UNIVERSITY OF AGDER

The performance of Value-at-Risk models on the OBX index

Dan Wiggo Jore

Supervisor
Steen Koekebakker

This Master's Thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education. However, this does not imply that the University answers for the methods that are used or the conclusions that are drawn.

University of Agder, 2013
Faculty of Economics and Social Sciences
Department of Economics and Business Administration

Abstract

The measuring of risk has become one of the main fields in finance during the last two decades. Value-at-Risk (VaR) has become one of the most important risk measures and is widely used for numerous applications. This thesis compares different approaches to VaR based on traditional methods such as Historical Simulation, Moving Average and Exponentially Weighted Moving Average as well as advanced approaches based on GARCH models. Comparison is done on the OBX index return data, which is the main benchmark index on the Oslo Stock Exchange. The performance of the different VaR models is evaluated with out of sample backtests over two periods of changing market conditions. The first period is the crisis period with high volatility and market uncertainty that covers the financial crisis in 2008. The second period is the post crisis period after the financial crisis that has more normal market conditions.

Our findings are that traditional VaR methods do not capture the risk of the OBX index. The models tend to underestimate the risk when the market goes through a crisis and generally perform poorly. Several of the VaR models based on GARCH dynamics perform quite well and overall the best model is the skew Student-t GARCH(1,1) which is not rejected in any backtest and therefore captures the risk in both the crisis period and the post crisis period. The model also outperforms sophisticated GARCH models that are able to capture asymmetries in volatility and power effects. The choice of error distribution for the GARCH models is also found to be very important. Changing the normal error distribution to the skew Student-t distribution significantly improves the forecasting performance of the GARCH models.

Acknowledgements

This thesis concludes my Master of Business Administration at the University of Agder (UIA). The learning environment at UIA has been great throughout the whole Master´s programme. All my fellow students and lecturers have been very supportive and encouraging.

I would like to thank my supervisor Professor Steen Koekebakker for his useful comments, remarks and guidance through the learning process of this master thesis. Especially his first hand experience with the R programming language has been very helpful and saved me from a lot of trouble.

I would also like to thank Silve-Linn Aase for reading through my thesis and for her comments and remarks.

Dan Wiggo Jore

Kristiansand

May 2013

Table of contents

1 Introduction

1.1 Background

Measuring of market risk has become one of the main fields in finance as it has become increasingly important for regulatory purposes and management decisions. Market risk arises from unexpected changes in market prices and can be classified into interest rate risk, exchange rate risk, equity risk, commodity risk and so on. The recent financial crisis in 2008 with the following fluctuations in equity, commodity and property prices emphasize the importance of correctly measuring the risk which corporations, individual investors and even countries are exposed to.

The standard measure of market risk is volatility and is defined as the standard deviation of returns. Volatility captures the risk of a financial asset if the returns are normally distributed because all the statistical properties of the normal distribution are described by the mean and standard deviation. However, it is generally known that financial returns are not normally distributed. Therefore, assuming that the returns are normally distributed could lead to an underestimation of the risk associated with a financial asset. Another common risk measure is the Value-at-Risk (VaR) approach, which is distribution independent. It focuses at estimating the potential loss given a probability level that losses are equal or exceed the VaR. Although the definition of VaR is broadly the same, there is no general consensus among either researchers or practitioners on how VaR should be calculated. This has lead to a development of an enormous amount of different models. There are two main methods for calculating VaR: Non-parametric and parametric. Most of the methods used are parametric volatility models that estimate the underlying distribution of an asset returns. The parametric volatility model is used to forecast the volatility over the risk horizon from which the VaR forecast can be obtained. There are many volatility models available today and the majority of these models are in the family of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. The GARCH models are able to quickly adapt to changing market conditions and can also capture the most common characteristics of financial return series.

The history of VaR started in the late 1970s and 1980s as major financial institutions needed tools to aggregate the total risk across the institution as a whole. As the institutions grew and became more complex, the need to accurately aggregate the risk became increasingly difficult. The institutions started to impose arbitrary restrictions that limited the traders and asset managers, which resulted in sub-optimal decision making since investment opportunities that could decrease the overall risk were passed. There was also little connection between the risk that was actually taken and the limits that were imposed. The best known system that evolved from this era was the RiskMetrics system developed by JP Morgan, which was intended to give the management a daily one-page report that summarized the overall risk of the company over the next day. This report was given to the management 4:15 each day after the closing of the markets and was therefore called the "4:15 report". The system was based on the VaR approach and was built on the traditional portfolio theory of standard deviations and correlations estimates between the returns of the different assets in their portfolio. Around 1990 the system was up and running internally in JP Morgan, and it was soon discovered that the system had very positive qualities that gave the management a tool to make more efficient investment decisions. In 1994, JP Morgan decided to make the RiskMetrics system available for free enabling outside investors and institutions to use the system and incorporate it as they wished. This lead to a rapid adoption and development of the VaR framework in the 1990s, which resulted in VaR being the risk measure of choice by most institutions today.

1.2 Previous research

There is a vast amount of research done on the forecasting performance of volatility and VaR models. A lot of theses studies provide results that contradict each other. Below we present two studies of this kind.

Hansen and Lunde (2005) published the paper "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?". They studied 330 different GARCH model for their ability to forecast one-day ahead conditional variance for an out of sample period of about 250 days. The models were applied to the DM/USD rate and the IBM stock price returns, and the models were tested for their forecasting performance between 1992/1993 for the DM/USD currency, and 1999/2000 for the IBM stock. They found that none of the sophisticated models performed any better than the standard GARCH model with normal error distribution for the DM/USD rate, while they found that models allowing for leverage effects could perform better for the IBM stock.

More recently, Ghalanos (2013a) published an article on his webpage www.unstarched.net where instead of asking if any model could beat the standard GARCH(1,1), he asked if anything does NOT beat the GARCH(1,1). He used a range of tests to make a comparison of different VaR models based on GARCH. The data used is the S&P500 index and the models are tested for one-day out of sample performance over 1500 days from 2007 to 2013. In his tests he found that the standard GARCH(1,1) was actually not hard to beat at all, in fact it was one of the worst performing models. He argues that the normality assumption of the standard GARCH(1,1) does not realistically capture the observed market movements.

1.3 Purpose

The main inspiration for this thesis has been drawn from the studies of Hansen and Lunde (2005) and Ghalanos (2013a). The purpose is to evaluate different VaR models on the OBX index return history, which is the main benchmark index on the Oslo Stock Exchange. Even if the studies presented above contradict each other to some extent, this could relate to differences in the assets that have been studied and the testing framework. Different assets have different price dynamics, and therefore it is likely that some models will perform better for certain assets and worse for others. With this in mind it is important to validate a risk model for the actual assets that it is supposed to be applied to. There are also very few studies done on Norwegian data that are freely available. This study will focus on the one-day forecasting performance of VaR models based on the non-parametric Historical Simulation method and the parametric Moving Average, Exponentially Weighted Moving Average and GARCH methods.

1.4 Outline

In the next chapter we will focus on the theory of financial returns and the most common characteristics that are present in financial returns. We will also identify these characteristics in the returns of the OBX index. The third chapter will cover volatility modeling and present the models that are used with their main properties. It will also cover in sample fit diagnostics for volatility models based on GARCH that is applied to our OBX index data. Chapter 4 gives the theory of VaR, how it is applied and the strengths and weaknesses of the VaR measure. Chapter 5 will cover the methodology that is used for evaluating the forecasting performance of VaR models with backtesting methods. The models are backtested over two time periods of the OBX index returns that have very different market dynamics. In chapter 6 we will discuss our empirical results. Finally, chapter 7 will present our most important findings and conclusions.

2 Data description and return characteristics

2.1 Financial returns and characteristics

Most financial analyses involve returns rather than prices of assets. Campbell, Lo, and MacKinlay (1997) give two main reasons why this is the case:

1. The return of an asset is a complete and scale-free summary of the investment opportunity.
2. Return series of assets have more attractive statistical properties and are easier to handle than asset price series.

There are several definitions of asset return, but we will use the continuously compounded returns in this thesis. One advantage of continuously compounded returns is that the return over a period of time is simply the sum of the single-period returns in the period. Handling of the time-series is also easier when using continuously compounded returns. The definition of continuously compounded return Y_t is:

$$Y_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log P_t - \log P_{t-1} \tag{2.1}$$

where log is the natural logarithm and P_t is the asset price series.

Extensive research of financial returns has shown some stylized characteristics that are present in almost all financial return series:

1. Non-normality
2. Volatility clustering

Non-normality

The non-normality of financial return series has been known since Mandelbrot (1963) studied the price of cotton. Today financial return series are generally regarded to be leptokurtic, also known as positive excess kurtosis, which means that they have fatter tails and excess peakedness at the mean. The fat tails implies that the market has more large and small return outcomes than one would expect if the returns were truly normal distributed. Excess peakedness indicates that there are more days when little occurs in the market than indicated by the normal distribution. There is also a wide range of literature showing that financial return distributions are not symmetric but are in fact skewed. Several studies have found that indices are usually negatively skewed and that individual assets are positively skewed, more recently studied by Albuquerque (2012).

There are two main methods for identifying the non-normality of financial returns: Graphical methods and statistical tests. With statistical tests you compare the observed returns with a base distribution, usually the normal distribution. Graphical methods use values predicted from some distribution and compares it to the observed returns. A typical statistical test for identifying non-normality is the Jarque-Bera test (Jarque & Bera, 1987), where the skewness and excess kurtosis of the sample distribution is significantly different from zero is tested. The Jarque-Bera test statistic is defined as:

$$JB = \frac{T}{6}\left(\widehat{skew}^2 + \frac{(\widehat{kurt} - 3)^2}{4}\right) \tag{2.2}$$

where the $\widehat{skew}$ and $\widehat{kurt}$ is the sample skewness and sample kurtosis. Reject the null hypothesis of normality if: $JB > \chi^2_{\alpha,2}$.

A typical graphical method used to identify non-normality is the Quantile-Quantile plot. Unlike the Jarque-Bera test, which can only test for non-normality, the QQ plot can be used to assess if a dataset has any specific distribution. The QQ plot compares the quantiles of a reference distribution to the quantiles of the sample data. If the sample data and the reference distribution are distributed approximately the same, the sample data should lie on the 45-degree reference line in the QQ plot. Plotting the observed returns in a histogram is also a common approach to check the distribution of the returns. If the returns are distributed according to a certain distribution, the histogram should follow the curve of the distribution.

Volatility clustering

The tendency that volatility in financial returns appears in bursts is called volatility clustering. High returns (positive or negative) are more likely to be followed by high returns, and low returns (positive or negative) are more likely to be followed by low returns. In other words the volatility tends to be high in some periods and low for other periods. For example in the mid 1990s volatility was very low in most markets, but it increased in the last part of the decade due to among other things, the Asian crisis. Since Engle (1982) published his work on the

Autoregressive Conditional Heteroskedasticity (ARCH) model, the volatility clustering has been widely considered a feature observed in most financial return series.

A standard graphical method to identify volatility clustering is the autocorrelation function (ACF). The ACF is used to measure how squared returns on one day are correlated with squared returns on the previous days. If the correlations are significant, we have strong evidence for volatility clustering in the data. The reason for studying squared returns is that they are proxies for volatility and are used in most volatility forecast models. The definition of ACF of squared returns: $\rho_{t,i} = \frac{E[(y_t^2 - \mathrm{E}(y_t^2))(y_{t-i}^2 - \mathrm{E}(y_{t-i}^2))]}{Var(y_t^2)}$ where $\rho_{t,i}$ is the ith autocorrelation.

A statistical method to test for volatility clustering is the Ljung-Box test, which is built on the ACF. It tests for overall randomness based on a number of lags. Under the null hypothesis all autocorrelations for the lags included are zero. The Ljung-Box Q statistic is defined as:

$$Q = T(T+2)\sum_{i=1}^{m}\frac{\rho_{t,i}^2}{T-i} \tag{2.3}$$

where m is the number of lags included. Reject the null of no autocorrelation if: $Q > \chi^2_{\alpha,m}$.

2.2 OBX return characteristics

We will now study the characteristics of the OBX index returns and identify the most important features that are present in financial time series. The complete historical data has been provided by the Oslo Stock Exchange and consists of the daily closing price from 02.01.1996 to 31.01.2013. The OBX index is based on the 25 most traded securities listed on the Oslo Stock Exchange and is a tradable index with ETFs and options available. The index is free-float adjusted along with its compositions semiannually. Figure 2.1. plots the OBX closing price and log returns. Looking at the plots we easily identify the financial crisis in 2008 with a sharp decline in the index value and very high volatility. We also see that there is a clear pattern of volatility clustering, as the volatility seems to appear in bursts.

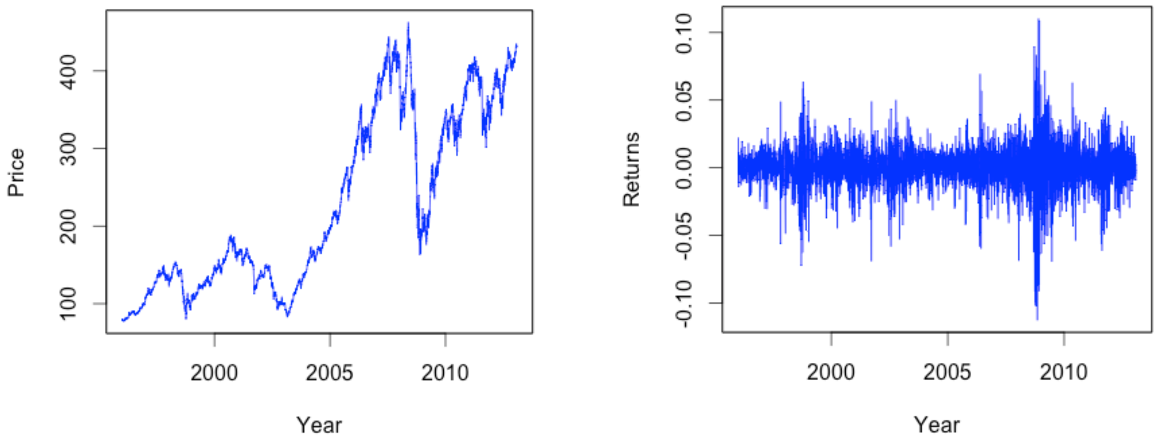

Figure 2.1. OBX index closing price and log returns 02.01.1996 - 31.01.2013

Table 2.1. OBX return statistics 02.01.1996 - 31.01.2013

Mean	0.040 %
Standard deviation	1.601 %
Min	-11.273 %
Max	11.020 %
Skewness	-0.543
Kurtosis	9.050
ACF returns, one lag	0.004
ACF squared returns, one lag	0.227
Jarque-Bera test	p = 0.000
Ljung-Box test returns, 20 lags	p = 0.018
Ljung-Box test squared returns, 20 lags	p = 0.000

From Table 2.1. we see that the returns of the OBX index have a daily mean of 0.04% and a daily volatility of 1.60%. Thus the daily mean is only about one-fortieth of daily volatility. The minimum and maximum values lies far from what the normal distribution could predict and the kurtosis is also very high indicating that the returns are leptokurtic. We also notice that the distribution of returns is negatively skewed. This is confirmed by the Jarque-Bera test that rejects the null hypothesis that the returns are normally distributed. A graphical examination of the returns in Figure 2.2. also supports our findings. In the normal QQ-plot we immediately detect an S-shape of the sample data. This indicates that the sample data has heavier tails than the normal distribution. The leptokurtic feature of the data is easily detected in the histogram in Figure 2.2. as the data lie outside the normal curve in the tails and in the center of the distribution. Taking all of the evidence presented above into consideration the returns of the OBX index is very unlikely to be normally distributed.

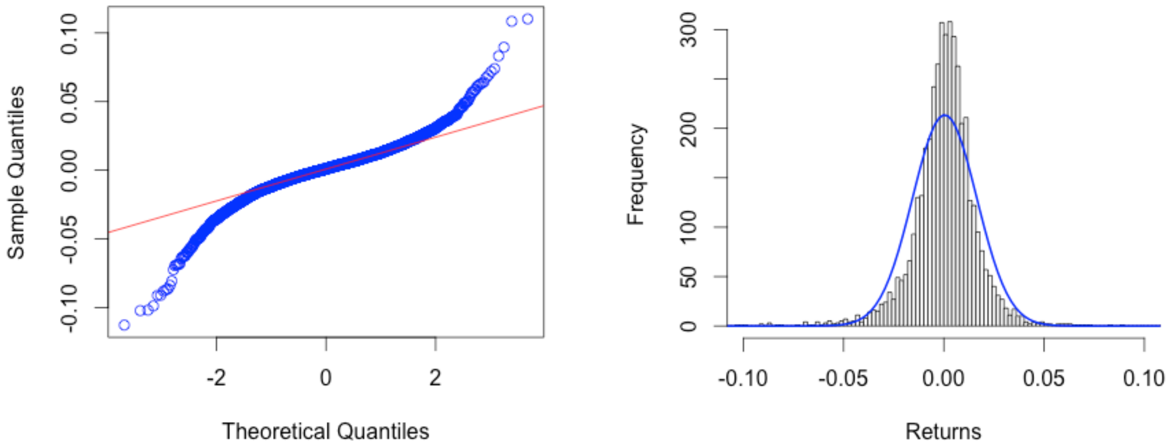

Figure 2.2. OBX log returns normal QQ-plot and histogram 02.01.1996 - 31.01.2013

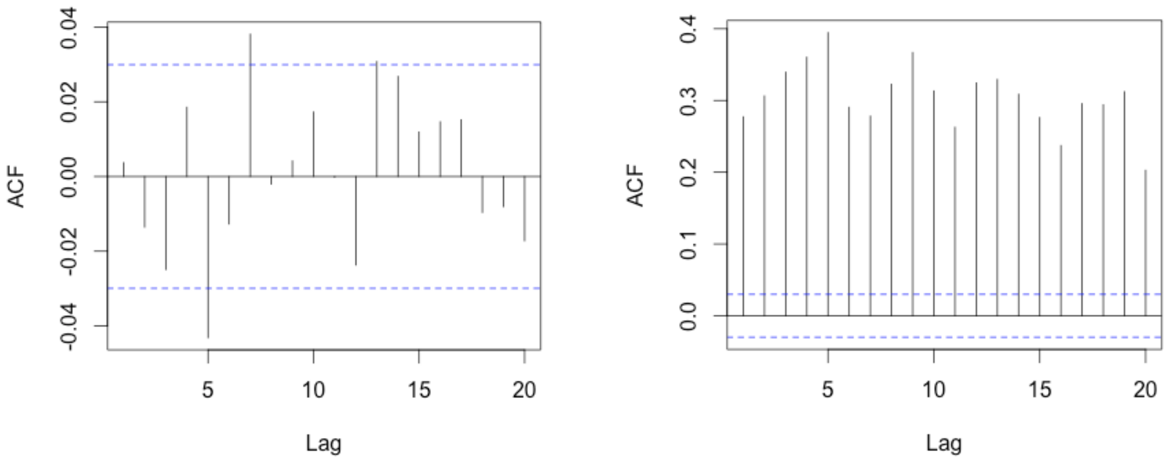

Figure 2.3. OBX ACF of returns and squared returns 02.01.1996 - 31.01.2013

From Table 2.1. we see that there is significant correlation between squared returns and lagged squared returns. The Ljung-Box test of no volatility clustering is also rejected. In Figure 2.3. the ACF of returns and squared returns are plotted. Most of the correlations of returns lie inside the confidence interval while all the correlations of squared returns lie outside the confidence interval. Overall we have found significant evidence that the OBX index has volatility clustering.

To evaluate the VaR models in this thesis we are testing the performance of the models in the period from 21.02.2005 to 11.02.2009 and the period from 12.02.2009 to 31.01.2013. The first period covers the financial crisis in 2008 and therefore represents market conditions that are extreme and is referred to as the crisis period. The second period is the period after the crisis period and represents market conditions that are more stable and is therefore referred to as the post crisis period. Table 2.2. prints some statistics from each period.

Table 2.2. OBX return statistics crisis and post crisis

Statistic	Crisis	Post crisis
Mean	-0.225 %	0.073 %
Standard deviation	3.433 %	1.660 %
Min	-11.273 %	-6.949 %
Max	11.020 %	7.189 %
Skewness	-0.381	-0.171
Kurtosis	4.710	4.587
ACF returns, one lag	0.024	-0.053
ACF squared returns, one lag	0.188	0.131
Jarque-Bera test	p = 0.000	p = 0.000
Ljung-Box test returns, 20 lags	p = 0.275	p = 0.305
Ljung-Box test squared returns, 20 lags	p = 0.000	p = 0.000

In Table 2.2. we see that the mean return in the crisis period is extremely low at -0.225%. This relates to an average yearly return of $-0.225\% * 250 = -56.3\%$. The volatility is also very high at 3.433%, which implies that the average yearly volatility is $3.433\% * \sqrt{250} = 54.3\%$. If we do the same calculations for the post crisis period we get an average yearly return of 18.3% and an average yearly volatility of 26.2%. We also notice that both periods have non-normal return characteristics and that they are rejected by the Ljung-Box test for no volatility clustering. Finding a risk model that is able to capture the risk in both these periods is challenging since the market dynamics is very different. We will have to find a model that is able to quickly adapt to new market conditions.

3 Volatility modeling

One of the most important developments in empirical finance has been the modeling of volatility, and it has been one of the main subjects for academics and practitioners for the past two decades. The reason is that volatility is often used as a standard measure of risk for financial assets, and plays an important role in asset allocation under the mean-variance framework. Also the pricing of options heavily relies on volatility forecasts as the Black and Scholes formula uses volatility as one of its main inputs. In most VaR applications volatility models have to be calculated to estimate a risk forecast.

The definition of volatility is the standard deviation of returns. A volatility forecast at time t, σ_t, is typically obtained from a statistical model, $\sigma(\cdot)$, that uses an estimation window, W_E, which contains a sample of historical observations of returns. The volatility forecast at time t is:

$$\sigma_t = \sigma\left(y_{t-1}, y_{t-2}, \dots, y_{t-W_E}\right) \tag{3.1}$$

For the volatility models that are in the class of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models the innovations in returns is driven by random shocks Z_t, which is an independent and identically distributed (IID) random variable with zero mean and unit variance. The return on day t is then defined as:

$$Y_t = \sigma_t Z_t \tag{3.2}$$

The distribution of Z_t is typically assumed to be normal, but it can also be changed to a distribution that will fit the data better. We will cover the Student-t distribution that will allow for higher kurtosis, and the skew Student-t distribution that will allow for higher kurtosis and more skewness than the standard normal process.

A central feature of volatility is that it is not directly observable, which makes modeling difficult. Volatility has to be inferred from the observable market prices by looking at the price movements from day to day. The presence of non-normality and volatility clustering makes volatility modeling even harder. This has lead to the development of a rich family of volatility models with different advantages and weaknesses. We need to separate volatility into two different concepts: Unconditional volatility and conditional volatility. Unconditional volatility is simply defined as the historical volatility. It is unconditional since it does not depend upon all the information that is available today and volatility forecasts will not change through time. Conditional volatility takes into account all the information that is available today. Since we know that volatility changes in time, and that financial return series have structures like volatility clustering, this can be used to improve forecasting. Conditional volatility models normally have an unconditional volatility, which the model will revert to as the forecasting horizon increases. This is a good feature for a volatility model, as it is generally known that volatility time series are mean reverting.

To ensure that the unconditional volatility is defined we usually impose a covariance stationarity condition for conditional volatility models. If a volatility model is not covariance stationary it will have highly undesirable properties. For instance, if we try to forecast volatility with a non-stationary volatility model, the volatility forecast will explode as the forecast horizon is increased. Volatility can also never by definition be negative, and a negative volatility forecast would also make little sense. Therefore it is usually imposed a non-negativity constraint on conditional volatility models to ensure that all future volatility forecasts will be positive.

We will assume that the mean return is zero. While this is obviously not precise, the daily mean is usually very close to zero, and compared to the daily volatility it is relatively insignificant. As presented in Table 2.1. the daily average the return of the OBX index is only 0.04% while the daily volatility is 1.6%. Therefore the mean can be safely ignored in most volatility forecast models without losing any significant forecasting power.

In the next sections we will first introduce simple volatility models, and then introduce conditional volatility models where the parameters are estimated by using maximum likelihood.

3.1 Moving Average

The easiest method to forecast volatility is to calculate the sample standard error from the return sample. We keep the sample size at the same level, and every day we add a new return to the sample we drop the oldest. This is the Moving Average (MA) model, also known as the historical volatility. There are no parameters to estimate in the MA-model and it is also very simple to use and understand. The MA model is defined as:

$$\hat{\sigma}_t^2 = \frac{1}{W_E}\sum_{i=1}^{W_E} y_{t-i}^2 \tag{3.3}$$

where $\hat{\sigma}_t^2$ is the volatility forecast for day t, y_t is the return on day t and W_E is the length of the estimation window.

A limitation of the MA model is that all the observations are equal weighted, or in other words the first observation has equal impact on the volatility forecast as the last observation. Since we know that financial return series have volatility clustering and that the last observations will indicate if we are in a high or low volatility period, equal weighting is problematic. Consequently the MA model is very sensitive to the estimation window length. If the estimation window is too long, the volatility forecasts will be very sluggish. If the estimation window is too short, the volatility forecasts will jump around.

3.2 EWMA

A further development of the MA model is the Exponentially Weighted Moving Average (EWMA) model that was originally developed by J.P. Morgan (1993) and was given the name RiskMetrics. The model is based on the MA model but it assigns higher weights to the latest observations. The weights exponentially decline into the past. The EWMA is defined as:

$$\hat{\sigma}_t^2 = \frac{1-\lambda}{\lambda(1-\lambda^{W_E})}\sum_{i=1}^{W_E} \lambda^i y_{t-i}^2 \tag{3.4}$$

which can be rewritten as the weighted sum of the previous period´s volatility forecast and squared returns:

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2 \tag{3.5}$$

The λ is known as the decay factor and is restricted to be between zero and one. The larger the decay factor is the smoother the volatility estimates will be. J.P. Morgan originally proposed that the decay factor should be set to 0.94 for daily returns.

A disadvantage of the EWMA model is that the decay factor is a constant, and if we have multiple assets in a portfolio it has to be the same for all assets. It is unreasonable to assume that all assets have the same decay factor and that it is constant in time. Unlike the GARCH models presented later, the EWMA model does not revert to the unconditional volatility as the forecast horizon increases. But EWMA is simple and can be implemented very easily compared to more sophisticated models. An advantage of EWMA is that as long as the estimation window length is not very short, the model is indifferent to the estimation window length. The weighting of the observations will make sure that very distant observations will not have a significant impact on the forecasts.

3.3 ARCH

Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model which was the beginning of a systematic framework for volatility modeling. This was the first model that was able to capture the volatility clustering typically observed in financial time series. The ARCH(p) model is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 \tag{3.6}$$

where p is the number of lagged squared returns included. ω and α_i are the model parameters that have to be estimated. Setting $p = 1$ results in the ARCH(1) model:

$$\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 \tag{3.7}$$

This states that the conditional volatility depends on a constant and one lagged squared return. It will therefore only pick up the dependence from the previous day squared return. The problem is that to capture all the dependence in the conditional volatility one would usually have to include a very high number of lags in the ARCH model. If we take a look at Figure 2.3. we see that there is significant autocorrelations for all the 20 first squared returns. We would have to include at least 20 lagged squared returns in the model to capture all the clustering in the OBX volatility. When more parameters are introduced the estimation will become trickier and the risk of estimating a negative volatility increases. The conditional volatility is changing, but the unconditional volatility is constant and is given by: $\sigma^2 = \frac{\omega}{1-\sum_{i=1}^{p}\alpha_i}$. To make sure that the unconditional volatility is defined we will impose a parameter restriction to ensure covariance stationarity.

ARCH(p) parameter restrictions

In the estimation of ARCH models usually it is imposed restrictions on the parameters:

1. To ensure non-negative conditional volatility forecast: $\forall\ i = 1, \dots, p,\ \alpha_i, \omega > 0$
2. To ensure covariance stationarity: $\sum_{i=1}^{p}\alpha_i < 1$

3.4 GARCH

The Generalized Autoregressive Conditional Heteroskedastic (GARCH) model was developed independently by Bollerslev (1986) and Taylor (1987) in the middle of the 1980s. It generalized the ARCH model by allowing the conditional volatility to be dependent upon its own lags. This resulted in the GARCH(p,q) model:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p}\alpha_i Y_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \tag{3.8}$$

where p is the number of lagged squared return included and q is the number of lagged volatilities included. ω, α_i and β_j are the model parameters that have to be estimated. Setting p=1 and q=1 results in the GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{3.9}$$

The GARCH(1,1) has only three parameters but it can be shown that the model can be rewritten as an ARCH(∞). This allows the GARCH model to be influenced by an infinite number of past squared returns when estimating the conditional volatility. The GARCH model is a more parsimonious model than the ARCH model and therefore it is less likely to breach the non-negative constraint. The unconditional volatility for the GARCH(1,1) is: $\sigma^2 = \frac{\omega}{1-\alpha_1-\beta_1}$.

GARCH(1,1) parameter restrictions

As for the ARCH model there are two parameter restrictions imposed:

1. To ensure non-negative conditional volatility forecast: $\omega, \alpha_1, \beta_1 > 0$
2. To ensure covariance stationarity: $\alpha_1 + \beta_1 < 1$

3.5 EGARCH

Standard GARCH models are able to capture volatility clustering and non-normal returns, but there are also other features that are typically found in financial time series. Black (1976) discovered, and later confirmed in many studies, that financial returns are likely to be negatively correlated with changes in volatility. That is, a negative shock to a financial return series is likely to increase volatility more than a positive shock of the same magnitude. This has been called the leverage effect since one explanation of its existence is that when a leveraged firms stock price fall, its debt to equity ratio increases. When the debt to equity ratio increases, the shareholders, who bear the residual risk of the firm, view their future cash flows to be more risky.

The Exponential GARCH (EGARCH) model was developed by Nelson (1991) and is a popular model that is able to incorporate the asymmetries in how volatility reacts to past returns. The general EGARCH(p,q) can be written as:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} (\alpha_i Z_{t-i} + \gamma_i(|Z_{t-i}| - \mathrm{E}|Z_{t-i}|)) + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2) \qquad (3.10)$$

Setting $p = 1$ and $q = 1$ results in the EGARCH(1,1) model:

$$\log(\sigma_t^2) = \omega + \alpha_1 Z_{t-1} + \gamma_1(|Z_{t-1}| - \mathrm{E}|Z_{t-1}|) + \beta_1 \log(\sigma_{t-1}^2) \qquad (3.11)$$

where γ_1 is the parameter that picks up the size effect of the asymmetries and α_1 picks up the sign effect.

An advantage of the EGARCH model compared to the standard GARCH model is that since it is the logarithm of the conditional volatility that is modeled, there is no need to impose non-negativity constraints because even if the parameters are negative the conditional volatility will still be positive. EGARCH is also attractive since it is closely related to continuous time finance. As shown by Ghalanos (2013b) the unconditional volatility of EGARCH(1,1) is: $\sigma^2 = \frac{\omega}{1-\beta_1}$.

EGARCH(1,1) parameter restrictions

The conditions for covariance stationarity was shown by Nelson (1991). The parameter restrictions are:

1. There is no need to impose any restrictions to ensure non-negative volatility forecast.
2. Covariance stationary if: $|\beta_1| < 1$

Engle and Ng (1993) proposed the sign and size bias tests for asymmetries in volatility. The tests are used to check if a symmetric model is adequate or if an asymmetric model is required. The sign and size bias test is a joint test that tests for sign bias, where positive and negative shocks have different impacts upon future volatility, and also size bias that tests if the magnitude of the shocks are important. The Engle and Ng sign and size bias test is usually applied to the standardized residuals of a GARCH model. The test is based on the regression:

$$\hat{z}_t^2 = \phi_0 + \phi_1 I_{\hat{z}_{t-1}<0} + \phi_2 I_{\hat{z}_{t-1}<0}\hat{z}_{t-1} + \phi_3 I_{\hat{z}_{t-1}\geq 0}\hat{z}_{t-1} + v_t \tag{3.12}$$

where v_t is an IID error term and I is the indicator function. If ϕ_1 is significant, it indicates that there is sign bias. If ϕ_2 or ϕ_3 is significant, it indicates that there is size bias. Under the null hypothesis of no asymmetries in volatility the joint test statistic is TR^2, where T is the number of observations and R^2 is the coefficient of determination from the regression. The test statistic will follow a χ^2 distribution with 3 degrees of freedom. Table 3.1. shows the sign and size bias test on the OBX returns. The tests show that there are significant asymmetries in the whole sample and the post crisis period. But in the crisis period there little evidence for asymmetries. This could indicate that an asymmetric model is not needed to capture the risk in the crisis period.

Table 3.1. OBX returns sign and size bias test

Statistic	Whole sample	Crisis	Post crisis
Engle and Ng sign and size bias test	p = 0.013	p = 0.733	p = 0.008

A graphical representation of the leverage effect is the news impact curve introduced by Pagan and Schwert (1990). The news impact curve plots the next day volatility that will arise from the return today given a specific volatility model. We have fitted the standard GARCH(1,1) and the EGARCH(1,1) to the whole sample and plotted two news impact curves of the models in Figure 3.1. From the plots we see that the symmetric GARCH model gives the same volatility forecast for a return of a given magnitude whatever its sign. In contrast, the asymmetric model will give a higher volatility forecast if the return is negative than if it is positive of the same magnitude.

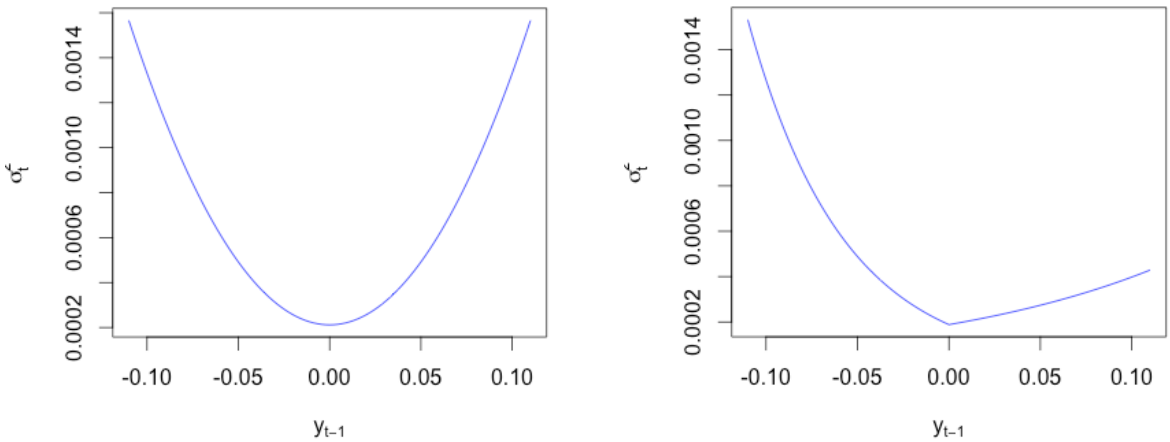

Figure 3.1. OBX News impact curves GARCH(1,1) and EGARCH(1,1) whole sample

3.6 gjrGARCH

A similar model to the EGARCH model is the gjrGARCH model that is also able to capture asymmetries in volatilty. It was developed by Glosten, Jagannathan, and Runkle (1993). The gjrGARCH is an extension of the standard GARCH model with an extra term added to model any asymmetries. The added term has an indicator function that takes the value 1 if the return is less or equal to zero. The general gjrGARCH(p,q) model can be written as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p}(\alpha_i Y_{t-i}^2 + \gamma_i I_{t-i} Y_{t-i}^2) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3.13}$$

Setting $p = 1$ and $q = 1$ results in the gjrGARCH(1,1):

$$\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 + \gamma_1 I_{t-1} Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.14}$$

where $I_{t-1} = 1$ if Y_{t-1} is negative. Otherwise $I_{t-1} = 0$. As shown by Glosten et al. (1993) the unconditional volatility of gjrGARCH(1,1) is: $\sigma^2 = \frac{\omega}{1-\alpha_1-\beta_1-0.5\gamma_1}$.

gjrGARCH(1,1) parameter restrictions

The non-negativity and covariance has been shown by Glosten et al. (1993).

1. Non-negativity constraint: $\omega, \alpha_1 > 0$, $\beta_1 \geq 0$ and $\alpha_1 + \gamma_1 \geq 0$
2. Covariance stationary if: $\gamma_1 < 2(1 - \alpha_1 - \beta_1)$

3.7 APARCH

Taylor (1986) revealed that the ACF of absolute returns usually have stronger autocorrelations than for squared returns. This has been found in a large number of financial time series and was named the Taylor effect. Ding, Granger, and Engle (1993) later discovered that for different financial time series power transformations of absolute returns could give even higher autocorrelations. Generally any transformation of the absolute returns that lead to stronger autocorrelations is known as power effects. When forecasting volatility it is rational to include any transformation of returns that will increase the predictability of the model.

The Asymmetric Power ARCH (APARCH) model was developed Ding et al. (1993). APARCH is able to capture asymmetries, power effects and other structures in the data. The APARCH model is a sophisticated model that nests a wide range of other GARCH models. The general APARCH(p,q) model can be written as:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i(|Y_{t-i}| - \gamma_i Y_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta \tag{3.15}$$

Setting $p = 1$ and $q = 1$ results in the APARCH(1,1):

$$\sigma_t^\delta = \omega + \alpha_1(|Y_{t-1}| - \gamma_1 Y_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta \tag{3.16}$$

The model fits asymmetries when $\gamma_1 \neq 0$ and power effects when $\delta \neq 2$. Ding et al. (1993) showed that the unconditional volatility of gjrGARCH(1,1) is: $\sigma^2 = \left(\frac{\omega}{1-\beta_1-\alpha_1\mathbb{E}(|Z|-\gamma_1 Z)^\delta}\right)^{2/\delta}$. For further derivation of $\mathrm{E}(|Z| - \gamma_1 Z)^\delta$ refer to Ding et al. (1993).

apARCH(1,1) parameter restrictions

1. Non-negativity constraint: $\omega > 0, \delta \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0$ *and* $-1 < \gamma_1 < 1$
2. Covariance stationary if: $\alpha_1 \mathrm{E}(|Z| - \gamma_1 Z)^\delta + \beta_1 < 1$

In Figure 3.2. we have plotted two ACF of squared returns and absolute returns. It is difficult to identify if there is any Taylor effect from the plots. By estimating the normal APARCH(1,1) model δ parameter along with its standard error on the OBX returns we found that the parameter is significant and is very unlikely to be equal to 2 for the whole sample and the post crisis period. But for the crisis period the parameter is rejected for significance. In other words there seems to be power effects in the whole sample and the post-crisis. Table 3.2. lists the parameters and standard errors.

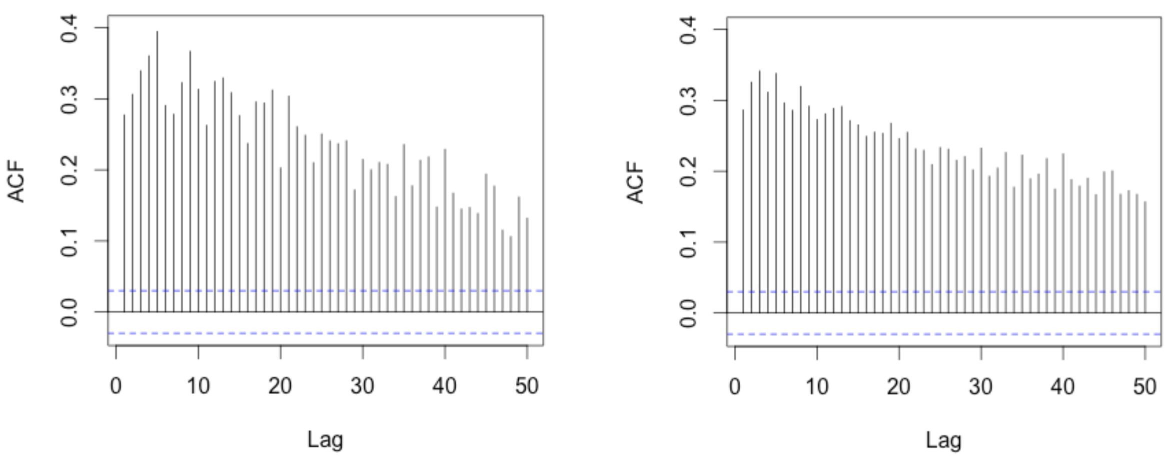

Figure 3.2. OBX ACF squared returns and absolute returns

Table 3.2. OBX normal APARCH(1,1) δ parameter estimation

Statistic	Whole sample	Crisis	Post crisis
δ parameter	1.137	0.663	1.252
Standard error	0.153	0.484	0.254

3.8 Error distributions

The random shocks Z_t can be modeled in different ways to make the GARCH models fit the data better. Extreme outcomes can be systematically underestimated if the GARCH models fail to capture the fat tails and skewness that are typically seen in financial return series. Z_t is usually assumed to be normal, but other distributions like the Student-t and skew Student-t distributions are also used. Even though a GARCH model with a normal error distribution can accommodate for return series that have fat tails and skewness, it is often the case that the observed returns has fatter tails and more skewness than the normal process allows for. Changing the normal error distribution to a Student-t or a skew Student-t distribution can help to improve the fit of the models and possibly offset the difficulty and cost of estimating additional parameters.

Normal distribution

The Normal distribution was originally used by Engle (1982) in the ARCH model for the error process $Z_t \sim N(0,1)$. The normal distribution is defined as:

$$f(y_t;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2}\right] \tag{3.17}$$

Any normal distribution can be rewritten as the standardized normal distribution scaled by the standard deviation, $Z_t = \frac{Y_t}{\sigma_t}$. The standardized normal distribution is defined as:

$$f(y_t;\mu,\sigma^2) = \frac{1}{\sigma}\left(\frac{exp\left[-\frac{1}{2}z_t^2\right]}{\sqrt{2\pi}}\right) = \frac{1}{\sigma}f(z_t) \tag{3.18}$$

Student-t distribution

Since it was observed that the normal GARCH usually had too thin tails for financial data, Bollerslev (1987) proposed the Student-t distribution for the error distribution to improve the GARCH model. $Z_t \sim t^*(0,1,v)$. The standardized Student-t distribution with zero mean and unit variance:

$$f(z_t;v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(1+\frac{z_t^2}{v-2}\right)^{-(v+1)/2} \tag{3.19}$$

where $\Gamma(\cdot)$ is the gamma distribution and v is the shape parameter. v has to be larger than 2 and as $v \to \infty$ we get back to the normal distribution. The lower v is the fatter the tails become.

Skew Student-t distribution

Fernández and Steel (1998) proposed an extension to the Student-t distribution to account for more skewness by adding a skewness parameter. Lambert and Laurent (2000, 2001) extended the work by Fernández and Steel (1998) to the GARCH framework by expressing the mean and variance of the density such that error process has zero mean and unit variance. $Z_t \sim st^*(0,1,\xi,v)$. The standardized skew Student-t distribution with zero mean and unit variance:

$$f(z_t;\xi,v) = \begin{cases} \frac{2}{\xi+\frac{1}{\xi}} sg[\xi(sz_t+m);v] & if\ z_t < -\frac{m}{s} \\ \frac{2}{\xi+\frac{1}{\xi}} sg[(sz_t+m)/\xi;v] & if\ z_t \geq -\frac{m}{s} \end{cases} \tag{3.20}$$

$$m = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi-\frac{1}{\xi}\right) \quad \text{and} \quad s^2 = \left(\xi^2+\frac{1}{\xi^2}-1\right)-m^2$$

where $g(\cdot\,;v)$ is the Student-t density and ξ is the asymmetry parameter. m and s^2 are the mean and the variance of the non-standardized skewed Student-t. Setting the asymmetry parameter to 1 will give the symmetric Student-t density. The asymmetry parameter $\xi > 0$ is defined such that the ratio of probability masses above and below the mean is: $\frac{P(z\geq 0|\xi)}{P(z\leq 0|\xi)} = \xi^2$

3.9 Estimation of GARCH models

The non-linear nature of volatility models makes estimation by linear regression methods not possible. There are many feasible methods available, but the common estimation method is by maximum likelihood. It uses the idea that if we have a sample of data and an assumption of the distribution of the data, maximum likelihood assigns the most probable parameters given our sample data by maximizing the likelihood function.

If we have an IID random variable following a parametric distribution with density $f(\cdot)$. We then draw a sample of size T from this distribution, we get $z = \{z_1, z_2, z_3, \ldots, z_T\}$. The joint density of drawing this exact sample given the parameters is: $f(z) = f(z_1)f(z_2)\ldots f(z_T)$. But in our application we want to estimate the parameters given our data. We define the likelihood function $\mathcal{L}(\theta; z)$ where θ is the parameters and z is our sample: $\mathcal{L}(\theta; z) = \prod_{t=1}^{T} f(z_t; \theta)$. Estimates of the parameters are then obtained by maximizing the likelihood function. Usually it is much easier to work with the log likelihood function because it is a sum rather than a product. The likelihood function is a monotonically increasing function, thus maximizing the log likelihood function will produce the same result as maximizing the likelihood function. The parameter estimates are defined as:

$$\hat{\theta}_{ML} = \arg\max_{\theta} \mathcal{L}(\theta; z) \ = \arg\max_{\theta} \mathit{log} \prod_{t=1}^{T} f(z_t; \theta) \tag{3.21}$$

The normal likelihood function

For a T sample of IID observations, the normal likelihood function can be derived as:

$$\begin{aligned} \mathcal{L}(\theta; y_t) &= \prod_{t=1}^{T} f(y_t; \theta) = \prod_{t=1}^{T} \frac{1}{\sigma} f(z_t) \\ &= \prod_{t=1}^{T} \frac{1}{\sigma} \frac{exp(-\frac{1}{2} z_t^2)}{2\pi} \end{aligned}$$

Obtaining the log likelihood:

$$\begin{aligned} \log\mathcal{L}(\theta; y_t) &= \sum_{t=1}^{T} \left(\log\frac{1}{\sigma} - \frac{1}{2} z_t^2 - \frac{1}{2}\log(2\pi) \right) \\ &= \sum_{t=1}^{T} \left(-\frac{1}{2}\log\sigma^2 - \frac{1}{2} z_t^2 - \frac{1}{2}\log(2\pi) \right) \\ &= -\frac{1}{2}\sum_{t=1}^{T} (\log\sigma^2 + z_t^2 + \log(2\pi)) \end{aligned}$$

The constants can be dropped since they will not have any impact on the estimation:

$$\log\mathcal{L}(\theta; y_t) = -\frac{1}{2}\sum_{t=1}^{T}(\log\sigma^2 + z_t^2) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\sigma^2 + \frac{y_t^2}{\sigma^2}\right) \quad (3.22)$$

To get the likelihood functions for the different GARCH models, simply replace the conditional volatility in the log likelihood function with the GARCH conditional volatility. For example, the normal GARCH(1,1) log likelihood function is:

$$\log\mathcal{L}(\theta; y_t) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log(\omega+\alpha_1 y_{t-1}^2 + \beta_1\hat{\sigma}_{t-1}^2) + \frac{y_t^2}{\omega+\alpha_1 y_{t-1}^2 + \beta_1\hat{\sigma}_{t-1}^2}\right) \quad (3.23)$$

The Student-t likelihood function

The Student-t likelihood function can be derived in a similar way as for the normal distribution. The Student-t log likelihood function is:

$$\begin{aligned}\log\mathcal{L}(\theta; y_t) = T\left(\log\Gamma\left(\frac{v+1}{2}\right) - \log\left(\frac{v}{2}\right) - \frac{1}{2}\log\big(\pi(v-2)\big)\right) - \\ \frac{1}{2}\sum_{t=1}^{T}\left(\log(\sigma_t^2) + (1+v)\log\left(1 + \frac{z_t^2}{v-2}\right)\right)\end{aligned} \quad (3.24)$$

The skew Student-t likelihood function

The skew Student-t likelihood function was derived by Lambert and Laurent (2001). The log likelihood function of the skew Student-t density is:

$$\begin{aligned}\log\mathcal{L}(\theta; y_t) = T\log\Gamma\left(\frac{v+1}{2}\right) - \log\left(\frac{v}{2}\right) - \frac{1}{2}\log\big(\pi(v-2)\big) + \log\left(\frac{2}{\xi + 1/\xi}\right) \\ +\log(s)) - \frac{1}{2}\sum_{t=1}^{T}\left(\log(\sigma_t^2) + (1+v)\log\left(1 + \frac{(sz_t + m)^2}{v-2}\xi^{-2I_t}\right)\right)\end{aligned} \quad (3.25)$$

$$I_t = \begin{cases} 1, & z_t \geq -\frac{m}{s} \\ -1, & z_t < -\frac{m}{s} \end{cases},$$

$$m = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \quad \text{and} \quad s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2$$

3.10 Maximum likelihood estimation

In practice it is very difficult, time consuming and sometimes impossible to obtain analytical solutions to maximum likelihood functions. Therefore computer algorithms called solvers or optimizers are generally used to estimate the parameters. When using an iterative algorithm it is not always the case that it will produce a sequence that converges to a solution. This can be due to the solution being too far from the starting values, causing the solver to need a lot of time to converge. Instead of trying to converge, the computer software gives up after a number of iterations. Finding the optimal solution can also be difficult as some maximum likelihood functions have multiple maxima. If the solver stops at a local maximum, the solution that has been estimated is not the optimal solution.

For simple GARCH models it is very rare to encounter the problems discussed above, but for more complicated models such as the APARCH model estimation issues are more likely. The more parameters a model has, the higher the risk for estimation issues. Standard techniques to reduce the risk of estimation issues are:

- Increasing the data set
- Setting randomly starting values
- Trying other solvers
- Specifying a simpler model with less parameters

This thesis utilizes an open source software package called R, which is an increasingly popular programming language used especially for data analysis. R provides a wide variety of statistical and graphical techniques and is also easily extended by the use of software packages from the R community. We use the "rugarch" package developed by Alexios Ghalanos, which is a powerful package for modeling univariate GARCH models. The package currently supports five solvers: "solnp", "gosolnp", "nlminb", "L-BGFS-U" and "nlopt". All the different solvers have their advantages and disadvantages, so if a model does not converge it could possibly converge if a different solver is used. The "rugarch" package also has a "hybrid" solver that automatically rotates among the solvers if it encounters a non-converged model. See reference manual or vignette for detailed information about the "rugarch" package and solvers. (Ghalanos, 2013c)

3.11 Model selection and tests

There are several methods to evaluate the fit of a GARCH volatility model. But the model selection should be done on the basis of what the intended use of the model is. If the intended use is to forecast volatility, the model should be selected by considering the out of sample forecasting performance. Even though a model is significantly better fitted in sample, it does not necessarily perform better out of sample than another model. Actually it is often the case that a more parsimonious model will outperform a more flexible model out of sample, even if the flexible model is significantly better in sample.

For in sample fit diagnostics there are several statistical methods available to evaluate the fit of a model. The first step is to check if the parameters of the estimated model are significant or not and then the residuals of the model can be analyzed. To choose among different models we can use likelihood ratio tests or information criteria to evaluate which GARCH model fit the data the best.

Analysis of residuals

If a GARCH model is correct, the residuals of the model should be IID and distributed according to the assumed conditional distribution. The fitted residuals are: $\hat{z}_t = \frac{y_t}{\hat{\sigma}_t}$. To check how well the model captures the data, the residuals can be tested if they are IID and follow the assumed conditional distribution. If we want to test if the residuals are IID we can use the Ljung-Box test and plot the ACF of residuals and squared residuals. We can test if the conditional distribution assumption is correct by checking if the moments of the distribution of residuals are according to the assumed process and by plotting QQ-plots of the residuals.

Likelihood ratio tests

A common approach to statistically test if a model fits the data better than another model is to use a likelihood ratio (LR) test. The test involves estimating two models, one restricted and one unrestricted. The maximized values of the log likelihood functions are then used to compare the two models. In theory, the unrestricted model will always have a greater maximized log likelihood function than the restricted model. The question is whether the difference in log likelihood is great enough to offset the error induced into the unrestricted model by having more parameters. Under the null hypothesis the unrestricted model is not significantly better than the restricted model. The test statistic asymptotically follows a Chi-squared distribution where m is the number of restricted parameters. The test statistic is defined as:

$$LR = 2(\log\mathcal{L}_u - \log\mathcal{L}_r) \sim \chi^2_m \tag{3.26}$$

where $\log\mathcal{L}_u$ is the loglikelihood of the unrestricted model and $\log\mathcal{L}_r$ is the loglikelihood of the restricted model. If $LR > \chi^2_{\alpha,m}$ ⇒ Reject the null. A problem with the LR test is that the models tested have to be nested models. A model is nested in another model if we can obtain the nested model by restricting one or more of the parameters in the unrestricted model. For example the ARCH(1) model is nested in the GARCH(1,1) since we can obtain the ARCH(1) model by setting $\beta_1 = 0$ in the GARCH(1,1) model. Actually the APARCH model nests the ARCH, GARCH and gjrGARCH, but not the EGARCH model. So we cannot use the LR test to check if the fit of an APARCH model is significantly better than the fit of an EGARCH model.

Information criterion

A more general approach to compare the fit of GARCH models is to use information criterion. There are several information criteria available which are all based on likelihood. An advantage of these criteria is that they do not depend on nested models. The first information criterion was the Akaike Information Criterion (AIC) developed by Akaike (1974). The AIC is defined as:

$$AIC = -\frac{2\log\mathcal{L}}{T} + \frac{2m}{T} \tag{3.27}$$

The first part of the equation measures the goodness of fit of the model to the data, and the second part penalizes the model by the number of parameters used. The second part is known as the penalty function and varies for different information criteria. Another information criterion function is the Schwarz-Bayesian Information Criterion (BIC) developed by Schwarz (1978).The BIC is defined as:

$$BIC = -\frac{2\log\mathcal{L}}{T} + \frac{m * \log(\mathrm{T})}{T} \tag{3.28}$$

The BIC penalizes the number of parameters used to a higher degree than AIC when the sample size is not small. This is due to the penalty for each parameter used in BIC is log(T) and 2 for AIC. Burnham and Anderson (2002) argue that AIC has theoretically advantages over BIC, but there is no evidence that AIC outperforms BIC in a real application.

3.12 OBX fit diagnostics

As previously discussed we will have to fit the volatility models in sample to be able to forecast the volatility one-day out of sample. The models will later be evaluated for out of sample performance in the crisis period and the post crisis period. This means that the first estimation windows of the models will be in the time frame before the crisis and the post crisis. Table 3.4. presents likelihood ratio tests of the first estimation window to the crisis period. We see that all the models that are nested in the skew Student-t APARCH are rejected, in other words the model is significantly better fitted than all the nested models. Looking at the information criterion results in Table 3.5. we see that the the AIC chooses the skew Student-t APARCH model and the BIC chooses the EGARCH model. This is because the BIC penalizes the number of parameters used higher and therefore chooses the APARCH. From these results we are not able to conclude if the skew Student-t APARCH or the skew Student-t EGARCH model has the better fit in the first estimation window before the crisis.

Table 3.4. LR tests unrestricted skew Student-t APARCH(1,1) 02.01.1996 - 21.02.2005

Restricted model	LR statistic	Restrictions	p-value
normal GARCH(1,1)	91,1	4	0,000
Student-t GARCH(1,1)	38,4	3	0,000
skew Student-t GARCH(1,1)	19,5	2	0,000
normal gjrGARCH(1,1)	70,6	3	0,000
Student-t gjrGARCH(1,1)	22,7	2	0,000
skew Student-t gjrGARCH(1,1)	7,8	1	0,005
normal APARCH(1,1)	56,9	2	0,000
Student-t APARCH(1,1)	12,7	1	0,000

Table 3.5. Information criterion & likelihood 02.01.1996 - 21.02.2005

Model	Conditional dist.	AIC	BIC	Loglikelihood
GARCH(1,1)	Normal	-6.1094	-6.1019	6995.193
GARCH(1,1)	Student-t	-6.1315	-6.1215	7021.534
GARCH(1,1)	Skew Student-t	-6.1389	-6.1264	7031.005
gjrGARCH(1,1)	Normal	-6.1175	-6.1075	7005.456
gjrGARCH(1,1)	Student-t	-6.1375	-6.1250	7029.39
gjrGARCH(1,1)	Skew Student-t	-6.1431	-6.1281	7036.831
EGARCH(1,1)	Normal	-6.1199	-6.1098	7008.172
EGARCH(1,1)	Student-t	-6.1388	-6.1262	7030.811
EGARCH(1,1)	Skew Student-t	-6.1437	**-6.1287**	7037.454
APARCH(1,1)	Normal	-6.1226	-6.1101	7012.307
APARCH(1,1)	Student-t	-6.1410	-6.1260	7034.391
APARCH(1,1)*	Skew Student-t	**-6.1457**	-6.1282	**7040.747**

Table 3.6. lists the likelihood ratio tests from the first estimation window leading up to the post crisis period. We get the same results as for the first estimation window to the crisis period. The skew Student-t APARCH model has the best fit and all the nested models are rejected. The information criterion results are listed in Table 3.7. and both the AIC and the BIC chooses the skew Student-t APARCH as the best fit.

Overall the best fitted model is the skew Student-t APARCH model. It will be interesting to see if the models also perform well out of sample. Even though the model have superior in sample fit the models could perform worse than the other models out of sample. The tests also only give a measure of the fit at one point in time. When the model is used for forecasting the estimation window will change and cover different market conditions, which could make other models perform better.

* During fitting of the skew Student-t APARCH(1,1) model we had some estimation problems that resulted in a local maximum solution with lower loglikelihood value than for nested models. Changing the starting value of the ω parameter solved this problem.

Table 3.6. LR tests unrestricted skew Student-t APARCH(1,1) 02.01.1996 - 12.02.2009

Restricted model	LR statistic	Restrictions	p-value
normal GARCH(1,1)	126,9	4	0,000
Student-t GARCH(1,1)	80,8	3	0,000
skew Student-t GARCH(1,1)	40,1	2	0,000
normal gjrGARCH(1,1)	77,1	3	0,000
Student-t gjrGARCH(1,1)	41,9	2	0,000
skew Student-t gjrGARCH(1,1)	9,2	1	0,002
normal APARCH(1,1)	66,6	2	0,000
Student-t APARCH(1,1)	31,8	1	0,000

Table 3.7. Information criterion & likelihood 02.01.1996 - 12.02.2009

Model	Conditional dist.	AIC	BIC	Loglikelihood
GARCH(1,1)	Normal	-5.9127	-5.9071	9726.369
GARCH(1,1)	Student-t	-5.9261	-5.9187	9749.435
GARCH(1,1)	Skew Student-t	-5.9378	-5.9286	9769.792
gjrGARCH(1,1)	Normal	-5.9272	-5.9198	9751.306
gjrGARCH(1,1)	Student-t	-5.9373	-5.9280	9768.896
gjrGARCH(1,1)	Skew Student-t	-5.9466	-5.9355	9785.24
EGARCH(1,1)	Normal	-5.9239	-5.9165	9745.923
EGARCH(1,1)	Student-t	-5.9348	-5.9256	9764.831
EGARCH(1,1)	Skew Student-t	-5.9439	-5.9328	9780.724
APARCH(1,1)	Normal	-5.9298	-5.9205	9756.558
APARCH(1,1)	Student-t	-5.9398	-5.9286	9773.958
APARCH(1,1)	Skew Student-t	**-5.9488**	**-5.9358**	**9789.834**

4 VaR

Value-at-Risk is a single statistical measure that tries to capture the risk of a loss on a trading portfolio as a result of typical market movements. Although there are many different approaches and no general consensus on how it should be calculated, the definition of VaR is generally quite similar in the financial literature. Jorion (2001, p. 22) defines VaR as:

"VaR summarizes the worst loss over a target horizon with a given level of confidence."

VaR is a quantile on the distribution of profit and loss. We indicate the profit/loss by a random variable Q and with a particular realization indicated by q. $Q = P_t - P_{t-1}$. More generally the profit/loss of a portfolio is: $Q = \vartheta Y$. Where ϑ is the portfolio value multiplied by the returns Y. The density of Q is denoted by $f_q(\cdot)$. VaR is then defined as:

$$p = \int_{-\infty}^{-VaR(p)} f_q(x)dx \tag{4.1}$$

Since VaR is a positive number we use -VaR(p) when we integrate the density of the profit/loss function. Figure 4.1. shows the VaR(1%) and VaR(5%) of a standard normal profit/loss function.

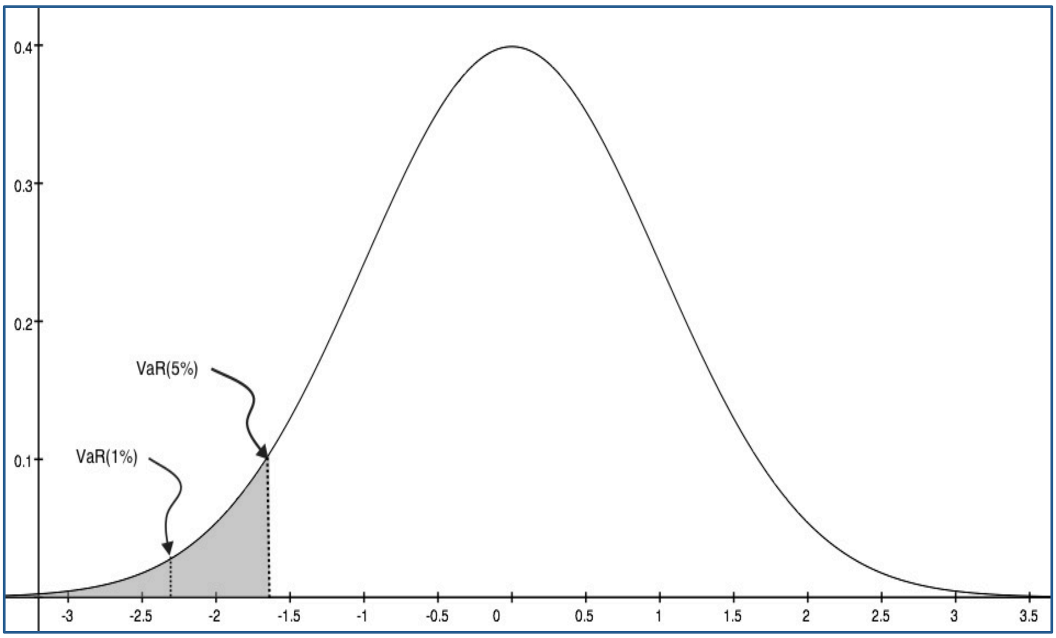

Figure 4.1. Standard normal profit/loss density and VaR

There are three main steps in VaR calculation:

1. Setting the probability p of losses exceeding VaR
2. Specifying the holding period
3. Identifying the probability distribution of the profit/loss function

The choice of probability *p* of losses exceeding VaR needs to be specified. The 1% and the 5% level are the most commonly used probability levels, but other levels are used for some applications. For instance, long-run risk analysis for pension funds may apply VaR calculations with even lover probability levels than the 1% level.

The time period over which losses are calculated is the holding period. In VaR calculation it is usually done at the one-day holding period, but it can be calculated for longer or shorter periods of time depending on the investment horizon. For example active investors like day traders have limited use of one-day VaR calculations as their portfolio might be liquidized by the end of the trading day. For these investors intraday VaR calculated from hour to hour makes more sense. Dowd (2005) suggests that an investor should use a holding period that equals the time needed to liquidize the portfolio.

The last step is to identify the probability distribution of the profit/loss function. Due to its simplicity and ease of use, the normal distribution is the most commonly used. The justification of using the normality assumption is the central limit theorem, but the theorem only applies to quantities and probabilities in the central mass of the density function. When dealing with VaR we are usually trying to calculate outcomes in the lower tail of the profit/loss function, thus the central limit theorem assumption makes little sense in this application. Another common distribution that is used is the Student-t distribution that has fatter tails and excess kurtosis. To accommodate for skewness in the distribution the skew Student-t distribution can be used.

VaR has many attractions as a risk measure. Dowd (2005) provides several important attractions that VaR has over traditional risk measures:

1. Provides a common consistent measure of risk across positions and risk factors.
2. Allows us to aggregate the risk of subpositions to an overall measure.
3. A holistic measure that takes full account of all risk factors.

4. A probabilistic measure that gives useful information on the probability of losing a certain amount.
5. Expressed as a simple and easily understood measure in terms a monetary value.

In the next sections we will show how VaR can be calculated by the non-parametric historical Simulation method and the parametric methods with conditional volatility.

4.1 Non-parametric approach

Historical Simulation (HS) is a non-parametric approach to VaR, and is considered the simplest method to estimate. It assumes that past price movements will continue into the future and that history will repeat itself. The VaR is computed by using the percentile of the empirical distribution corresponding to the chosen confidence level.

VaR at probability p is the negative $(T * p)$th value of the sorted return vector. To get the VaR of a portfolio simply multiply with the portfolio value.

In the simplest form each historical portfolio return is weighted the same, which makes it sluggish in adapting to structural changes in the volatility. But in the absence of structural changes the historical simulation method performs quite well compared to other methods. This is because it is less sensitive to outliers in the latest observations and it is has less estimation error than the parametric models. The number of observations included in the empirical distribution is the window size. The size is essential as it will have a big impact on the estimation. If the window size is small, the impact of the latest observations will be higher, resulting in greater movements in the historical simulation. Daníelsson (2011) recommends that the window size should be at least $3/p$. If we want to calculate VaR for a probability level 0.01 we will need at least 300 observations, which is slightly above one trading year.

The main advantage of historical simulation is that we do not have to make any assumptions of the distribution of returns. Although parametric models can also incorporate the known heavy tails and skewed distribution for returns, fitting the parameters can be very difficult. HS also uses the observed data directly and is therefore not subject to estimation error. HS is straightforward to understand, not only by risk managers but also for people without in depth knowledge of risk management. It is easy to calculate, the only thing required is the portfolio return time series. There is no need to calculate parameters such as variance and covariance. This makes it simple to implement in any business or organizational set up.

Maybe the most severe shortcoming of HS is that it assumes that the future will be like the past, and that the financial return series holds all the possible outcomes of the future risk. This results in a sluggish performance when market conditions change. The lacking ability to pick up on sudden changes in market risk is problematic and leads to underestimation of VaR when market risk increases and overestimation of VaR when market risk decreases.

4.2 Parametric approach

Parametric approaches to VaR are based on estimating the underlying distribution of returns and then obtain risk forecasts from the estimated models. In the multivariate case the first step is usually to forecast the covariance matrix, hence the parametric approach is often referred to as the variance-covariance method. Since the parametric method is based on estimating some distribution of the return data to obtain a VaR forecast, estimation error becomes a problem. The more parameters a model has the more complexity is added and model risk also becomes a concern.

We will cover some of the typical volatility models used to forecast parametric VaR, which are the MA, EWMA and GARCH models that have previously been described. We will also use different error distributions for the GARCH models such as the normal distribution, Student-t distribution and the skew Student-t distribution.

4.2.1 VaR for continuously compounded returns

If we use continuously compounded returns then:

$$
\begin{aligned}
p &= \Pr\big(P_t - P_{t-1} \leq -VaR(p)\big) \\
&= \Pr\big(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p)\big) \\
&= \Pr\big(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p)\big) \\
&= \Pr\left(\frac{Y_t}{\sigma} \leq \left(-\frac{VaR(p)}{P_{t-1}} + 1\right)\frac{1}{\sigma}\right)
\end{aligned}
$$

Since the distribution of standardized residuals (Y_t/σ) can be denoted by $F(\cdot)$, $-VaR(p)/P_{t-1} \leq 1$ and significance level $\gamma(p) = F_y^{-1}(p)$, we have:

$$VaR = -(e^{F_y^{-1}(p)\sigma} - 1)P_{t-1}$$

The VaR for holding one unit of asset with price 1, when $F_y^{-1}(p)\sigma$ is small, is then given by:

$$VaR(p) = -\sigma\gamma(p) \tag{4.2}$$

4.2.2 VaR with normal error distribution

The error distribution was not specified above. If we assume that the errors are normally distributed and $\Phi(\cdot)$ is the standardized normal distribution. The VaR is then easily calculated as:

$$VaR(p) = -\sigma\gamma(p) = -\sigma\Phi^{-1}(p) \tag{4.3}$$

where σ is the next day volatility forecast and $\Phi^{-1}(p)$ is the quantile of the standardized normal distribution. If we want to calculate the 1% VaR then p=0.01 and the VaR is: $VaR(1\%) = -\sigma\Phi^{-1}(0.01) = 2.3264\sigma$

The normal distribution has zero skewness, zero excess kurtosis and thin tails. Even if the error distribution is normally distributed, the GARCH model will still be able to model the actual returns to have heavy tails, excess kurtosis and skewness. However, with a normal error distribution the return distribution is still often unable to fully accommodate for the heavy tails, excess kurtosis and skewness that are typically seen in financial return series. Changing the error distribution to be Student-t or skew Student-t distributed can help to make the GARCH models to fit the return series better.

4.2.3 VaR with Student-t error distribution

If we assume that the errors are Student-t distributed and $t^*(\cdot)$ is the standardized Student-t distribution. The VaR is then calculated as:

$$VaR(p,v) = -\sigma\gamma(p,v) = -\sigma t^*(p,v)$$
$$t^*(p,v) = \frac{t(p,v)}{\sqrt{\frac{v}{v-2}}} \tag{4.4}$$

where σ is the forecast of the next day volatility and $t^*(p,v)$ is the standardized Student-t quantile. For example if p=0.01 and v=4, then the VaR is:

$$VaR(1\%) = -\sigma * \frac{t(0.01,4)}{\sqrt{\frac{4}{4-2}}} = -\sigma * \frac{-3.746947}{\sqrt{2}} = 2.6495\sigma$$

The Student-t distribution converges to the normal distribution as the degrees of freedom gets larger. Therefore, allowing for Student-t distributed errors can be seen as a generalization of the normal process which at the same time allows for fatter tails and excess kurtosis when it is needed. But the Student-t distribution still has zero skewness. If we need a process that allows for more skewness, we can use the skew Student-t distribution.

4.2.4 VaR with skew Student-t error distribution

Assuming that the errors are skew Student-t distributed and $st^*(\cdot)$ is the standardized skew Student-t distribution, the VaR is then calculated as:

$$VaR(p,v,\xi) = -\sigma\gamma(p,v,\xi) = -\sigma * st^*(p,v,\xi)$$

$$st^*(p,v,\xi) = \frac{st(p,v,\xi) - m}{s}, \tag{4.5}$$

$$m = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \quad and \quad s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2$$

where σ is the forecasted volatility and $st^*(p,v)$ is the standardized skew Student-t quantile. For example if $p = 0.01$, $v = 4$ and $\xi = 0.8$ (negative skewness), then the VaR is:

$$VaR(1\%) = -\sigma\frac{st(0.01,4,0.8) - m}{s} = -\sigma\frac{-3.51792 - (-0.31819)}{1.04940} = 3.0491\sigma$$

The skew Student-t distribution allows the GARCH models to capture more skewness than they normally can do. But the estimation can become more difficult since we have to estimate more parameters. Especially for the advanced GARCH models such as APARCH, estimation issues are more likely to be a problem when the skew Student-t is used as error distribution.

4.2.5 Implementing VaR forecasts with time dependent volatility

Moving Average

The simplest way to implement a VaR forecast with time-dependent volatility is to use the Moving Average (MA) volatility model. We simply calculate the sample standard deviation of our return series and insert this into the VaR formula for normally distributed returns. The one-day ahead VaR forecast is:

$$VaR(p) = -\hat{\sigma}\Phi^{-1}(p) \tag{4.6}$$

EWMA

It takes a bit more effort to implement a VaR forecast with the EWMA since we somehow have to specify the volatility on the first day σ_1. By setting the σ_1 to some random number it will induce some error into the model. This can be reduced by estimating σ_1 by calculating the sample standard deviation of returns of the first 30 days. Then σ_t can be estimated by running the model through the data. The one-day ahead volatility forecast is then finally calculated: $\hat{\sigma}_{t+1}^2 = (1-\lambda)y_t^2 + \lambda\hat{\sigma}_t^2$. The one-day ahead normal VaR forecast is:

$$VaR(p) = -\sqrt{(1-\lambda)y_t^2 + \lambda\hat{\sigma}_t^2}\Phi^{-1}(p) \tag{4.7}$$

GARCH

In a similar way as with the EWMA model, the GARCH models can be used to forecast volatility. But first we have to estimate the parameters of the model by maximum likelihood. If we are using the standard GARCH(1,1) the volatility of day t is: $\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2$. Leading the model by one period gives the one-day ahead volatility forecast: $\hat{\sigma}_{t+1}^2 = \omega + \alpha_1 y_t^2 + \beta_1\hat{\sigma}_t^2$. Finally, the one-day ahead normal VaR forecast is:

$$VaR(p) = -\sqrt{\omega + \alpha_1 y_t^2 + \beta_1\hat{\sigma}_t^2}\Phi^{-1}(p) \tag{4.8}$$

VaR forecasts can be implemented with each of the GARCH models described earlier in a similar way. Simply estimate the model parameters and obtain the volatility forecast. Then calculate the quantile of the assumed conditional distribution given the probability level of the VaR model. By multiplying the volatility forecast with the quantile figure we obtain the VaR forecast. As with the EWMA model we have to specify the volatility on the first day σ_1. The default implementation of the "rugarch" package is to set σ_1 to the entire sample standard deviation.

4.3 The impact of the mean

Previously we have argued that we can safely assume that the expected return is zero when calculating the volatility. Assuming that the mean is zero simplifies the calculations since we do not have to specify the mean. But how will this impact the calculation of VaR? Rewriting the general parametric VaR formula to take the mean into account: $VaR(p) = -\sigma\gamma(p) - \mu$. Where μ is the expected daily return. When calculating VaR there is usually a statistical uncertainty of more than 10%, therefore the calculated VaR will only be significant to one digit. The expected daily return will for most applications be smaller than this, so assuming the mean is zero will have an insignificant impact on the daily VaR calculation.

Under the assumption that the returns are IID, the mean will aggregate at the rate of time and the volatility will aggregate at the square root of time. The T-period VaR is: $VaR(p) = -\sqrt{T}\sigma\gamma(p) - T\mu$. We find that for longer time horizons the impact of the mean becomes relatively larger compared to the volatility. When calculating VaR for longer time horizons, the mean could become significant and should then be specified. But usually VaR is calculated for a horizon of up to 10 days and can therefore be assumed to be zero.

4.4 VaR criticism

A drawback of VaR is that it does not give any information about the loss that can occur beyond the calculated VaR level. Actually, the VaR estimate only tells us what the maximum loss is if a tail event does not occur. VaR therefore provides the "best of worst case scenarios" and will systematically underestimate the potential losses given a probability level. This can lead to some very unfavorable outcomes, as a potential investment opportunity can appear to be more desirable than they really are.

VaR has also been criticized for not being a coherent risk measure. In their article "Thinking Coherently" Artzner, Delbaen, Eber, and Heath (1997) wrote the first formal mathematical study on financial risk measures. They studied which properties a risk measure should have in order to be a sensible and useful risk measure, and they identified four axioms that have to be fulfilled. If a risk measure does fulfill these axioms the risk measure is said to be coherent.

If X and Y represent two portfolios´ profit/loss functions, and $\varphi(\cdot)$ is a measure of risk over a chosen horizon, the risk measure $\varphi(\cdot)$ is said to be coherent if it satisfies the axioms:

1. Monotonicity: If $Y \geq X \Rightarrow \varphi(Y) \geq \varphi(X)$
2. Positive homogeneity: $\varphi(cX) = c\varphi(X)$ for $c > 0$
3. Translational invariance: $\varphi(X + c) = \varphi(X) - c$
4. Subadditivity: $\varphi(X + Y) \leq \varphi(X) + \varphi(Y)$

The first axiom, monotonicity, states that if risk X never exceeds the risk of Y, the risk measure of Y should always be larger than the risk measure of X. If positive homogeneity holds, risk is directly proportional to the value of the portfolio. For example, the risk of holding 10 shares of some stock should have 10 times the risk of holding one share of the same stock. The translational invariance axiom holds if adding cash to the portfolio, or some other risk free asset, reduces the risk with this exact cash amount. Subadditivity is the last axiom that ensures that diversification results in reduced risk, which is fundamental in investment theory. If subadditivity does not hold we could get to an irrational conclusion that diversification is actually bad for an investment, and that putting all the wealth in one asset might be a good risk management decision.

The first three axioms are intended to rule out awkward outcomes in risk measures. In the follow-up paper by Artzner, Delbaen, Eber, and Heath (1999) it was proven that VaR measures generally satisfies the first three axioms, but not subadditivity. Although they showed that subadditivity holds when the returns are normally distributed, they could not find that subadditivity would always hold in practice. Later Daníelsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2012) studied the subadditivity of VaR further and found that VaR is subadditive when the tails of the return distribution are not super fat. Assets like equities, exchange rates and commodities do rarely have tails that are so fat that subadditivity is violated. Even though VaR seems to be coherent in most cases, we cannot generalize that VaR is a coherent measure of risk.

Another weakness of VaR is that it is very easy to manipulate. A financial institution will find it very easy to move the quantiles of the profit and loss distribution around to change the VaR. For example Daníelsson (2002) has shown how easy it is to manipulate VaR with put options to deliver any VaR level desired. This will result in a lower VaR for the calculated probability level, but the VaR for almost all other levels will increase.

Based on the criticism above alternative risk measures have been developed such as Expected Shortfall (ES), which theoretically is a better risk measures than VaR. ES estimates the expected loss when losses exceed VaR and has the advantage of being subadditive. But ES is calculated with a greater estimation error than VaR and it is much harder to backtest. To backtest ES it has to be compared with the output from a model, while VaR can be compared with the actual outcome. Extreme Value Theory (EVT) has more recently been developed to calculate VaR in economics, though it has been widely used in engineering for many decades. EVT is considered to improve risk forecasting especially for calculating VaR at a lower level than the 1% probability level, but EVT is complicated and very challenging to implement.

Even though the inadequacies of VaR are well documented and alternative risk measures have been developed, VaR is still the risk measure of choice by financial institutions. The reason is that VaR has been incorporated internally in the institutions and is used as a standard measure for regulatory reporting. As long as there is a demand for VaR calculations, the challenge will be to make VaR models that reflect the true risk as accurately as possible.

5 Backtesting

It is important to validate a risk model before it is put to practical use, and its performance should regularly be evaluated. To validate that VaR models calculate the future risk appropriately, the models are tested with out of sample quantitative methods that are known as backtests. A quote from a risk manager underlines the importance of backtesting VaR models:

> *"VaR is only as good as its backtest. When someone shows me a VaR number, I don't ask how it is computed, I ask to see the backtest."* (Brown, 2008, p. 20)

Backtesting aims to take ex ante VaR forecasts from a risk model and compare them with ex post realized returns. A VaR violation is said to have occured when losses exceed the forecasted VaR. For example if we backtest a 1% VaR model we would expect to have a VaR violation every 100 day on average. So if we backtest over a 1000 day time period we expect to get only 10 violations. On the other hand, if we backtest a 5% VaR model over the same time period we expect to get 50 violations.

A good VaR model should produce around the expected number of VaR violations when it is backtested. If the model produces violations that exceed the expected number of violations the model underestimates the risk, and if the model produces less than the expected number of violations it overestimates the risk. In theory, the violations of a backtested risk model should also be evenly spread out in time and not be clustered together. Even if a backtested model produces the expected number of violations, all the violations could come in a very short time-period. This indicates that the VaR model is unable to capture the changes in market volatility and underestimates the risk in periods of high market volatility. If a VaR model fails the backtesting, the assumptions and parameters of the model should be carefully examined.

5.1 Estimation window and testing window

The number of observations used to forecast risk is the estimation window W_E. W_T is the testing window that is the historical data which the risk is forecasted over. The entire sample T is the sum of the estimation window W_E and the testing window W_T. The backtest is performed using a rolling window method where the testing window is rolled over almost the entire data sample. The estimation window and forecast horizon is constant during the whole rolling process. Figure 5.1. illustrates how the rolling backtest method works.

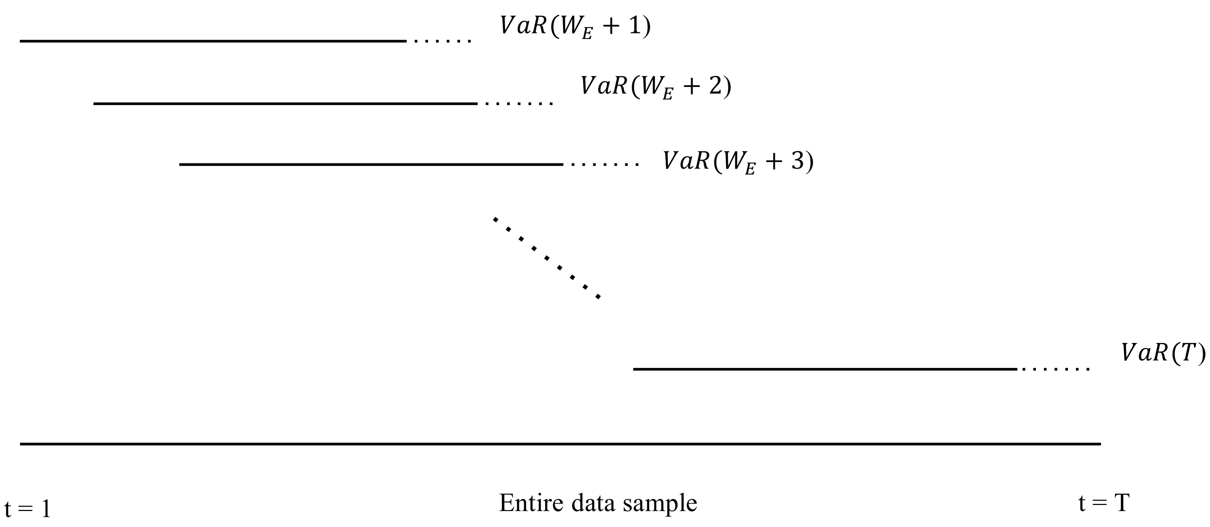

Figure 5.1. Rolling windows with VaR forecasts

VaR violations are rare events, which makes it difficult to analyze the result of a backtest if the testing window is too small. For example, if the testing window is 100 days and we backtest a VaR(1%) model we would only expect to get one violation. This makes it impossible to reject or accept a well specified model since the sample of violations will be too small. To get reliable testing results Daníelsson (2011) recommends that the expected number of VaR violations should be at least 10. Hence the testing window of a VaR(1%) will have to be at least 1000 days. The longer the testing window is, the more powerful the results will be. The size of the estimation window depends on the VaR model and the probability level. Different forecasting models will need different amounts of data to perform well. Of the models described earlier, the MA and EWMA do not need as much data as historical simulation. However, GARCH models need a lot of data. It could also be interesting to use different sizes for the estimation window for HS and MA models since the performance for these models is highly affected by the size of the estimation window.

5.2 VaR violations and violation ratio

A quick and easy method to evaluate if a risk model is acceptable is to calculate the violation ratio. If we get a violation ratio of 1, the number of violations of the backtested model equals the number of expected violations. A violation ratio larger than one indicates that the model underestimates the risk, while a violation ratio less than 1 indicate that the model overestimates the risk. First we define the VaR violations:

$$\eta_t = \begin{cases} 1 & \text{if } y_t \leq -VaR_t \\ 0 & \text{if } y_t > -VaR_t \end{cases} \tag{5.1}$$

where v_1 is the count of $\eta_t = 1$ and v_0 is the count of $\eta_t = 0$, which is defined as: $v_1 = \sum \eta_t$ and $v_0 = W_T - v_1$. The violation ratio is:

$$VR = \frac{\text{Observed number of violations}}{\text{Expected number of violations}} = \frac{v_1}{p * W_T} \tag{5.2}$$

Because there is some degree of uncertainty in VaR calculations we will accept models that have violation ratios close to 1. Daníelsson (2011) recommends that risk models which produce violation ratios in the range 0.8 to 1.2 can be accepted as good risk models and models that produce violation ratios below 0.5 or above 1.5 should be rejected. This can only be seen as a rule of thumb method as the bounds theoretically should shrink with larger testing windows. For a more formal approach, the number of violations can be statistically tested if they are significantly different from the expected number of violations.

The violation ratio only provides an evaluation of the actual number of violations versus the expected number of violations. To check if the violations are clustered it could be helpful to plot a backtesting chart with both the predicted profit/loss and the actual profit/loss plotted. A violation has occurred each time the actual profit/loss exceeds the predicted profit/loss. But it is difficult to detect clustering of violations by only studying the backtesting chart and therefore statistical methods are more appropriate to use. In the next section we will show the statistical methods that can be used to check the significance of backtests.

5.3 Statistical tests

A more formal approach to backtesting is to use statistical methods. The violations in a backtest are a sequence of ones and zeroes that are Bernoulli distributed with a probability level p. If the violations reflect the assumed risk level the probability level of the Bernoulli variable should be equal to the VaR probability level. This is known as the unconditional coverage property. The Bernoulli variable of the violations should also be independent to ensure that the violations do not cluster, which is known as the independence property. The statistical testing is divided into three tests: Unconditional coverage test, independence test and joint test of conditional coverage.

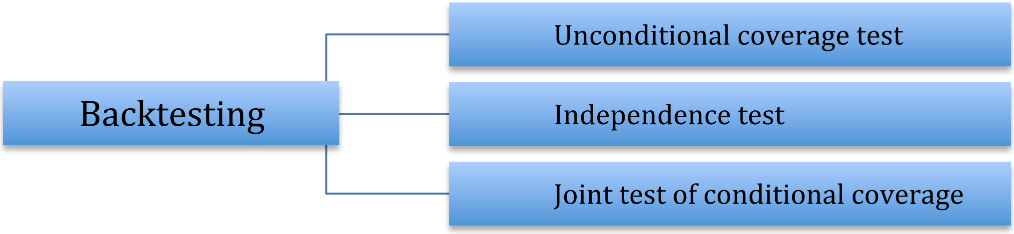

Figure 5.2. Overview statistical backtesting methods

When we perform statistical backtesting we use the standard hypothesis paradigm. We specify a null hypothesis that we want to test, and an alternative hypothesis that is accepted if the null is rejected. But we also have to select a significance level for the tests. However, there is a problem with choosing the significance level as there is a tradeoff between the possibility of rejecting a correct model and accepting an incorrect model, respectively known as type 2 and type 1 errors. Ideally, the significance level should be chosen to balance out the probability of getting a type 1 error and a type 2 error. In practice the significance level is set to a 5% level, which implies that the null hypothesis is rejected only if the evidence against it is reasonably strong.

5.3.1 Unconditional coverage test

Kuiper's unconditional coverage test is based on testing if the number of VaR violations is significantly different from the expected number of violations (Kupiec, 1995). The null hypothesis of the VaR violations is: $H_0 : \eta \sim B(p)$. Where B is the Bernoulli distribution. The Bernoulli density is: $(1-p)^{1-\eta_t}(p)^{\eta_t}, \quad \eta_t = 0, 1$. The probability p is estimated by: $\hat{p} = \frac{v_1}{W_T}$. Kupiec showed how the test could be performed by a likelihood ratio test. The test statistic is:

$$LR_{UC} = 2 \log \frac{(1-\hat{p})^{v_0}(\hat{p})^{v_1}}{(1-p)^{v_0}(p)^{v_1}} \sim \chi_1^2 \tag{5.3}$$

If $LR_{UC} > \chi_{\alpha,1}^2 \Rightarrow$ Reject the null. A shortcoming of the Kupiec test is that it does not take into account how large the discrepancy between the VaR forecast and the actual loss is. Consequently, a VaR model will pass the Kupiec test if it produces violations within the acceptable range of the test, even if it provides very poor VaR forecasts of the losses larger than VaR. Alternative backtests like the loss function based backtest developed by Lopez (1999) takes the magnitude of VaR violations into account, but will not be covered in this thesis.

5.3.2 Independence test

To detect clustering of violations, a test for independence of the violations was developed by Christoffersen (1998). The test is based on comparing the probability of two consecutive violations to follow each other versus the probability of a non-violation to be followed by a violation. The two probabilities should be equal if the violations are independent. If we let n_{ij} be the number of days that state j occurred after state i occurred the day before, where the states refers to states of violations or non-violations. Table 5.1. illustrates the outcomes.

Table 5.1. Contingency table independence test

	$\eta_{t-1} = 0$	$\eta_{t-1} = 1$	
$\eta_t = 0$	n_{00}	n_{01}	$n_{00} + n_{01}$
$\eta_t = 1$	n_{10}	n_{11}	$n_{10} + n_{11}$
	$n_{00} + n_{10}$	$n_{01} + n_{11}$	N

We then define π_{ij} as the probability of observing state j if the state on the day before was i. The probabilities are estimated by: $\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}}$, $\hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}}$, $\hat{\pi} = \frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}$. Under the null hypothesis of independent violations, the probabilities π_{01} and π_{11} are equal. Christoffersen showed how the test could be performed by a likelihood ratio test. The test statistic is:

$$LR_{IND} = 2\log\frac{(1-\hat{\pi})^{n_{00}+n_{01}}\hat{\pi}^{n_{10}+n_{11}}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}} \sim \chi_1^2 \qquad (5.4)$$

If $LR_{IND} > \chi^2_{\alpha,1}$ $\Rightarrow$ Reject the null. A problem with the independence test is that it only has power to detect clustering in the form of violations occurring in pairs. For example, the test will not detect if the probability of a violation today is dependent upon if there was a violation two days ago, which is also a breach of the independence property.

5.3.3 Joint test of conditional coverage

Christoffersen (1998) also proposed a joint test for unconditional coverage and independence that is known as a conditional coverage test. The test is based on combining the two likelihood ratio tests of unconditional coverage and independence described above. Under the null hypothesis of correct coverage and independence, the test statistic is:

$$LR_{CC} = LR_{UC} + LR_{IND} \sim \chi_2^2 \tag{5.5}$$

If $LR_{CC} > \chi^2_{\alpha,2}$ ⇒ Reject the null. The joint test allows us to test for both coverage and independence at the same time, but the test loses power to detect if a VaR model only satisfies one of the properties. Therefore, the individual tests for coverage end independence should always be performed if the model is not rejected by the joint test.

5.4 Selecting risk models based on backtests

Choosing a VaR model should ultimately be based on the performance during backtesting. It is possible that two different VaR models perform the same during backtesting, yet they could have major differences in their forecasts. Daníelsson (2011) suggests that VaR models with the least amount of variations in their VaR forecasts are preferred. Therefore, calculating the standard deviation of the VaR forecasts can be helpful to choose between VaR models that perform the same during backtesting.

5.5 Backtesting results

All the models are backtested over 1000 days in each of the crisis period and the post crisis period. We have also backtested the models over both periods to get more reliable total results. Backtesting the models for the complete period should in theory give the sum of the backtests from the individual periods, but since the estimation windows are different for the two testing periods we have to backtest the models for the complete period. The GARCH models parameters are refitted every 100 days. For HS and MA the estimation windows are also changed to check if different window sizes can improve the VaR forecasts. Since the GARCH model is regarded as superior to the ARCH model, we have excluded to checking the performance of the ARCH model. All the models that are passed both for coverage and independence in the backtesting results are highlighted.

5.5.1 Scenario 1 - Crisis

Table 5.2. VaR 1% Exceedance 21.02.2005 - 11.02.2009

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	10	21	2,10	0,002	0,075	0,002
GARCH(1,1)	Student-t	10	17	1,70	0,043	0,031	0,012
GARCH(1,1)	Skew Student-t	10	13	1,30	0,362	0,157	0,243
gjrGARCH(1,1)	Normal	10	25	2,50	0,000	0,151	0,000
gjrGARCH(1,1)	Student-t	10	21	2,10	0,002	0,075	0,002
gjrGARCH(1,1)	Skew Student-t	10	14	1,40	0,231	0,013	0,022
EGARCH(1,1)	Normal	10	30	3,00	0,000	0,068	0,000
EGARCH(1,1)	Student-t	10	25	2,50	0,000	0,151	0,000
EGARCH(1,1)	Skew Student-t	10	19	1,90	0,011	0,370	0,026
APARCH(1,1)	Normal	10	25	2,50	0,000	0,151	0,000
APARCH(1,1)	Student-t	10	21	2,10	0,002	0,075	0,002
APARCH(1,1)	Skew Student-t	10	16	1,60	0,079	0,023	0,016

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
HS	2289	10	29	2,90	0,000	0,009	0,000
MA	2289	10	51	5,10	0,000	0,004	0,000
EWMA	2289	10	23	2,30	0,000	0,109	0,001
HS	1000	10	29	2,90	0,000	0,056	0,000
MA	1000	10	54	5,40	0,000	0,002	0,000
HS	300	10	25	2,50	0,000	0,151	0,000
MA	300	10	44	4,4	0,000	0,002	0,000

Table 5.2. gives the backtesting results from the crisis period for VaR(1%). The skew Student-t GARCH(1,1) model is the only model that passed the backtesting both for coverage and independence. It is interesting to see that the EGARCH and the APARCH models that are chosen as the best in sample fit in Table 3.4. and Table 3.5. do not perform as well out of sample. We also notice that the GARCH models are all improved by changing the error distribution to the Student-t and the skew Student-t distribution. All the traditional VaR models based on HS, MA and EWMA perform poorly. They underestimate the risk and are not able to adapt to the extreme market conditions.

Table 5.3. VaR 5% Exceedance 21.02.2005 - 11.02.2009

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	50	69	1,38	0,009	0,059	0,006
GARCH(1,1)	Student-t	50	69	1,38	0,009	0,059	0,006
GARCH(1,1)	Skew Student-t	50	63	1,26	0,069	0,054	0,030
gjrGARCH(1,1)	Normal	50	64	1,28	0,051	0,064	0,027
gjrGARCH(1,1)	Student-t	50	67	1,34	0,019	0,105	0,017
gjrGARCH(1,1)	Skew Student-t	50	62	1,24	0,093	0,045	0,032
EGARCH(1,1)	Normal	50	74	1,48	0,001	0,132	0,002
EGARCH(1,1)	Student-t	50	75	1,50	0,001	0,068	0,001
EGARCH(1,1)	Skew Student-t	50	65	1,30	0,037	0,076	0,024
APARCH(1,1)	Normal	50	70	1,40	0,006	0,070	0,004
APARCH(1,1)	Student-t	50	72	1,44	0,003	0,097	0,003
APARCH(1,1)	Skew Student-t	50	63	1,26	0,069	0,138	0,064

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
HS	2289	50	99	1,98	0,000	0,000	0,000
MA	2289	50	100	2,00	0,000	0,000	0,000
EWMA	2289	50	63	1,26	0,069	0,054	0,030
HS	1000	50	104	2,08	0,000	0,000	0,000
MA	1000	50	100	2,00	0,000	0,000	0,000
HS	300	50	80	1,60	0,000	0,000	0,000
MA	300	50	81	1,62	0,000	0,000	0,000

Table 5.3. gives the backtesting results from the crisis period for VaR(5%). A few models perform quite well. The skew Student-t gjrGARCH(1,1) model produces the most exact exceedance, but it is rejected for independence. The skew Student-t GARCH(1,1), normal gjrGARCH(1,1), skew Student-t APARCH(1,1) and EWMA are passed both for coverage and independence. We also notice that the GARCH models are not improved by changing the error distribution to the Student-t distribution and for some models even perform worse. This is probably because the VaR(5%) does not lie as far out in the tail region of the return distribution as the VaR(1%). But the skew Student-t distribution improves all the GARCH models. All of the simple VaR models based on HS and MA are rejected for both coverage and independence. As for the VaR(1%) level they underestimates the risk and fail to adapt to the new market conditions.

5.5.2 Scenario 2 - Post crisis

Table 5.4. VaR 1% Exceedance 12.02.2009 - 31.01.2013

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	10	16	1,60	0,079	0,471	0,166
GARCH(1,1)	Student-t	10	13	1,30	0,362	0,559	0,556
GARCH(1,1)	Skew Student-t	10	9	0,90	0,746	0,686	0,875
gjrGARCH(1,1)	Normal	10	17	1,70	0,043	0,443	0,096
gjrGARCH(1,1)	Student-t	10	13	1,30	0,362	0,559	0,556
gjrGARCH(1,1)	Skew Student-t	10	9	0,90	0,746	0,686	0,875
EGARCH(1,1)	Normal	10	16	1,60	0,079	0,471	0,166
EGARCH(1,1)	Student-t	10	11	1,10	0,754	0,621	0,842
EGARCH(1,1)	Skew Student-t	10	9	0,90	0,746	0,686	0,875
APARCH(1,1)	Normal	10	18	1,80	0,022	0,416	0,053
APARCH(1,1)	Student-t	10	12	1,20	0,538	0,589	0,715
APARCH(1,1)	Skew Student-t	10	9	0,90	0,746	0,686	0,875

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
HS	3289	10	5	0,50	0,079	0,823	0,208
MA	3289	10	18	1,80	0,022	0,329	0,046
EWMA	3289	10	20	2,00	0,005	0,366	0,013
HS	1000	10	0	0,00	0,000	1,000	0,000
MA	1000	10	5	0,50	0,079	0,823	0,208
HS	300	10	4	0,40	0,030	0,858	0,094
MA	300	10	8	0,80	0,510	0,719	0,755

Table 5.4. gives the backtesting results from the post crisis period for VaR(1%). All the models with skew Student-t error distribution perform very well. We notice that the models adapt to the more tranquil market conditions very quickly. Changing the error distribution of the GARCH models to the Student-t and the skew Student-t distribution improves the performance of all the models. Some of the traditional VaR models are also passed for both coverage and independence but the results seems to be spurious. Most of the models seems to overestimate the risk and are influenced by the high volatility in the previous period.

Table 5.5. VaR 5% Exceedance 12.02.2009-31.01.2013

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	50	61	1,22	0,122	0,301	0,178
GARCH(1,1)	Student-t	50	61	1,22	0,122	0,301	0,178
GARCH(1,1)	Skew Student-t	50	55	1,10	0,475	0,157	0,285
gjrGARCH(1,1)	Normal	50	62	1,24	0,093	0,274	0,134
gjrGARCH(1,1)	Student-t	50	64	1,28	0,051	0,224	0,071
gjrGARCH(1,1)	Skew Student-t	50	53	1,06	0,666	0,194	0,392
EGARCH(1,1)	Normal	50	60	1,20	0,159	0,728	0,349
EGARCH(1,1)	Student-t	50	60	1,20	0,159	0,728	0,349
EGARCH(1,1)	Skew Student-t	50	54	1,08	0,566	0,549	0,709
APARCH(1,1)	Normal	50	62	1,24	0,093	0,274	0,134
APARCH(1,1)	Student-t	50	64	1,28	0,051	0,544	0,124
APARCH(1,1)	Skew Student-t	50	56	1,12	0,393	0,469	0,534

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
HS	3289	50	55	1,10	0,475	0,252	0,402
MA	3289	50	51	1,02	0,885	0,374	0,667
EWMA	3289	50	56	1,12	0,393	0,933	0,691
HS	1000	50	73	1,46	0,000	0,554	0,000
MA	1000	50	19	0,38	0,000	0,370	0,000
HS	300	50	32	0,64	0,005	0,019	0,001
MA	300	50	36	0,72	0,033	0,044	0,014

Table 5.5. gives the backtesting results from the post crisis period for VaR(5%). All the GARCH models perform well and are passed for both coverage and independence. We notice that some of the models are performing worse by changing the error distribution to the Student-t distribution, but all models are improved with the skew Student-t distribution. All the traditional VaR models with large estimation windows are also passed for coverage and independence.

5.5.3 Scenario 3 - Complete period

Table 5.6. VaR 1% Exceedance 21.02.2005 - 31.01.2013

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	20	39	1,95	0,000	0,226	0,000
GARCH(1,1)	Student-t	20	31	1,55	0,022	0,093	0,018
GARCH(1,1)	Skew Student-t	20	23	1,15	0,510	0,266	0,434
gjrGARCH(1,1)	Normal	20	44	2,20	0,000	0,346	0,000
gjrGARCH(1,1)	Student-t	20	37	1,85	0,001	0,185	0,001
gjrGARCH(1,1)	Skew Student-t	20	25	1,25	0,279	0,038	0,065
EGARCH(1,1)	Normal	20	47	2,35	0,000	0,124	0,000
EGARCH(1,1)	Student-t	20	39	1,95	0,000	0,226	0,000
EGARCH(1,1)	Skew Student-t	20	28	1,40	0,090	0,409	0,169
APARCH(1,1)	Normal	20	43	2,15	0,000	0,319	0,000
APARCH(1,1)	Student-t	20	36	1,80	0,001	0,167	0,002
APARCH(1,1)	Skew Student-t	20	26	1,30	0,197	0,045	0,059

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind p-value	Joint p-value
HS	2289	20	31	1,55	0,022	0,320	0,045
MA	2289	20	69	3,45	0,000	0,001	0,000
EWMA	2289	20	43	2,15	0,000	0,001	0,000
HS	1000	20	29	1,45	0,058	0,320	0,101
MA	1000	20	59	2,95	0,000	0,000	0,000
HS	300	20	29	1,45	0,058	0,071	0,033
MA	300	20	52	2,60	0,000	0,000	0,000

Table 5.6. gives the backtesting results from the complete period for VaR(1%). The skew Student-t GARCH(1,1) perform the best. We also see that the skew Student-t EGARCH(1,1) and the skew Student-t APARCH(1,1) models that were chosen as the best in sample fit in Table 3.5. and Table 3.7. perform well, but they are outperformed by the skew Student-t GARCH(1,1). All the GARCH models are improved by changing the error distribution to the Student-t and the skew Student-t distribution. The traditional VaR models based on HS are not rejected for shorter estimation windows, but does not perform well compared to the best GARCH models.

Table 5.7. VaR 5% Exceedance 21.02.2005-31.01.2013

Model	Conditional dist.	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
GARCH(1,1)	Normal	100	130	1,30	0,003	0,368	0,009
GARCH(1,1)	Student-t	100	130	1,30	0,003	0,368	0,009
GARCH(1,1)	Skew Student-t	100	114	1,14	0,160	0,323	0,228
gjrGARCH(1,1)	Normal	100	126	1,26	0,010	0,451	0,028
gjrGARCH(1,1)	Student-t	100	129	1,29	0,004	0,546	0,014
gjrGARCH(1,1)	Skew Student-t	100	116	1,16	0,109	0,375	0,187
EGARCH(1,1)	Normal	100	132	1,32	0,002	0,424	0,005
EGARCH(1,1)	Student-t	100	133	1,33	0,001	0,278	0,003
EGARCH(1,1)	Skew Student-t	100	118	1,18	0,072	0,248	0,102
APARCH(1,1)	Normal	100	133	1,33	0,001	0,453	0,004
APARCH(1,1)	Student-t	100	135	1,35	0,001	0,515	0,002
APARCH(1,1)	Skew Student-t	100	117	1,17	0,089	0,402	0,166

Model	Estimation window	Expected	Actual	Ratio	Kupiec p-value	Ind. p-value	Joint p-value
HS	2289	100	151	1,51	0,000	0,000	0,000
MA	2289	100	141	1,41	0,000	0,000	0,000
EWMA	2289	100	119	1,19	0,058	0,144	0,057
HS	1000	100	127	1,27	0,008	0,000	0,000
MA	1000	100	119	1,19	0,058	0,000	0,000
HS	300	100	112	1,12	0,227	0,000	0,000
MA	300	100	117	1,17	0,089	0,000	0,000

Table 5.7. gives the backtesting results from the complete period for VaR(5%). We get very similar results as for the VaR(1%) backtest results for the complete period. All the GARCH models with skew Student-t error distribution are passed for both coverage and independence. The best model is once again the skew Student-t GARCH(1,1). As for the backtesting results in the crisis period and the post crisis period we see that the models are not improved by changing the error distribution to the Student-t distribution, but the skew Student-t distribution improves all the GARCH models. Of the traditional VaR models only the EWMA is passed for both coverage and independence.

6 Discussion

None of the traditional VaR models based on HS, MA and EWMA are able to capture the risk appropriately for both the different VaR levels and the different periods that are tested. If we look at the individual testing periods we see that all the models underestimates the risk in the crisis period. Although the models are passed for both coverage and independence in some of the testing periods, generally they do not perform well when they are compared to the GARCH models.

When working with the data analysis we did some preliminary testing and backtested the models over whole years, and consequently excluded the data from 2013. The results from this testing showed that the more advanced GARCH models performed significantly better compared to the final test results in this thesis. It shows how sensitive the backtesting framework is and how dependent the results are on the estimation window and testing window. In the end we decided to include the data from 2013 and backtest the models over 1000 days, which provided the results in this thesis. Without going too deep into the reason why this is the case, it could give an insight to why there are such different results in studies of VaR models. If the dynamics of the financial returns are constantly changing, different models will perform better in some periods and other models will perform better in other periods. With this in mind it would make sense to use a model that is able to capture the most important features that is found in the financial returns. Even though other features in the financial returns could be significant at some times, the added estimation error and complexity of modeling these features is probably not offset by the modeling ability.

The return statistics in Table 2.2. from the crisis and the post-crisis periods show that in both periods we have volatility clustering and non-normally distributed returns. There are also found some asymmetries in volatility as listed Table 3.1. and power effects as listed in Table 3.2., but as confirmed by the backtesting these structures do not improve the forecasting ability of the models. Overall the skew Student-t GARCH(1,1) model perform the best. Considering that the model is not rejected for either coverage or independence in any of the backtests it seems that the model is capable to capture the risk of the OBX-index even in periods with high volatility and changing market dynamics. Figure 6.1. plots the VaR forecasts from the skew Student-t GARCH(1,1) model for both the VaR(1%) level and the VaR(5%) level for the complete backtesting period. All the VaR exceedances are plotted with red marks, and from a quick visual inspection the exceedances appear to be evenly spread out in time and are not clustered together.

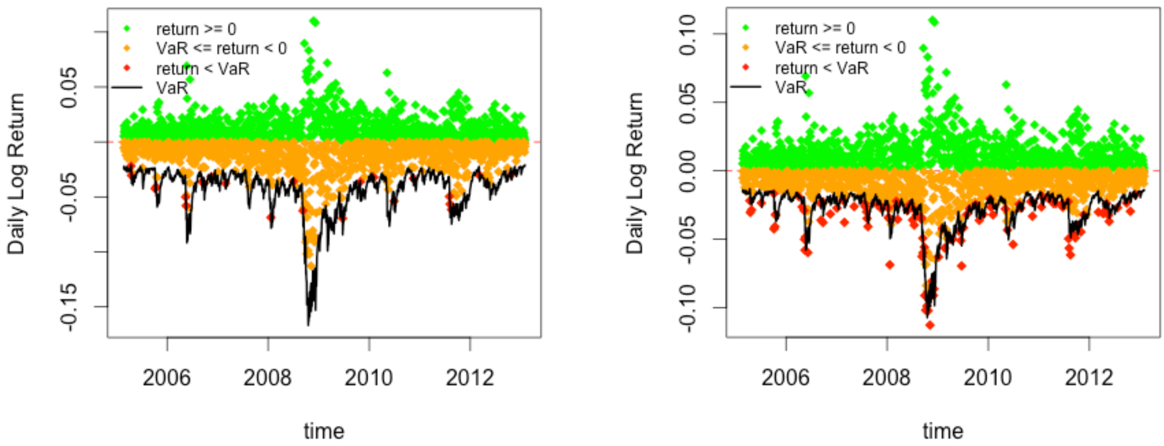

Figure 6.1. Backtesting plots skew Student-t GARCH(1,1) VaR(1%) VaR(5%) 2005-2013

These results are valid for the OBX-index for the time period that is tested. But it is very likely that the skew Student-t GARCH(1,1) model will perform well for future out of sample forecasting of the OBX-index. New models and testing methods are regularly developed and could improve the risk forecasting significantly in the future. In recent years the introduction of high frequency trading has changed the market conditions and there will certainly be changes in the financial markets in the years to come. Therefore the VaR model should always be up for validation and the model selection is not a one-time effort but a continuous process.

Comparing our findings with the results that Hansen and Lunde (2005) found in their study show that we get some of the same results. The standard GARCH model is not outperformed by any of the advanced models that can capture asymmetries or power effects. Contradictory to the findings of Hansen and Lunde, we have found that using the heavy tailed Student-t distribution improves the VaR forecasts especially for the 1% VaR level. But we also see that some of the more advanced GARCH models perform worse at the 5% VaR level. However, our best model, the skew Student-t GARCH(1,1), generally perform better when the error distribution is Student-t distributed. Hansen and Lunde did not study the possibility of skew distribution, which we have also found to significantly improve the VaR forecasts. Ghalanos (2013a) found that almost any other GARCH model beats the standard normal GARCH model. But if we take a closer look on his ranking of the models we notice that all the best ranked models have error distributions that are non-normal, while all the worse ranked models have error distributions that are normal. It seems that the choice of error distribution is very important and maybe more significant than the choice of GARCH model. Comparing our finding with the findings of Ghalanos we have also found that a non-normal error distribution significantly improves our VaR forecasts. Overall, our findings are not the same as either Hansen and Lunde or Ghalanos, but somewhere in-between.

We have also found that a good in sample fit does not guarantee that the models will perform better out of sample. If we take a look at the in sample fit for the crisis period in Table 3.5. we see that the skew Student-t APARCH(1,1) and the skew Student-t EGARCH(1,1) models are chosen by AIC and BIC respectively. But these models perform significantly worse out of sample compared to the more parsimonious skew Student-t GARCH(1,1) model.

For future research it could be interesting to study other asset classes in Norway such as individual stocks, currencies and bonds. We would probably get other results since the characteristics of different asset classes vary greatly. Since we have found that changing the error distribution significantly improves the forecasting ability of our models, it would also be interesting to study other error distributions such as the Johnson's Reparametrized SU distribution and the Normal Inverse Gaussian distribution which were frequently among the best performing models in Ghalanos (2013a). There is also a wide range of other GARCH models that have not been described or studied in this thesis that could possibly improve the VaR forecasts.

7 Conclusion

In this thesis we have studied the out of sample performance of one-day VaR models based on HS, MA, EWMA and GARCH models. The comparison has been made on the OBX-index, which is currently the main benchmark index on the Oslo Stock Exchange. To examine the performance in different market conditions the models have been backtested over two time periods and finally across both periods. Our findings can be summarized as follows:

- Based on the backtesting results of coverage and independence we have found that none of the traditional VaR models based on HS, MA and EWMA perform well. The models underestimate the risk in the crisis period and generally perform poorly in comparison to the GARCH models.
- The VaR model based on the skew Student-t GARCH(1,1) model is overall the best model. It is not rejected for either coverage or independence in any of the tests, and is therefore able to capture the risk well in periods of high volatility as well as in periods of more normal market conditions.
- Compared to the standard GARCH model, the advanced GARCH models that are able to model power effects and asymmetries are not found to generally improve the VaR forecasts. The modeling capabilities of these models do not offset the additional estimation error of fitting additional parameters.
- Changing the error distribution of the GARCH models to the skew Student-t distribution, which allows for more skewness and fatter tails than the standard GARCH model can, significantly improves the models. The normal error distribution is not able to capture the leptokurtic and skewness characteristics that are observed in the OBX index returns.

8 References

Akaike, Hirotugu. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6), 716-723.

Albuquerque, Rui. (2012). Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. *Review of Financial Studies, 25*(5), 1630-1673.

Artzner, Philippe, Delbaen, Freddy, Eber, Jean-Marc, & Heath, David. (1997). Thinking Coherently: Generalised scenarios rather than VAR should be used when calculating regulatory capital. *RISK-LONDON-RISK MAGAZINE LIMITED-, 10*, 68-71.

Artzner, Philippe, Delbaen, Freddy, Eber, Jean‐Marc, & Heath, David. (1999). Coherent measures of risk. *Mathematical finance, 9*(3), 203-228.

Black, Fischer. (1976). *Studies of Stock Price Volatility Changes.* Paper presented at the Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics.

Bollerslev, Tim. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics, 31*(3), 307-327.

Bollerslev, Tim. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542-547.

Brown, A. (2008). Private Profits and Socialized Risk – Counterpoint: Capital Inadequacy. *Global Association of Risk Professionals*(June/July 08 issue).

Burnham, Kenneth P, & Anderson, David R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*: Springer Verlag.

Campbell, John Y, Lo, Andrew Wen-Chuan, & MacKinlay, Archie Craig. (1997). *The econometrics of financial markets*: princeton University press.

Christoffersen, Peter F. (1998). Evaluating interval forecasts. *International economic review*, 841-862.

Daníelsson, Jón. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance, 26*(7), 1273-1296.

Daníelsson, Jón. (2011). *Financial risk forecasting : the theory and practice of forecasting market risk, with implementation in R and Matlab*. Chichester: John Wiley.

Daníelsson, Jón, Jorgensen, Bjørn N, Samorodnitsky, Gennady, Sarma, Mandira, & de Vries, Casper G. (2012). Fat tails, VaR and subadditivity. *Journal of econometrics*.

Ding, Zhuanxin, Granger, Clive WJ, & Engle, Robert F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance, 1*(1), 83-106.

Dowd, Kevin. (2005). *Measuring market risk* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.

Engle, Robert F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, Robert F, & Ng, Victor K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance, 48*(5), 1749-1778.

Fernández, Carmen, & Steel, Mark FJ. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association, 93*(441), 359-371.

Ghalanos, Alexios. (2013a). Does anything NOT beat the GARCH(1,1)? Retrieved 15.02.2013, from http://www.unstarched.net/2013/01/07/does-anything-not-beat-the-garch11/

Ghalanos, Alexios. (2013b). *Introduction to the rugarch package*.

Ghalanos, Alexios. (2013c). *rugarch: Univariate GARCH models. R package version 1.0-16.*

Glosten, Lawrence R., Jagannathan, Ravi, & Runkle, David E. (1993). *On the relation between the expected value and the volatility of the nominal excess return on stocks Federal Reserve Bank of Minneapolis, Research Department staff report 157*

Hansen, Peter R, & Lunde, Asger. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of applied econometrics, 20*(7), 873-889.

J.P. Morgan. (1993). RiskMetrics Technical Manual.

Jarque, Carlos M, & Bera, Anil K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.

Jorion, Philippe. (2001). *Value at risk: the new benchmark for managing financial risk*. New York: McGraw-Hill.

Kupiec, Paul. (1995). Techniques for verifying the accuracy of risk measurement models. *THE J. OF DERIVATIVES, 3*(2).

Lambert, Philippe, & Laurent, Sébastien. (2000). Modelling skewness dynamics in series of financial data. *Institut de Statistique*.

Lambert, Philippe, & Laurent, Sébastien. (2001). Modelling Financial Time Series Using GARCH-type Models with a Skewed Student Distribution for the Innovations, Institut de Statistique, Université Catholique de Louvain: Discussion Paper 0125.

Lopez, Jose A. (1999). *Regulatory evaluation of value-at-risk models.* Paper presented at the Journal of Risk.

Mandelbrot, Benoit. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, 36*(4), 394-419. doi: 10.2307/2350970

Nelson, Daniel B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Pagan, Adrian R, & Schwert, G William. (1990). Alternative models for conditional stock volatility. *Journal of econometrics, 45*(1), 267-290.

Schwarz, Gideon. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Taylor, Stephen J. (1986). *Modelling financial time series* (Vol. 113): World Scientific.

Taylor, Stephen J. (1987). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting, 3*(1), 159-170.