Uppsala University

Department of Statistics

Master Thesis

Value at Risk on the Swedish stock market

Author:

Anton Ringqvist

Supervisor:

Patrik Andersson

UPPSALA UNIVERSITET

June 2014

Abstract

Managing and quantifying market risks has become key today for investors, financial institutions, regulators and other parties. This master thesis investigates several models that estimate the financial risk measure Value at Risk (VaR) with the objective to find the best model for the Swedish stock market. Using 1-day forecasted VaR at 95% and 99% level the following VaR models are compared: Basic Historical Simulation (HS), age weighted HS (AWHS), volatility weighted HS (VWHS) using a GARCH model, Normal VaR and t-distributed VaR. The study is performed on the Swedish stock exchange data OMXS and on the single stock series Boliden for the years 2005-2013. Running a backtest of the models it is found that the VWHS, where the volatility is modelled with a GARCH$(1,1)$ model, estimates 1-day 95% and 99% VaR most accurately on the Swedish stock market and is therefore preferred to the other models.

Keywords: Value at Risk, GARCH, OMX, Market Risk

Acknowledgement

I would like to thank my master thesis supervisor Patrik Andersson, Uppsala University, for all the great support and advices during the work with this thesis. I also want to thank Lars Forsberg and Fan Y. Wallentin at Uppsala University.

Contents

List of Figures

List of Tables

Abbreviations

VaR	**V**alue **at** **R**isk
P/L	**P**rofit & **L**osses
HS	**H**istoric **S**imulation
i.i.d	**i**ndependently & **i**dentically **d**istributed
VWHS	**V**olatility **W**eighted **H**istorical **S**imulation
AWHS	**A**ge **W**eighted **H**istorical **S**imulation
CDF	**C**umulative **D**istribution **F**unction
PDF	**P**robability **D**istribution **F**unction
GARCH	**G**eneralized **A**uto **R**egressive **C**onditional **H**eteroscedasticity
AIC	**A**kaike **I**nformation **C**riterion
HQC	**H**annan **Q**uinn information **C**riterion
ML	**M**aximum **L**ikelihood
EACF	**E**xtended **A**uto **C**orrelation **F**unction
ARMA	**A**uto **R**egressive **M**oving **A**verage
CC	**C**onditional **C**overage

1. Introduction

1.1 Background

The word "risk management" is a term that has exploded in popularity over the last ten years. Of all measures and methods incorporated under this term, *Value at Risk* (VaR) can be seen as the main measure of market risk developed to quantify financial market risk in the early 1990s. At that time the world suffered from the still ongoing economic recession that led to severe bank losses and defaults. The conclusion of these disasters was that immense amounts of money could be lost in the lack of market risk supervision and management. This led financial institutions and regulators to adopt a new risk framework, the VaR.

The probably most famous pioneer of the market risk systems was the RiskMetrics system developed by JP Morgan in 1989. They used VaR defined as the maximum likely loss over the next trading day based on standard portfolio theory and estimates of standard deviations and correlations between different financial instruments. With time the financial industry developed other models, not only based on parametric assumptions about the portfolio but also based on historical time series as well as Monte Carlo simulations.

In 1993 the non-profit international organization for financial issues "G-30" established the recommendations for managing derivatives where VaR as a risk measure played a substantial role. The Basel committee on Banking Supervision also came out with the well-known Basel rounds I,II and III in 1988, 2004 and 2011 where the latter two stress the use of VaR as market risk measure when maintaining the amount of regulatory capital a financial institution must hold.

While the suggestions are many to use VaR for market risk control, the question *what* type of VaR model performs the best remains unanswered.

1.2 Purpose of the thesis

The primarily goal of this thesis is to find the best VaR model out of a number of candidates for the Swedish stock market. Using 1-day forecasted VaR at 95% and 99% level the following VaR models are compared: Basic Historical Simulation (HS), age weighted HS (AWHS), volatility weighted HS (VWHS) using a GARCH model, Normal VaR and t-distributed VaR. The study is primarily performed on the Swedish stock exchange data OMXS for the years 2005-2013. The models will additionally be tested on the single stock series Boliden under the same time period, in order to allow for more volatility and to see how well the models behave. In the selection of the best model a backtesting analysis is carried out.

Specifically, this thesis seeks to answer the following research questions:

- Do VaR models using historical simulation outperform parametric models like the Normal VaR and t-distributed VaR?
- Do VaR estimates improve when the volatility is modelled with a GARCH model? If so, what order of the GARCH model and what underlying distribution should be used?
- Among the historical simulation models, the Normal and the t-distributed model, which one is the most reliable in estimating VaR?

The aim is to answer these questions in order to give a direction of what model to use when estimating VaR on the Swedish stock market.

1.3 Previous research

The amount of studies on VaR and its approaches has been extensive ever since financial risk modelling became popular. Beder (1995) made some early work on VaR modelling where she evaluated eight of the most common VaR models, historical simulation being one of them. The results showed that for the same portfolio the VaR estimations fluctuated heavily between different models, and Beder concluded that the outcome is much dependent on model specification and underlying assumptions.

Näsström (2003) use different generalized autoregressive conditional heteroskedasticity (GARCH) models (first proposed by Engle (1982)) to estimate VaR for different stock series and the Stockholm stock exchange index, OMX. He concludes among others that the standard GARCH$(1,1)$ performs well in estimating VaR. Schmidt and Duda (2009) searched in their master thesis to find the best VaR model out of some parametric and non-parametric models applied to three indices. In their backtesting analysis based on 250 observations they found that the Conditional Autoregressive VaR (CAViaR) introduced by Engle & Manganelli (2004) yielded the most accurate estimates for 1-day-95% VaR but was less accurate for 1-day-99% VaR. They also find an improvement in the estimation when volatility is modelled with a GARCH$(1,1)$ model.

1.4 Delimitations

There are plenty of models suggested for modelling VaR. Although it would be preferred to compare all of them in the search for the best model this thesis will only investigate the models mentioned in Section 1.2.

This thesis will not consider expected tail loss (ETL), also called Average VaR. This measure can be advantageous when the shape of the tail in the loss distribution is curly and other than slowly decaying as in the normal distribution. Expected tail loss is also a more conservative measure than VaR. In the same way extreme value theory (EVT), which has become popular when modelling extreme deviations from the median, will not be investigated in this thesis.

1.5 Outline

The structure of the thesis includes a theory section, a results section where the empirical findings are presented, and a conclusion. The theory section covers the theory of VaR, the different types of models and distributions used as well as the concept of backtesting. This follows by an empirical section where the backtesting is implemented on the models and an accompanying discussion. The final section is a sum-up of the thesis and a conclusion of what models to prefer when measuring VaR on the Swedish stock market.

2. Method

2.1 Data

The primary OMXS dataset contains a sample of 2262 daily closing price returns of the Stockholm stock exchange index OMXS from January 3rd, 2005 to December 31st, 2013. The index includes all stocks noted on the nordic exchange in Stockholm and was renamed from SAX to OMXS in October 3rd, 2005. The index basis starts on December 29th, 1995 with the value 100. Boliden is a Swedish metal company working mainly with export of zinc, copper and precious metals to the European market. The stock, with ISIN number SE0000869646, is noted on the Stockholm stock exchange. The dataset contains a sample of 2262 daily closing prices in SEK with the same length and dates as the OMXS one mentioned above. All data was collected from Thomson Reuters Datascope.

2.2 VaR estimation

The first five years of the datasets, January 3rd 2005 to 30th December 2009, containing 1256 observations are used to estimate VaR for the following day, January 4th 2010. The five year window is then shifted one day forward up until the last observation December 31st, 2013. This yields 1006 VaR estimates which will then be compared in the backtesting analysis with the 1006 actual returns on these days.

3. Theory

3.1 Introduction to Value at Risk

Value at Risk (henceforth abbreviated to VaR) is a financial risk measure used to calculate the potential maximum loss one might lose with a certain probability over a given time horizon. Let us illustrate this with an example. Imagine we have a one year time series of daily profits and losses (P/L) for a company. If we would plot all these P/L in a histogram we would probably end up with a bell shaped curve as the one displayed in Figure 3.1. There we see most of the days have returns around zero, and only a very few days have very extreme losses and profits respectively. If we would want to look at the 1-day 95% VaR here we would make a cut right between the bottom 5% and the top 95% of the observations. In Figure 3.1, which here happens to be a standard normal distribution, this refers to the value -1.645, meaning our 1-day 95% VaR is 1.645 for this time period. Note the positive sign of VaR which represents the loss of 1.645. In brief, if we want to calculate VaR we always need to specify two things:

- A holding or horizon period, being the period of time over which we measure the profit or loss of a portfolio. This is usually a 1-day or 10-day horizon but can also be weekly or monthly.
- A confidence level α, which will represent the probability of not getting a worse loss than our VaR. The confidence level is often set to 95% or 99% but can assign any fraction between 0 and 1.

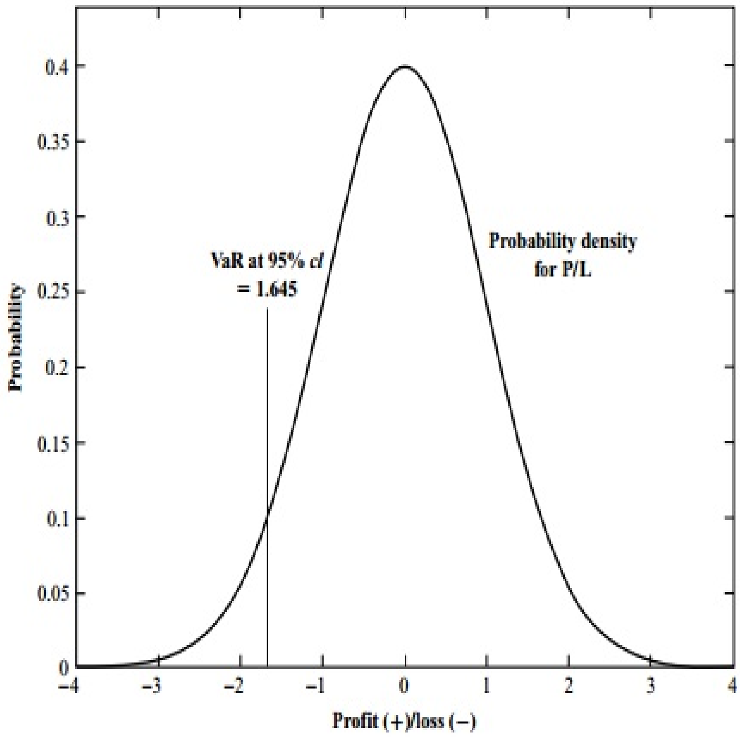

FIGURE 3.1: *VaR at a 95% confidence level (cl). Source: Dowd, 2013.*

A 1-day 95% VaR of let us say USD 100 million means in 95 out of 100 days we expect to lose not more than 100 million. Another way to look at it is to say in 5 out of 100 days we will lose 100 million or more. Based on this information a company, financial institution or investor can get a better overview of how risky the portfolio is and adjust the risk appetite accordingly.

VaR can also be expressed for long and short positions[1]. Let us say at the time t we want to estimate the risk of a financial position for the next k periods. Let $\delta V(k)$ denote the value change of the position from time t to $t+k$, α the confidence level, and let the CDF of $V(k)$ be $F_k(x)$. VaR for a long position can then be expressed as

$$1 - \alpha = Pr[\delta V(k) \leq VaR] = F_k(VaR) \tag{3.1}$$

[1]A long position in financial terms means the buying of a security such as a stock with the expectation that the asset will rise in value. A short position means the sale of a borrowed security with the expectation that the asset will fall in value.

and for a short position

$$1 - \alpha = Pr[\delta V(k) \geq VaR] = 1 - Pr[\delta V(k) \leq VaR] = 1 - F_k(VaR). \qquad (3.2)$$

Note that for the long position the left tail of $F_k(x)$ is of interest and VaR typically represents a negative value, while for the short position the right tail is interesting and VaR is a positive value (which for a short position means a loss). (Tsay, 2005)

Going forward in this thesis, as we will only handle pure P/L in terms of index returns we will not discuss positions any further, and VaR will be a positive value representing a loss in the left tail of the distribution.

3.2 Criticism to VaR

Although VaR is a good tool for measuring and monitoring market risk it has its limitations. First of all VaR is often criticised for not saying anything about the absolute worst loss, i.e. we don't know how much more money we can lose beyond our VaR. Another drawback when measuring historical VaR is that the chosen in-sample time period needs to reflect the future in a good way, an assumption that may not hold in many cases (Jorion, 2009). Dowd (2013) also stresses the danger in relying blindly on VaR when the measure is in general too blunt to provide a complete presentation of the risks. VaR will also differ from model to model depending on what framework and assumptions you rely on, something that makes the measure even more uncertain. As Taleb (1997) expresses it: "Youre worse off relying on misleading information than on not having any information at all. If you give a pilot an altimeter that is sometimes defective he will crash the plane. Give him nothing and he will look out the window."

3.3 VaR models

Many different approaches to model VaR have been proposed over the last decades. In this thesis we will delimit ourselves to compare five models. The first two are the Basic Historical Simulation (HS) and the age weighted HS (AWHS). These models do not make any assumption about from what specific probability distribution the returns come from. The third model is the volatility weighted HS (VWHS), which models volatility with a GARCH-model. For the distribution of the innovations in the volatility equations three distributions will be tested: the normal distribution, the t-distribution and a skewed version of the t-distribution. The last two models assume the log-returns are normal and t-distributed respectively. Note that for simplicity, going forward when we speak of returns r we actually mean the daily log-returns $\log(\frac{V_t}{V_{t-1}})$ where V_t is the OMXS index value (or the Boliden closing price) at time t and V_{t-1} the index value the previous day. Indeed for financial institutions and other VaR-consuming parties it is of interest to measure the risks in pure P/L or returns, not log-returns. However, since this thesis primarily focuses on finding the best VaR model using backtesting the scale of the observations is less important, and one can always get back to the returns from the log-returns by a simple transformation. With this said, going forward we will work with the log-returns denoted with r and for a justification of this see Appendix B.3.

3.3.1 Basic historical simulation

The most naive and straight forward model is the basic historical simulation, already mentioned in the introduction. Imagine we have a time series of market returns r starting back in time at r_1 and up to r_{t-1} which was yesterday. If we want to calculate VaR for today t we simply take the $(1-\alpha)$-percentile $p_{1-\alpha}$ of the returns r and calculate

$$VaR = p_{1-\alpha}(r_{t-1}, r_{t-2}, ..., r_1). \tag{3.3}$$

In other words, if our series contains 100 daily observations our 1-day 95% VaR will be the value of the sixth biggest loss. (Dowd, 2013)

While HS is easy to implement, it has its drawbacks. The main problem is the assumption that the market returns are IID, something which is often far from reality. In fact, the market returns often have autocorrelation where the market return today is partly dependent on yesterday's return. (Ding et al., 1993)

3.3.2 Volatility-weighted historical simulation

To solve the autocorrelation problem Hull and White (1998) propose to filter the market returns with the volatility changes from the period the historical data covers. This derives from the observation that the distribution of a market variable scaled by an estimate of its own volatility often approximately becomes stationary, meaning that the historical simulation could profit from this filtering when estimating VaR. If the market volatility was 5% three months ago while it is rather 2% today, using the observations two months ago to estimate tomorrow's market changes will clearly overstate it. In the same way the reverse holds.

Let us denote tomorrow's time point with T. Let r_t be the historical return at time t where $t < T$, σ_t^2 denotes the historical volatility estimate for time t and σ_T^2 the volatility estimate for tomorrow. If we assume the probability distribution of $\frac{r_t}{\sigma_t}$ is stationary then we can replace each r_t by r_t^* where

$$r_t^* = \sigma_T \frac{r_t}{\sigma_t}. \tag{3.4}$$

By substituting all the historical returns with the volatility updated ones we can then estimate VaR using the series of r_t^*. (Hull and White, 1998)

How do we estimate the volatility σ_t? Hull and White suggest a Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model. This model will be further explained in Section 3.5.

3.3.3 Age-weighted historical simulation

While the basic historical simulation assigns the same weight to every observation back in time in the sample period, it could be of greater interest to put more weight on recent observations and less to the older ones. Boudoukh et al. (1998) suggest to exponentially weight each of the K most recent returns according to

$$r^w_{t-K+1} = r_{t-K+1}\frac{1-\delta}{1-\delta^K}\delta^{K-1} \tag{3.5}$$

where δ is a constant between 0 and 1 reflecting the exponential decay, r the historical return and r^w the age weighted historical return. Using this method Boudoukh et al. (1998) find that the age weighting yields significantly better VaR estimates than unweighted HS.

There are several advantageous with this method. First of all a suitable choice of δ can make the estimates better at handling recent large losses (and profits) since they will be weighted more heavily than under basic HS. Age-weighted HS also reduces the distortion effect from events back in time that are unlikely to happen again, since the weights go towards zero the longer back in time one gets. Moreover, with age-weighting it is possible to let the sample period grow over time with every new observation and in this sense not lose valuable information, as well as preventing big jumps in VaR from happening as an extreme observation leaves the time window (since that observation would have a weight close to zero). On the other hand age-weighting has been found to still produce insufficient VaR estimates when volatility changes (Pritsker, 2001). Furthermore the method also reduces the effective sample size. (Dowd, 2013)

3.3.4 Normal distributed VaR

If the assumption of normality holds for the returns (recall we actually model the log-returns as normal) VaR can be calculated by estimating the two first moments μ and σ^2 with $\bar{x}$ and s^2. The estimation method can be either Least Squares regression or maximum likelihood[2]. VaR at confidence level α is then retrieved by taking

$$VaR(\alpha) = -\bar{x} + s\Phi_z(\alpha) \tag{3.6}$$

where $\Phi_z(\alpha)$ is the standard normal variate corresponding to the chosen confidence level α. If the confidence level is 95% then $\Phi_z(\alpha) = -1.65$ and if the confidence level is 99% then $\Phi_z(\alpha) = -2.33$. (Dowd, 2013)

Empirics show that market returns typically have fatter tails then those of a normal distribution (Huang and Lin, 2004). Hence a model like the t-distribution, which is similar to the normal in its shape but allows for excess kurtosis, could yield better estimates.

3.3.5 t-distributed VaR

The Student-t distribution was developed as a probability distribution for normally distributed random variables that did not involve the population variance σ^2 (Newbold et al., 2007). The single parameter ν represents the number degrees of freedom, which in turn controls the kurtosis κ[3]

We choose the degrees of freedom ν so that it fits the empirical kurtosis according to

$$\nu = \frac{4\kappa - 6}{\kappa - 3} \tag{3.7}$$

given that $\nu \geq 5$. In the risk measurement purpose we work with a t-distribution where we use our empirically found sample mean $\bar{x}$, variance s^2 and kurtosis κ. These sample moments can be estimated using Maximum Likelihood. Given these parameter estimates we can now calculate VaR via

$$VaR_t(\alpha) = -\bar{x} + s\sqrt{\frac{(\nu - 2)}{\nu}}t_{\nu,\alpha} \tag{3.8}$$

[2]both estimation methods yield best linear unbiased estimators, (BLUE).

[3]The kurtosis, or the steepness of a distribution, is the normalized fourth central moment and measures the weight in the tails relative to the center. Population kurtosis for the normal distribution is 3. (Newbold et al., 2007)

where t is the variate corresponding to the chosen degrees of freedom ν and confidence level α. (Dowd, 2013)

As the number of degrees of freedom ν gets large the t-distribution converges to the normal distribution, as can be seen in Figure 3.2. The figure also shows the fatter tails for the t-distributions with lower degrees of freedom, allowing for a higher probability of extreme observations.

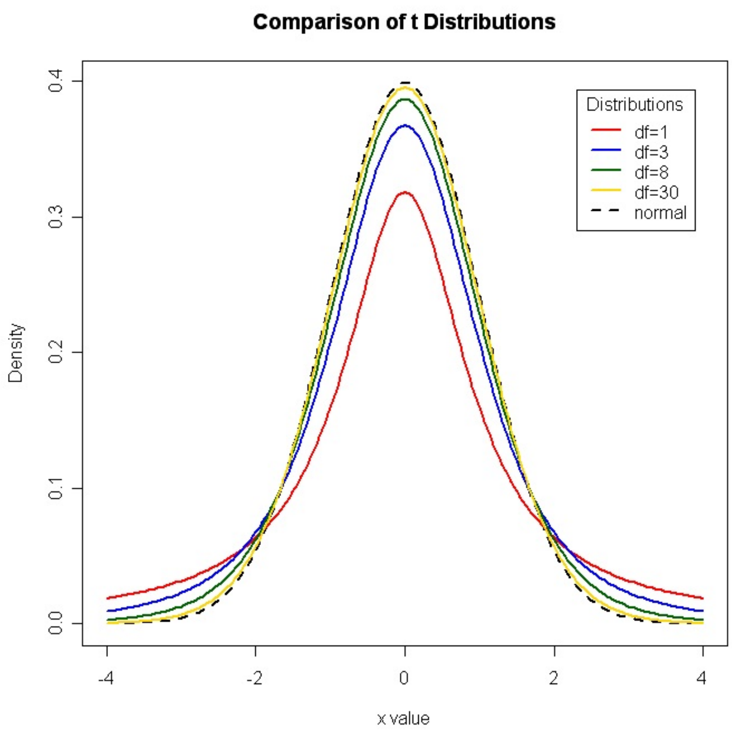

Figure 3.2: *t-distributions with different degrees of freedom "df" displaying the convergence to the normal distribution as df*$\rightarrow \infty$.

Figure 3.3 additionally displays how VaR at 95% and 99% varies with the degrees of freedom.

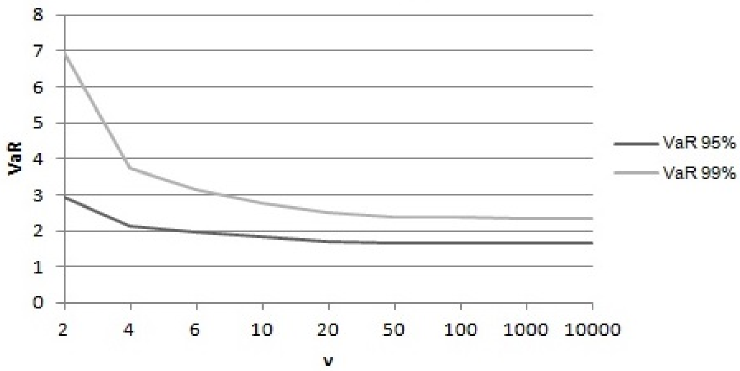

FIGURE 3.3: *t-VaR at different degrees of freedom.*

3.3.6 The skewed t-distribution

Empirics often show that the distribution of daily financial returns may not only be fat-tailed but also skewed with a larger probability mass in one tail and a smaller mass in the other one. In order to capture this behaviour a probability distribution of this asymmetric form is needed. Hansen (1994) introduces a generalized t-distribution with density function

$$f(x \mid \nu, \lambda) = \begin{cases} bc\Big(1 + \frac{1}{\nu-2}\big(\frac{bz+a}{1-\lambda}\big)^2\Big)^{-(\nu+1)/2}, & \text{if } z < -a/b \\ bc\Big(1 + \frac{1}{\nu-2}\big(\frac{bz+a}{1+\lambda}\big)^2\Big)^{-(\nu+1)/2}, & \text{if } z \geq -a/b \end{cases} \tag{3.9}$$

where $\nu > 2$ is the degrees of freedom, λ the skewness parameter with range $-1 < \lambda < 1$, and a, b and c defined as

$$a \equiv 4\lambda c \frac{\nu-2}{\nu-1}, \qquad b^2 \equiv 1 + 3\lambda^2 - a^2, \qquad c \equiv \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)}.$$

If the skewness parameter λ is close to -1 the distribution is skewed to the left and vice versa. If λ equals 0 the distribution collapses to the symmetrical, traditional t-distribution. Figure 3.4 shows some examples of how the distributions look like for different values of λ.

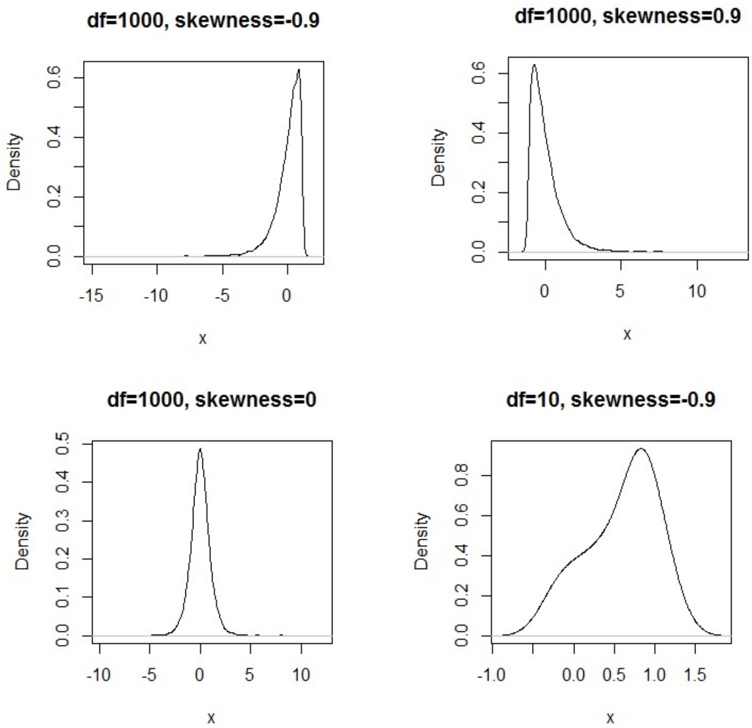

Figure 3.4: *Generalized t-distributions with skewness demonstratitng the different tail heavinesses.*

3.4 Disadvantages of Normality

The normal distribution provides an easy way to model market returns as Stahl et al. (2006) report when they conclude even though single components of a trading portfolio can be non-normal the top aggregation of a financial institution's positions usually produces symmetric, normal-looking distributions. However, there are two strong arguments against using normality as an assumption for market returns. Firstly, as already mentioned, financial returns don't fully seem to behave normal and often underestimate VaR due to the fatter tails (Manganelli and Engle, 2001). For example, Hendricks (1996) evaluates VaR on foreign exchange rates concluding that "Actual 99th percentiles for the foreign exchange portfolios considered in this article tend to be larger than the normal distribution would predict.". Secondly, since there are no boundaries of what values an observation can take in the normal distribution losses can theoretically get bigger than the total value of our investment, which is not realistic. (Dowd, 2013)

3.5 ARCH and GARCH models

Already in 1963 the Polish mathematician Benoit Mandelbrot noticed that market returns behaved in a clustered way where "...large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes...". (Mandelbrot, 1963). Such findings led Engle (1982) to propose the autoregressive conditional heteroscedasticity (ARCH) model where the changing variance of a time series is captured. Let us denote the conditional variance, or the *conditional volatility* of our return series r_t with $\sigma^2_{t|t-1}$ where subscript $t-1$ symbolises that the condition is upon returns at time $t-1$. Moreover, the squared return r_t^2 is an unbiased estimator of $\sigma^2_{t|t-1}$. We can then express the ARCH model as a regression model where the conditional volatility is the response variable and the lagged squared returns the covariates. An ARCH(1) model would hence look like

$$r_t = \sigma_{t|t-1}\varepsilon_t \tag{3.10}$$

$$\sigma^2_{t|t-1} = \omega + \alpha r^2_{t-1} \tag{3.11}$$

where α and ω are parameters to be estimated and $\{\varepsilon_t\}$ a sequence of i.i.d random variables with zero mean and unit variance where ε_t is independent of $r_{t-j}, j = 1, 2, ...$ Engle (1982) also proposed a more generalised version of the above mentioned equation, the ARCH(q) model where q lags of the squared returns are included according to

$$\sigma^2_{t|t-1} = \omega + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \ldots + \alpha_q r^2_{t-q}. \tag{3.12}$$

Here, q is called the "ARCH order". In addition to this, Bollerslev (1986) and Taylor (1986) further developed the GARCH model where p lags of the conditional variance are included. The GARCH(p, q) model hence looks like

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \ldots + \beta_p \sigma^2_{t-p|t-p-1} + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \ldots + \alpha_q r^2_{t-q} \quad (3.13)$$

where q is the ARCH order, p the GARCH order and α, β and ω are positive parameters to be estimated. (Cryer and Chan, 2008)

3.5.1 GARCH$(1, 1)$ model

The simplest GARCH model is the GARCH$(1, 1)$ model:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 r^2_{t-1} \quad (3.14)$$

with the parameter restriction $\alpha_1 + \beta_1 < 1$ in order for the weak stationarity condition to hold.[4] If we set $E(r^2_{t-1}) = \sigma^2_t = \sigma^2_{t-1} = \sigma$ in Equation 3.14 and solve for σ we get the average, *unconditional* variance:

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}. \quad (3.15)$$

The GARCH$(1, 1)$ model is according to Dowd (2013) as well as Jorion (2009) easy to apply since it uses a small number of parameters, and it is often found to fit the data well. Hansen and Lunde (2005) also don't find evidence that more sophisticated ARCH models outperform a GARCH$(1, 1)$ in their comparison of 330 ARCH models.

In Equation 3.14, a high β-value implies a clustering, persisting volatility that takes time to change, and a high α-value means that volatility quickly reacts to market moves. Estimation normally yields $\alpha < 0.25$ and $\beta > 0.7$. (Dowd, 2013)

One drawback of the GARCH model is that it is nonlinear and one needs to use maximum likelihood and numerical optimization in order to estimate the parameters. This require us to specify the distribution of the error term ε_t seen in the basic ARCH specification,

[4] For a derivation of this condition see Appendix A.1.

Equation 3.10. A common assumption is to say that the error terms are either normally distributed or t-distributed, and the two ML estimations are presented below.

3.5.1.1 GARCH$(1,1)$ ML estimation

Recall the GARCH$(1,1)$ equation for the conditional variance

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 r^2_{t-1} \tag{3.16}$$

where we seek to estimate α, β and ω. Here $t \geq 2$ and the initial value $\sigma_{1|0}$ is set to the unconditional variance $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$. If the error term ε_t in Equation 3.10 is normally distributed the conditional pdf of the returns is

$$f(r_t|r_{t-1},\ldots,r_1) = \frac{1}{\sqrt{2\pi\sigma^2_{t|t-1}}} exp[-r^2_t/(2\sigma^2_{t|t-1}]. \tag{3.17}$$

If the time series reaches from $t = 1, 2, \ldots, T$ the joint pdf can be written as

$$f(r_T,\ldots,r_1) = f(r_{T-1},\ldots,r_1) f(r_T|r_{T-1},\ldots,r_1). \tag{3.18}$$

Taking the logs of Equation 3.18 yields the log-likelihood function

$$\mathcal{L}(\omega,\alpha,\beta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{T}\Big(\log(\sigma^2_{t-1|t-2}) + r^2_t/\sigma^2_{t|t-1}\Big) \tag{3.19}$$

and by maximizing Equation 3.19 numerically (since there is no closed-form solution) estimates of ω, α and β can be obtained.

In the same way we can obtain the log-likelihood function for a GARCH$(1,1)$ model if ε_t is assumed to follow a t-distribution with ν degrees of freedom:

$$\mathcal{L}(\omega,\alpha,\beta) = T\log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}\right) - \frac{1}{2}\sum_{i=1}^{T}\log\sigma^2_{t-1|t-2} \\ - \frac{v+1}{2}\sum_{i=1}^{T}\log\left(1+\frac{r_t^2}{\sigma^2_{t|t-1}(\nu-2)}\right) \quad (3.20)$$

where Γ is the gamma function and ν the degrees of freedom determined by Equation 3.7.

3.5.2 Selection of GARCH model: Information criteria

An information criteria is used to evaluate how good a specified model fits a dataset. Based on this information one can select what model to use. Although there are many different information criterion some have received more attention. For models with GARCH effects Javed and Mantalos (2013) suggest to use the Akaike Information Criteria (AIC) for GARCH models of higher (p,q)-dimension, and Hannan-Quinn information criterion (HQC) for low dimensional models. AIC is defined as

$$AIC = -2\log(\iota) + 2k \quad (3.21)$$

where ι denotes the maximized value of the likelihood function and k the number of parameters in the model. In choosing between several competing models according to AIC one should select the model with lowest AIC value. The first term of Equation 3.21 measures the model's goodness of fit while the second is the *penalty function* that penalizes models with more parameters, preventing model overfitting. (Akaike, 1974)

The HQC is close to the AIC and is given by

$$HQC = -2\log(\iota) + 2k\log(\log(T)) \quad (3.22)$$

where ι denotes the maximized value of the likelihood function, k the number of parameters in the model and T the number of observations. Like AIC the lowest HQC value implies the best model fit. (Hannan and Quinn, 1979)

3.5.3 Selection of GARCH model: Extended Autocorrelation Function

Another way to determine what GARCH model to choose given a dataset is to consider the connection between an autoregressive moving average (ARMA) model and a GARCH model, where a GARCH(p, q) for the returns is the same as an ARMA($max(p, q), p$) model for the *squared* returns.[5] This gives us the advantage of using identification techniques for ARMA models in our search for a suitable GARCH model.

Tsay and Tiao (1984) propose an iterative identification method using the extended autocorrelation function (EACF). First, we start by assuming we have an ARMA process of order k and j for our returns r, where the true dimensions are p and q. The autoregressive residuals $w_{t,k,j}$ can then be expressed as

$$w_{t,k,j} = r_t - \tilde{\phi}_1 r_{t-1} - \ldots - \tilde{\phi}_k r_{t-k} \tag{3.23}$$

where $\tilde{\phi}_i$ are the estimated AR coefficients given the initial ARMA(k, j) specification. For $k = p$ and and $j \geq q$, $\{w_{t,k,j}\}$ will approximately be an MA(q) model, and all theoretical autocorrelations of lag $q + 1$ or greater equal zero. If $k > p$ we will overfit the model and this will increase the MA dimension for $\{w_{t,k,j}\}$ by $min[k - p, j - q]$. By summarizing these conditions in a table we end up with something similar to Figure 3.5 where the element in row k and column j is denoted with a $\times$ if the lag $j + 1$ sample correlation of $w_{t,k,j}$ is significantly different from zero. If not different from zero, the element will be populated with a 0. This will yield a pointy triangle pattern made of 0:es in the table where the upper-left peak signifies the ARMA-order p and q. (Tsay and Tiao, 1984)

[5] To see this refer to Section A.1.

While the theoretical pattern in Figure 3.5 clearly indicates an ARMA$(1,1)$ model a sample EACF can often be hard to interpret.

AR/MA	0	1	2	3	4	5	6	...	j
0	x	x	x	x	x	x	x	x	x
1	x	0	0	0	0	0	0	0	0
2	x	x	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0
6	x	x	x	x	x	x	0	0	0
...	x	x	x	x	x	x	x	0	0
k	x	x	x	x	x	x	x	x	0

FIGURE 3.5: *Theoretical extended autocorrelation function for an ARMA*$(1,1)$ *model.*

3.6 Backtesting VaR

A VaR model is only accurate if the proportion of observations falling outside VaR equals or is close to 1 minus the confidence level, $\rho = 1 - \alpha$. This is tested by backtesting the model.

Let us say we have T days of daily returns r in our time series. We then denote the number of *backtesting exceptions*, i.e. the number of times VaR is exceeded in our sample, with X. If we measure the 1% of the worst observations ρ (i.e. the left tail), or equivalently the 1-day 99% VaR, for a total of T days we can compute the failure rate $\frac{X}{T}$. Optimally we want $\frac{X}{T} \to \rho$ as T increases. This procedure gives us a way to decide how accurate a VaR model is; if our observed $\frac{X}{T}$ is substantially higher than ρ we say the model underestimates the risk and if $\frac{X}{T}$ is much lower the model is conservative. ρ is often called the nominal size while $\frac{X}{T}$ is called the empirical size. (Jorion, 2009)

How do we decide if a model's estimates are *too* far off from what we expect? Since the event of a backtesting exception can be seen as success or failure over T trials with probability ρ we can say the number of exceptions X follows a binomial distribution:

$$f(x) = \binom{T}{x} \rho^x (1-\rho)^{T-x} \tag{3.24}$$

where $E(x) = \rho T$ and $V(x) = \rho(1-\rho)T$. Using the central limit theorem[6] we can approximate the binomial distribution by the normal distribution as T gets large via

$$z = \frac{x - \rho T}{\sqrt{\rho(1-\rho)T}} \sim N(0,1) \tag{3.25}$$

where z is standard normal. Imagine we have $\rho = 0.02$ and $T = 1000$, meaning we expect to see $0.02 \times 1000 = 20$ exceptions. Under the null-hypothesis of $H_0 : \rho = 0.02$ versus the alternative hypothesis $H_1 : \rho \neq 0.02$ we would reject H_0 if $z \leq -2.33$ or if $z \geq 2.33$ (with 1% in each tail). A rejection here is interpreted as an indication of an inaccurate model (given the dataset) where the model under- or overestimates risk. (Jorion, 2009)

3.6.1 Kupiec test

Kupiec (1995) suggests a test[7] in line with the procedure above but where the cut-off regions are calculated using the log-likelihood ratio:

$$LR_K = -2\log((1-\rho)^{T-X}\rho^X) + 2\log((1-(X/T))^{T-X}(X/T)^X) \tag{3.26}$$

which is asymptotically $\chi^2(1)$-distributed under the null of ρ being the true exception probability. We reject the null-hypothesis if $LR_K \geq F^{-1}_{\chi^2_1}(1-\alpha_K)$ where $F^{-1}_{\chi^2_1}(Pr)$ is the inverse of the χ^2-CDF with 1 degree of freedom at the probability Pr, and α_K the significance level of the test. Table 3.2 displays some calculated regions for number of backtesting exceptions X allowed without rejecting the null.

[6] For an explanation of the central limit theorem see Appendix A.2.
[7] also called the Proportion of Failure-test, or the POF-test.

Probability level ρ	VaR confidence level α	$T = 252$ days	$T = 510$ days	$T = 1000$ days
0.01	99%	$X < 7$	$1 < X < 11$	$4 < X < 17$
0.05	95%	$6 < X < 20$	$16 < X < 36$	$37 < X < 65$
0.10	90%	$16 < X < 36$	$38 < X < 65$	$81 < X < 120$

Table 3.2: *Nonrejection regions for number of backtesting exceptions X. Source: (Kupiec, 1995)*

The Kupiec test is called an *unconditional coverage* model since it doesn't regard time variation in the data. It is however of interest to investigate if the exceptions are randomly spread over the time series or if they tend to cluster. If the latter is the case an unconditional coverage model will be invalid and one should instead use a *conditional coverage* model(Dowd, 2013). One of these models is developed by Christoffersen (1998) and will be presented in the following section.

3.6.2 Christoffersen's test of conditional coverage

The conditional coverage (CC) test developed by Christoffersen (1998) involves two tests of which the first, unconditional test coincides with the Kupiec test described earlier. The second test is a *test of independence* which tests serial independence between the backtesting exceptions. Let T_{ij} be the number of days where state j occurred while it was in state i on the previous day, where t, i assign 1 if a backtesting exception was observed and 0 if not. Moreover, let π_i denote the probability of observing an exception given the previous day's state i. Table 3.3 shows how the conditional expected T_{ij}-values can be calculated. The relevant test statistic is then

$$LR_{ind} = -2\log\left((1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}\right) + 2\log((1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}\right) \quad (3.27)$$

where LR_{ind} is the log-likelihood ratio for the independence test with null-hypothesis $H_0 : \pi = \pi_0 = \pi_1 = (T_{01} + T_{11})/T$, i.e. that the probability of getting an exception on a specific day is independent of previous days. Christoffersen (1998) show that this test statistic asymptotically is χ^2-distributed with 1 degree of freedom. Recall that the Kupiec test of unconditional coverage also was $\chi^2(1)$-distributed, making the joint test be asymptotically $\chi^2(2)$-distributed[8]:

$$LR_{cc} = LR_K + LR_{ind} \xrightarrow{d} \chi^2(2). \tag{3.28}$$

Hence we reject accuracy of exception probability ρ and independence of the exceptions if $LR_{cc} \geq F^{-1}_{\chi^2_2}(1-\alpha_{cc})$ where $F^{-1}_{\chi^2_2}(Pr)$ is the inverse of the χ^2-CDF with 2 degrees of freedom at the probability Pr, and α_{cc} the significance level of the test.

		Previous day	
		No exception	Exception
Current day	No exception	$T_{00} = T_0(1-\pi_0)$	$T_{10} = T_1(1-\pi_1)$
	Exception	$T_{01} = T_0(\pi_0)$	$T_{11} = T_1(\pi_1)$
	Total	T_0	T_1

Table 3.3: *Exception table with expected number of exceptions. T_{01} for example corresponds to the expected number of days where one day had an exception while there was no exception on the previous day.*

3.6.3 Statistical power of VaR backtests

The power of a statistical test is the probability that the test leads to rejection of the the null hypothesis H_0 when H_0 is false, i.e.

$$power = P(reject\ H_0 \mid H_0\ is\ false). \tag{3.29}$$

Figure 3.6 shows a typical power curve for the test of a parameter θ where we have $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. (Wackerly et al., 2008)

[8] Since the sum of $\chi^2(a)+\chi^2(b)$ equals $\chi^2(a+b)$.

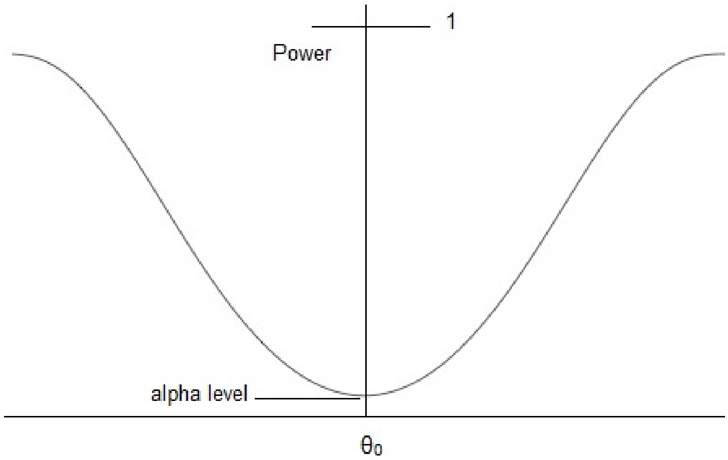

Figure 3.6: *Typical power curve for the test of a parameter θ where we have $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.*

Nordbo et al. (2012) investigate the power for the VaR Geometric backtest (which is an extension of Christoffersen's test) and discover among other that the more amount of backtesting data the higher power of the test. Hence they come up with minimum sample sizes being 750 when backtesting 95% VaR and 1000 for 99% VaR, showing that the power of the test shrinks when we test a smaller part of the tail. They also conclude backtesting with only 100 or 250 observations has too low power to be sufficient.

4. Results

In this results section the output of the 1-day 95% and 99% VaR backtesting analysis of the OMXS and Boliden data will be presented. The calculations were performed in the statistical software R. For the GARCH estimations the R package fGarch was used[9]. For the volatility-weighted models the $GARCH(1,1)$ parameters are reestiamted for every step in the shifting time series window, i.e. 1006 times.

For the Kupiec's test and Christoffersen's test a significance level of 5% will be used.

4.1 Descriptive statistics of the OMXS data

Table 4.4 provides descriptive statistics for the full dataset of daily OMXS log-returns starting at January 3rd, 2005 and ending at December 31st, 2013. A graph of the full dataset is also available in Appendix B.1.

Series: OMXS Sample: 03-01-2005 to 31-12-2013	
Observations	2262
Mean	0.0002701840
Median	0.0009229750
Maximum	0.0863063620
Minimum	-0.0738152080
Range	0.1601215700
Std. dev.	0.0142410794
Skewness	-0.06394976
Kurtosis	7.506046
Jarque-Bera	1915.236
p-value	0.00000000

Table 4.4: *Descriptive statistics for the Stockholm Stock Exchange index series OMXS ranging from January 3rd, 2005 to December 31st, 2013.*

[9] For a full documentation see http://cran.r-project.org/web/packages/fGarch/fGarch.pdf

4.2 GARCH choice

Applying the extended autocorrelation function on the OMXS in-sample yields Figure 4.7 which suggests a GARCH(1,1) model for the log-returns, although it is not fully clear. Table 4.5 further shows AIC and HQC values (where the lowest value suggests the best model) as well as number of non-significant parameters for different GARCH(p,q) models. Recall the lower AIC and HQC value the better model fit we should have. Although AIC and HQC values are slightly lower for some models than for the GARCH(1,1) model, the GARCH(1,1) is the only model with all its parameters significant. Hence a GARCH(1,1) model seems to be most appropriate and will therefore be used going forward.

AR/MA	0	1	2	3	4	5	6	7	8
0	x	x	x	x	x	x	x	x	x
1	x	0	x	x	x	0	0	0	x
2	x	0	x	0	x	x	0	x	x
3	x	x	x	0	x	0	0	x	0
4	x	x	x	x	x	0	0	x	x
5	x	0	x	x	0	0	0	x	x
6	x	0	0	x	0	0	0	x	x
7	x	x	0	x	x	x	x	x	0

FIGURE 4.7: *Extended autocorrelation function for the in-sample, indicating an ARMA$(1,1)$ model and hence a GARCH(1,1) model.*

	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)	GARCH(2,3)	GARCH(3,2)	GARCH(3,3)
AIC	-5.605	-5.605	-5.609	-5.611	-5.610	-5.610	-5.609
HQC	-5.600	-5.599	-5.603	-5.603	-5.601	-5.601	-5.598
z	0	1	1	1	4	3	5

TABLE 4.5: *Different GARCH(p,q) models with their respective AIC and HQC values as well as number of non-significant parameters z.*

4.3 Backtesting results of the OMXS

Table 4.6 presents the VaR backtesting results on the OMXS data for the tested models: Basic historical simulation (BHS), Age weighted historical simulation (AWHS), VaR under Normal distribution as well as t-distribution, and volatility weighted historical simulation (VWHS) using a $GARCH(1,1)$ model under normal, t- and skewed t-distribution. Values in bold in the table for the three tests are cases where the null-hypotheses of a good model fit can't be rejected on the 5% significance level.

1-day 95% VaR	BHS	AWHS	Normal	t	VWHS (Normal)	VWHS (t)	VWHS (skewed t)
# of BT exceptions X	24	71	24	54	51	53	53
# of X:es after an X	5	14	5	11	5	5	5
Empirical size $\frac{X}{T}$ in %	2.386	7.058	2.386	5.368	5.070	5.268	5.268
Kupiec test (p-value)	0.000	0.005	0.000	**0.597**	**0.920**	**0.699**	**0.699**
Independence test (p-value)	0.000	0.000	0.000	0.000	**0.155**	**0.204**	**0.204**
Joint test CC (p-value)	0.000	0.000	0.000	0.000	**0.363**	**0.414**	**0.414**
1-day 99% VaR	BHS	AWHS	Normal	t	VWHS (Normal)	VWHS (t)	VWHS (skewed t)
# of BT exceptions X	5	0	9	14	10	10	10
# of X:es after an X	0	0	0	1	0	0	0
Empirical size (%)	0.497	0.000	0.895	1.392	0.994	0.994	0.994
Kupiec test (p-value)	**0.076**	0.000	**0.732**	**0.239**	**0.985**	**0.985**	**0.985**
Independence test (p-value)	**0.823**	NaN	**0.687**	**0.185**	**0.654**	**0.654**	**0.654**
Joint test CC (p-value)	**0.201**	NaN	**0.870**	**0.207**	**0.904**	**0.904**	**0.904**

Table 4.6: *Summary table of the 1-day 95% and 99% backtesting analysis of the OMXS data. Values in bold symbolize p-values above the 5% significance level, i.e. non-rejection of the null-hypothesis of a good model fit. The critical χ^2-value for which we reject H_0 if the test statistic exceeds it is 3.8415 for the Kupiec and independence test and 5.9915 for the joint test, corresponding to a 5% significance level with 1 df and 2 df respectively.*

First of all, table 4.6 clearly shows how the volatility weighted GARCH models perform better than the other models for the 95% VaR. The empirical sizes are all close to 5% and since they pass the independence test the backtesting exceptions seem spread out independent from each other. It is hard to say what underlying distribution is to prefer for the GARCH model: the model under normality has the smallest deviance between empirical and nominal size (5.070% vs 5%) but the two models under t-distribution and skewed t-distribution score lower in the joint test due to a what seems to be a higher degree of independence between the exceptions. However one should here prefer the GARCH model under normality here since its empirical size is the closest one to the nominal size while it still passes the independence test. Regarding the identical results between GARCH under t and skewed t it seems like the skewness parameter is of small or no importance with respect to the VaR estimate, although the skewness estimate in the first step (as an example) is 1.006 and clearly significant. Nevertheless, the choice between a GARCH(1,1) under normality, t or skewed t seems to have little impact on the overall result. Turning to the other four models none of them passes Christoffersen's test of conditional coverage (the joint test). The normal distribution clearly underestimates VaR detecting only 24 exceptions where the target is close to 50. This is expected since market returns tend to follow a more fat-tailed distribution leading the normal distribution to understate the risk. The t-distribution, which do have fatter tails, actually passes the Kupiec test with the empirical size close to the nominal one but the independence test undoubtedly rejects independence between the backtesting exceptions. Hence the underlying assumption of i.i.d returns of the OMXS series does not seem valid. Both the BHS and AWHS are rejected in the three tests.

The backtesting results of the 99% VaR show that all models apart from the AWHS pass all the tests. This could mean all the models are fairly good at estimating the 1% of the left distribution tail, and it is rather the shape of the 2nd-5th percentile that is hard to capture. On the other hand, as explained in Section 3.6.3, when we test the 1% tail a minimum of 1000 observations is suggested and this is just what we have. This means the power of the test could still be too weak to correctly reject the null hypotheses, and in order for us to get more convincing results a larger amount of backtesting data is required. Moreover, the fact that more models pass the joint test for 99% VaR than for 95% VaR can be explained by the independence test where a smaller number of exceptions makes it easier not to occur after each other.

Although all models pass apart from AWHS, which fails to detect a single exception, the VWHS models are still performing better than the other yielding empirical sizes very close to the nominal size 1%. Again, there is no clear winner between the VWHS since the three models yield exactly 10 exceptions ordered in the same pattern which provide identical results in the joint test. Among the other four models the normal distribution gives surprisingly good results with an empirical size of 0.89%, providing better accuracy than the t-distribution which overstates the risk. Perhaps the 1%-tail of the OMXS returns is thinner looking more like the normal distribution while the overall 5%-tail is thicker.

4.4 Backtesting results of Boliden

As a secondary base for the study the backtesting analysis was also performed on the Boliden stock data to see how well the models handle more volatile returns. Table 4.7 provides descriptive statistics for the log-returns starting at January 3rd, 2005 and ending at December 31st, 2013. Here we see how the Boliden data is more volatile with a standard deviation of 0.032 compared to 0.014 for the OMXS. Appendix B.2 also displays a graph of the data.

Series: Boliden	
Sample: 03-01-2005 to 31-12-2013	
Observations	2262
Mean	0.0005885152
Median	0.0006253721
Maximum	0.2151113796
Minimum	-0.1914125285
Range	0.4065239081
Std. dev.	0.0320685599
Skewness	0.2146396
Kurtosis	8.603349
Jarque-Bera	2976.585
p-value	0.00000000

TABLE 4.7: *Descriptive statistics for the Boliden stock close prices in SEK ranging from January 3rd, 2005 to December 31st, 2013.*

Table 4.8 presents the VaR backtesting results of the Boliden data. As in the OMXS case we will also here be using a $GARCH(1,1)$ model for consistency (here we do not test which order fits the data the best, but instead assume the $(1,1)$-model is the best just like in the OMXS case). Values in bold in the table for the three tests are cases where the null-hypotheses of a good model fit can't be rejected on the 5% significance level.

1-day 95% VaR	BHS	AWHS	Normal	t	VWHS (Normal)	VWHS (t)	VWHS (skewed t)
# of BT exceptions X	18	53	16	34	49	51	51
# of X:es after an X	2	9	2	4	5	5	5
Empirical size $\frac{X}{T}$ in %	1.789	5.268	1.590	3.380	4.871	4.970	4.970
Kupiec test (p-value)	0.000	**0.698**	0.000	0.012	**0.850**	**0.965**	**0.965**
Independence test (p-value)	0.039	0.001	0.023	0.029	**0.116**	**0.135**	**0.135**
Joint test CC (p-value)	0.000	0.005	0.000	0.004	**0.285**	**0.326**	**0.326**
1-day 99% VaR	BHS	AWHS	Normal	t	VWHS (Normal)	VWHS (t)	VWHS (skewed t)
# of BT exceptions X	1	0	4	10	11	12	12
# of X:es after an X	0	0	0	1	0	0	0
Empirical size (%)	0.099	0.000	0.398	0.994	1.093	1.193	1.193
Kupiec test (p-value)	0.000	0.000	0.029	**0.985**	**0.769**	**0.551**	**0.551**
Independence test (p-value)	**0.964**	NaN	**0.858**	**0.084**	**0.622**	**0.590**	**0.590**
Joint test CC (p-value)	0.001	NaN	**0.090**	**0.225**	**0.848**	**0.724**	**0.724**

TABLE 4.8: *Summary table of the 1-day 95% and 99% backtesting analysis of the Boliden data. Values in bold symbolize p-values above the 5% significance level, i.e. non-rejection of the null-hypothesis of a good model fit. The critical χ^2-value for which we reject H_0 if the test statistic exceeds it is* 3.8415 *for the Kupiec and independence test and* 5.9915 *for the joint test, corresponding to a* 5% *significance level with 1 df and 2 df respectively.*

In general the Boliden backtesting results are more or less in line with the OMXS results. Again the volatility-weighted models produce better VaR estimates than the other models, both in terms of empirical sizes and independent backtesting exceptions. This time the VWHS-models with a $GARCH(1,1)$ under t and skewed t seem to be slightly better for 95% VaR while the $GARCH(1,1)$ under normality is to prefer for 99% VaR. Once again it is hard to say what underlying distribution is the better, the choice between the three doesn't seem to be critical to the accuracy of the VaR estimates. Of the remaining models the AWHS passes the Kupiec test but fails the independence test for 95% VaR, while none of the other models perform well here. For 99% VaR all models except for the BHS and AWHS pass the joint test, again showing that the power of the 1-day 99% VaR backtest might be too low to reject false hypotheses, and that an even bigger sample size might be needed.

5. Conclusion

The objective of this thesis has been to find the best VaR model out of a number of candidates for the Swedish stock market. By estimating 1-day forecasted VaR at 95% and 99% level on the Swedish stock exchange data OMXS as well as on the more volatile single stock series Boliden, the following VaR models were compared: Basic Historical Simulation (HS), age weighted HS (AWHS), volatility weighted HS (VWHS) using a GARCH model, Normal VaR and t-distributed VaR.

We started out with the three research questions:

- Do VaR models using historical simulation outperform parametric models like the Normal VaR and t-distributed VaR?
- Do VaR estimates improve when the volatility is modelled with a GARCH model? If so, what order of the GARCH model and what underlying distribution should be used?
- Among the historical simulation models, the Normal and the t-distributed model, which one is the most reliable in estimating VaR?

From the backtesting analysis of the OMXS and Boliden we can first conclude that it is hard to say if the Normal and t-distribution are better at estimating VaR than the BHS and AWHS. All four models are fairly blunt with empirical sizes far from the nominal sizes and with what it seems like dependent backtesting exceptions. For the 99% VaR most of these four models pass the Kupiec and Christoffersen tests which would motivate us to claim they are good models, but here we rather refer to the fact that the statistical power of correctly rejecting a false model is lower for the 99% VaR than for the 95% VaR.

The most accurate 95% and 99% VaR estimations, for both the OMXS data and for the more volatile Boliden data, came from the VWHS models where volatility was modelled with a GARCH model. Here we received empirical sizes close to the nominal levels while at the same time the backtesting exceptions seemed spread out and independent from each other. Hence we conclude the VWHS model where volatility is modelled with GARCH seems the best for modelling VaR at the Swedish market, at least for the studied time period. When looking at the order of GARCH(p, q) model we found that the standard GARCH$(1, 1)$ model seems to work the best at modelling volatility on the Swedish market. Among the three underlying distributions in the VWHS models; the Normal, the t and the skewed t-distribution, the choice between them seems to have little impact on the VaR estimations on the Swedish market.

Since the model fitting of the 99% VaR was problematic due to the lower statistical power of the tests it would be interesting for future studies to run VaR backtesting with a bigger out-of-sample, i.e. bigger than four years of data. This could then enforce the models' capacity of correctly rejecting wrong hypothesis. It would also be useful to complement this thesis's result with other alternative financial risk measures like expected shortfall, which is more focused on the shape of the loss-tail distribution, or extreme value theory, which concentrates on the most extreme observations, in order to improve market risk quantification on the Swedish market.

6. References

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

T. Beder. Var: Seductive but dangerous. *Financial Analysts Journal*, 51(5):12–24, 1995.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

J. Boudoukh, M. Richardson, and R. Whitelaw. The best of both worlds: a hybrid approach to calculating value at risk. *Risk*, May(11):64–67, 1998.

P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39 (4):841862, 1998.

J. D. Cryer and K.-S. Chan. *Time Series Analysis With Applications in R.* Springer, New York, second edition, 2008.

Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

K. Dowd. *Measuring Market Risk.* Wiley, Chichester, second edition, 2013.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

A. Gut. *An Intermediate Course in Probability.* Springer, New York, second edition, 2009.

E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, 41(2):190–195, 1979.

B. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 3(35):705–730, 1994.

P. R. Hansen and A. Lunde. A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of Applied Econometrics*, 20(7):873889, 2005.

D. Hendricks. Evaluation of value-at-risk models using historical data. *FRBNY Economic Policy Review*, 2(1), 1996.

Y. C. Huang and B. . Lin. Value-at-risk analysis for taiwan stock index futures: Fat tails and conditional asymmetries in return innovations. *Review of Quantitative Finance and Accounting*, 22(2):79–95, 2004.

J. Hull and A. White. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, Fall(1):5–19, 1998.

F. Javed and P. Mantalos. Garch-type models and performance of information criteria. *Taylor and Francis*, 42(8):1917 – 1933, 2013.

P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill Professional, New York, third edition, 2009.

P. Kupiec. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84, 1995.

B. Mandelbrot. Value-at-risk analysis for taiwan stock index futures: Fat tails and conditional asymmetries in return innovations. *The Journal of Business*, 36(4):394–419, 1963.

S. Manganelli and R. F. Engle. Value at risk models in finance. *EUROPEAN CENTRAL BANK WORKING PAPER SERIES*, Working paper(75), 2001.

J. Näsström. Volatility modelling of asset prices using garch models. *Master Thesis: Linköping University*, 2003.

P. Newbold, W. L. Carlson, and B. Thorne. *Statistics for Business and Economics*. Pearson Education, New Jersey, sixth edition, 2007.

N. Nordbo, T. Roynstrand, and V. Strat. Evaluating power of value-at-risk backtests. *Master Thesis: Norwegian University of Science and Technology*, 2012.

M. Pritsker. The hidden dangers of historical simulation. *Board of Governors of the Federal Reserve System*, 2001.

H. Schmidt and M. Duda. Evaluation of various approaches to value at risk. *Master Thesis: Lund University*, 2009.

G. Stahl, C. Wenn, and A. Zapp. Backtesting beyond the trading book. *Journal of Risk*, 8(Winter):1–16, 2006.

N. Taleb. Against var. *Derivatives Strategy*, 2(April):21–26, 1997.

S. Taylor. *Modelling Financial Time Series*. Wiley, Chichester, 1st edition, 1986.

R. S. Tsay. *Analysis of Financial Time Series, 2nd Edition*. Wiley-Interscience, New Jersey, 2nd edition, 2005.

R. S. Tsay and C. G. Tiao. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, 79(385):84–96, 1984.

D. Wackerly, W. Mendelhall, and R. Scheaffer. *Mathematical Statistics with Applications*. Thomson, Duxbury, seventh edition, 2008.

Appendix A

Proofs and formulas

A.1 Weak stationarity condition for a GARCH model

Recall the GARCH(p, q) model:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \ldots + \beta_p \sigma^2_{t-p|t-p-1} + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \ldots + \alpha_q r^2_{t-q}. \tag{A.1}$$

If we now let $\eta_t = r_t^2 - \sigma^2_{t|t-1}$ we can show that $\{\eta_t\}$ is a serially uncorrelated sequence with zero mean, and uncorrelated with past squared returns. Substituting $\sigma^2_{t|t-1} = r_t^2 - \eta_t$ into our GARCH model above gives

$$r_t^2 = \omega + (\beta_1 + \alpha_1) r^2_{t-1} + \ldots + (\beta_{max(p,q)} + \alpha_{max(p,q)}) r^2_{t-max(p,q)} + \eta_t - \beta_1 \eta_{t-1} - \ldots - \beta_p \eta_{t-p} \tag{A.2}$$

which shows that a GARCH(p, q) model for the returns is the same as an ARMA$(max(p, q), p)$ model for the *squared returns*.

Now, if we assume the return process is weakly stationary, taking expectations on both sides of the above formula and solving for σ^2 yields

$$\sigma^2 = \omega + \sigma^2 \sum_{i=1}^{max(p,q)} (\beta_i + \alpha_i) \tag{A.3}$$

and further

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{max(p,q)} (\beta_i + \alpha_i)} \tag{A.4}$$

which will only be a finite variance if

$$\sum_{i=1}^{max(p,q)} (\beta_i + \alpha_i) < 1. \tag{A.5}$$

Applying this to the GARCH(1,1) model implies $\alpha_1 + \beta_1 < 1$.

A.2 Central Limit Theorem

Let $X_1, X_2 \ldots, X_n$ be n i.i.d random variables where $E(X) = \mu$ and $V(X) = \sigma^2$. Let $\bar{X_n} = \sum X_i/n$ be the sample mean. Then, as $n \to \infty$, $\bar{X_n}$ approaches the normal distribution with $E(X) = \mu$ and $V(X) = \frac{\sigma^2}{n}$,

$$\bar{X_n} \to \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \tag{A.6}$$

and as a result it follows that

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{A.7}$$

where z is a standardized, normal variable. For a full proof of the central limit theorem see Gut (2009).

Appendix B

Tables and empirical findings

B.1 Graph of the OMXS log-returns

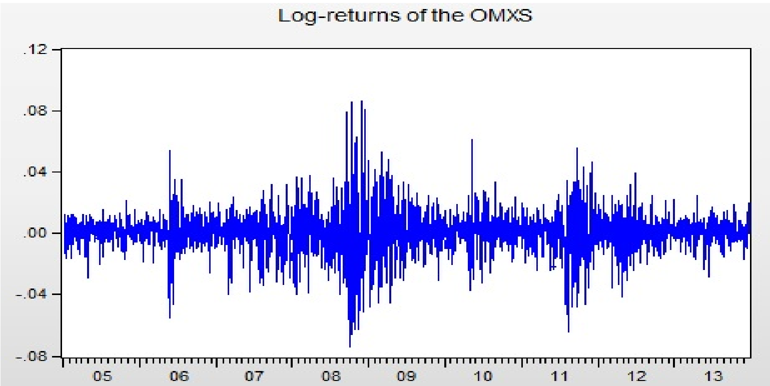

FIGURE B.1: *Graph of the dataset containing daily log-returns of OMXS from Januari 3rd, 2005 to December 31st, 2013.*

B.2 Graph of the Boliden log-returns

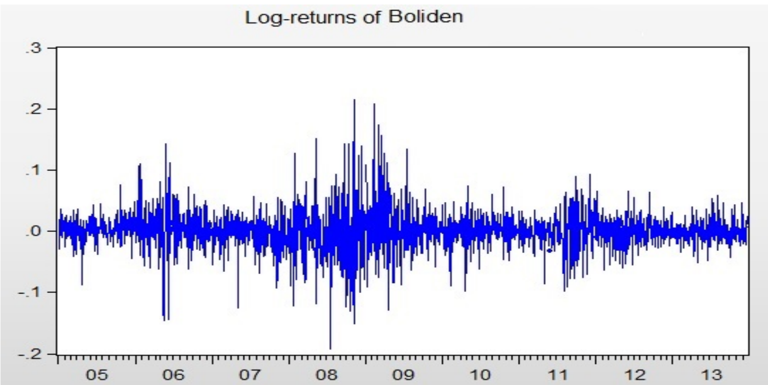

FIGURE B.2: *Graph of the Boliden dataset containing daily log-returns from Januari 3rd, 2005 to December 31st, 2013.*

B.3 Log-returns justification

Figure B.3 below shows the histogram of the returns $r = \frac{V_t}{V_{t-1}}$ (V_t being the OMXS index value at time t) where we see a quite big kurtosis of 6.2 and a positive skewness of 0.18, indicating non-normality. The Jarque-Bera test of normality also clearly rejects the null of normality (test statistic of 545 with p-value 0.00). After taking the log of the return series the kurtosis falls slightly closer to 3 and the skewness is now close to zero (as in the normal distribution). This gives support for our return data being close to log-normally distributed. Hence the log-return transformation $log(\frac{V_t}{V_{t-1}})$ will be used for the series.

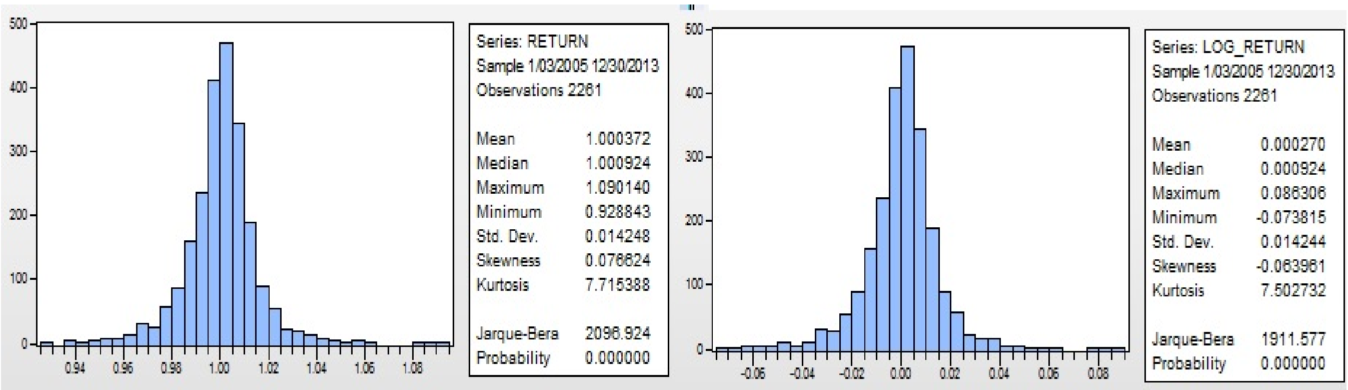

FIGURE B.3: *Histograms for the return series (to the left) and the log-return series (to the right).*