 **(0.001)**	7.838 **(0.020)**	0.288 (0.591)
Unconditional EVT	14.026 **(0.001)**	17.134 **(0.000)**	67.745 **(0.000)**	18.182 **(0.000)**	11.446 **(0.001)**
0.99 Quantile					
Conditional Normal	8.710 **(0.013)**	8.708 **(0.013)**	12.293 **(0.000)**	15.444 **(0.000)**	5.096 **(0.024)**
Conditional t	2.345 (0.310)	0.892 (0.640)	0.803 **(0.370)**	2.072 (0.355)	1.979 **(0.159)**
Conditional EVT Normal	0.873 (0.646)	1.192 (0.551)	0.988 (0.320)	2.449 (0.294)	0.626 (0.429)
Conditional EVT t	0.835 (0.659)	0.629 (0.730)	1.538 (0.215)	2.072 (0.355)	4.077 **(0.043)**
Unconditional EVT	0.775 (0.679)	0.615 (0.735)	34.294 **(0.000)**	20.975 **(0.000)**	18.696 **(0.000)**
0.995 Quantile					
Conditional Normal	28.278 **(0.000)**	26.719 **(0.000)**	18.63 **(0.000)**	29.396 **(0.000)**	8.596 **(0.003)**
Conditional t	3.105 (0.212)	0.235 (0.889)	0.17 (0.68)	0.967 (0.617)	0.178 (0.673)
Conditional EVT Normal	0.312 (0.856)	0.432 (0.806)	0.305 (0.581)	3.978 (0.137)	1.339 (0.247)
Conditional EVT t	0.866 (0.649)	0.268 (0.875)	0.174 (0.677)	4.483 (0.106)	1.339 (0.247)
Unconditional EVT	0.312 (0.856)	6.008 **(0.05)**	22.16 **(0.000)**	0.967 (0.617)	9.438 **(0.002)**

* The table includes all the calculated LR_{CC} test statistics ($\chi^2(2)$ distributed) and their associated p-value. Bold numbers indicate rejection of the null hypothesis.

Table 5: Results from the conditional coverage test

As mentioned in section 5.5.2 there is one aspect that needs to be further investigated. Since even if a model on average is correct it is necessarily not the case that the degree of coverage is correct for smaller subperiods, remembering that daily financial data is clustered. In other words, if the model is correctly specified and fully accounts for the heteroscedasticity in the negative log return series then the exceedances of VaR should not appear in clusters. The test statistic for the conditional coverage test in equation (11) in the same section is constructed by summing the test statistics from the unconditional coverage test and the test of independence. The results for the conditional coverage tests will thus be similar but not identical to those obtained in the unconditional coverage testing.

In table 5 we can once again see that the unconditional EVT model performs poorly at all significance levels and especially at the 0.95 level where all models are rejected. The conditional t model is rejected twice at the two lower significance levels but apart from that there are no considerable differences from previous results. The two conditional EVT models perform best overall, in particular the EVT model with innovations modeled by a standard normal distribution with only one rejection.

7 Discussion

The conditional coverage test implemented in section 6 and used in order to evaluate whether the exceedances of the estimated VaR limits appears in clusters do suffer from a substantial shortcoming. The framework of the test results in that it only manages to consider independency between two adjacent exceedances, or only independency in one step. Thereby it fails to detect whether the exceedances are clustered for longer time periods. Because of this one adequate complement to this analysis could be to examine if the conditional duration test proposed by Christoffersen and Pelletier (2004) would give support for the our results.

As seen in figure 6 and 8 the majority of the estimated AR(1) parameters in our models are not significant. We chose this approach in order for our results to be comparable with those obtained by McNeil & Frey. However, since the expected value of the log returns are close to zero we have reasons to believe that this analysis could be carried out without the autoregressive modeling of the mean in the different conditional models. This comment may be a trifle but we find it difficult to justify the inclusion of the autoregressive component unless it contributes to improved VaR estimates in relation to a reduced time series model (which of course could be examined). A further analysis could include a more rigid analysis when specifying the models before using the POT model for the innovations. By examining the tails of the standardized residuals from from the AR(1)-GARCH(1,1) model fitted with MLE in figure 3, we can see that the upper tail might possibly be heavier than the lower. Perhaps an asymmetric model such as the EGARCH would be more appropriate for the estimation of VaR for some data sets?

Even though our results speaks in favor for a conditional approach when estimating VaR (our best performing models where indeed conditional models) the answer is not completely univocal if we are to remember that our worst performing model was the conditional normal. Therefore another extension to this analysis could be to divide the data sets where one subsample could be used for model fitting and the other for VaR estimation. By doing so it would be possible to incorporate the variance resulting from the estimation of the model parameters which can be incorporated in the VaR estimates. This might possible lead to different conclusions in the comparison of unconditional and conditional models.

8 Conclusion

The main purpose of this thesis was to demonstrate how time series models whose innovations have been modeled by the application of extreme value theory can be used in risk management, or more specifically when estimating the one day ahead VaR for various confidence levels. We restricted our analysis to comparing four time series models namely an AR(1)-GARCH(1,1) model based on the assumption that the innovations either followed a standard normal or t distribution. These were then compared to two equivalent models whose innovations where modeled using the Peak-Over-Threshold method (POT method). In order to enable comparisons between conditional and uncdonditional approaches, an unconditional model were included by calculating VaR estimates by using the GPD estimates fitted on the series of negative log returns.

These models were then evaluated using four backtesting techniques: The exact binomial test, the independence test, the unconditional and the conditional coverage test. The results from these backtests show that the AR(1)-GARCH(1,1) model with innovations modeled by the POT method outperforms the other models. These results hold for all confidence levels but in particular for the EVT model with standard normal innovations but the results become more clear at the higher significance levels since it is where the other models generally have their worst performance. The results also indicate that a conditional approach is preferable to an unconditional one. This conclusion is based on the fact that our unconditional EVT model is one of the two worst performing models and we have no real reason to believe that other unconditional approaches such as Historical Simulation would alter this statement. Worth mentioning is that these results are in line with those of McNeil & Frey whom also estimates Expected Shortfall in their paper inspiring us to write this thesis.

9 References

[1] Acerbi, C. & Tasche, D. (2002), 'Expected Shortfall: A Natural Coherent Alternative to Value at Risk', *Economic Notes* Vol. 31, pages 379-388.

[2] Angelidis, A., Benos, A. & Degiannakis, S. (2003), 'The Use of GARCH Models in VaR Estimation', *Statistical Methodology* Vol. 1, pages 105-128.

[3] Bali, T. G. (2003), 'An Extreme Value Approach to Estimating Volatility and Value at Risk ', *The Journal of Business* Vol. 76, pages 83-108

[4] Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* Vol. 31, pages 307-327.

[5] Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* Vol. 39, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance, pages 841-862.

[6] Christoffersen, P. F. & Pelletier, D. (2004), 'Backtesting Value-at-Risk: A Duration-Based Approach', *Journal of Financial Econometrics* Vol. 2, pages 84-108.

[7] Christoffersen, P. F. (2012), *Elements of financial risk management*, second edn, Academic Press.

[8] Danielsson, J. & de Vires, C. G. (2000), 'Value-at-risk and extreme returns', *Annales d'economie et de Statistique* Vol. 60, pages 236-269.

[9] de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: an introduction* , Springer.

[10] Embrechts, P. , Kluppelberg, C. & Mikosch, T. (1997), *Modeling extremal events for insurance and finance*, Springer.

[11] Engel, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimatesof the variance of united kingdom ination', *Econometrica* Vol. 50, pages 987-1007.

[12] Engel, R. F. (2001), 'GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics', *Journal of Economic Perspective* Vol. 15, pages 157-168.

[13] Gençay, R., Selçuk, F. & Ulugülyağci, A. (2003), 'High volatility, thick tails and extreme value theory in value-at-risk estimation', *Insurance: Mathematics and Economics* Vol. 33, pages 337-356.

[14] Jorison, P. (2007), *Value at risk: the new benchmark for managing financial risk*, third edn, McGraw-Hill.

[15] Kjellson, B. & Koskinen Rosematrin, S. (2012), 'The Best of Two Worlds: Combining Conditional Volatility Models with Extreme Value Theory to Calculate Value at Risk', *LUP Student Papers*.

[16] Kuester, K., Mistnik, S. & Paolella, M. S. (2006), 'Value-at-Risk Prediction: A Comparison of Alternative Strategies', *Journal of Financial Econometrics* Vol. 4, pages 53-89.

[17] McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk meassures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* Vol. 7, pages 271-300.

[18] McNeil, A. J. , Frey, R. & Embrechts, P. (2005), *Quantitative risk management: concepts, techniques, and tools*, Princeton University Press.

[19] Tsay, R. (2010), *Analysis of financial time series*, third edn. John Wiley & Sons.

10 Appendix

10.1 Additional figures & tables

In figure 8 we are able to observe that the estimated parameters constituting the AR(1) components (the expected conditional mean MU and the autoregressive coefficient AR1 in (5)) tend to not be significant to the same extent as the estimated parameters which consitutes the GARCH(1,1) components (the estimated intercept OMEGA and parameters ALPHA1 and BETA1 in (7)). This applies to all included data sets and strengthens our conjecture that the analysis could have been carried out using a reduced model.

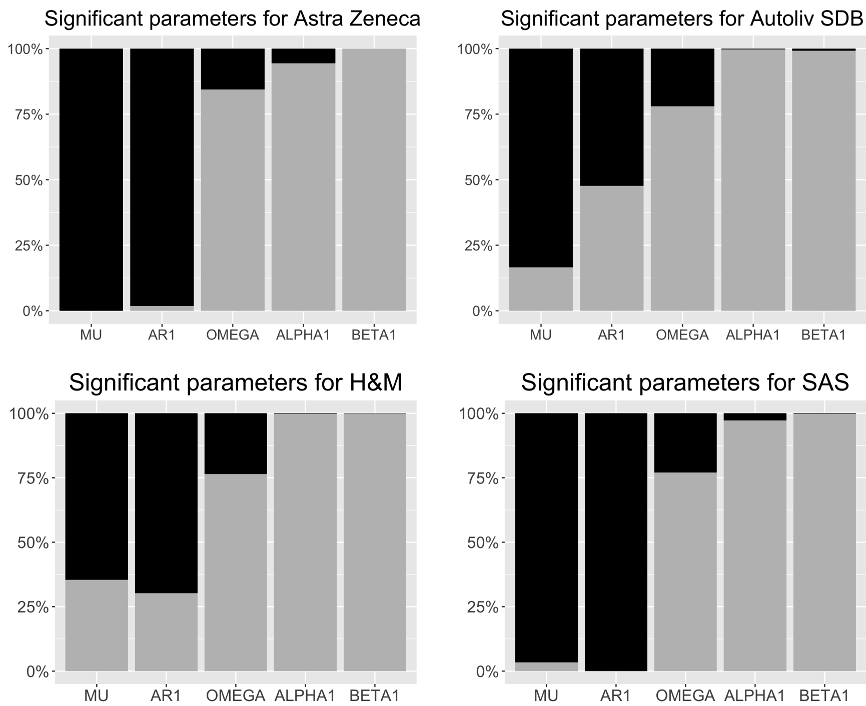

Figure 8: Illustration of the percentage of significant estimated parameters in the AR(1)-GARCH(1,1) model with normal conditional distribution for the remaining data sets.

In figure 9 below we illustrate the estimated parameters over time when fitting an AR(1)-GARCH(1,1) model with normal innovations using a rolling window of 500 and 1000 observations to our data of negative log returns for Danske Bank. Note that in the figure we follow the notation in section 4.2.1 and 4.2.2 where Phi = $\hat{\phi}$, Alpha = $\hat{\alpha}_1$, Beta = $\hat{\beta}_1$, the estimated intercepts are not included since they are all close to zero. In both figures we are able to see that the same model parameters does not seem to apply for the whole dataset since the parameter estimates change over time. Furthermore we can see that when we expand the rolling window to 1000 observations the parameters fluctuate less, especially the estimated β_1 parameters. This leads us to believe that reducing the window size would not lead to further insights regarding the VaR forecasts.

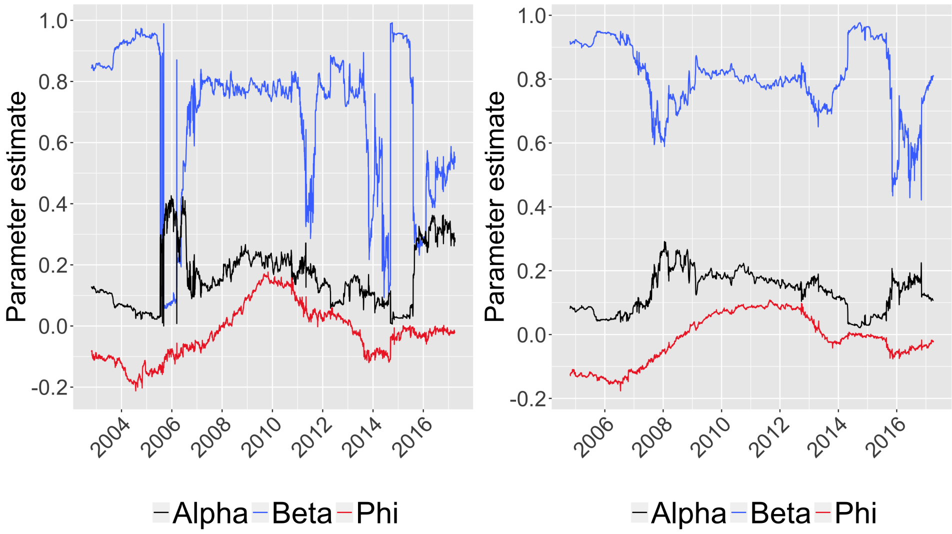

Figure 9: Parameter estimates over time from the AR(1)-GARCH(1,1) model with normal innovations using a window of 500 observations (Left) and 1000 observations (Right)

In figure 10 below we illustrate the difference between the one day ahead forecasts of VaR from the AR(1)-GARCH(1,1) model with normal innovations with those of the corresponding model with innovations modeled by the POT method for all levels of significance. As mentioned in section 6 the VaR estimates from the conditional EVT model is generally higher than those of the normal distribution, which especially applies at the higher levels of significance. This is clearly visible since the difference is negative for most (if not all) values in the case for the 0.99 and 0.995 confidence level.

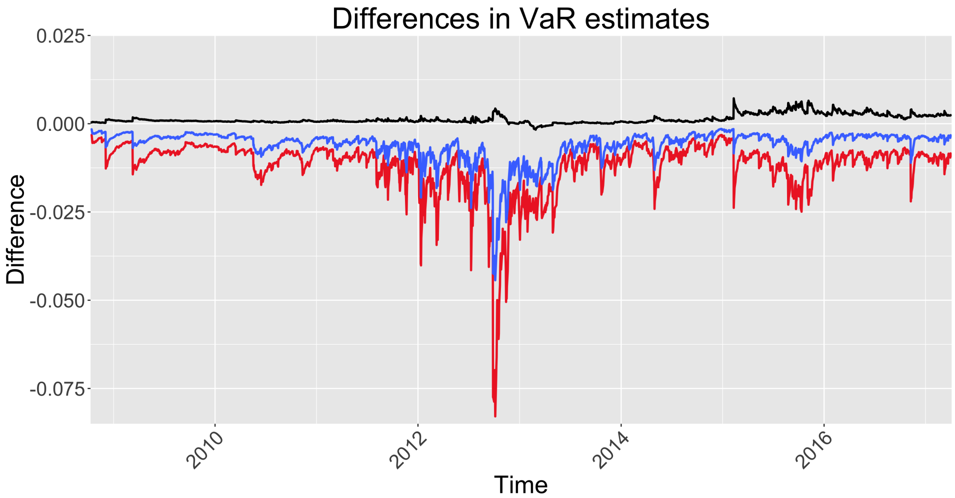

Figure 10: Illustration of the difference of the one day ahead forecasts of VaR resulting from the AR(1)-GARCH(1,1) model with normal innovations and the corresponding model with innovations modeled by the POT method for significance level 0.95 (see black line), 0.99 (see blue line) and 0.995 (see red line). As seen in the figure the differences are negative to a greater extent for the higher significance levels.

On the following page we list the backtest results for the unconditional coverage test and the test of independence in table 6 and 7.

	Astra Zeneca	Autoliv SDB	Danske Bank	H&M	SAS
Total amount of predictions	3524	4006	3124	5761	2962
0.95 Quantile					
Conditional Normal	3.979 **(0.046)**	1.100 (0.294)	0.716 (0.397)	17.801 **(0.000)**	8.408 **(0.004)**
Conditional t	2.044 (0.153)	2.182 (0.140)	6.417 **(0.011)**	0.239 (0.625)	3.004 (0.083)
Conditional EVT Normal	0.199 (0.656)	0.966 (0.326)	0.917 (0.338)	0.914 (0.339)	0.000 (0.993)
Conditional EVT t	0.564 (0.453)	0.707 (0.401)	2.542 (0.111)	1.568 (0.21)	0.169 (0.681)
Unconditional EVT	0.000 (0.988)	0.707 (0.401)	1.841 (0.175)	1.424 (0.233)	0.069 (0.793)
0.99 Quantile					
Conditional Normal	8.676 **(0.003)**	8.696 **(0.003)**	10.599 **(0.001)**	11.474 **(0.001)**	3.944 **(0.047)**
Conditional t	2.039 (0.153)	0.109 (0.741)	0.098 (0.754)	0.198 (0.657)	1.619 (0.203)
Conditional EVT Normal	0.087 (0.768)	0.849 (0.357)	0.239 (0.625)	0.919 (0.338)	0.091 (0.763)
Conditional EVT t	0.213 (0.644)	0.093 (0.760)	0.698 (0.403)	0.198 (0.657)	0.188 (0.664)
Unconditional EVT	0.016 (0.898)	0.022 (0.882)	9.626 **(0.002)**	1.469 (0.226)	6.136 **(0.013)**
0.995 Quantile					
Conditional Normal	27.973 **(0.000)**	26.417 **(0.000)**	17.837 **(0.000)**	24.374 **(0.000)**	8.099 **(0.004)**
Conditional t	2.748 (0.097)	0.054 (0.816)	0.025 (0.874)	0.586 (0.444)	0.045 (0.831)
Conditional EVT Normal	0.106 (0.745)	0.189 (0.664)	0.119 (0.73)	2.68 (0.102)	1.093 (0.296)
Conditional EVT t	0.614 (0.433)	0.046 (0.829)	0.009 (0.924)	3.263 (0.071)	1.093 (0.296)
Unconditional EVT	0.106 (0.745)	4.323 **(0.038)**	3.879 **(0.049)**	0.586 (0.444)	8.099 **(0.004)**

* The table includes all the calculated LR_{UC} test statistics ($\chi^2(1)$ distributed) and their associated p-value. Bold numbers indicate rejection of the null hypothesis.

Table 6: Results from the Unconditional coverage test.

	Astra Zeneca	Autoliv SDB	Danske Bank	H&M	SAS
Total amount of predictions	3524	4006	3124	5761	2962
0.95 Quantile					
Conditional Normal	0.970 (0.325)	0.226 (0.635)	10.179 **(0.001)**	5.827 **(0.016)**	1.343 (0.247)
Conditional t	0.986 (0.321)	0.679 (0.410)	14.506 **(0.000)**	6.041 **(0.014)**	0.394 (0.530)
Conditional EVT Normal	2.205 (0.138)	0.019 (0.891)	9.622 **(0.002)**	4.843 **(0.028)**	0.024 (0.877)
Conditional EVT t	1.775 (0.183)	0.005 (0.944)	9.176 **(0.002)**	6.270 **(0.012)**	0.120 (0.729)
Unconditional EVT	14.025 **(0.000)**	16.427 **(0.000)**	65.904 **(0.000)**	18.182 **(0.000)**	11.377 **(0.001)**
0.99 Quantile					
Conditional Normal	0.035 (0.852)	0.011 (0.915)	1.693 (0.193)	3.970 **(0.046)**	1.151 (0.283)
Conditional t	0.306 (0.580)	0.783 (0.376)	0.705 (0.401)	1.874 (0.171)	0.360 (0.548)
Conditional EVT Normal	0.785 (0.375)	0.343 (0.558)	0.748 (0.387)	1.530 (0.216)	0.535 (0.465)
Conditional EVT t	0.622 (0.430)	0.536 (0.464)	0.840 (0.359)	1.874 (0.171)	3.888 **(0.049)**
Unconditional EVT	0.758 (0.384)	0.593 (0.441)	24.667 **(0.000)**	19.506 **(0.000)**	12.56 **(0.000)**
0.995 Quantile					
Conditional Normal	0.306 (0.580)	0.302 (0.583)	0.793 (0.373)	5.022 **(0.025)**	0.497 (0.481)
Conditional t	0.357 (0.550)	0.181 (0.670)	0.145 (0.704)	0.380 (0.537)	0.133 (0.715)
Conditional EVT Normal	0.206 (0.650)	0.243 (0.622)	0.186 (0.666)	1.298 (0.254)	0.245 (0.620)
Conditional EVT t	0.252 (0.616)	0.221 (0.638)	0.165 (0.685)	1.220 (0.269)	0.245 (0.620)
Unconditional EVT	0.206 (0.650)	1.685 (0.194)	18.281 **(0.000)**	0.380 (0.537)	1.339 (0.247)

* The table includes all the calculated LR_{IND} test statistics ($\chi^2(1)$ distributed) and their associated p-value. Bold numbers indicate rejection of the null hypothesis.

Table 7: Results from the independence test

10.2 Autocorrelation function

When working with time series the concept of correlation is generalized into autocorrelation. The correlation coefficient between X_t and $X_{t-\ell}$ is called the lag-ℓ autocorrelation of X_t and is denoted throughout this thesis as ρ_ℓ. Note that since we assume that our log return series are weakly stationary ρ_ℓ will be a function of the lags only. The autocorrelation function ACF can mathematically be expressed as:

$$\rho_\ell = \frac{Cov(X_t, X_{t-\ell})}{\sqrt{Var(X_t)Var(X_{t-\ell})}} = \frac{Cov(X_t, X_{t-\ell})}{Var(X_t)} \tag{12}$$

Where the last equality holds due to the fact that X_t is assumed to be weakly stationary. For further details, the reader is referred to Tsay (Tsay 2010, pp. 31).

10.3 Ljung-Box test

The Ljung-Box test is used on order to test the null hypothesis H_0 against the alternative hypothesis H_A shown below:

$$\begin{aligned} H_0 &: \rho_1 = \rho_2 = \ldots = \rho_p = 0 \\ H_A &: \rho_i \neq 0, \text{ for some } i \in (1, 2, \ldots, p) \end{aligned}$$

The test statistic is formulated as:

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

Where N is the length of the time series and $\hat{\rho}_\ell$ is the autocorrelation coefficient at lag ℓ. The decision rule is to reject H_0 if $Q(m) > \chi^2_\alpha$, where χ^2_α denotes the $100(1-\alpha)$th percentile of a chi-squared distribution with m degrees of freedom. In fact, the choice of m may affect the performance of the test statistic. Previous studies suggest that the choice $m \approx \ln(N)$ provide better power performance (with an exception for time series submitted to seasonal tends) (Tsay 2010, pp. 32-33). Since we are analyzing 5 time series that are assumed not to be influenced by a seasonal trend, we will follow this convention.

10.4 ML estimation

When we are to specify or formulate a likelihood function we can practically always assume that the observations are independent and identically distributed. This makes the computation of the likelihood substantially simpler since it can thereby be expressed as a product sum. However, the essence of time series analysis is to quantify serial dependence over time, which in other words means that the analysis is based on the fact that the observations are not independent. This gives grounds for an approach utilizing conditioning when analyzing the serial dependency of $\{x_1, x_2, \ldots, x_T\}$, a random sample of (negative) log-returns X including T observations.

This makes it possible to formulate the joint probability density function as:

$$f(x_t, x_{t-1}, \ldots, x_1; \theta) = f(x_t | x_{t-1}, \ldots, x_1; \theta) \ldots f(x_2 | x_1; \theta) \cdot f(x_1; \theta)$$

By considering the observations $\{x_1, x_2, \ldots, x_T\}$ as fixed the joint density functions parameter θ is the only parameter(s) which are allowed to vary freely. This means that the general likelihood function $L(\theta; X)$ may be expressed as:

$$L(\theta; X) = \prod_{t=1}^{T} f(x_t | x_{t-1}, \ldots, x_1; \theta).$$

In practice, it is often more convenient to work with the logarithm of the likelihood function called the log-likelihood or the more recently established term loglihood function:

$$l(\theta; X) = \sum_{t=1}^{T} \log f(x_t | x_{t-1}, \ldots, x_1; \theta).$$

In this thesis we make great use of the AR(1)-GARCH(1,1) model, and therefore we think the derivation of the maximum likelihood estimates of the GARCH(p,q) model is motivated. For a similar derivation in the case for the parameters of the ARCH(p) model see Tsay (Tsay 2010, pp. 120).

If we are to find an expression for the parameter vector θ including the estimated parameters in a GARCH(p,q) model (presented in section 4.2.2) with a further assumption that the conditional distribution of the time series z_t is standard normal. Then we may write the joint conditional density function as:

$$f(z_t|z_{t-1},\dots,z_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}}$$

Knowing that $\epsilon_t = \sigma_t z_t$, we may formulate the conditional likelihood function of ϵ_t as:

$$f(\epsilon_t|\epsilon_{t-1},\dots,\epsilon_0) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\epsilon_t^2}{2\sigma_t^2}}$$

Therefore the conditional loglihood function for the parameter vector θ is

$$\begin{aligned} l(\theta;\epsilon_{t-1},\epsilon_{t-2},\dots,\epsilon_0) &= \sum_{t=q+1}^{T} \log f(\epsilon_t|\epsilon_{t-1},\dots,\epsilon_0) \\ &= \sum_{t=q+1}^{T} \log \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\epsilon_t^2}{2\sigma_t^2}} \\ &= \sum_{t=q+1}^{T} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2} \right] \end{aligned}$$

The ML-estimates are then obtained by differentiating the conditional loglihood function with respect to the parameter(s) of interest, and then set equal to zero (simultaneously).

10.5 An illustrating example by means of simulation

In order to facilitate for those readers who are not familiar with time series analysis and Value at Risk we decided to explicitly show how to estimate the one-day ahead 95% VaR with an AR(1)-GARCH(1,1) model with normally distributed innovations. In this brief example we simulate 2000 observations from a corresponding model by using a `for` loop along with the `garchSim()` command from the `fGarch` package, and iterate this 50 times. At the same time we examine whether our model manages to produce VaR estimates that holds the expected amount of observations.

When simulating we fix the model parameters so that ω is close to zero, $\alpha = 0.08$, $\phi = 0.10$, $\beta = 0.91$. Since we simulate from the same type of model that will be applied, we expect that the observed coverage rate will be close to the theoretical one.

We plot the results stored in the vector `cover_vec` in figure 11 and the reader may produce their own results by implementing the code that follows after loading the relevant packages.

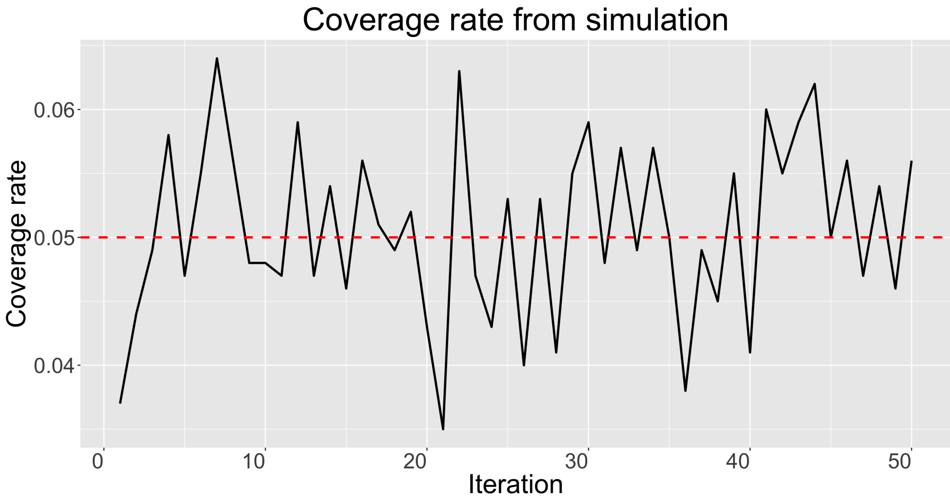

Figure 11: Results of coverage rates from the AR(1)-GARCH(1,1) model with normal innovations fitted on simulated data.

As seen in figure 11 the coverage proportions received from our time series model is close to the theoretical one which is illustrated by the dashed red horizontal line. We conclude this example by preforming a one sample t-test where the test and alternative hypothesis may be formulated as:

$$H_0 : \text{Estimated coverage proportions} = 0.05$$
$$H_1 : \text{Estimated coverage proportions} \neq 0.05$$

We calculate the observed test statistic t_{obs} by constructing the following ratio using the sample size n, the sample mean $\bar{X}$, the sample standard deviation s and the expected value stipulated in the null hypothesis μ.

$$\begin{aligned} t_{obs} &= \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1) \\ &= \frac{0.05066 - 0.05}{0.00692/\sqrt{50}} \approx 0.67487 \sim t(50-1) \end{aligned}$$

The corresponding p-value is 0.503 which means that we may not reject the null hypothesis at any conventional significance level. This is reassuring since the code used for model making in the previous analysis closely resembles the code presented below.

```
k <- 50                    # We chose to run the loop 50 times
nObs <- 2000               # Since we simulate 2000 obs.
window <- 1000             # The rolling window of 1000 obs.
quantileLevel <- 0.95      # Chosen VaR coverage level

# Empty vector for storage of VaR estimates.
VaR_vec <- rep(0, (nObs-window))
# Empty vector for storage of coverage rates.
cover_vec <- rep(0,k)

for (j in 1:k){
  # Specs for simulated data.
  spec = garchSpec(model = list(omega = 1e-6, alpha = 0.08, beta = 0.91,
  ar = 0.10), cond.dist = "norm")
  # Simulate 2000 data points.
  data_sim <- c(garchSim(spec, n = nObs, n.start = 1000))

  for (i in 1:(nObs-window)){
    # The rolling window of 1000 observations.
    data_insert <- data_sim[(i):(i+ 999)]
    # Fitting an AR(1)-GARCH(1,1) model with normal cond.dist.
    fitted_model <- garchFit(~ arma(1,0) + garch(1,1), data_insert,
                             trace = FALSE, cond.dist = "norm")
    # One day ahead forecast of conditional mean and standard deviation.
    prediction_model <- predict(fitted_model, n.ahead = 1)
    mu_pred <- prediction_model$meanForecast
    sigma_pred <- prediction_model$standardDeviation
    # Calculate VaR forecast
    VaR_vec[i] <- mu_pred + sigma_pred*qnorm(quantileLevel)

    if ((nObs-window)-i != 0){
      print(c("Countdown, just",((nObs-window) - i),"iterations left"))
    } else {
      print(c("Done!"))
    }
  }

  # Exctract only the estiamtes ralated to the forecasts.
  compare_data_sim <- data_sim[(window + 1):length(data_sim)]
  # Create an empty vector used in order for storage of VaR exceedances.
  hit <- rep(0,length(VaR_vec))
  # Count the amount of exceedances.
  for (i in 1:length(VaR_vec)){
    hit[i] <- sum(VaR_vec[i] <= compare_data_sim[i])
  }
  # Calculate the covered proportions for the k:th run of the loop.
  cover_vec[j] <- sum(hit)/length(hit)
}
```