

Forecast Precision of Value at Risk: An Evaluation of ARCH-Type Models

Jacob Lindberg (22765)

Department of Finance

&

Alexander Matsson (23026)

Department of Economics

Supervisors: Cristian Huse and Rickard Sandberg*

Date submitted: 16 May 2016

Date examined: 1 June 2016

Discussant: Patrik Edberg

Examiner: Laurent Bach

Abstract

Over the recent years, value at risk has become an industry standard for measuring downside market risk. This thesis aims to give a thorough differentiation between the different types of models used to estimate value at risk. It shows what gains to be achieved going from a symmetric to an asymmetric model as well as a few other tweaks that can be done to potentially improve the estimates. We provide a detailed comparison of several different kinds of approaches coupled with a thorough backtesting procedure giving a precise evaluation of every alternative approach. Our observations show the importance of choosing a model with regards to its end use and the importance of not adhering too much to convention. Moreover, we show that the asymmetric models prove superior in the analysed stock indices though it is hard to predict the winning model beforehand.

JEL Classification Codes: C22, C52, G17, G32

Keywords: value at risk, volatility modelling, GARCH models, backtesting, asymmetric effect

*We would like to thank Christian Huse and Rickard Sandberg for giving us important guidance throughout the project and for pushing us to do more than we thought we could. All remaining shortcomings are our own.

Contents

1	Introduction	6
1.1	Literature Review	7
1.2	Research Question	7
2	Theory and Methodology	7
2.1	Time Series Concepts	7
2.1.1	Market Returns	8
2.1.2	Information Set and Conditionality	8
2.1.3	Autocorrelation Function	8
2.1.4	Stationarity	8
2.1.5	Volatility Clustering	9
2.1.6	White Noise	9
2.2	Theory for Analysing the Data	10
2.2.1	Test of Stationarity using ADF-tests	10
2.2.2	Test Normality using Q-Q Plots	11
2.2.3	Test of Autocorrelation using ACF Plots	11
2.3	Value at Risk	12
2.3.1	VaR in Regulations	12
2.3.2	Defining and Calculating VaR	13
2.3.3	Backtesting VaR	15
2.4	Modelling Conditional Volatility	17
2.4.1	Symmetric GARCH models	18
2.4.2	Asymmetric GARCH models	19
2.4.3	Maximum Likelihood Estimation	23
2.4.4	Information Criteria	23
2.4.5	Rolling Window Estimation	24
2.4.6	Mean Squared Error	26
2.5	Model Diagnostics	26

3	Data Analysis	26
3.1	Exploratory Data Analysis	27
3.1.1	Stationary of Prices and Returns	27
3.1.2	Descriptive Statistics and Q-Q Plots	28
3.1.3	ACF Plots of Return	29
3.1.4	ACF Plots of Squared Return	30
3.1.5	Conclusion of EDA	31
3.2	GARCH Model Competition	31
3.3	Model Comparison of Winner and Benchmark	32
3.3.1	Fitting the GARCH Model and Estimated Coefficients	32
3.4	VaR Backtesting	35
3.4.1	Tabulated Results	35
3.4.2	VaR Forecast Plot	36
3.4.3	Impact of Asymmetry	36
3.4.4	MSE	39
3.4.5	Conclusion of VaR Backtesting	39
3.5	Model Diagnostics	39
4	Conclusions	42
4.1	Future Research	43
5	References	44
A	Appendix	48
A.1	Distribution Assumptions	48
A.2	Complete Results from the Backtesting Procedure	49

Abbreviations

ACF Autocorrelation Function

ADF Augmented Dickey-Fuller

ARCH Autoregressive Conditional Heteroskedasticity

CDF Cumulative Distribution Function

cov Covariance

GARCH Generalized Autoregressive Conditional Heteroskedasticity

MLE Maximum Likelihood Estimation

model(m,s)-dist. where **model** stands for the model name (e.g. sGARCH), **m,s** stands for how many lags that are incorporated into the model (e.g. 2,1), and **dist.** stands for what distribution is assumed for the residuals (e.g. N). For example sGARCH(2,1)-N.

N Normal Distribution

NIG Normal-Inverse Gaussian Distribution

PDF Probability Density Function

sstd Skewed Student's t-Distribution

std Student's t-Distribution

VaR Value at Risk

var Variance

Q-Q plot Quantile-Quantile Plot

1 Introduction

The recent financial crisis has shown that there are substantial risks in finance and it has spurred interest into measuring and managing these risks. Financial risk is divided into four different categories: market risk, liquidity risk, credit risk and operational risk (Jorion, 2007:22-27). The most well-understood of these risks is market risk, which is the exposure that an investor has to changes in the underlying market prices of its investments. It is usually approximated by historically estimated volatility, however, this is not an optimal way of estimating market risk. For example, volatility effectively treats above average return and below average return the same, even though (for a long position) price increases above the average is something we want to achieve. For this reason a new measure of market risk called value at risk (VaR) was developed and gained a lot of momentum when J.P. Morgan published RiskMetrics in 1994 (Longerstaey and Spencer, 1996) and has since then been widely implemented in the world of finance. Not only is it used within financial institutions as a risk measurement tool as well as for risk management but also among regulators to enforce capital requirements that are dependent on a firm's specific risk exposure. Regulators also give reprimands to institutions that fail to correctly estimate their VaR.

Why should the state meddle in a bank's affairs? Let's look at one example where financial regulation of a bank's risk is desirable. The risk of bank runs, which causes premature liquidation of a bank's long-term assets, has in many countries been alleviated by the state in the form of deposit insurances (Diamond and Dybvig, 1983, showed that this could alleviate the risk of bank runs). When banks have this insurance neither their investors nor their customers need to worry as much about economic downfall since the state will step in and save the day. This is where a moral hazard problem emerges because banks get an incentive to take on higher risk which in turn increases the state's risk of having to pay for the party (see Kareken and Wallace, 1978). To alleviate this problem the state needs to accurately monitor and regulate a bank's risk-taking.

VaR's popularity has a lot to do with its simplicity. The measure is a single value that captures a company's exposure to market risk. VaR answers the question: how much might the company lose due to adverse price movements in the market under normal circumstances? Estimating VaR, however, is not a straightforward process and a lot of decisions have to be made regarding what models to be used. In this thesis we will provide a thorough differentiation between the main alternatives.

The data set that we are going to test the models on is the capitalisation-weighted stock market index OMXS30 which is traded on the Stockholm Stock Exchange. It consists of the 30 most frequently traded stocks in Sweden in terms of turnover. The complete data set consists of observed daily closing prices from the beginning of January 2004 to the end of March 2016 which corresponds to 3076 unique observations, all of which were downloaded from Yahoo Finance. We apply 64 different combinations of models and assumptions about the distribution of the residual which we use to estimate VaR. When presenting the different distributions and models we also give our predictions on which type of model that should provide the best estimates based on previous theory and observations (e.g. leptokurtic distributions usually fit financial data well and models that account for

the leverage effect should give better predictions).¹ After having run all the estimations we evaluate why the winning model was superior in this data set and compare it to the benchmark model as well as to our predictions. Lastly we do a robustness test to see if the results are generalisable so that similar models prevail for different indices.

1.1 Literature Review

There is a substantial literature that compares different ways of estimating value at risk, for example see Kuester et al. (2006), Manganelli and Engle (2001) and Giot and Laurent (2004). To contribute to this extensive literature we will evaluate the forecast precision of 64 different combinations of ARCH-type models and assumptions about the residual distribution, much like what Hansen and Lunde (2005) did when they compared several ARCH-type models with regards to their forecast performance of conditional volatility. This thesis, on the other hand, compares the different models with regards to the model's forecast precision of VaR. We also provide the reader with a robustness test for different assets. The evaluations of the estimates are done in an out-of-sample setting.

1.2 Research Question

This approach can be summarised in two research questions.

1. Which ARCH-type model and assumption of the residual distribution best predicts the one-day-ahead 95% VaR and is this what is expected with regards to conventional theory?
2. Are the results robust to different underlying assets?

2 Theory and Methodology

In this section the theoretical framework applied in this thesis for measuring market risk will be introduced together with the procedure followed to execute the analysis. To begin with, a few cornerstone definitions for time series are introduced. Following that, VaR is introduced as well as ways of estimating it through nonparametric methods and parametric methods. Two methods for evaluating VaR estimates through backtesting are presented. In the last subsections methods for exploratory data analysis as well as methods for model diagnostics are specified.

2.1 Time Series Concepts

This subsection introduces a few of the keystone concepts needed for an analysis of a financial time series.

¹However, there is uncertainty inherent in the analysis and sometimes the only differences between the models are minor tweaks, making an attempt to pinpoint the exact model or distribution *ex ante* bound to be unsuccessful. We will, therefore, mostly rely on inductive reasoning.

2.1.1 Market Returns

The continuously compounded return (log return) at time t is defined as

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

where P_t is the price of an asset at time t . The log returns are used over the arithmetic returns because of the time additive properties as well as the advantageous statistical properties. Furthermore, because we are not building a portfolio of different assets we need not worry about the drawback that log returns are not portfolio additive. (Tsay 2013:4f)

VaR is usually expressed as a positive value. We therefore define the relevant loss random variable as

$$L_t = -(1 + r_t)\mathcal{P}_t \quad (2)$$

where \mathcal{P}_t is the portfolio price such that losses are recorded as positive numbers while profits are recorded as negative losses and as such denoted as negative numbers.

2.1.2 Information Set and Conditionality

In this thesis we will be doing estimations for t that are conditional on the information set \mathcal{F}_{t-1} that is available at $t - 1$ where $\mathcal{F}_{t-1} = \{r_1, \dots, r_{t-1}\}$. This means that since they are conditional they depend on the outcomes inside the information set. We will estimate the conditional mean $\mu_t = E[r_t | \mathcal{F}_{t-1}]$ and the conditional variance $\sigma_t^2 = \text{var}(r_t | \mathcal{F}_{t-1}) = E[(r_t - \mu_t)^2 | \mathcal{F}_{t-1}]$. Because we mostly deal with conditional estimations we call the conditional variance only variance whilst if it is unconditional that will be explicitly specified.

2.1.3 Autocorrelation Function

In a stationary time series r_t we have that the correlation between r_t and its k^{th} lag is

$$\rho_k = \frac{\text{cov}(r_t, r_{t-k})}{\sqrt{\text{var}(r_t) \text{var}(r_{t-k})}} = \frac{\text{cov}(r_t, r_{t-k})}{\text{var}(r_t)} = \frac{\gamma_k}{\gamma_0} \quad (3)$$

for $k \in \mathbb{N}$, due to the fact that $\text{var}(r_t) = \text{var}(r_{t-k})$ in a stationary time series. In Equation (3) γ_k is the coiance between r_t and its k^{th} lag and γ_0 is the coiance between r_t and its 0^{th} lag, i.e. its variance. (Tsay, 2013:45-47)

2.1.4 Stationarity

A time series is strictly stationary when the joint distribution is time invariant. For the time series models applied in this thesis, however, a weaker form of stationarity is sufficient called weak stationarity (or sometimes covariance stationarity). Because only this weaker form is needed we will

in this thesis refer to weakly stationarity by simply stationarity. For a time series process, x_t to be weakly stationary the following three properties must be fulfilled:

- i $E[x_t] = \mu$
- ii $E[x_t - \mu]^2 = \gamma_0 < \infty$
- iii $\text{cov}(x_t, x_{t+k}) = \gamma_k$

for all $t \in \mathbb{Z}$ and $k \in \mathbb{N}$. This means that a time series is weakly stationary when the mean, the variance and the autocovariance structure are time invariant. (Tsay 2013:40-43; Brooks 2002:230f)

2.1.5 Volatility Clustering

Many asset return series go through periods of relatively low volatility while other periods exhibit relatively high. For example, the returns of S&P 500 had a period, after the crisis initiated by the dot-com bubble had stabilised, that was relatively calm---especially from 2004---up until the the unrest of 2007 and the Great Recession which initiated another period of high volatility. Noted by Mandelbrot (1963) "large changes tend to be followed by large changes---of either sign---and small changes by small changes" implying that absolute returns or squared returns exhibit positive autocorrelation. This is what has come to be referred to as volatility clustering.

2.1.6 White Noise

A white noise (WN) process is a process with a mean zero and no covariance between its values at different times. More formally: let z_t be a time series, then

$$z_t \sim WN(0, \sigma_z^2)$$

if and only if

$$\begin{cases} \gamma_k &= 0, k > 0 \\ \gamma_0 &= \sigma^2, k = 0 \end{cases} \quad (4)$$

for $k \in \mathbb{N}$ where σ_z^2 is a constant. This means that it cannot be predicted other than a best guess at the mean. A white noise process can follow different distributions such as the Gaussian distribution or the Student's t-distribution. (Brooks, 2002:232-234; Tsay, 2010:36)

The assumptions we will impose on the white noise term in this thesis are the following distributions: normal distribution, Student's t-distribution, skewed Student's t-distribution and normal-inverse Gaussian distribution, abbreviated N, std, sstd and NIG respectively. The distributions are discussed and their probability density function's (PDF) are listed in the appendix section A.1.

Even though return series often exhibit leptokurtosis this does not necessarily have to be translated to a fat tail assumption for the white noise term. Jorion (2007:221f) explains that a data set with fatter tails does not need to have been drawn from a true distribution that exhibits fat tails

but rather that it could be explained by autocorrelation in the data set causing the observations to be temporarily drawn from a distribution exhibiting an increased likelihood of what would have been viewed as outliers in a normal distribution (the same argument can be applied to peakiness around the mean). In many data sets, both of these explanations (a true distribution with fatter tails and autocorrelation) are likely to be part of the answer. Although, this makes *ex ante* (prior to applying the volatility model) inference about which distribution best fits the residuals difficult.

2.2 Theory for Analysing the Data

To decide which model that is appropriate for a specific data set a thorough analysis of the data is needed. Besides from analysing descriptive statistics we will also use a test of stationarity, the Q-Q plots and most importantly the plots of the autocorrelation function (ACF). The Q-Q plots and ACF are also used to evaluate the models *ex post*. These concepts are introduced in the following subsections.

2.2.1 Test of Stationarity using ADF-tests

If the times series $\{r_t\}$ is stationary then it reverts to its mean. If that's the case then small values tend to followed larger values, and vice versa. In other words the level that the series is at will be a good predictor of the change in the next period---it's autoregressive. The augmented Dickey-Fuller (ADF) test is used to determine whether the the time series exhibit stationarity or if there is some structural break in the data set (Dickey and Fuller, 1979).

The ADF regression is

$$\Delta r_t = \alpha + \beta t + \gamma r_{t-1} + \sum_{k=1}^p \delta_k \Delta r_{t-k} + \epsilon_t$$

where ϵ_t is the regression error term, α is a constant, and β the coefficient on the time trend t . We test

$$H_0 : \gamma = 0$$

against the one-sided alternative $H_1 : \gamma < 0$. The test statistic

$$DF_{obs} = \hat{\gamma} / SE(\hat{\gamma})$$

is compared with $DF_{critical}$. We draw $DF_{critical}$ from a table where the type I error is 5 percent.² If $DF_{obs} < DF_{critical}$ then the null is rejected at a 5 percent significance level and we would conclude that no unit root is present.

²This value is obtained automatically by statistical software, although there exists critical values for the Dickey-Fuller t-distribution. For example, if the sample size is 500 and a test at 5 percent significance level the value is 3.42.

The test can be understood by noting that if the series is integrated then the lagged level of the series, y_{t-1} , are not of any help (in the presence of the lagged changes Δr_{t-k}) in the change in return i.e. the left hand side Δr_t . If it is not of any help then $\gamma = 0$ and the null hypothesis is not rejected.

2.2.2 Test Normality using Q-Q Plots

Besides inspecting the descriptive statistics and histogram plots of the frequency of returns we use Q-Q plots to determine if the data is normally distributed or if it exhibits fatter tails. In a normal Q-Q plot the assumed distribution is plotted using the expected values creating a straight line with the theoretical probabilities on the y-axis and the sample quantiles on the x-axis. If the observed data follows this theoretical distribution then the observed data points and the straight line should coincide reasonably well. If the data exhibits fatter tails than a normal distribution then the observed data points should deviate above the straight line in the lower end of the distribution and below the straight line in the upper end of the distribution. The Q-Q plot can also be constructed so that the straight line follows some other distribution (e.g. the Student's t-distribution).

Sometimes a formal test is used such as the Jarque-Bera test but we prefer Q-Q plot because our main objective is to verify whether the data has fat tails and will thus only rely on this to evaluate the data's distribution.

2.2.3 Test of Autocorrelation using ACF Plots

To evaluate the autocorrelation function of the returns and squared returns a few tests need be introduced.

Test of Autocorrelation in the Returns

To determine if data is autocorrelated or not we use an ACF plot, see for example Figure 4 on page 30. The autocorrelation function (ACF) displays how the different lags are correlated with the current observation such that the 0th lag is perfectly correlated (because it is the current observation) and the following correlations are displayed by bars above the respective lag. That is, we plot ρ_k from Equation (3) for different lags. In the ACF plot the dotted lines are for $-1/T \pm 1.96/\sqrt{T}$ where T is the sample size and $1.96 = \lambda_{0.95}$. If a lag is above the dotted line it is statistically significantly different from zero at the 5 percent level.

The Ljung-Box Test

For a more formal test of the autocorrelation the Ljung-Box test is used (Ljung and Box, 1978). The null hypothesis of the test is

$$H_0 : \gamma_k = 0 \quad \text{for } k = 1, \dots, K$$

against the alternative hypothesis that there exists a k such that $\gamma_k \neq 0$. The appropriate test statistic is

$$Q(K) = T(T+2) \sum_{k=1}^K \frac{\gamma_k^2}{T-l} \quad (5)$$

where T is the number of observations in the data set. Under the null hypothesis $Q(K) \sim \chi^2(K)$. We will be using a confidence region of 95%. (Hull 2012:508f)

The Student's t-test

To formally test whether the mean is significantly different from zero a one sample t-test is applied, whereby we test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. The relevant test statistic is $T_{obs} = \hat{\mu}/SE(\hat{\mu})$ where $SE(\cdot)$ is the standard error and $\hat{\mu}$ is the sample mean of our returns. We reject the null if the observed test statistic $t_{obs} > t_{critical}$ where the critical value is drawn from a Student's t-distribution.

Test of Autocorrelation in the Squared Returns

If, on the other hand, the squared returns are autocorrelated then large absolute movements are followed by large absolute movements the following days. This would be call for some model of the conditional volatility that can account for this. To evaluate this, the same procedure is applied for the squared returns as was applied for the returns.

2.3 Value at Risk

The recent century has seen many a company suffer under sudden bursts of financial instability amplified by unsatisfactory risk measurement and management. Developed after the derivatives infused financial crisis during the beginning of the 1990s, VaR has grown into one of the most prominent ways of measuring and managing risk in a comprehensible and transparent way. It gives an answer to how much an institution might lose under normal conditions in the form of "*the worst loss over a target horizon that will not be exceeded with a given level of confidence*", Jorion (2007:viii). VaR can be used to estimate the risk in a specific portfolio of assets as well as for the total market risk faced by a financial institution. For a comprehensive and extensive outline of VaR see Jorion (2007) and Duffie and Pan (1997).

2.3.1 VaR in Regulations

In order to force financial institutions to be better prepared for severe financial losses, regulators enforce capital requirements on these institutions. Lately, in order to make these requirements reflect the underlying risk that a specific institution faces, these capital requirements are being based on calculations that take the firm's underlying risk into consideration. These risk-based models are generally based on VaR measurements conducted by the institutions themselves.

One such regulation is the Basel Accords developed by Basel Committee on Banking Supervision (BCBS). It is currently adopted by many countries, including the United States (Federal Reserve, 2013) and the European Union (Dierick et al., 2005). It allows for the capital requirement based on

market risk to depend on an internally measured estimate of a company's downside risk such as VaR.³ This puts pressure on the ability to verify the validity of a financial institution's measurement of their market risk. This is something regulators are able to do with VaR through backtesting---which will be discussed in section 2.3.2. Failure to meet the validity requirements will result in reprimands for the financial institution making it important both for society at large as well as for the financial institution that the VaR is not underestimated. Moreover, if financial institutions overestimate their VaR they are not allocating capital to maximise capital returns making pinpointing the VaR estimate important both to keep financial institutions from bankruptcy as well as allowing them to utilise the full potential of their capital. (Basel Committee on Banking Supervision, 2011)

2.3.2 Defining and Calculating VaR

In this subsection a definition of value at risk is introduced together with two alternative approaches to estimating it, historical simulation and the modelling approach. See Tsay (2013:327-385) or Jorion (2007:105-115) for a more detailed explanation of VaR estimation and its definition.

Defining VaR

VaR is the worst loss---which can be seen as a loss random variable at time t denoted L_t ---over the target horizon (in this thesis set to one day) that will not be exceeded with the given confidence level $1 - p$, where $p \in (0, 1)$ is a small probability, usually 5% or 1%.⁴ As mentioned in 2.1.1, losses recorded in L_t will be denoted as positive numbers whereas profits are seen as negative losses. VaR is thus concerned with the upper tail quantile of $1 - p$ (or 0.95 and 0.99 respectively for the given probabilities mentioned earlier) of the loss distribution. Mathematically this can be expressed

$$\text{VaR}_{1-p} = \inf\{x | F_t(x) \geq 1 - p\} \quad (6)$$

where $F_t(x)$ denotes the cumulative distribution function (CDF) of L_t . Thus we have

$$P[L_t \leq \text{VaR}_{1-p}] \geq 1 - p$$

which is in line with what we specified earlier in this paragraph. Figure 1 displays the VaR loss in the PDF of the continuous variable L_t with an arbitrarily chosen distribution. From this it can clearly be seen that VaR is the value such that the probability to observe a bigger loss is exactly equal to 5%.

To estimate VaR it is important to assume that the underlying portfolio is constant or to construct the portfolio so that it does not change over time because otherwise the underlying returns

³In the latest revision of these regulations from January 2016, however, the BCBS recommend the use of conditional value at risk (referred to in the revision as expected shortfall) because of its ability to account for tail risk. It had previously not been recommended due to the difficulties associated with backtesting. This is also why we choose to focus on VaR in this thesis because of the need for a precise and thoroughly developed backtesting procedure. (Basel Committee on Banking Supervision, 2016)

⁴We will choose $p=0.05$ because it gives us more violations in the backtesting procedure to base the statistical tests on.

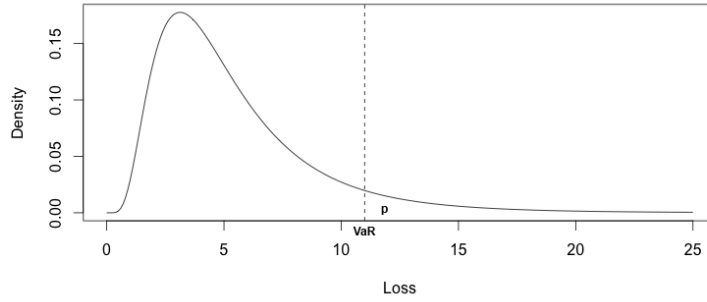


Figure 1: VaR in the PDF of a loss random variable L_t .

might exhibit non-stationarity. This is because the probability distribution of potential losses would change over time as some components or weights of the portfolio change.

In this thesis we will estimate the VaR for one day ahead. This is also what is commonly done in practice due to the difficulty of estimating it further ahead. To extend the VaR estimate N days into the future one might use the square root of time rule in which the N – day $\text{VaR} = 1\text{ – day VaR} \cdot \sqrt{N}$. This is also allowed by the Basel accords. Suffice it to say, the one-day-ahead forecast of VaR is the relevant form that is of interest.

Historical Simulation

One way of estimating a VaR is through historical simulation (Hull 2012:474-478). This is executed by extracting the historical returns from the window of estimation and using them to simulate the PDF of the losses. The $1 - p^{\text{th}}$ quantile is then extracted and used as the $1 - \text{day VaR}_p$ estimate.⁵ This approach is non-parametrical as no assumptions regarding the distribution or its parameters need be facilitated. The main drawback of this approach is that it gives equal weights to every observation ignoring the fact that more recent observations might carry more information on the future volatility than older ones do. More specifically we discussed in section 2.1.5 that squared returns tend to be autocorrelated. Thus, given that the time series exhibits this empirical property, some model that can account for this autocorrelation is bound to be superior.⁶

Modelling Approach

To account for the autocorrelation in the squared return one might try a parametric modelling approach through some weighting scheme, shown in Tsay (2013:327-385) and Jorion (2007:110-113).

⁵VaR is usually expressed in the home currency and therefore needs to be multiplied by the portfolio price \mathcal{P}_t .

⁶Kuester et al. (2006) compared unconditional models (such as historical simulation) with conditional models (such as the GARCH model) and concluded that only the conditional models gave satisfactory forecast results. We will therefore not include historical simulation in our analysis and instead focus on the comparison between the more appropriate symmetric models and the asymmetric models.

The formula for calculating VaR is

$$\text{VaR}_t = (\hat{\mu}_t + z_{0.95}\hat{\sigma}_t)\mathcal{P}_{t-1} \quad (7)$$

where $\hat{\mu}_t$ is the estimated mean in the rolling sample \mathcal{F}_{t-1} , $z_{0.95}$ is the $1 - p^{\text{th}}$ quantile in the white noise distribution where $p=0.05$, σ_t is the estimated volatility using the information set \mathcal{F}_{t-1} and \mathcal{P}_{t-1} is the portfolio value at $t - 1$. In Equation (7) what often proves to be most interesting and difficult to estimate is the volatility $\hat{\sigma}_t$, atleast for financial data with short intervals, and the whole of section 2.4 is dedicated to it. In this thesis we will assume that $\mathcal{P}_{t-1} = 1$ such that the forecast 0.95^{th} quantile in the loss distribution is the portfolio's VaR. This is at no loss of generality and it is done for easier notation. To understand this equation we break it down into it's individual parts.

The model for the return at time t is

$$r_t = \hat{\mu}_t + \varepsilon_t \quad (8)$$

where ε_t is some innovation (or shock) that moves the return away from $\hat{\mu}_t$. This innovation can be modelled by

$$\varepsilon_t = \hat{\sigma}_t z_t. \quad (9)$$

Here, $\hat{\sigma}_t$ could be based purely on the historically estimated variance or it could be forecast using some weighting scheme and forecast procedure, introduced under section 2.4, which can enable the volatility to be conditional on the information set up until $t - 1$. This can allow for volatility clustering to be present in the model and hence allow the model to adapt to periods of increased volatility. The random variable z_t is usually assumed to follow the standardised normal distribution.

2.3.3 Backtesting VaR

To ensure that the estimated VaR complies with reality the predictions need to be evaluated *ex post* using a backtesting procedure. To begin with, we first have to specify a framework for how to treat an observed loss that is in excess of the VaR estimate. One such approach is to define a time series of such violations as an indicator variable I_t for the forecast VaR at time t , made at $t - 1$, as

$$I_t = \begin{cases} 1, & \text{if } L_t > \text{VaR}_{1-p,t} \\ 0, & \text{if } L_t \leq \text{VaR}_{1-p,t} \end{cases}. \quad (10)$$

The variable I_t can then be used to count every time the observed loss is greater than the $(1 - p)$ VaR estimate which we call a violation.

For an effective VaR measure, two properties must be satisfied (Christoffersen, 1998):

1. the probability of observing a violation in the data set should be p ---or sufficiently close to p unconditionally on the information set available at time $t - 1$. More formally, $E[I_t] = p$.

2. the probability of observing a VaR violation is not dependent on violations observed during the previous days. Hence, VaR violations should not come distributed in clusters. This can be expressed as $E[I_t | \mathcal{F}_{t-1}] = p, \forall t$, where \mathcal{F}_{t-1} denotes the information set available at time $t - 1$. As opposed to property 1, this property concerns the efficiency of the estimate conditional on the information set at $t - 1$.

To test these properties we will be using the Kupiec test for the first property and Christoffersen's test for the second property referred to as the unconditional coverage (uc) test and the conditional coverage (cc) test respectively.⁷

Kupiec's Test of Unconditional Coverage

The uc test (Kupiec, 1995) evaluates whether the observed ratio between violations and the number of out-of-sample estimates corresponds to the level of significance specified in the VaR calculations, e.g. with a significance level of $p = 0.05$ we should expect to observe that 5% of the out-of-sample estimates are in excess of the VaR estimate. Let T be the number of out-of-sample estimates and N be the number of violations. Then the null hypothesis of the test is

$$H_{0,cc} : p = 0.05$$

against $H_1 : p \neq 0.05$. The appropriate test statistic is

$$LR_{uc} = -2 \ln [(1 - p)^{T-N} p^N] + 2 \ln \left[\left(1 - \frac{N}{T}\right)^{T-N} \left(\frac{N}{T}\right)^N \right]. \quad (11)$$

Under the null hypothesis LR_{uc} is asymptotically distributed as $\chi^2(1)$. The test is for a 95% confidence region and thus, with one degree of freedom, to reject the null hypothesis we would need to observe $LR_{uc} > 3.841$.

Christoffersen's Test of Independence

The test of independence developed by Christoffersen (1998) evaluate the conditional efficiency that a violation today does not give any information whether there will be a violation or not the following day. Christoffersen states that it is not enough for a VaR model to simply get the number of violations correct but also to account for clustered returns so that the VaR violations will not experience any clustering. For the VaR estimate to exhibit conditional coverage (i.e. nonclustered violations) the probability of observing a violation should not be affected by whether there was a violation the day before or not. To test if the observed distribution is significantly different from this, Christoffersen developed a hypothesis test where the null hypothesis is

$$H_{0,ind} : \pi = \pi_0 = \pi_1$$

⁷The two tests are summarised by Jorion (2007:143-152).

against the alternative hypothesis $\pi_0 \neq \pi_1$ where π_i is the probability of observing a violation at time t if $I_{t-1} = i$ was observed the previous day and π is simply the unconditional probability of observing a violation. The test statistic is

$$LR_{ind} = -2 \ln[(1 - \pi)^{(T_{00}+T_{10})} \pi^{(T_{01}+T_{11})}] + 2 \ln[(1 - \pi_0)^{T_{00}} \pi_0^{T_{01}} (1 - \pi_1)^{T_{10}} \pi_1^{T_{11}}] \quad (12)$$

where T_{ij} is the number of days where outcome $I_t = j$ followed outcome $I_{t-1} = i$ the day before. The test statistic LR_{ind} is asymptotically distributed as $\chi^2(1)$ under the null hypothesis. Once again we reject the null hypothesis if $LR_{ind} > 3.841$ with a 95% confidence level.

One important critique of Christoffersen's test is that it only treats two consecutive violations as a cluster and hence ignores violations that occur with a few days between them. Even though violations that are a few days apart might be a consequence of clustering.

Conditional Coverage Test

To test both unconditional and conditional coverage at the same time we use the cc test in which

$$H_{0,cc} : p = \pi_0 = \pi_1.$$

Christoffersen (1998) proposed the combined test statistic

$$LR_{cc} = LR_{uc} + LR_{ind} \quad (13)$$

where LR_{cc} under the null hypothesis is distributed as $\chi^2(2)$. For a 95 percent confidence level we reject the null hypothesis if $LR_{cc} > 5.991$.

2.4 Modelling Conditional Volatility

If the time series exhibit autocorrelation and an equal weighting scheme is used to forecast volatility then if a VaR violation is observed we are more likely to observe a VaR violation the following day. This is because we know that large values tend to be followed by large values---the volatility clusters together as discussed under Section 2.1.5---and this would not be accounted for in the model that puts equal weights on all observations---i.e. making the violations cluster together as well. These models are "dumb" in the sense that, even though we as researchers know that the risk is higher during these limited periods of high volatility, the models do not alter their risk estimates accordingly. For this reason several volatility models that could adapt to these clusters and hence account for them in their estimates were developed.

Originally developed by Engle (1982), the autoregressive conditional heteroscedastic (ARCH) model laid the foundation for many different conditional volatility models. The most prominent of these is the standard generalized autoregressive conditional heteroscedastic (sGARCH) model put forth by Bollerslev (1986). For an introduction to the subject see Tsay (2010:109-173). The R package that is mainly used in this thesis is the *rugarch* package by Ghalanos (2015).

2.4.1 Symmetric GARCH models

In Equation (8) on page 15 we see that the returns are modelled by $r_t = \mu_t + \varepsilon_t$. If we let $\varepsilon_t = \sigma_t z_t$ as in formula (9) then ε_t follows a sGARCH(1,1) process if the conditional variance σ_t^2 can be described by

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (14)$$

under the constraints $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$. Oftentimes, it can be assumed that $\mu_t = 0$ which makes $\varepsilon_{t-1}^2 = r_{t-1}^2$.⁸ The variable σ_{t-1}^2 is the previous day's estimation of the conditional variance.

The innovation ε_t in Equation (8) is the deviation from the mean that cannot be explained by the mean equation. It is therefore sometimes referred to as the residual of the mean equation. In this thesis, however, the mean equation will simply be a constant. For some time series the mean equation needs to be modelled (for example using an ARMA model) and this is common for monthly or quarterly data. Because we will be using daily data the return is close to zero and the mean equation, therefore, does not need to be modelled.

In Equation (14), α_1 and β_1 are the respective weights on the previous day's innovation, ε_{t-1}^2 , and the estimation of the previous day's conditional variance, σ_{t-1}^2 . There is, however, a third weight γ that is put on the long-run unconditional variance V_L . These have been simplified to ω such that

$$\omega = \gamma V_L. \quad (15)$$

This simplified form is usually used for estimating the parameters, i.e. ω , α_1 and β_1 . Because the weights must sum to unity such that

$$\gamma + \hat{P} = 1,$$

where, for the (1,1) model, $\hat{P} = \alpha_1 + \beta_1$ (which is called persistence and is a measure of the extent that a relatively high volatility today will affect the forecast volatility in the future) must sum to less than one, which is what we specified under the constraint $\alpha_1 + \beta_1 < 1$, for the weight on γ to take on a positive value needed for a stationary process.⁹ (Hull, 2012:502f)

In the analysis we will also use the concept of half-life to facilitate further differentiation between the models. It is defined as the number of days it takes for half of the reversion back to the unconditional volatility V_L to complete. Mathematically it is expressed as,

$$\text{half-life} = \frac{-\ln 2}{\ln \hat{P}} \quad (16)$$

⁸This is often a reasonable assumption when the conditional volatility is estimated for a short window---say an hour or a day. This is because the mean usually is small for these short intervals, especially relative to the standard deviation (Hull 2012:479f). However, this assumption can be formally tested using a t-test.

⁹If $\gamma \leq 0$ then the process will not experience mean reversion which is when the level of variance returns to normal levels after experiencing some shock, i.e. when $\sigma_t^2 > V_L$. Hence, the variance will not have a drift towards the long-run variance, V_L (which is unconditional on previous innovations), and will instead move entirely conditionally on past innovations (past innovations are persistent) or even move away from the mean, V_L , if the weight is less than zero.

where \hat{P} is the persistence parameter.

To allow for more of the previous days to affect the current one-day-ahead estimation of the conditional volatility, Equation (14) can be extended to an sGARCH(m,s) model by changing it to

$$\sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (17)$$

under the constraints $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{max(m,s)} (\alpha_i + \beta_j) < 1$. Persistence is now given by

$$\hat{P} = \sum_{i=1}^m \alpha_i + \sum_{j=1}^s \beta_j.$$

For most circumstances, however, many argue that the sGARCH(1,1) model is appropriate because it incorporates earlier days' observations in σ_{t-1}^2 which is the volatility forecast for $t-1$ made with the information available at $t-2$ which includes the volatility forecast for $t-2$ made with the information available at $t-3$ and so on. The model therefore includes prior innovations. In this thesis we will test a (2,1) model to hopefully improve on the forecast precision through a more precise effect of the innovation for $t-2$.

The Component sGARCH

Lee and Engle (1999) developed the component sGARCH (csGARCH). They acknowledged the fact that volatility exhibited a long-run memory (i.e. high persistence) and therefore tried to more accurately account for this by decomposing the conditional volatility into a short-run and a long-run trend. They further motivated this by looking at the movements in the implied volatility term structure of equities, interest rates and foreign exchange markets and noting that, for shorter maturities, the volatility of implied volatility was higher than for longer maturities. By letting q_t denote the long-run trend then the csGARCH(m,s) model can be written as

$$\sigma_t^2 = q_t + \sum_{i=1}^m \alpha_i (\varepsilon_{t-i}^2 - q_{t-i}) + \sum_{j=1}^s \beta_j (\sigma_{t-j}^2 - q_{t-j}) \quad (18)$$

where the long-run trend is

$$q_t = \omega + \rho q_{t-1} + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2).$$

Persistence and unconditional volatility are given by the same formula as in the GARCH(m,s) model. This means that the expected volatility in the very long-run will converge to the same unconditional volatility as the GARCH(m,s) model.

2.4.2 Asymmetric GARCH models

The symmetric sGARCH models hold a distinct disadvantage, they fail to account for the asymmetric effects in financial time series of returns on volatility. Observed by Black (1976), volatility tends to be affected more by returns below expectation and less by returns above expectation (something

that has come to be known as the leverage effect).¹⁰ This section introduces several models that tries to alleviate the symmetric constraint of the sGARCH models. The main differences between these models are mainly that they try to model asymmetry in different ways.

The asymmetric treatment of returns above or below expectation can be displayed for the different models in a news impact curve (Pagan and Schwert, 1990). It shows the effect that a particular innovation will have on the forecast volatility in a specific model. A symmetric model will treat return above and below expectation equally (and will therefore be symmetric around the expected value) whilst an asymmetric model may have different effects depending on if the return is above or below (and might therefore be distributed asymmetrically around the expected value).

The Exponential GARCH model

Nelson (1991) developed an asymmetric model called the exponential GARCH (EGARCH). The variance follows an EGARCH(m,s) process if

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^m (\alpha_i z_{t-i} + \gamma_i (|z_{t-i}| - E|z_{t-i}|)) + \sum_{j=1}^s \beta_j \ln(\sigma_{t-j}^2) \quad (19)$$

where $\omega, \alpha_i, \gamma_i, \beta_j$ are parameters to be estimated and z_t is the white noise term. By inspecting the equation above we note that α_i captures the sign effect and γ_i captures the size effect---this is important since it's a major distinction from the sGARCH model.

The persistence is given by

$$\hat{P} = \sum_{j=1}^s \beta_j$$

in the EGARCH model which differs from persistence in the sGARCH model. The unconditional variance is given by $V_L = \frac{\omega}{1-\hat{P}}$. (Tsay, 2013:215-221)

Previous research has shown that increasing the number of parameters rarely beats a (1,1) model. We will, however, test a (2,1) for the same reason as for the sGARCH model.

The GJRGARCH model

Another asymmetric model is the GJRGARCH model developed by Glosten et al. (1993). In this model the conditional variance is modelled by

$$\sigma_t^2 = \omega + \sum_{i=1}^m (\alpha_i \varepsilon_{t-i}^2 + \gamma_i I_{t-i} \varepsilon_{t-i}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (20)$$

¹⁰This leverage effect stems from that the markets tend to react differently to positive news and negative news. The most prominent explanation for this is that when prices depreciate, companies' debt-to-equity ratio appreciate. Hence, the companies become more leveraged causing volatility to increase. (Aït-Sahalia et al. 2013)

where α_i and β_j are as in the sGARCH model, I_{t-i} is an indicator variable and γ_i captures the leverage effect. The indicator variable is

$$I_{t-i} = \begin{cases} 1, & \text{if } \varepsilon_{t-i} \leq 0 \\ 0, & \text{if } \varepsilon_{t-i} > 0 \end{cases}$$

and thus γ_i scales the effect of a negative innovation.

The persistence parameter is here given by

$$\hat{P} = \sum_{i=1}^m \alpha_i + \sum_{i=1}^m \gamma_i \kappa + \sum_{j=1}^s \beta_j, \quad (21)$$

where $\kappa = E[I_{t-i} z_{t-i}^2]$ which is the probability of drawing an observation below zero in the distribution.

Asymmetric Power ARCH

The models introduced until now are based on autocorrelation in the squared returns. Taylor (1986) observed that it was more common for the autocorrelations in the absolute returns to be bigger than in the squared returns, suggesting that a model that is based on autocorrelation in the absolute returns might possess superior prediction performance. A model that can account for this Taylor effect as well as asymmetric effects was developed by Ding et al. (1993) called asymmetric power ARCH (apARCH). It models the conditional volatility as

$$\sigma_t^\delta = \omega + \sum_{i=1}^m \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^s \beta_j \sigma_{t-j}^\delta \quad (22)$$

where α_i and β_j are as in the sGARCH model, γ_i is the leverage parameter and $\delta \in \mathbb{R}^+$ is the asymmetric power parameter which is a Box-Cox transformation of σ_t .¹¹ What makes this model stand out is that it actually contains the sGARCH model as a special case (when $\delta = 2$ and $\gamma_i = 0, \forall i$) as well as the gjrGARCH (when $\delta = 2$). The model thus should allow for more fine tuning than those two models.

The persistence parameter is given by

$$\hat{P} = \sum_{i=1}^m \alpha_i \kappa_i + \sum_{j=1}^s \beta_j$$

where $\kappa_i = E(|z| - \gamma_i z)^\delta$ which is the expected value of z_t under the Box-Cox transformation of the term with γ_i .

¹¹Box and Cox (1964) introduced the power transformation function called Box-Cox transformation.

The unconditional variance is given by

$$V_L = \left(\frac{\omega}{1 - \hat{P}} \right)^{2/\delta}.$$

AVGARCH

Taylor (1986) and Schwert (1990) discusses the absolute value GARCH (AVGARCH) model where the volatility is modelled as

$$\sigma_t = \omega + \sum_{i=1}^m \alpha_i \sigma_{t-i} (|z_{t-i} - \eta_{2i}| - \eta_{1i} (z_{t-i} - \eta_{2i})) + \sum_{j=1}^s \beta_j \sigma_{t-j} \quad (23)$$

where shifts and rotations in the news impact curve are modelled with η_{2i} and η_{1i} respectively and $|\eta_{1i}| \leq 1$.

Persistence in the model is given by

$$\hat{P} = \sum_{i=1}^m \alpha_i \kappa_i + \sum_{j=1}^s \beta_j$$

where $\kappa_i = E(|z_{t-i} - \eta_{2i}| - \eta_{1i} (z_{t-i} - \eta_{2i}))$ which effectively is the expected value of the standardised residual z_t under the Box-Cox transformation. The unconditional variance is given by

$$V_L = \left(\frac{\hat{\omega}}{1 - \hat{P}} \right)^2.$$

NGARCH

Higgins and Bera (1992) developed the nonlinear GARCH (NGARCH) model. The conditional volatility follows a NGARCH process when

$$\sigma_t^\delta = \omega + \sum_{i=1}^m \alpha_i \sigma_{t-i}^\delta (|z_{t-i}|)^\delta + \sum_{j=1}^s \beta_j \sigma_{t-j}^\delta \quad (24)$$

where $\delta \in \mathbb{R}^+$ is the asymmetric power parameter. The non-negative restriction on δ makes sure that the conditional variance is defined for all innovations.

Persistence in the NGARCH model is given by

$$\hat{P} = \sum_{i=1}^m \alpha_i \kappa_i + \sum_{j=1}^s \beta_j$$

where $\kappa_i = E(|z_{t-i}|)^\delta$ which is the expected value of z_t under the Box-Cox transformation and the unconditional variance is given by

$$V_L = \left(\frac{\hat{\omega}}{1 - \hat{P}} \right)^{2/\delta}.$$

NAGARCH

Engle and Ng (1993) introduced another nonlinear GARCH model called the nonlinear asymmetric GARCH (NAGARCH) model. In this model the variance is given by

$$\sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i \sigma_{t-i}^2 (|z_{t-i} - \eta_{2i}|)^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (25)$$

where shifts in the news impact curve is captured in η_{2i} .

Persistence is again given by the same formula as for AVGARCH but now κ_i is given by

$$\kappa_i = E(|z_{t-i} - \eta_{2i}|)^2$$

and the unconditional variance is given by

$$V_L = \left(\frac{\hat{\omega}}{1 - \hat{P}} \right).$$

2.4.3 Maximum Likelihood Estimation

To estimate the GARCH parameters we use maximum likelihood estimation (MLE). Let $\theta = (\omega, \alpha_1, \beta_1)$ be the parameters, and let $r = (r_1, \dots, r_T)$ be the returns data. The likelihood function is defined as

$$\mathcal{L}(\theta; r) = \prod_{t=1}^T f(r_t | \theta).$$

The log-likelihood function $\ell(\theta; r)$ is the natural log of the likelihood function i.e. $\ell(\theta; r) = \ln \mathcal{L}(\theta; r) = \sum_{t=1}^T \ln f(r_t | \theta)$ and we use $\ell(\theta; r)$ because it is easier to work with in practice.

When calculating the maximum likelihood estimate $\hat{\theta}_{ML}$ we want to maximize the likelihood-function \mathcal{L} given the data. But because the log-likelihood function $\ell = \ln \mathcal{L}$ is monotonically increasing function of \mathcal{L} we can---instead of maximizing the likelihood function---maximize the log-likelihood function. This yields the same estimates but is faster computationally. Denoting the maximum likelihood estimate $\hat{\theta}_{ML}$ we define

$$\hat{\theta}_{ML} = \arg \max_{\theta} \ell(\theta; r)$$

where r is the data and $\ell(\theta; r)$ is the log-likelihood function.

2.4.4 Information Criteria

The parameters are estimated such that the Likelihood function is maximized, but the likelihood is not merely a tool to give us the estimated parameters but is also useful in comparing two or more models in seeing which of them better fits the data. The value of the likelihood functions for the compared models are reported in the data analysis under section 3.3 but in order to compare models with a different number of parameters there are a few information criteria frequently used such as

Akaike (AIC), Bayesian (BIC), Shibata (SIC) and Hannan-Quinn (HQIC). We will use AIC and HQIC because they compliment each other well. The definitions are

$$\text{AIC} = 2p - 2\hat{\ell} \quad (26)$$

and

$$\text{HQIC} = -2\hat{\mathcal{L}} + 2p \ln(\ln T) \quad (27)$$

where p is the number of estimated parameters in the model, $\hat{\ell}$ is the log-likelihood value, $\hat{\mathcal{L}}$ is likelihood value and T is the sample size. A reason for picking one model over another, if they perform similarly, is to pick the one with the lowest information criteria. As seen from the formulas AIC and HQC penalizes a model with many parameters but rewards a high likelihood.

In this thesis, information criteria are used to determine whether one model fits better in the estimation window than another. We include these because a model should not only perform well in the forecasting window but also be sound according to statistical theory.

2.4.5 Rolling Window Estimation

t^*	t	$r^{(1)}$	$r^{(2)}$	$\hat{\theta}^{(n)}$	$\hat{\sigma}_t$	$\hat{\mu}_t$	VaR_t	rr	rv	I_t
1	$1 - w_E$	r_{1-w_E}								
2	$2 - w_E$	r_{2-w_E}	r_{2-w_E}							
\vdots	\vdots	\vdots	\vdots							
w_E	0	r_0	r_0							
$w_E + 1$	1		r_1	$\theta_1^{(1)}$	$\hat{\sigma}_1$	$\hat{\mu}_1$	VaR_1	r_1	$(r_1 - \hat{\mu}_1)^2$	0 or 1
$w_E + 2$	2			$\theta_2^{(2)}$	$\hat{\sigma}_2$	$\hat{\mu}_2$	VaR_2	r_2	$(r_2 - \hat{\mu}_2)^2$	0 or 1
\vdots	\vdots			\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$w_E + w_F$	w_F			$\theta_{w_F}^{(w_F)}$	$\hat{\sigma}_{w_F}$	$\hat{\mu}_{w_F}$	VaR_{w_F}	r_{w_F}	$(r_{w_F} - \hat{\mu}_{w_F})^2$	0 or 1

Table 1: Rolling Window Estimation

Divide the whole data set into two windows: the estimation window, denoted w_E , and the forecasting window, denoted w_F . The estimation window goes from $t^* = 1$ to $t^* = n(w_E)$ and the forecasting window goes from $t^* = n(w_E) + 1$ to $t^* = n(w_E) + n(w_F)$. where $n(w_E)$ and $n(w_F)$ denotes the length of each window. In our sample the length of the windows are $n(w_E) = 2076$, going from 2 January 2004 to 28 March 2012, and $n(w_F) = 1000$, going from 29 March 2012 to 1 April 2016.¹² For easier notation we define t so that the forecasting window starts at $t = 1$ so do note that $t \neq t^*$. Forecasting of the VaR estimates is done using rolling window estimation summarised in Table 1 in which we use w_E and w_F instead of $n(w_E)$ and $n(w_F)$ for easier notation.

¹²The reason these intervals for w_E and w_F have been chosen is because: (1) When estimating w_E we want a good balance between getting enough observations to obtain precise estimates whilst only using relevant data. (2) When estimating w_F we want to make sure that we have enough observations to get statistically reliable results when evaluating the backtesting procedure.

Define θ to be the parameter vector so for a GARCH-n $\theta = (\omega, \alpha, \beta)$ and for a GARCH-std $\theta = (\nu, \omega, \alpha, \beta)$. In this explanatory example the parameter vector will be updated every day. The reestimation of the parameter vector θ is done to keep them up-to-date which improves forecasting precision. The first parameter vector is denoted $\theta_1^{(1)}$ where the subscript 1 stands for t and the superscript (1) stands for the first estimation of parameters.¹³ The estimation is done using the data available up until t which is the information set \mathcal{F}_{t-1} . The parameters are estimated using MLE as described in section 2.4.3.

Volatility is forecast using the parameters $\hat{\theta}_1^{(1)}$ together with the volatility formula that is being used, see section 2.4.1 and 2.4.2. For the first day $t = 1$ the innovation ε_0 is, as always, set to the previous day's innovation. The previous day's estimated volatility $\hat{\sigma}_0$, however, is treated differently the first time and is set to the sample standard deviation in w_E .

The mean is forecast using the sample mean. Using our notation $r^{(1)}$ from the table we have that $\hat{\mu}_1 = \text{mean}(r^{(1)})$ and more generally $\hat{\mu}_t = \text{mean}(r^{(n)})$.

VaR is forecast using the mean forecast, the volatility forecast and the quantile from the assumed distribution for the innovations as in Equation (7) where $\mathcal{P}_{t+1} = 1$ giving us

$$\text{VaR}_t = \hat{\mu}_t + z_{0.95} \hat{\sigma}_t$$

where $z_{0.95}$ is the quantile from the assumed distribution for the innovations ε_t .¹⁴ In this thesis we will for example assume that $z \sim t(\nu)$ where ν is estimated via MLE separately for both models.¹⁵

The realised return at time t , rr_t , is revealed only after we have forecast VaR_t so for example we use data $r^{(1)}$ to estimate VaR_1 then wait until markets close the following day and only after the markets have closed rr_1 is revealed to us. The indicator variable I_t takes on the value 1 if $rr_t > \text{VaR}_t$, i.e. if the loss at day t is bigger than our stated VaR_t .

After all this have been done (parameter estimation, sigma forecast, mean forecast, VaR calculation and violation calculation) the window is rolled forward one day---and the procedure is repeated again but this time using the data set $r^{(2)}$. This goes on until the end of the forecasting window and then we have $n(w_F) = 1000$ different VaR estimates and their violations which is ready to be backtested in order to answer the question of which model is the most reliable for forecasting value at risk in our out-of-sample setting.

¹³As a side note, if we update the parameters every second day the column $\hat{\theta}^{(n)}$ would read $\hat{\theta}_1^{(1)}, \hat{\theta}_2^{(1)}, \hat{\theta}_3^{(2)}, \hat{\theta}_4^{(2)}, \hat{\theta}_5^{(3)}, \dots, \hat{\theta}_{1000}^{(500)}$. In our analysis, however, the parameter vector will only be updated every 100 days for the first ranking procedure and every 20 days for the more thorough model comparison. We do this because the parameters do not change a lot on a daily basis and this also increase the speed of computation.

¹⁴The careful reader will note that this estimation and forecast is done while knowing all the data up until the forecasting window while not knowing (ignoring) the information that comes after. If you will, the forecast could be said to be done after the markets close but before they open the next day.

¹⁵For example using R we get $qt(1 - 0.05/2, df = 9) = 2.262$ whereas the quantile for normal distribution is $qnorm(1 - 0.05/2) = 1.96$.

2.4.6 Mean Squared Error

Another model selection tool is the mean squared error (MSE). It is used to determine how close the volatility estimates are to the realised volatility. The mathematical formula is

$$MSE = \frac{SSE}{n(w_F)} = \frac{1}{n(w_F)} \sum_{t=1}^{n(w_F)} (\hat{\sigma}_t - \sigma_t)^2 \quad (28)$$

where $n(w_E)$ is the number of estimations, SSE is sum of squared errors, $\hat{\sigma}_t$ is the estimated volatility and σ_t is the realised volatility. MSE has value when choosing a model for estimating, for example, a financial institutions VaR because it can help them to not take a model that overestimates the risk. This is important because, as we mentioned in section 2.3.1, overestimating the risk hinders the financial institution from allocating capital efficiently.

2.5 Model Diagnostics

In the GARCH model we impose a theoretical distribution on the white noise term z_t , for example a normal or Student's t-distribution. In order to check if this assumption is fulfilled or not we plot the standardized residuals $\tilde{\epsilon}_t$ to check the adequacy of the mean equation and $\tilde{\epsilon}_t^2$ to check the adequacy of the volatility equation. If they appear as white noise then the assumption is fulfilled. The standardized residuals are calculated as

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t} \quad (29)$$

where $\hat{\epsilon}_t = r_t - \hat{\mu}_t$ according to the GARCH model equation (8). Upon comparing the white noise term $z_t = \epsilon_t/\sigma_t$ and the formula for $\tilde{\epsilon}_t$ we see that it is indeed logical to evaluate the assumed distribution of z_t by plotting $\tilde{\epsilon}_t$. This is also done for the $\tilde{\epsilon}_t^2$.

To evaluate whether the standardized residuals seem to be white noise we plot the autocorrelation function, in which there should be no apparent autocorrelation, as well as the Q-Q plot, where the standardized residual should follow the assumed distribution. This is done in the Data Analysis under section 3.5.

3 Data Analysis

We begin with an exploratory data analysis to evaluate how the returns can be modelled given our research question. We then fit all 64 models to OMXS30 in the estimation window. We use these 64 models to forecast one-day ahead VaR in the forecasting window and these forecasts are then backtested using the uc and cc tests. Using the performance of the models with regards to backtesting, we pick the winning model and compare it to the benchmark model.

We need to be very careful about the distinction between the estimation window w_E and the forecasting window w_F . The estimation window is used in the exploratory data analysis as well as

to estimate the first parameters and the forecasting window is used in VaR testing in section 3.4 and onwards.

Lastly, we do a robustness test of our findings to other stock indices.

3.1 Exploratory Data Analysis

In this section the following steps will be performed:

- (1) prices are plotted together with returns and stationarity is tested using the ADF-test,
- (2) descriptive statistics are tabulated and the normal Q-Q plot of the return series is created together with a histogram displaying the frequency of returns,
- (3) ACF of returns is plotted together with the Ljung-Box test to evaluate the autocorrelation in returns,
- (4) test if the mean return is statistically different from zero using a t-test,
- (5) ACF of squared returns is plotted together with the Ljung-Box test to evaluate the autocorrelation in squared returns,
- (6) test if the mean squared return is statistically different from zero using a t-test, and
- (7) conclude what type of model should prove superior using statistical and financial theory.

From these steps we draw the following conclusions. From step (1) we conclude that returns are stationary. From step (2) we see that the distribution of returns is more leptokurtic than a normal distribution. However, drawing any conclusions regarding the best distribution assumption for the residuals from this is dubious because, as we discussed in subsection 2.1.6, the model that gets applied could "pick up" these tail events, thus rendering any fat-tail assumption superfluous. Furthermore, the distribution exhibit low but noticeable skewness making inference regarding appropriate skewness assumption difficult as well. From step (3) we can see some autocorrelation in the returns from the ACF plot and the LB test confirms this. However, step (4) shows that the return series is not statistically different from zero---hence we need not model the mean equation. Moving on to squared returns we see from step (5) that squared returns exhibit clear autocorrelation and this time the t-test in step (6) show that squared returns are statistically significantly different from zero---hence we need to model the variance equation. In step (7) we conclude that an asymmetric ARCH-type model is likely to perform best.

3.1.1 Stationary of Prices and Returns

In Figure 2 we plot the prices and returns for OMXS30 from 2 January 2004 to 28 March 2012 referred to as the window of estimation w_e . Prices see a big dip during the Great Recession and in line with what Schwert found in his article from 1989---that volatility clustering tend to be more

pronounced during economic recessions---we see a tendency for volatility clustering to become more prominent during this period.

To be able to estimate the conditional volatility models in this thesis we need the data set to be stationary. According to a Dickey Fuller-test the price series is not stationary (which also clearly can be seen from the plot) whilst the return series is stationary. We can therefore model the return series.

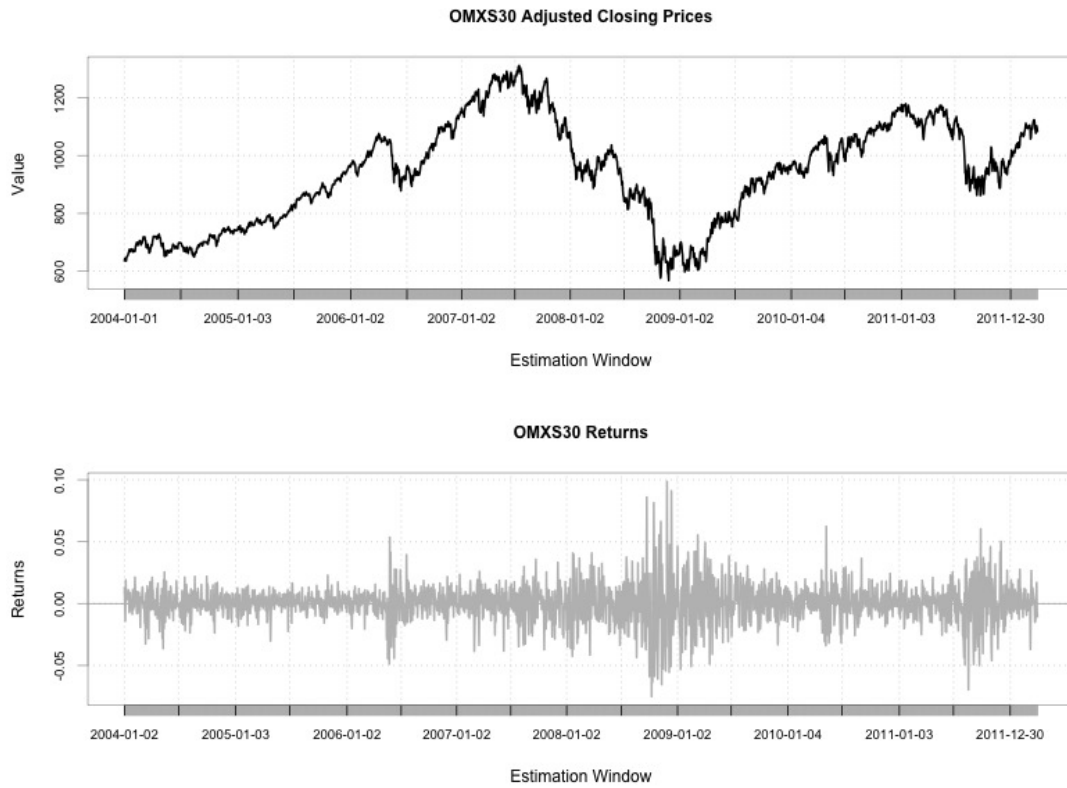


Figure 2: Adjusted closing prices (above) and returns (below) for OMXS30.

3.1.2 Descriptive Statistics and Q-Q Plots

The mean return is 0.0003, very close to zero, as expected since we have daily data and there is no time for a return to occur. This is confirmed by the t-test which shows that the mean is not statistically significantly different from zero as the test has a p-value of 0.4462.

The minimum value of the daily returns is -0.0751 occurred on 2008-10-06 and the maximum value is 0.0987 occurred on 2008-11-24. The first quantile and third quantile are -0.0070 and 0.0078 respectively. These values, combined with the lower part of Figure 2 describe the return series.

The returns have a skewness of 0.0544 and more importantly $kurtosis = 7.3103$ which is 4.3103 higher kurtosis than the normal distribution. This higher kurtosis is observed in the left part of Figure 3. We could use these values to test if the data is normal using a Jarque-Bera test but we

rely on the Q-Q plot instead because of the difficulties in trying to determine an exact distribution for the residuals *ex ante*.

In Figure 3, we can see that the distribution is leptokurtic. In the histogram displaying the frequency of returns (left) we see that the distribution is more peaked at the mean than normality. From the normal Q-Q plot (right) we see that the return series exhibits fatter tails as is evident from its deviation above the straight line in the lower tail and its deviation below the straight line in the upper tail.

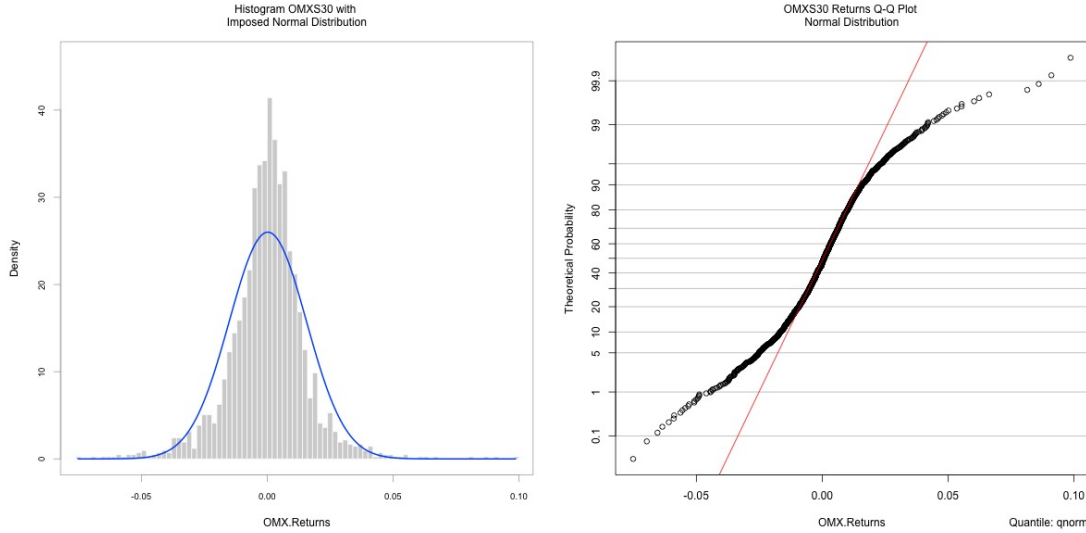


Figure 3: The histogram (left) shows that the distribution is leptocurtic and the Q-Q plot (right) shows that the return distribution has fatter tails than normality.

So what can be inferred from this? We see from the GARCH-model on page 15 that $r_t = \hat{\mu}_t + \hat{\sigma}_t z_t$ where z_t is assumed to follow some distribution. Does this mean that z_t should be assumed to follow a fat tailed distribution like the t-distribution? What needs to be remembered is that z_t is the distribution of the standardised residual (not the raw return) making them highly dependent on the specific volatility model applied. Our data shows clear kurtosis, suggesting that the assumption of a Student's t-distribution for the white noise term might be more appropriate than that of normality. Although, this is provided that the volatility model does not pick up all of the tail events and the peakiness around the mean. However, as we discussed in section 2.1.6, some of the tail events are likely to be picked up by our volatility models but not all, making the assumption of a Student's t-distribution still quite probable to be superior to normality for most of the models but we cannot say for sure *ex ante* to applying the volatility model.

3.1.3 ACF Plots of Return

In Figure 4 we display two ACF plots. On the y-axis we have the value of the ACF, see γ_k in equation (3), and on the x-axis we have the number of lags k . The left plot is the ACF of the return series and the right plot is the ACF of the squared returns. We use the ACF plot to see if

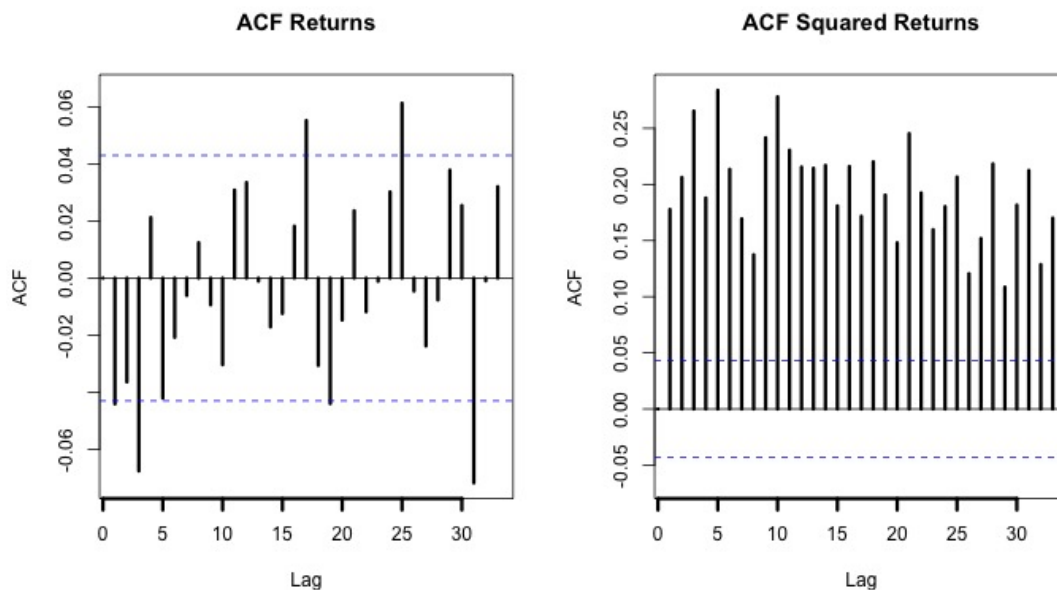


Figure 4: ACF plots for returns and squared returns.

the time series are autocorrelated or not. If γ_k lies within the dotted lines for many values of k then the data is not serially correlated, otherwise it is. The number of "many" is up to researcher. Formal statistical tests may be used---in this case the Ljung-Box test---and we do use this test in conjunction with the visual ACF plots to thoroughly evaluate whether the series are autocorrelated or not. Although, we consider the ACF plot to be a more useful tool than the Ljung-Box test and rely more on the plots.

By looking at the left plot in Figure 4 we see that there is some autocorrelation, though not very distinct. The Ljung-Box test give a p-value of 0.0045 which indicates that the lags are indeed autocorrelated and this would indicate the need to model the mean equation. However, the t-test that was conducted in subsection 3.1.2 did not show that the mean of the return series was statistically significantly different from zero leading us to conclude that the mean is too small to have any significant effect on the results. Because of this, we will not model the mean equation μ_t with an AR(p) or ARMA(p,q) model but instead use a constant mean μ , which for most practical purposes can be thought of as zero. We will therefore sometimes refer to innovations and returns interchangeably.

3.1.4 ACF Plots of Squared Return

By looking at the right part of Figure 4 we see that squared returns exhibit strong autocorrelation according to the ACF plots and this is verified by a p-value of less than $2.2 \cdot 10^{-16}$ for the Ljung-Box test. A t-test also confirms that the squared returns are indeed statistically significantly different from zero since the p-value is zero. Hence there exists ARCH effects. This is an argument for using a GARCH model.

Furthermore, the Ljung Box-test display a p-value close to zero so we draw the same conclusions as the ACF plots (namely that squared returns are correlated) thus the formal tests point in the same direction as the plots---we should use a GARCH model for our data.

3.1.5 Conclusion of EDA

What distribution should be assumed for the white noise term? According to our EDA we deem it more probable that a fat-tailed distribution best fits the residual assumption (though this is very uncertain). We also conclude that there is no need to model the mean equation because it is not statistically significantly different from zero, therefore we assume it to be constant and, for all practical purposes, essentially zero. There is, however, strong autocorrelation in the squared returns giving apparent reason that a GARCH-type model should be used to model the conditional volatility.

The asymmetric models are likely to perform better than the symmetric models because of the leverage effect. Exactly which one, however, is difficult to say because the main differences between the models are minor tweaks to how the asymmetric treatment is modelled.

3.2 GARCH Model Competition

In this section we do a rolling fit of all the 64 GARCH models. From these models we calculate their individual VaR estimates and backtest these estimates. Based on the backtesting results we rank the models from best to worst.

From the rolling window procedure¹⁶ each GARCH model produce 1000 VaR forecasts. These forecasts are backtested using cc test and uc test. The models are then ranked according to their p-value from the cc test from high to low.¹⁷ Results from the backtesting are summarised in Table 2. This table shows the top 10 rows from Table 6 in Appendix section A.2.

A clear result from Table 2 is that asymmetric models perform much better than symmetric models, because there are no symmetric models in the table (which is what we expected beforehand). More precisely, the AVGARCH, eGARCH and apARCH models perform well---as do models with sstd and NIG distributed innovations (even though the winner has the assumption of normality). The winning model having the normal assumption shows the difficulty of trying to predict the best residual assumption *ex ante*.

For this data set there were two winning models, AVGARCH(2,1)-N and eGARCH(2,1)-std. We have done a thorough analysis of both models but because they do not differ in a meaningful way we will only present the former here.

¹⁶As a detail, in the rolling window procedure the parameters are refitted every 100 days. One model did not converge, this was the NAGARCH(2,1)-sstd.

¹⁷If two models have the same p-value from the cc test we rank them based on the uc test's p-value. These p-values are referred to as cc-p and uc-p. The reason for ranking on the p-value is that we wish to accept the null hypothesis that the violations are not clustered and that the number of violations are correct with respect to VaR's confidence level. The higher the p-value the farther away we are from rejecting the null. Hence this is a reasonable sorting criteria.

The remaining part of this thesis will compare the winning model AVGARCH(2,1)-N from this table with a benchmark model and discuss how they differ. We will also mention why it's important in risk measurement.

	cc-p	uc-p	actual
AVGARCH(2,1)-N	0.949	0.885	51
eGARCH(2,1)-std	0.949	0.885	51
eGARCH(2,1)-N	0.904	0.884	49
apARCH(1,1)-sstd	0.847	0.770	48
apARCH(2,1)-sstd	0.847	0.770	48
AVGARCH(2,1)-sstd	0.847	0.770	48
apARCH(1,1)-nig	0.847	0.770	48
apARCH(2,1)-nig	0.847	0.770	48
AVGARCH(2,1)-nig	0.847	0.770	48
AVGARCH(1,1)-std	0.845	0.566	54

Table 2: Top 10 models from the backtesting evaluation.

3.3 Model Comparison of Winner and Benchmark

In this section we compare a benchmark model with the winning model from Table 2. Thus the two models to be compared are:

1. the winning model AVGARCH(2,1)-N also called *winner*, and
2. the benchmark model sGARCH(1,1)-N also called *benchmark*.

Firstly, we motive our choice of benchmark model as well as perform a fit and comparison of the two models in the estimation window. Secondly, we compare the VaR backtesting results and VaR forecast plots from these models and discuss the results. The VaR forecast plots are accompanied by the very important news impact curve. These two plots highlights the key differences that distinguishes the winner from the benchmark. Lastly, we assess model fit and assumptions using diagnostical plots of the standardised residuals.

3.3.1 Fitting the GARCH Model and Estimated Coefficients

Why do we use an sGARCH(1,1)-N as the benchmark? In choosing our benchmark we have to decide on three things: Firstly, the family: standard, E-, GJR-, AV-, N- etc. We chose the sGARCH because it is heavily used as well as being the first GARCH model to be developed. Secondly, the number of lagged innovations: either (1,1) or (2,1) model. We chose (1,1) because it is the standard choice. Thirdly, the assumed distribution for the innovation: normal, std, sstd, and NIG. We chose a normal distribution because our winning model AVGARCH-N have a normal distribution and this make the comparison more appropriate.

The winner and benchmark models are fitted to the data from the estimation window. The results are seen in Table 3 and Table 4. Table 3 shows the unconditional mean, unconditional

variance, persistence, halflife, likelihood, and information criteria. Table 4 shows rounded estimated coefficients, their standard errors, the t values, and p-values. Note that this is going to be the first set of parameters $\hat{\theta}_1^{(1)}$ in our rolling window analysis, see details in section 2.4.5. These estimates are used in the calculation of VaR_1 . We will analyse these first estimated coefficients and some important facts are summarised in two separate bullet lists commenting the two tables.

	AVGARCH(2,1)-N	sGARCH(1,1)-N
Unc. Mean	0.000082	0.000749
Unc. Variance	0.000225	0.000223
Persistence	0.980318	0.988307
Halflife	35	59
Likelihood	6160	6109
AIC	-5.926	-5.881
HQIC	-5.917	-5.878

Table 3: Coefficient Comparison

The information in Table 3 is commented on in the bullet points below.

- The information criterions AIC and HQIC are very close to each other but they slightly favour winner over benchmark.
- Recall that *halflife* measures how long it takes for half of the shock to disappear, see details in section 2.4.1. Because *halflife* for winner is lower than for benchmark, we know that the winning model will go back to its unconditional variance faster after a shock. (This difference in inertia is also reflected in the the calculated persistence where $Persistence^{winner} < Persistence^{benchmark}$ so these two estimates show the same thing but *halflife* have a nicer economic interpretation.)
- The unconditional mean is lower in winner than in benchmark.
- The unconditional variance is about the same for both models. This means that in the long run both models revert back to the same mean variance.

The information in Table 4 is commented on in the bullet points below.

- We see that β_1 is close to 1 and α_1 close to zero for both models---as is expected with financial data.
- The α_1 are pretty close to each other, 7/10 vs 8/10. For benchmark α_1 is not significant whereas for winner it is.
- There is only an α_2 for winner because it is a (2,1) model. It is the weight put on the innovation that occurred two days before. It is natural to expect a lower β_1 when α_2 is present because ε_{t-2} is also present in σ_{t-1} .

AVGARCH(2,1)-N				
	Estimate	Std. Error	t value	p-value
ω	0.0003	0.00	10.50	0.000
α_1	0.0705	0.00	42.97	0.000
α_2	0.0491	0.00	16.24	0.000
β_1	0.8575	0.01	129.11	0.000
η_{11}	1.0000	0.10	10.38	0.000
η_{12}	0.6595	0.02	26.76	0.000
η_{21}	0.3518	0.04	8.05	0.000
η_{22}	-2.2701	0.04	-52.37	0.000
sGARCH(1,1)-N				
	Estimate	Std. Error	t value	p-value
ω	0.0000	0.00	0.36	0.721
α_1	0.0883	0.05	1.71	0.088
β_1	0.9000	0.06	14.61	0.000

Table 4: Coefficients and robust standard errors.

- The η values are not discussed in detail, but rather seen as the values that create the shift and rotation in the news impact curve that is presented later on.

By inserting the estimated parameters into the AVGARCH formula, Equation (23) on page 22, we get the estimated model

$$\begin{aligned}
\sigma_t &= \omega + \alpha_1 \sigma_{t-1} (|z_{t-1} - \eta_{21}| - \eta_{11}(z_{t-1} - \eta_{21})) + \\
&\quad + \alpha_2 \sigma_{t-2} (|z_{t-2} - \eta_{22}| - \eta_{12}(z_{t-2} - \eta_{22})) + \beta_1 \sigma_{t-1} = \\
&= 0.0003 + 0.0705 \sigma_{t-1} (|z_{t-1} - 0.3518| - 1.0000(z_{t-1} - 0.3518)) + \\
&\quad + 0.0491 \sigma_{t-2} (|z_{t-2} + 2.2701| - 0.6595[z_{t-2} + 2.2701]) + 0.8575 \sigma_{t-1}.
\end{aligned}$$

Because we allow AVGARCH(2,1) to model the innovations more precisely---through treating the $t - 1$ lag and $t - 2$ lag individually---it is better at forecasting VaR in our out-of-sample setting. However, this does not necessarily have to be the case because there could be some change between w_E and w_F that could make a (1,1) model better. Remember that η_{2i} shifts the news impact curve and η_{1i} rotates the news impact curve.

It is worth noting that the coefficients are not the same throughout our entire forecasting window, because we do a rolling window procedure where the parameters are updated. Updating the parameters does not only change the coefficients but also the standard errors and thereby the p-values. Thus the low p-values in Table 3 should be taken with a grain of salt---it is only for the first set of parameters $\hat{\theta}^{(1)}$. After the specified time newly available data is used to reestimate the parameter vector and we get the second set of parameters $\theta^{(2)}$ and so on. If there is a more volatile period then some parameter values might become insignificant, since during volatile periods the standard errors increase since we get less certain about our estimates. This is one reason why models

are not as suitable during periods of high volatility. Therefore, during crisis, we should trust our models less.

Do note that we have done a new fit for the benchmark model and the winning model. The reason we do a new rolling window estimation (instead of keeping the old one that was used when the models competed against each other) is so that we can now refit the parameter vectors every 20 days¹⁸ allowing for more precise estimates.¹⁹

As a last note, we do the fit using data from the estimation window. The purpose of using only the estimation window, and not the entire sample is that we want to reserve the forecasting window to test our model---peeking at the forecasting window beforehand would be statistical cheating.

3.4 VaR Backtesting

In this section we show why the AVGARCH(2,1)-N outperforms the benchmark according to VaR backtesting, described in 2.3.3. We also plot the VaR forecasts and calculate the MSE.

3.4.1 Tabulated Results

Since we use a five percent VaR and a forecasting window of length 1000 the expected number of violations is $0.05 \cdot 1000 = 50 = \textit{expected}$ regardless of which model is used. The actual violations are $\textit{actual} = \sum_{t=1}^{n(w_F)} I_t$ in other words the sum of I_t in the forecasting window. The farther away *actual* is from *expected* the worse the model is. If $\textit{actual} < \textit{expected}$ the model is said to overestimate risk, and if $\textit{actual} > \textit{expected}$ the model is said to underestimate risk. Overestimating the risk is bad for the bank because it leads to unnecessary capital reserves and it is bad for the society because resources are not used efficiently. Moreover, underestimating the risk is bad for the bank because it leads to punishments from regulators and, in the worst case scenario, may even force a liquidation of the bank which can be very costly for society at large. For these reasons the best models have *actual* very close to *expected*.

The unconditional coverage test (uc test) tells us if the number of violations are correct. The null may be framed as $H_{0,uc} = \textit{number of violations are correct}$. The conditional coverage test (cc test) tells us if violations are clustered and if the number of violations are correct. The null may be framed as $H_{0,cc} = \textit{violations are not clustered and number of violations are correct}$. For details of these test see section 2.3.3. We apply the cc test as well as the uc test and summarise the findings in Table 5. In this table it is obvious that winner outperforms benchmark with respect to VaR backtesting.

Given the formulation of the null hypothesis of the cc test and the uc test, we wish to “accept” the null hypothesis (or more formally “fail to reject”). When both models succeed at the hypothesis testing we can decide which one is the better by looking at which model have the highest p-value since the higher the p-value the farther away we are from rejecting the null. Using this idea, winner

¹⁸A refit every day would be excessive, the parameters do not change that much.

¹⁹It was impractical to get these precise estimates before because we fitted 64 models so we did a refit every 100 days. Now, however, we now only need to do it for two models this is computationally possible to refit more often.

	expected	actual	cc-p	uc-p
Winner AVGARCH(2,1)-N	50	51	0.949	0.885
Benchmark sGARCH(1,1)-N	50	65	0.104	0.037

Table 5: Results from VaR backtesting. Winner is better than benchmark, based on actual number of violations as well as the p-values from cc test and uc test.

is better than benchmark with respect to VaR backtesting because the p-value for the uc test and cc test is higher than for benchmark. (If given the choice to interpret either uc-p or cc-p we would chose cc-p since the null of the latter tests clustering as well as unconditional coverage.)

Even someone who is not familiar with interpreting p-values should notice that winner outperforms benchmark with respect to VaR backtesting simply by comparing the actual number of violations and noting that 51 is closer to *expected* = 50 than 65 is. This is important because with 15 more violations than what is expected VaR is not a reliable measure of risk. This would lower a banks readiness to severe economic downturn and would also be call for reprimands from the regulators.

3.4.2 VaR Forecast Plot

In Figure 5 we plot the forecast value at risk together with the realised return rr_t as well as marking the violations with a circle. Hence, if a violation have occurred at time t then this day is marked with a circle on the horizontal axis. An inspection of how the circles are spread out can be considered a (crude) visual test if the violations are independent. Figure 5 is the most important figure in this thesis.

The VaR forecast VaR_t varies over time, and when returns are more volatile the VaR forecast increases---this behaviour is evident from the GARCH-formula by the relationship between $\hat{\sigma}_t$ and r_{t-1} . The forecast VaR for the two models have similar properties because the lines exhibit similar patterns. But the winning model adapts faster to changing circumstances, as can be seen from the volatile period towards the end of the sample, and it incorporates the asymmetric effects, as can be seen during periods of large losses. That the winning model adapts faster may be good for estimating risk precisely but it makes it difficult for the bank to allocate capital correctly since the risk changes so often.

3.4.3 Impact of Asymmetry

Financial data often exhibit an asymmetric effect, as discussed in section 2.4.2 on page 19. The winning model AVGARCH(2,1)-N is able to incorporate this asymmetry into its treatment of innovations. Something that the benchmark model is unable to do. The asymmetric effects are illustrated in the news impact curve in part (a) of Figure 6 which is the second most important figure in this thesis.

Given a *positive* innovation $\varepsilon_{t-1} > 0$ we have that $\hat{\sigma}_t^{winner} < \hat{\sigma}_t^{benchmark}$. In other words when there is good news the winning model blows up volatility *less* than the benchmark model does. This is desired because sufficiently large positive returns does not give cause for as much an increase in

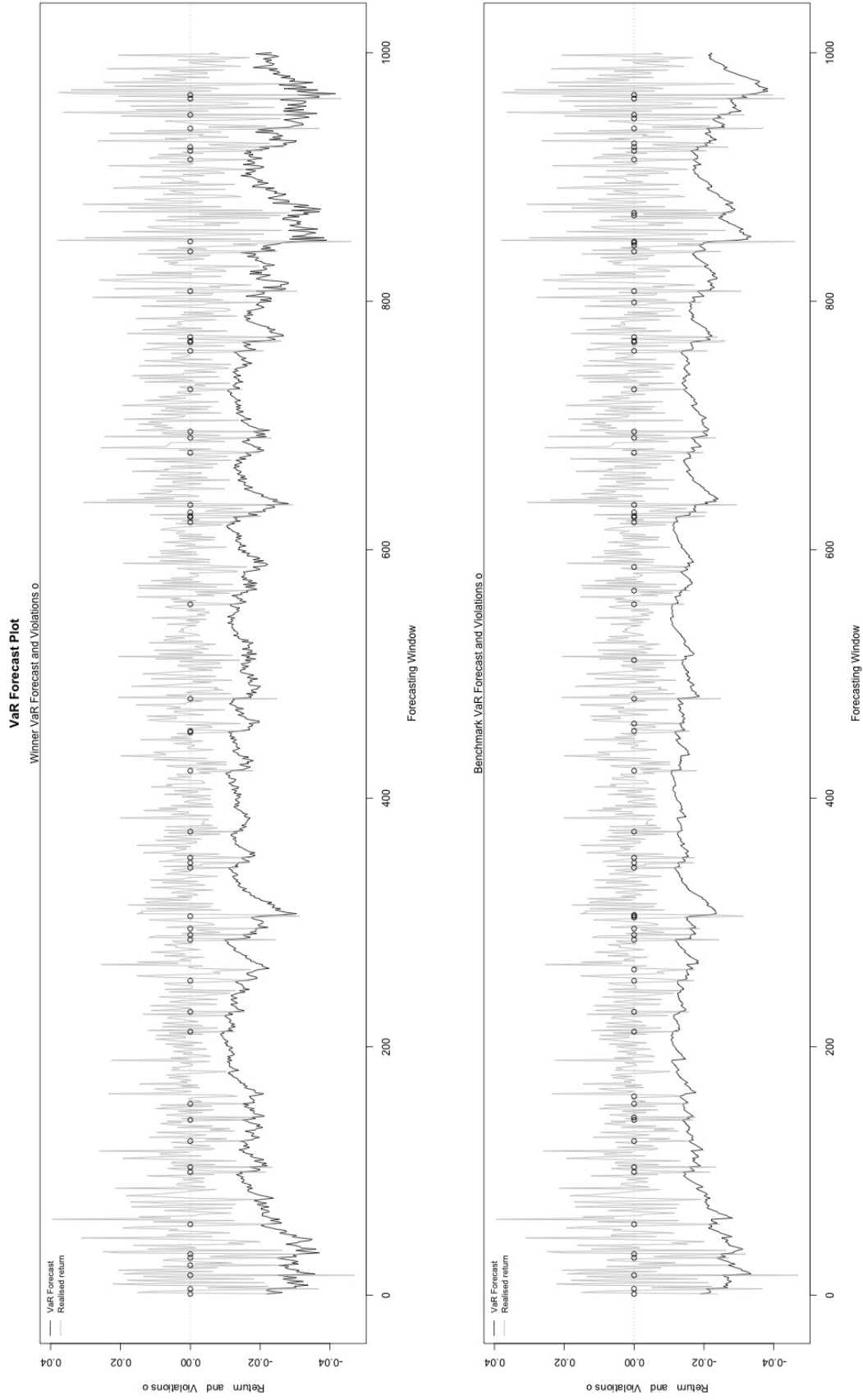


Figure 5: VaR Forecast Plot. Violations and realised return for both the winning model and the benchmark model. Violations are marked with a circle as a visual indication of the violation.

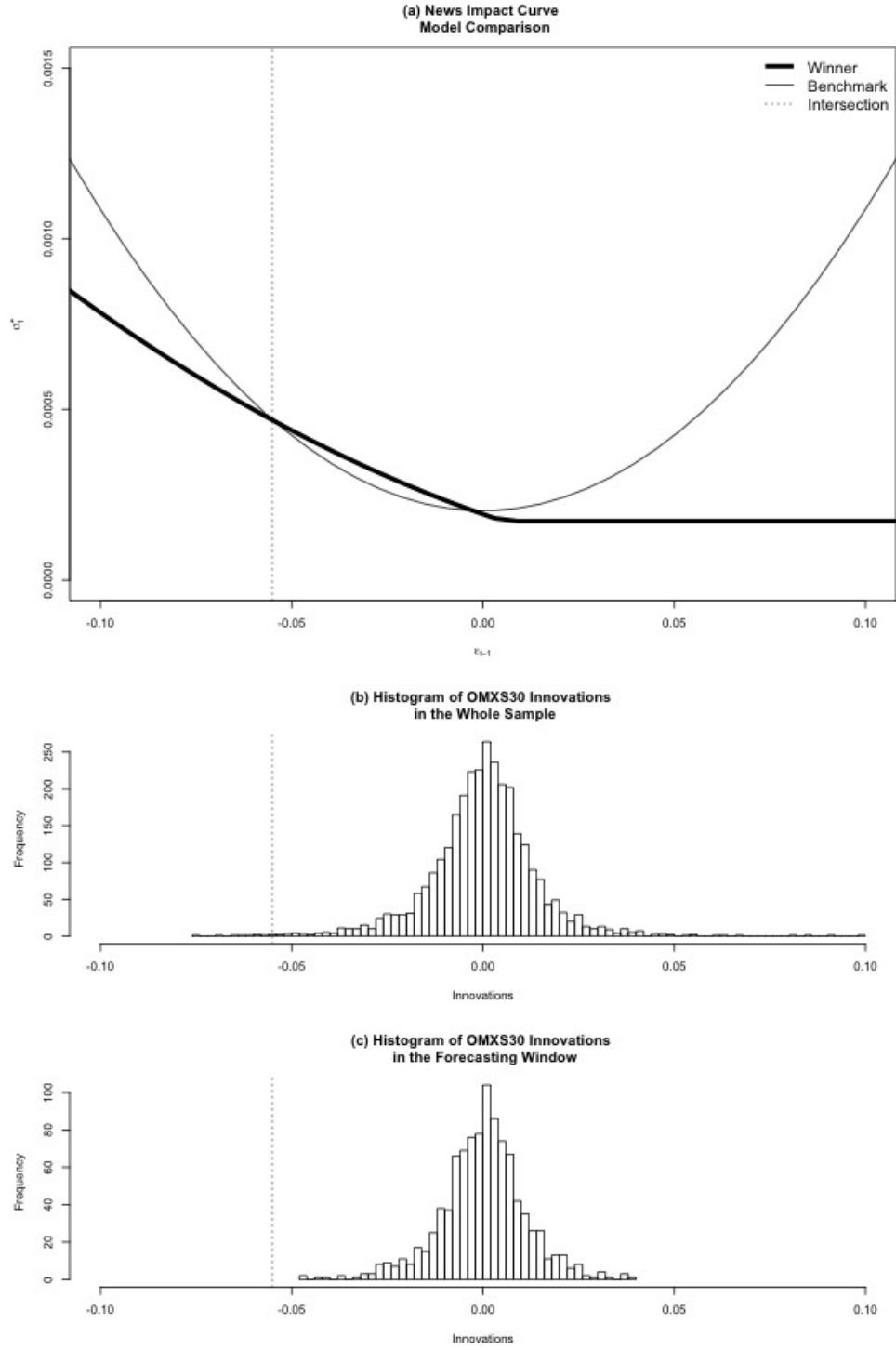


Figure 6: News impact curve for both models. The winning model increase the forecast volatility $\hat{\sigma}_t$ more when the shock ε_{t-1} is negative.

risk. Avoiding to blow up the volatility forecast when times are good will improve the MSE which makes the capital allocation more efficient.²⁰

Given a *negative* innovation $\varepsilon_{t-1} < 0$ we have that $\hat{\sigma}_t^{winner} > \hat{\sigma}_t^{benchmark}$. In other words when there is bad news the winning model affects the volatility forecast *more* than the benchmark model does. As discussed under subsection 2.4.2 this is done because volatility tends to increase more when prices fall due to the leverage effect.

3.4.4 MSE

The Mean Squared Error (MSE) is a measure of how tight the estimated VaR line is to the returns. A low MSE measure will be shown in the VaR plot as having very little white space between the gray realised return and the black VaR line. It matters to banks because of capital requirements. MSE can be thought of as a proxy of how much money is put in the reserve.

MSE is calculated from data in the forecasting window just like the cc test and uc test is. The cc test is used to see if the actual number of violations correspond to the expected and if these violations are independent or clustered. By contrast, MSE is used to see how tight the VaR forecast is to the realised return. The lower the MSE the better the model is.

The calculated values are $MSE^{winner} \approx 4.6 \cdot 10^{-8}$ and $MSE^{benchmark} \approx 4.9 \cdot 10^{-8}$ and the fraction is

$$\frac{MSE^{winner}}{MSE^{benchmark}} \approx 0.94$$

telling us that the the winner model have a 6 percent lower MSE than the benchmark.

3.4.5 Conclusion of VaR Backtesting

There are three distinct reasons why winner is more appropriate than benchmark. (1) The cc test and VaR forecast plots suggest violations are less clustered in winner. (2) The actual violations are closer to the expected for winner. (3) MSE is lower for winner.

3.5 Model Diagnostics

In this section we show that the standardised residuals are close to white noise, hence the model assumptions are fulfilled.

As described in section 2.5 we would like to conclude that the standardised residuals

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

²⁰The inequality is a minor simplification of the truth so we give the whole story in this footnote. The careful reader observes that when ε_{t-1} is sufficiently negative (e.g less than -0.055) then in fact $\hat{\sigma}_t^{winner} > \hat{\sigma}_t^{benchmark}$ but as illustrated in the histograms in Figure 6 (b) and (c). It is rarely the case that ε_{t-1} is sufficiently negative in the full sample, as seen by histogram (b)---to be precise it occurs 9 times. And it never occurs in the forecasting window, as seen by histogram (c).

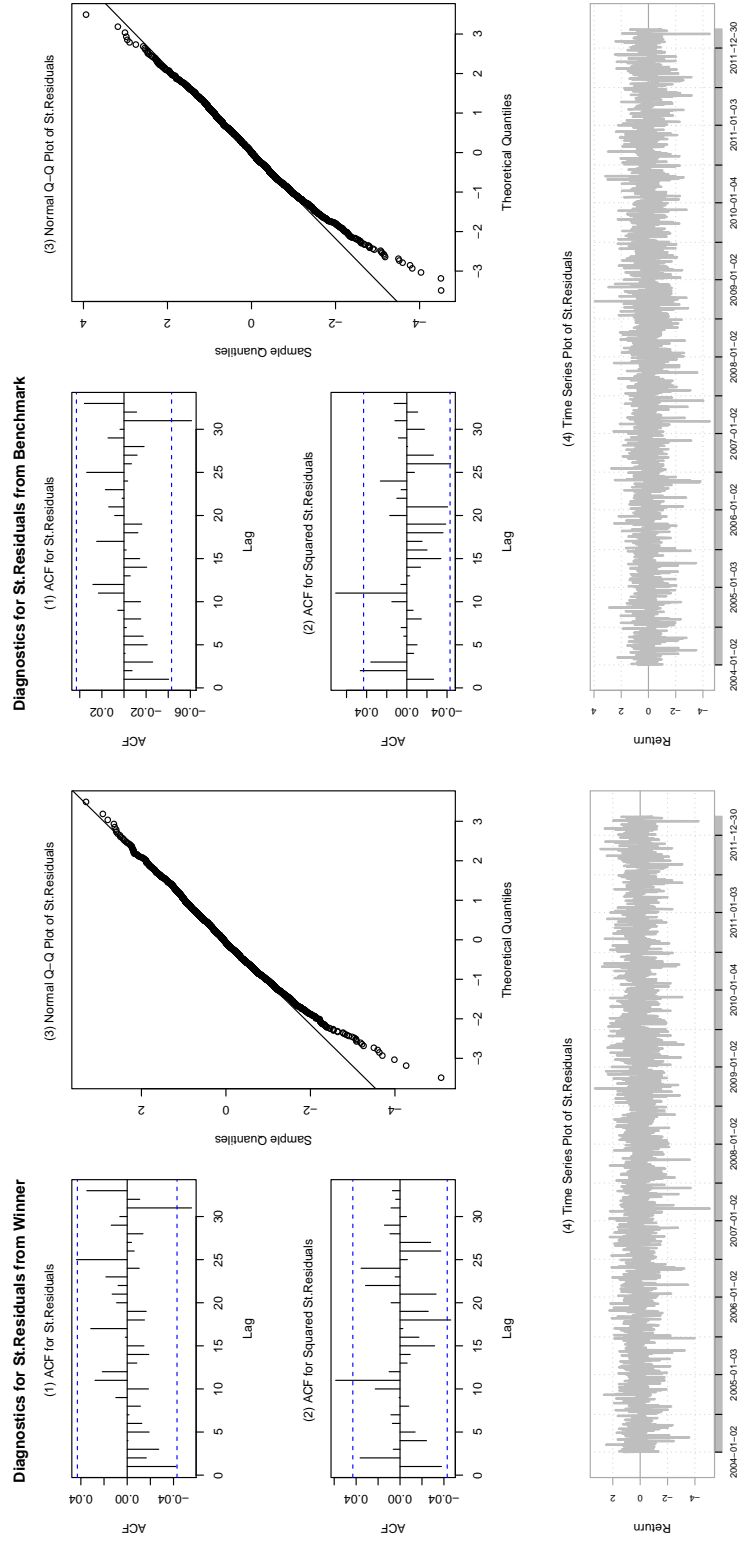


Figure 7: Winner (left) and Benchmark (right) model assumption check where St. Residual stands for standardised residuals. Assumptions (1) to (4) are fulfilled for both models.

follow a white noise process. To be able to say that $\tilde{\epsilon}_t$ follows a white noise process we want the following assumptions to be fulfilled:

- (1) uncorrelated $\tilde{\epsilon}_t$,
- (2) uncorrelated $\tilde{\epsilon}_t^2$,
- (3) $\tilde{\epsilon}_t \sim N(0, 1)$, and
- (4) $\tilde{\epsilon}_t$ looks like white noise.

Assumption (1) is needed because then the mean equation is good. In this thesis we have used a constant (rather than an ARMA model) for the mean. The residuals from this mean equation are then modelled by a GARCH model. The assumption is tested by an ACF plot of the standardised residuals. Autocorrelation in $\tilde{\epsilon}_t$ is not present if the value of γ_k for most lags k lies within the dotted lines in an ACF plot. It can also be tested formally using a Ljung-Box test.

Assumption (2) is needed because then the volatility equation is good. It is tested the same way as the first assumption but using $\tilde{\epsilon}_t^2$ instead of $\tilde{\epsilon}_t$. If there is no correlation there is no ARCH effects left in the residuals, indicating that the GARCH model was successful in removing the ARCH effect we observed in Figure 4 on page 30.

Assumption (3) is needed because when fitting a GARCH model we have assumed a certain distribution (for the winning model we assumed a normal distribution). If that assumption is wrong then the model assumptions are not met. It is tested by plotting the theoretical quantiles of the assumed distribution against the sample quantiles of $\tilde{\epsilon}_t$ in a Normal Q-Q Plot. If the theoretical distribution is similar to the sample distribution it will be a straight line.

Assumption (4) is that the standardised residuals should *look* like white noise, which is of course needed in order for us to draw the conclusion that we consider it to *be* white noise. This is seen visually from drawing a time series plot of $\tilde{\epsilon}_t$. In the best of worlds you would compare this to plots with simulated white noise and see if they look alike, but for brevity we skip this.

In Figure 7 these assumptions are evaluated for the winner AVGARCH(2,1)-N and the benchmark sGARCH(1,1)-N. Both plots look very similar. More importantly, both models fulfil the four assumptions reasonably well though there is a small deviation in the the Q-Q plots making it not a perfect fit. Especially the lower tails deviate from the normal distribution. If one cares about a 1 percent VaR or a conditional VaR then it would be more appropriate to use another distribution or perhaps model the tails using extreme value theory. However, this thesis is concerned with the 5 percent VaR so the departure from the normal distribution below the 5th percentile does not affect the the estimate making it not as sensitive to the deviation observed here.

4 Conclusions

In this thesis we differentiate between several potential models when it comes to VaR forecast precision and rank them accordingly using conventional backtesting methods in an out-of-sample setting. The winning model, AVGARCH(2,1)-N, is considerably better than the benchmark, sGARCH(1,1)-N. The benchmark model has 30% more violations than what is expected making it unsatisfactory as a risk measure. The winning model only has 2% more violations than what is expected---making it a reliable risk measure for the analysed time series. See Figure 5 for an overview of how the models perform throughout the forecasting window.

Previous literature and common conception is that the GARCH(1,1) model is an appropriate model for forecasting conditional variance. We have shown the importance of selecting a model with regards to its end application and purpose---in our case to forecast VaR---and the importance of evaluating the models with regard to this same end purpose. What is usually assumed more is that the (1,1) models are enough so that any more developed models are often not considered. In our analysis, 7/10 of the 10 best models are (2,1) models, suggesting that this more developed form proved better when it comes to forecast VaR in our data set.

What makes the winning AVGARCH(2,1)-N model outperform the sGARCH(1,1)-N model is primarily its asymmetric treatment of positive and negative innovations, seen in Figure 6. It is apparent that OMXS30 exhibit a leverage effect in its market reactions to price changes and this can be picked up by the winning model. Something that the benchmark model is unable to do. Not only does this make the benchmark model underestimate the risk after negative price movements---which makes us observe more violations---but this also makes the benchmark model overestimate the risk after positive price movements---which leads to bad capital allocation efficiency (which is captured in the lower MSE for the winning model). For this data set the asymmetric models prove substantially better than the symmetric ones.

To test the robustness of these results we apply the same method to other stock market indices and find that the asymmetric models clearly outperform throughout the data sets. Although, there are slight variations between the specific asymmetric models that prevail. We do not, however, find that the (2,1) models prove considerably better in the other indices.

These observations show the importance of choosing a model with regards to its end use and purpose as well as the importance of not adhering too much to convention. We also demonstrate that asymmetric models outperform symmetric models and that this finding also holds for different indices.

4.1 Future Research

We see four potential areas for future research that would be of interest. Because we limit our analysis to stock indices it would be interesting to see whether the asymmetric models also outperform when analysing other asset classes such as bonds, foreign exchange and commodities.

Another important analysis that would be of interest to look into is the critique of Christoffersen's test that it only treats consecutive violations as clusters and not violations with one or two days between them, discussed under subsection 2.3.3. This test is one that we rely heavily on and perhaps some other backtesting evaluation method would yield a different result.

Because of the growing importance of conditional VaR---especially after the latest revision of the Basel Accords---it would be interesting to do an evaluation of the different models' forecast precision with respect to their conditional VaR backtesting performance.

Model uncertainty has been ignored in this thesis. We have estimated GARCH parameters with MLE and used these estimates to calculate volatility forecasts. The estimated parameters, however, are uncertain as can be seen from their standard errors. A practitioner needs to keep model uncertainty in mind. One way to demonstrate this uncertainty in a richer way is to not only report the standard errors but also (1) plot the parameters over time with their standard error bands, (2) plot the density of the parameter estimates (3) do a simulation with some given set of parameters and compute the RMSE. This could potentially be used to further differentiate between the models.

5 References

Articles

- Aït-Sahalia, Y., Fan, J. and Li, Y., 2013, The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency, *Journal of Financial Economics* 109(1), 224-249.
- Barndorff-Nielsen, O., 1978, Hyperbolic Distributions and Distributions on Hyperbolae, *Scandinavian Journal of Statistics* 5(3), 151-157.
- Black, F., 1976, Studies of Stock Price Volatility Changes, *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
- Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31(3), 307-327.
- Bollerslev, T., 1987, A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *The Review of Economics and Statistics* 69(3), 542-547.
- Christoffersen, P. F., 1998, Evaluating Interval Forecasts, *International Economic Review* 39(4), 841-862.
- Diamond, D. W. and Dybvig, P., 1983, Bank Runs, Deposit Insurance, and Liquidity, *Journal of Political Economy* 91(3), 401-419.
- Dickey, D. A., Fuller, W. A., 1979, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* 74(366a), 427-431.
- Ding, Z., Granger, C. W. J. and Engle, R. F., 1993, A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* 1(1), 83-106.
- Engle, R. F., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50(4), 987-1008.
- Engle, R. F. and Ng, V. K., 1993, Measuring and Testing the Impact of News on Volatility, *The Journal of Finance* 48(5), 1749-1778.
- Fernandez, C. and Steel, M. F. J., 1998, On Bayesian Modeling of Fat Tails and Skewness, *Journal of the American Statistical Association* 93(441), 359-371.

- Giot, P. and Laurent, S., 2004, Modelling daily Value-at-Risk using realized volatility and ARCH type models, *Journal of Empirical Finance* 11(3), 379-398.
- Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* 48(5), 1779-1801.
- Hansen, P. R. and Lunde, A., 2005, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, *Journal of Applied Econometrics* 20(7), 873-889.
- Harvey, C. R. and Siddique, A., 1999, Autoregressive Conditional Skewness, *The Journal of Financial and Quantitative Analysis* 34(4), 465-487.
- Higgins, M. L. and Bera, A. K., 1992, A Class of Nonlinear ARCH models, *International Economic Review* 33(1), 137-158.
- Kareken, J. H. and Wallace, N., 1978, Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition, *The Journal of Business* 51(3), 413-438.
- Kuester, K. and Mittnik, S. and Paolella, M. S., 2006, Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics* 4(1), 53-89.
- Kupiec, P. H., 1995, Techniques for Verifying the Accuracy of Risk Measurement Models, *The Journal of Derivatives* 3(2), 73-84.
- Ljung, G. M. and Box, G. E. P., 1978, On a measure of lack of fit in time series models, *Biometrika* 65(2), 297-303.
- Mandelbrot, B., 1963, The Variation of Certain Speculative Prices, *The Journal of Business* 36(4), 394-419.
- Nelson, D. B., 1991, Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59(2), 347-370.
- Pagan, A. R. and Schwert, G. W., 1990, Alternative Models For Conditional Stock Volatility, *Journal of Econometrics* 45(1-2), 267-290.
- Praetz, P. D., 1972, The Distribution of Share Price Changes, *The Journal of Business* 45(1), 49-55.

Schwert, G. W., 1989, Why Does Stock Market Volatility Change Over Time?, *The Journal of Finance* 44(5), 1115-1153.

Schwert, G. W., 1990, Stock Volatility and the Crash of '87, *The Review of Financial Studies* 3(1), 77-102.

Books

Brooks, C., 2002, *Introductory Econometrics for Finance*, The Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

Engle, R. F. and Lee, G. G. J., A Long-Run and Short-Run Component Model of Stock Return Volatility, 475-486, in Engle, R. F. and White, H. (eds.), 1999, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Oxford University Press, Oxford, United Kingdom.

Hull, J. C., 2012, *Options, Futures, and Other Derivatives*, 8th ed., Pearson Education Limited, Essex, United Kingdom.

Jorion, P., 2007, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed., The McGraw-Hill Companies, Inc., Two Penn Plaza, New York.

Taylor, S. J., 1986, *Modelling Financial Time Series*, John Wiley & Sons, Inc., New York City, New York.

Tsay, R. S., 2010, *Analysis of Financial Time Series*, 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey.

Tsay, R. S., 2013, *An Introduction to Analysis of Financial Data With R*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Electronic Resources

Basel Committee on Banking Supervision, 2011, *Revisions to the Basel II market risk framework*, Bank of International Settlements, URL: <http://www.bis.org/publ/bcbs193.pdf> (accessed 2016-05-05).

Basel Committee on Banking Supervision, 2016, *Minimum capital requirements for market risk*, Bank of International Settlements, URL: <http://www.bis.org/bcbs/publ/d352.pdf> (accessed 2016-05-05).

Dierick, F., Pires, F., Scheicher, M. and Spitzer, K. G., 2005, *The New Basel Capital Framework and Its Implementation in the European Union*, European Central Bank, Occasional Paper Series 42, URL: <https://www.ecb.europa.eu/pub/pdf/scpops/ecbocp42.pdf?819df1d197f718a070e14b8418e98aac> (accessed 2016-05-05).

Duffie, D. and Pan, J., 1997, *An Overview of Value at Risk*, Massachusetts Institute of Technology, URL: <http://web.mit.edu/junpan/www/ddjpa.pdf> (accessed 2016-01-22).

Federal Reserve, 2013, *Basel Regulatory Framework*, Board of Governors of the Federal Reserve System, URL: <http://www.federalreserve.gov/bankinfo/basel/default.htm> (accessed 2016-05-05).

Ghalanos, A., 2015, *Introduction to the rugarch package.*, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf (accessed 2016-05-05).

Longerstaey, J. and Spencer, M., 1996, *RiskMetrics---Technical Document*, Morgan Guaranty Trust Company and Reuters Ltd, New York, URL: http://pascal.iseg.utl.pt/~aafonso/eif/rm/TD4ePt_2.pdf (accessed 2016-04-30).

Manganelli, S. and Engle, R. F., 2001, *Value at Risk Models in Finance*, European Central Bank, Working Paper Series 75, URL: <http://www.euro.zentral-bank.eu/pub/pdf/scpwps/ecbwp075.pdf> (accessed 2016-05-01).

A Appendix

A.1 Distribution Assumptions

The Normal Distribution

The normal (or Gaussian) distribution is one that is commonly assumed for the distribution of returns. If z_t is normally distributed then its PDF is

$$f(z_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \cdot \frac{(z_t - \mu)^2}{\sigma^2}\right) \quad (30)$$

and when z_t follows a standardized normal distribution we have that

$$E|z_t| = \sqrt{2/\pi}. \quad (31)$$

The validity of the normality assumption can often be questioned when it comes to financial return series because they often exhibit leptokurtosis, i.e. the distribution has fat tails and is more peaked at the mean (Praetz, 1972; Brooks, 2002:179f).

The Student's t-Distribution

To allow for an assumption that matches the leptokurtic distribution that financial data often exhibit we are going to evaluate the assumption of the Student's t-distribution. This was also the distribution that Bollerslev proposed in his article from 1987. If z_t follows a t-distribution then it has the following density function

$$f(z_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\beta\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{(z_t - \alpha)^2}{\beta\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \quad (32)$$

where $\Gamma(\cdot)$ is the gamma function, ν is the distribution's degrees of freedom, and α and β are the location and scale parameters respectively. When z_t follows a standardized Student's t-distribution we have that

$$E|z_t| = \frac{2\sqrt{\nu-2}\Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}. \quad (33)$$

Note that as $\nu \rightarrow \infty$ the fraction tends to $\sqrt{2/\pi}$ which mimics the fact that a t-distribution tends to a Normal distribution when $\nu \rightarrow \infty$.

The Skewed Student's t-Distribution

In many cases, the true distribution is not symmetrically distributed around the mean but instead skewed positively or negatively. Previous literature has found that returns often exhibit some skewness (Harvey and Siddique, 1999 and the references therein) and that skewed distribution assumptions often are superior in estimating VaR, for example see Kuuster et al. (2006). If it is assumed that the distribution is symmetric when the true distribution is skewed then the model becomes subject to skewness risk. To alleviate this we will assume that the WN term

is distributed as the skewed Student's t-distribution (Fernandez and Steel, 1998). If z_t follows a skewed Student's t-distribution then its PDF is

$$f(z_t|\xi) = \frac{2}{\xi + \xi^{-1}} [f(\xi z_t)H(-z_t) + f(\xi^{-1}z_t)H(z_t)] \quad (34)$$

where $\xi \in \mathbb{R}^+$ is the skew parameter and $H(\cdot)$ is the Heaviside function. When the distribution is symmetric $\xi = 1$. This means that if we assume this distribution then the model can account for distributions that are skewed as well as symmetric. Because financial data often exhibit skewness combined with the flexibility of this assumption makes it likely that the skewed Student's t-distribution will prove superior to the normal and Student's t-distribution.

The Normal-Inverse Gaussian Distribution

Another distribution that can account for fat tails and skewness is the normal-inverse Gaussian (NIG) distribution (Barndorff-Nielsen, 1978). The PDF of the NIG distribution is

$$f(z_t) = \frac{\alpha \delta K_1 \left(\alpha \sqrt{\delta^2 + (z_t - \mu)^2} \right)}{\pi \sqrt{\delta^2 + (z_t - \mu)^2}} \exp(\delta \gamma + \beta(z_t - \mu)) \quad (35)$$

where δ is a scale parameter, K_1 is the modified third kind Bessel function, α denotes the tail heaviness, β denotes the asymmetry parameter and $\gamma = \sqrt{\alpha^2 - \beta^2}$.

A.2 Complete Results from the Backtesting Procedure

