Predicting Uncertainty in Financial Markets

-An empirical study on ARCH-class models ability to estimate Value at Risk

Authors: Henrik Rundberg & Arvid Nybrant
Abstract

Value at Risk has over the last couple of decades become one of the most widely used measures of market risk. Several methods to compute this measure have been suggested. In this paper, we evaluate the use of the GARCH(1,1)-, EGARCH(1,1)- and the APARCH(1,1) model for estimation of this measure under the assumption that the conditional error distribution is normally-, t-, skewed t- and NIG-distributed respectively. For each model, the 95% and 99% one-day Value at Risk is computed using rolling out-of-sample forecasts for three equity indices. These forecasts are evaluated with Kupiec’s test for unconditional coverage test and Christoffersen’s test for conditional coverage. The results imply that the models generally perform well. The APARCH(1,1) model seems to be the most robust model. However, the GARCH(1,1) and the EGARCH(1,1) models also provide accurate predictions. The results indicate that the assumption of conditional distribution matters more for 99% than 95% Value at Risk. Generally, a leptokurtic distribution appears to be a sound choice for the conditional distribution.

**Keywords:** VaR, GARCH, Volatility Forecasting, Backtesting, Conditional Heteroscedasticity.
Acknowledgements

We would like to give a special thanks to our supervisor Dr. Lars Forsberg for his guidance and valuable thoughts during the work with this thesis.
# Index

1. Introduction ................................................................................................................. 1  
2. Theoretical Framework ................................................................................................. 3  
   2.1 Value at Risk ........................................................................................................... 3  
   2.2 Autoregressive Conditional Heteroscedasticity Models ........................................... 4  
   2.3 Conditional Error Distributions ............................................................................... 7  
3. Earlier Research ............................................................................................................ 10  
4. Method .......................................................................................................................... 11  
   4.1 Backtesting Procedure ............................................................................................. 11  
   4.2 Delimitations and Limitations ................................................................................. 13  
5. Data ............................................................................................................................... 14  
   5.1 Data and Descriptive Statistics ................................................................................. 14  
   5.2 ARCH-Effects ......................................................................................................... 16  
6. Empirical Results ......................................................................................................... 17  
   6.1 Model Evaluation .................................................................................................... 17  
   6.2 Evaluation Discussion .............................................................................................. 20  
7. Conclusions ................................................................................................................... 21  
8. Suggestions for further research .................................................................................. 22  
9. References ..................................................................................................................... 23  
10. Appendix ...................................................................................................................... 25
1. Introduction

Managing risk is an essential task for everyone involved in financial markets. Taking risks is associated with potential benefits, but also with the possibility of losing capital when markets head in the wrong direction. Thus, participants in financial markets need to keep their exposure to risk at a manageable level to avoid substantial losses. Time and again, history has shown that this is a difficult task which requires serious attention. There are several examples where excessive risk exposure has led to disastrous consequences which ended careers as well as financial institutions\(^1\). Proper risk management therefore plays a vital role in keeping the risk/reward-balance and thereby to avoid catastrophic consequences.

Several measures have been proposed to help financial actors control their exposure to market risk. One of these measures is Value at Risk (VaR) which was introduced in the early 90s. Due to its simplicity and ease of implementation, it has been adopted by both regulators and practitioners. The measure can loosely be defined as: “the worst loss over a target horizon that will not be exceeded with a given level of confidence.” (Jorion, 2007). As an example, the 99% one day Value at Risk for a portfolio is defined as the value such that a loss greater than it only occurs in one out of 100 days.

The computation of the measure is flexible, and several approaches with varying complexity can be used. However, to accurately model the risk, several aspects of financial market behavior needs to be considered. For example, empirical research indicates that the volatility of financial time series tends to cluster and vary over time (Mandelbrot, 1963). This implies that risk is not constant over time, and therefore needs to be taken into consideration. Otherwise, the risk will be wrongly assessed.

In practice, time-varying risk is often modeled with volatility models which can capture these dynamics. A commonly used group of volatility models is ARCH models. These describe the evolvement of volatility over time. They have therefore become popular in risk management and provide a simple yet effective way to compute Value at Risk. This class of models started with the seminal work of Engle (1982) and Bollerslev (1986). Since then, a large number of ARCH models have been developed to account for different “stylized facts” observed in financial markets (Engle, 2004). While some of the models are highly complex and versatile, this does not necessarily make them better. In general, there is a tradeoff between complexity and usefulness. A complex model might take several aspects into account but fail to provide practically significant better predictions than a simple model. If this is the case, the simpler model should be preferred.

\(^1\) A further elaboration on these won’t be provided in this text. However, for the interested reader we refer to chapter 2 of Jorions (2007) Value at Risk.
As mentioned, ARCH models provide a simple yet effective way to compute Value at Risk. However, choosing a particular model is easier said than done. From a theoretical point of view, a model’s predictive ability is expected to increase the better it can replicate empirical dynamics. This motivates the use of models which takes important aspects of financial markets into account. This study therefore aims to evaluate three such ARCH models on their ability to estimate Value at Risk. The considered models are the standard GARCH model, the asymmetric EGARCH model and the generally specified APARCH model (Bollerslev, 1986; Ding, 1993; Nelson, 1991). Further, each of these models is evaluated under the assumption that the conditional error distribution is normal, t-distributed, skewed t-distributed and normal inverse Gaussian. In total, this yields a combination of 12 models. The overall purpose of this study can then loosely be condensed to the following question:

**Research question:** Which of the considered models can provide accurate Value at Risk estimates?

This paper is structured as follows: First, the theoretical framework is presented where Value at Risk and ARCH-class models are described. After the theoretical ground has been covered, some earlier research on these model’s ability to estimate Value at Risk is presented. In the method section which follows, the backtesting procedure is described together with some comments on the limitations of the study. Next, the data and its statistical properties are covered. Finally, the results are presented and discussed before a conclusion is provided. Non-key results are provided in the Appendix.
2. Theoretical Framework

2.1 Value at Risk

Managing risk is an essential aspect for everyone involved in financial markets. Maintaining an appropriate risk level can prevent excessive risk exposure and thereby help financial actors avoid substantial losses. One of the most popular measures for risk management is Value at Risk (VaR). It was primarily created to assess market risk but can also be applied to other types of risk as well (Jorion, 2007).

The 100(1-\(\alpha\)) % Value at Risk can be described as the maximal loss of a position over a given holding period with a probability (1-\(\alpha\)) (Tsay, 2010). Figure 1 illustrates this conceptually for an arbitrary distribution. The 95% VaR is chosen such that the left tail probability is 5%. Hence, with 95% probability, the loss won’t exceed the Value at Risk. Or equally, in one out of twenty days.

![Figure 1. 95% VaR for an arbitrary distribution of returns.](image)

The \(VaR_{1-\alpha}\) can be formally defined in probabilistic terms as

\[
\alpha = \Pr[r_t < VaR_{1-\alpha}],
\]

where \(\alpha\) is defined for \(0 < \alpha < 1\) and \(r_t\) denotes the return for a position over a fixed holding period. The probability that \(r_t\) is below \(VaR_{1-\alpha}\) is equal to \(\alpha\). Hence, the Value at Risk is just a quantile in the return distribution. Thereby, another more intuitive way to express the risk measure can be stated with the inverse cumulative return distribution (CDF) of a position

\[
VaR_{1-\alpha} = F^{-1}(\alpha).
\]

If the distribution of returns is known and continuous, the \(VaR_{1-\alpha}\) can simply be computed as the \(\alpha\)-quantile of the CDF. However, in practice the distributional properties of returns are unknown and likely to vary over time. Therefore, the distributional properties need to be estimated so that the VaR can be computed. Predicting this measure ultimately boils down to forecasting the distributional properties of returns.
2.2 Autoregressive Conditional Heteroscedasticity Models

It is generally acknowledged that financial time series does not exhibit constant volatility over time. Instead, periods of large changes in asset prices tend to be followed by large changes and periods of small changes tend to be followed by small changes as noted by Mandelbrot (1963). This is known as volatility clustering. From a risk perspective, this would imply that risk changes over time. It would therefore be inappropriate to assume that the volatility of equity-returns is constant since the risk will be underestimated in periods of high volatility and overestimated in periods of low volatility. To accurately assess the risk, volatility clustering needs to be considered. This motivates the use of volatility models which can capture this dependence. In this study, three models which belong to the ARCH-class are considered. In this section, these models are presented.

2.2.1 The ARCH/GARCH Model

ARCH models have gained much popularity over the last couple of decades. They have been found to describe market volatility quite well and the research on these is extensive (Teräsvirta, 2009). This class of models started with the work of Robert Engle (1982). He suggested that the variance of the error term could be modeled as a function of lagged squared error terms in what Engle called the ARCH model. To describe this model, one can first consider a time series process of returns written as

\[ r_t = E(r_t | F_{t-1}) + \varepsilon_t \]

and

\[ \varepsilon_t = z_t \sigma_t. \]

(3)

(4)

The deterministic part \( E(r_t | F_{t-1}) \) is the expected value given the information set \( F_{t-1} \). It can be described by a time series process such as ARFIMA model. However, for daily data on stocks, it might be plausible to assume that the expected value is zero\(^2\). This would imply that returns are random and thus equivalent to the error term \( \varepsilon_t \). Under the ARCH-framework, the error term is further written as \( z_t \sigma_t \) where \( \sigma_t \) is the conditional standard deviation and \( z_t \) is an i.i.d. process with mean zero and unit variance which follows a specified distribution (This is treated further in section 2.3).

In the ARCH(\( q \)) model, Engle proposed that the conditional variance of the error term could be modeled as a linear function of lagged squared residuals as

\[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2. \]

(5)

\(^2\) This will be the working assumption throughout the entire study. The exclusion of a mean process will be further motivated in section 3.
The model can capture time-varying volatility but has some shortcomings. To avoid a negative conditional variance, some restrictions of the coefficients are required. For higher order ARCH models these conditions might be difficult to control for. This is problematic since a higher order structure often is required to describe the underlying process. To avoid these problems, Bollerslev (1986) introduced the Generalized ARCH(GARCH) model where the conditional variance is a linear function of its own lags and squared lagged residuals. The general GARCH(p,q) model is defined as

\[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \]  

where the following restrictions are sufficient but not necessary to ensure non-negative conditional variance: \( \omega > 0, \alpha_i \geq 0 \) for \( i = 1,2,\ldots,q \) and \( \beta_j \geq 0 \) for \( j = 1,2,\ldots,p \). For the model, \( \alpha_i \) measures the responsiveness to lagged squared residuals while \( \beta_j \) measures the impact of lagged conditional variance. If the condition \( \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \) is satisfied the GARCH(p,q)-process will be weakly stationary and have the unconditional variance \( \frac{\omega}{1-(\sum_{i=1}^{q} \alpha_i+\sum_{j=1}^{p} \beta_j)} \). Several methods have been suggested to estimate ARCH-class models. Since these models include non-linear relationships and distributional assumptions, the procedure is relatively complex and usually involves computational methods. Commonly Maximum Likelihood Estimation or QMLE is used (Xekalaki and Degiannakis, 2010).

Despite its popularity, the GARCH model has some shortcomings. There is mainly one “drawback” that have been highlighted. The model assumes that volatility responds symmetrically to positive and negative shocks. However, as Fisher Black (1976) noted in his seminal work, a negative stock-return tends to increase the volatility more than a positive return of the same magnitude. This is commonly referred to as asymmetric volatility or the leverage effect. Since the GARCH model is unable to respond asymmetrically to shocks, it might ignore a critical aspect of the underlying process.

2.2.2 The Exponential GARCH model

Several extensions of the Bollerslev’s GARCH model have been suggested to account for the leverage effect. One of the most widely used extensions is the Exponential GARCH model (EGARCH) proposed by Nelson (1991). It includes a parameter which can capture sign-effects of previous shocks. The EGARCH(p,q) models is defined as

\[ \ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} (\alpha_i z_{t-i} + \gamma_i (|z_{t-i}| - E|z_{t-i}|)) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2), \]  

(7)
where $z_{t-i} = \frac{\xi_{t-i}}{\sigma_{t-i}}$ the coefficient $\alpha_i$ captures the sign effect and $\gamma_i$ the magnitude effect. Another advantage of the model is that it models the log variance instead of the level. Consequently, it does not require parameter constraints to ensure positive conditional heteroscedasticity since the transform to level guarantees it will be non-negative. The assumption of conditional distribution affects the variance directly in the model since $E[z_{t-i}]$ is included in the variance equation. For a specified distribution this term will simplify to a constant given as

$$E[z] = \int_{-\infty}^{\infty} |z| f(z) \, dz.$$  \hspace{1cm} (8)

### 2.2.3 The Asymmetric Power ARCH model

The GARCH model of Bollerslev specifies the conditional variance as a function of lagged squared residuals. For financial time series, squared returns often show strong serial dependence. However, the absolute value of returns generally exhibits even stronger serial dependence. This is called the Taylor-effect (Taylor, 1986). A model which incorporate this “stylized fact” is the asymmetric power ARCH model (APARCH) of Ding (1993). Instead of the squared residual, the model uses a power transformation of the absolute value. Like the EGARCH model it also includes a parameter to allow for leverage effects. The general APARCH(p,q) model is defined as

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i (|\xi_{t-i}| - \gamma_i \xi_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta,$$  \hspace{1cm} (9)

where the following restrictions are sufficient but not necessary to ensure non-negative conditional variance: $\omega > 0, \delta \geq 0, -1 < \gamma_i < 1, \alpha_i \geq 0$ for $i = 1,2,..q$ and $\beta_i \geq 0$ for $j = 1,2,..p$. As can be seen in the model, the coefficient $\gamma_i$ captures the sign effect while $\alpha_i$ captures the magnitude effect. The coefficient $\delta$ is the size of the power transform.

Due to its specification, the APARCH model can be considered a general model and provides a flexible way to model the behavior of volatility. It “contains” 7 special cases where it simplifies to other well-known ARCH models for fixed values of parameters. For example, $\delta = 2$ and $\gamma_i = 0$ the model simplifies to the GARCH model. Some other well-known special cases are the ARCH, the TGARCH, the GJR-GARCH & the NARCH model (Ding, 1993).
2.3 Conditional Error Distributions

The distributional assumption of $z_t$ is an important part of the ARCH-model specification. This innovation is commonly assumed to be normally distributed. However, empirical research often suggests that the conditional distribution of returns is unlikely to be normal. It might instead be more reasonable to assume that the conditional distribution exhibits fat tails and possibly skewness (Xekalaki and Degiannakis, 2010). In the context of Value at Risk estimation, this assumption plays a significant role. Since VaR essentially boils down to forecasting quantiles, it is crucial that the specified distribution accurately can describe the tail behavior of the empirical distribution. Otherwise, the risk will be wrongly assessed. In this study, four distributions are considered for $z_t$. These are presented in the section below, and their relation to the distributional properties of returns is briefly explained.

2.3.1 The Normal Distribution

The normal distribution is widely used in statistics. The assumption of normality often brings several advantages. In general, it makes the mathematics simpler and it becomes relatively easy to work with probabilities and quantiles. For a standardized normally distributed variable $z_t$ the probability density function is defined as

$$f_{\text{NORM}}(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}}, \quad z_t \in \mathbb{R}. \quad (10)$$

Initially, Engle (1982) assumed that the conditional error distribution of the ARCH model is normal. Research has shown that this assumption works relatively well but has some inadequacies. Under this assumption, standardized returns are often found to have excess kurtosis and sometimes skewness. This has led to the suggestion of several other conditional error distributions which takes these aspects into account. Often these are found to fit empirical data better (Lindström et al., 2015; Xekalaki and Degiannakis, 2010).
2.3.2 The Student´s t-Distribution

In his article in 1987, Bollerslev pointed out that the models which assume a normal conditional error distribution do not seem to fully capture the leptokurtosis found in unconditional returns. Instead he argued that t-distribution should be used. The t-distribution has a shape parameter that decides the kurtosis. It can therefore account for leptokurtosis in the conditional distribution (Bollerslev, 1987). The probability density function for a standardized t-distributed variable \( z_t \) is defined as

\[
f_{STD}(z_t|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}} \left(1 + \frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad z_t \in R,
\]

where \( \Gamma(v) \) is the gamma function which is defined as: \( \Gamma(v) = \int_0^\infty x^{v-1} e^{-x}dx \), \( v \) is the degrees of freedom where the density is defined for \( 2 < \nu \leq \infty \). The degrees of freedom decide the shape of the distribution. For small value of \( \nu \), the distribution will be leptokurtic. When \( \nu \to \infty \) the t-distribution converges to the normal distribution.

2.3.3 The Skew t-Distribution

While the standard t-distribution can be used to introduce leptokurtosis in the conditional error distribution, it has a potential shortcoming. It cannot account for skewness. Several skewed extensions of the t-distribution have been proposed. This study adopts a skewed version of the t-distribution using the method of Fernandez and Steel (1998). This distribution can account for skewness as well as thick tails. The probability density function for a variable \( z_t \) which follows the standardized skew t-distribution is defined as

\[
f_{SSST}(z_t|\nu, \xi, \vartheta, \bar{\omega}) = \begin{cases} \frac{2}{\xi + \frac{1}{\xi}} \vartheta f_{STD}[\xi(\vartheta z_t + \bar{\omega})|\nu] & \text{if } z_t < -\bar{\omega}, \\
\frac{2}{\xi + \frac{1}{\xi}} \vartheta f_{STD}[\xi(\vartheta z_t + \bar{\omega})|\nu] & \text{if } z_t \geq -\bar{\omega}, 
\end{cases}
\]

where \( z_t \in R, f_{STD} \) is the density function of the standardized t-distribution, \( \nu \) is the degrees of freedom and \( \xi \) is the skewness parameter. The parameters \( \vartheta \) and \( \bar{\omega} \) are defined as

\[
\bar{\omega} = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{(\nu-2)}}{\sqrt{\pi \Gamma\left(\frac{\nu}{2}\right)}} \left(\frac{\xi - 1}{\xi}\right) \quad \text{and} \quad \vartheta^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - \bar{\omega}^2.
\]

The density is defined for \( 0 < \xi \leq \infty \). For a value of \( \xi = 1 \) the density is symmetric, and the distribution in (12) simplifies to a standardized t-distributed variable.
2.3.4 The Normal Inverse Gaussian Distribution

Another distribution which can account for both skewness and leptokurtosis is the normal inverse Gaussian (NIG) distribution (Barndorff-Nielsen, 1997). It is relatively new but has become popular in finance and econometrics (Lindström et al., 2015). It is commonly used for volatility modeling. The probability density function for a normal inverse Gaussian distributed variable $z_t$ is defined as

$$f_{NIG}(z_t | \alpha, \beta, \mu, \delta) = a(\alpha, \beta, \mu, \delta) q\left(\frac{z_t - \mu}{\delta}\right)^{-1} K_1(\delta a q\left(\frac{z_t - \mu}{\delta}\right)) \exp(\beta z_t), \quad z_t \in \mathbb{R},$$

and

$$a(\alpha, \beta, \mu, \delta) = \pi^{-1} \alpha \exp(\delta \sqrt{(\alpha^2 - \beta^2) - \beta \mu}). \quad q(x) = \sqrt{1 + x^2},$$

where $K_1$ is the modified Bessel function of third order and index 1. $\alpha$ and $\beta$ can be interpreted as shape parameters where $\alpha$ decides tail heaviness and $\beta$ decides asymmetry in the distribution. $\delta$ decides and scale and $\mu$ the location. These parameters are expected to satisfy $0 \leq |\beta| \leq \alpha$ and $0 < \delta$ for the density to be well defined.

For ARCH-modelling, the distribution is commonly reparametrized so that it can be used more conveniently. One way to do so is to let $\tilde{\alpha} = \delta \alpha$ and $\tilde{\beta} = \delta \beta$. The density function can then be expressed as

$$f_{NIG}(z_t | \tilde{\alpha}, \tilde{\beta}, \mu, \delta) = \frac{\tilde{\alpha}}{\pi \delta} \exp\left(\tilde{\alpha}^2 - \tilde{\beta}^2\right) + \frac{\tilde{\beta}(z_t - \mu)}{\delta} q\left(\frac{z_t - \mu}{\delta}\right)^{-1} K_1(\tilde{\alpha} q\left(\frac{z_t - \mu}{\delta}\right)).$$

Advantageously, under this form the distribution parameters $\tilde{\alpha}$ and $\tilde{\beta}$ remains invariant to location scale changes i.e. NIG($\tilde{\alpha}, \tilde{\beta}, \mu, \delta$) is a location-scale family (Barndorff-Nielsen, 1997; Jensen and Lunde, 2001). However, this is only one parametrization and several versions can be used in practice. See Eberlein and Prauss (2000) for further information on some of these and their scale properties.
3. Earlier Research

The research on ARCH models and VaR is vast. In this section the reader will be provided with a short but informative background on the empirical research on the topic. Generally, studies try to identify the best risk model. The specification of the model boils down to the conditional mean process and the error term-process. While the conditional mean occasionally is modeled as an ARMA-process, it is commonly found to lack practical significance. For example, Angelidis (2004) and Jánský (2011) compare ARCH-type models specified with and without conditional mean for VaR-estimation. They both argue that the inclusion of a mean model does not improve predictive ability and only adds complexity to the models.

For the error term process, both the specification of ARCH model and conditional distribution is of importance. The optimal choice of ARCH model is widely debated and has generated a lot of discussions. Hansen and Lunde (2005) evaluates a broad set of models using daily returns. They find that models which include a leverage term are superior for volatility forecasting of stock returns. Angelidis (2004) compare the use of GARCH, EGARCH, and TGARCH for VaR predictions using daily returns of equity indices. Their results show that no model is superior to another. Contradictory, Jánský (2011) find that GARCH models generally perform better than EGARCH and TGARCH models. This type of mixed results is relatively common. Generally, it appears as the “best” ARCH model seems to vary depending on choice of data, the period of examination and the method which is used.

The research regarding conditional error distributions tend to be somewhat more coinciding. Several studies find support for the idea of a fat-tailed distribution works better for VaR predictions. Both Angelidis (2004) and Jánský (2011) compares the conditional normality assumption with leptokurtic distributions and find that the latter generally works better, especially for Value at Risk with higher confidence-levels. Both studies find that the t-distribution performs best. Giot and Laurent (2003) evaluate the use of a skewed t-distribution. They find that it better fits the data than its symmetric counterpart. Likely because of its ability to account for skewness in the empirical distribution. The research on the topic generally suggest that the use of a fat-tailed distribution is suitable for Value at Risk estimation. Accounting for skewness might also be important.
4. Method

4.1 Backtesting Procedure

A risk model is only useful if it provides an accurate assessment of risk. To determine the accuracy of the models, each model is put through an evaluation to see if it is a “good” or a “bad” model. To be considered “good” it should correctly estimate the risk over time while adapting to changes in volatility. This section presents the backtesting procedure\(^3\) used to determine which models are good and which are not. First, the approach of generating VaR-estimates is described. Second, the evaluation procedure of these estimates is presented.

4.1.1 Forecast Procedure

Several approaches can be used to evaluate the goodness of fit of a model. The measure of fit should ultimately be chosen in consideration to the purpose of the model. Under realistic circumstances, a risk model is used to predict the risk in a future period. For this reason, the performance of the models is evaluated using out-of-sample fit in this study. This allows evaluation of the models under somewhat realistic circumstances. One could also include in-sample fit. However, this does not say how well the models fit data outside the estimation window\(^4\). For this reason, such measures are disregarded.

To evaluate the models, one day ahead out-of-sample forecasts of volatility are first computed. These volatility estimates are produced with a rolling estimation scheme with a moving estimation window. For each model, the 1000 most recent observations are first excluded from estimation. This is called the forecast window. If the total number of observations is T, the remaining T-1000 observations are called the first estimation window. These observations are used to estimate the first model and compute a forecast for the next period t+1. Since a moving window is used, the estimation window moves one step forward and include a new observation while the oldest observation in the sample is excluded. The model is re-estimated and a forecast for period t+2 is produced. This procedure is repeated until the forecast for period t+1000 has been produced which leads to a total of 1000 forecasts of volatility. These estimates are in their turn used to compute the 95% and 99% Value at Risk.

\(^3\) The R-package rugarch is used to conduct the backtesting procedure described in this section (Ghalanos, 2018).

\(^4\) As an example, Jánský (2011) compares models in-sample and out-of-sample fit. Results show that these two does not correspond. The models with highest AIC and LL does not produce the best VaR-estimates.
4.1.2 Unconditional Coverage

The evaluations of each model are based on two tests for its VaR-estimates. One for unconditional coverage and another for conditional coverage. These are based on the number of times the loss exceeds the predicted Value at Risk. This information can be summarized with an indicator function as

\[ I_t = \begin{cases} 1, & \text{if } r_t < VaR_{t|t-1}, \\ 0, & \text{Otherwise.} \end{cases} \]  

(17)

When the function takes on value 1 it is denoted as a “hit”. The sum of hits over \( T \) time periods can be written as \( N = \sum_{t=1}^{T} I_t \). The hit ratio is further defined as \( \frac{N}{T} = HR \) which also is called the empirical size. For a perfectly calibrated model the expected empirical size is equivalent to the nominal size \( \alpha \). Otherwise the model will either overestimate or underestimate the risk. To statistically test if these two are equal, Kupiec (1995) developed the unconditional coverage test. For Kupiec’s test the hypotheses are formally defined as: \( H_0: HR = \alpha \) and \( H_1: HR \neq \alpha \). Under the null hypothesis, the number of hits is binomially distributed as \( N \sim Bin(T, \alpha = p) \). Consequently, the test statistic can be written as

\[ LR_{UC} = 2 \ln \left( \left( 1 - \frac{N}{T} \right)^{T-N} \left( \frac{N}{T} \right)^N \right) - 2 \ln \left( 1 - \alpha \right)^{T-N} \alpha^N. \]  

(18)

The test statistic is asymptotically \( \chi^2 \)-distributed with one degree of freedom under the null hypothesis. We will therefore reject the null with 95% confidence level if \( LR_{uc} > 3.841 \). If the empirical size is either too large (risk is underestimated) or to small (risk is overestimated) the test will reject the null hypothesis. This implies that the model used to generate the VaR-estimates does an insufficient job and therefore can be classified as a “bad” model. As Kupiec (1995) put it, the statistical power of the test is generally poor. It is therefore important to have a large number of observations to reach an adequate statistical power.

4.1.3 Independence and Conditional Coverage

Kupiec’s test allows one to evaluate whether the observed number of hits is significantly different from the expected hit count. A shortcoming of the test is that it does not take independence of hits into account. It is possible that a model generates the expected number of hits, but these are not independent across time. This can happen if a model adjusts slowly to level shifts in volatility. To test for independence of hits across time, Christoffersen (1998) proposed an independence test with the null hypothesis \( H_0: \pi = \pi_0 = \pi_1 \) and \( H_1: \pi_0 \neq \pi_1 \), where \( \pi_i \) is the probability of a hit conditional on state \( i=I_{t-1} \). To test this property, Christoffersen suggested the following statistic

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

(19)
where $T_{ij}$ is defined as the number of periods with state $j = I_t$ conditional on state $i = I_{t-1}$ in the previous period. The test statistic is asymptotically $\chi^2$-distributed with one degree of freedom under the null hypothesis of independence.

To test for the joint assumption of the right empirical size and serial independence, Christoffersen (1998) proposed the conditional coverage test. It will reject a model asymptotically if it either produces the wrong empirical size, clustered exceedances or both. The test statistic is constructed as

$$LR_{cc} = LR_{ind} + LR_{uc}. \quad (20)$$

The $LR_{cc}$-statistic is $\chi^2$-distributed with two degrees of freedom under the null hypothesis. We will therefore reject the null with 95% confidence level if $LR_{cc} > 5.991$. Similar to Kupiec’s test this test needs to have a large number of observations to ensure sufficient statistical power.

### 4.2 Delimitations and Limitations

In order to deal with time constraints, some delimitations have been made. First, it is possible that higher order volatility models might describe the volatility better. However, models of order $p=q=1$ are often found to be sufficient and does not necessarily perform worse than those of higher order (Hansen and Lunde, 2005). Therefore, higher order models are disregarded. Another critical delimitation regards the forecast horizon. Only one day ahead forecasts are used in this study. Consequently, the results cannot appropriately be generalized to longer time horizons. It is possible that some models might better capture the memory property of volatility better than others. This might be the case even if they don’t perform as well on a daily forecast horizon. Another thing that deserves to be pointed out is that no attempt to rank the models is made. The goal of this study is only to assess which models accurately can estimate VaR for the data and which cannot.
5. Data

In this study, the considered models are evaluated on three stock indices. The data is collected from Yahoo Finance and consists of closing prices of daily frequency from January 1, 2009 to December 31, 2017. Each index series has been transformed to logarithmic returns\(^5\). In this section, the data is presented, and the statistical properties of the return series are examined.

5.1 Data and Descriptive Statistics

The first index is the Swedish OMXS30. It consists of the 30 most traded stocks on the Stockholm Stock Exchange and is a market-weighted price index. The returns of the index are displayed in Figure 2.

![Figure 2. Returns for OMXS30 from January 2009 to December 2017.](image)

The second index considered is the German DAX-index. It consists of the 30 largest companies listed on the Frankfurt Stock Exchange in terms of market capitalization and book volume. The index is market weighted. The returns of the index are displayed in Figure 3.

![Figure 3. Returns for DAX from January 2009 to December 2017.](image)

The third index considered is the Financial Times Stock Exchange 100 Index (FTSE100). It consists of the 100 largest companies listed on the London Stock Exchange in terms of market capitalization. The index is calculated using market weightings. The returns of the index are displayed in Figure 4.

![Figure 4. Returns for FTSE100 from January 2009 to December 2017.](image)

\(^5\) The logarithmic returns are computed as \(r_t = 100 \times [\ln(p_t) - \ln(p_{t-1})]\). These approximately corresponds to percentage changes in the price. (Tsay, 2010)
From Figure 2-4, it can be observed that large changes in the returns appear to cluster. The return series also exhibits mean-reverting behavior. Table 1 which displays the descriptive statistics for the unconditional returns, shows that the mean value of each index is very close to 0. The returns also tend to be slightly negatively skewed and leptokurtic. The Jarque-Bera test strongly rejects the null hypothesis of normality for each series. However, it is important to point out that this does not necessarily imply that the conditional distribution of returns is not normal. Both volatility clustering and asymmetric volatility can cause excess kurtosis and skewness respectively even if the conditional distribution of returns is normal (Engle, 2004).

**Table 1. Descriptive statistics for the returns-series.**

<table>
<thead>
<tr>
<th>Index</th>
<th>Obs.</th>
<th>Mean</th>
<th>St.Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB-Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMXS30</td>
<td>2262</td>
<td>0.036</td>
<td>1.274</td>
<td>-0.256</td>
<td>6.390</td>
<td>1111.30 (p&lt;0.01)</td>
</tr>
<tr>
<td>DAX</td>
<td>2286</td>
<td>0.042</td>
<td>1.313</td>
<td>-0.247</td>
<td>5.489</td>
<td>615.52 (p&lt;0.01)</td>
</tr>
<tr>
<td>FTSE100</td>
<td>2272</td>
<td>0.023</td>
<td>1.033</td>
<td>-0.176</td>
<td>5.701</td>
<td>704.69 (p&lt;0.01)</td>
</tr>
</tbody>
</table>

Note: P-values for Jarque-Bera statistics are displayed within the parentheses.

The ACF-plots for the returns of each index is displayed in Figure 5 (*PACFs can be found in Appendix Figure 8*). The blue dashed lines show a 95% confidence band for what would be expected if returns are serially uncorrelated. It appears to be some statistically significant autocorrelations in the OMXS30 returns. However, they are noticeably small and unlikely to be of practical significance. The corresponding plot for the FTSE100 show an even weaker correlation pattern and the plot for DAX show an absence of statistically significant autocorrelations.

**Figure 5. Autocorrelation function(ACF) for return series of the three indices.**
5.2 ARCH-Effects

As mentioned earlier, equity returns tend to be serially uncorrelated or at most show some minor serial correlation. However, these often display higher order dependence such as volatility clustering. A simple way to determine if this type of dependence is present in a series is to examine the squared and absolute returns. ACF-plots for both these transformations is shown in Figure 6 and 7. *(PACFs can be found in the Appendix Figure 10).*

![Figure 6](image)

**Figure 6.** Autocorrelation function(ACF) for squared returns for the three indices.

The figures indicate that there is strong dependence in the quadratic and absolute value of the returns since the spikes greatly exceed the 95% confidence band. Similarly, Ljung-Box Q test and the ARCH-LM test verifies this finding *(see Appendix Table 5).* If one looks closely at Figure 6 and 7 it can be observed that the autocorrelations is somewhat higher for absolute returns than for squared returns. As mentioned in section 2.2.3 this is known as the Taylor-effect.

![Figure 7](image)

**Figure 7.** Autocorrelation function(ACF) for the absolute value of returns for the three indices.

To summarize, the data clearly corresponds with several stylized facts. The unconditional return distributions are leptokurtic and negatively skewed. The data also show strong signs of ARCH-effects and the Taylor-effect is clearly distinguishable from the autocorrelations from each index. Given the broad selection of models and conditional distributions, it is expected that at least some of these should be able to capture the underlying process well.
6. Empirical Results

6.1 Model Evaluation

In this section, the results from the backtesting procedure are presented. The results are given for each index separately. For each of these, the models are presented together with the results from the tests explained in section 4.1. A general discussion of the results is provided in section 6.2.

6.1.1 OMXS30

The results for the OMXS30 index returns are displayed in Table 2. As can be seen, the majority of the models do not reject the null hypothesis of correct coverage for each VaR-level. The only models that fail the tests are the EGARCH-SSTD and EGARCH-NIG models which are rejected for the 95% as well as the 99% VaR level. In both these cases, they overestimate the risk.

Table 2. Results from the backtesting procedure with OMXS30.

<table>
<thead>
<tr>
<th>Model</th>
<th>Hit ratio</th>
<th>L_{UC}</th>
<th>L_{CC}</th>
<th>Hit ratio</th>
<th>L_{UC}</th>
<th>L_{CC}</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH-N</td>
<td>4.9%</td>
<td>0.021</td>
<td>0.174</td>
<td>1.3%</td>
<td>0.831</td>
<td>1.173</td>
</tr>
<tr>
<td></td>
<td>(0.884)</td>
<td>(0.917)</td>
<td></td>
<td>(0.362)</td>
<td>(0.556)</td>
<td></td>
</tr>
<tr>
<td>GARCH-STD</td>
<td>5.1%</td>
<td>0.021</td>
<td>0.085</td>
<td>0.7%</td>
<td>1.016</td>
<td>1.114</td>
</tr>
<tr>
<td></td>
<td>(0.885)</td>
<td>(0.958)</td>
<td></td>
<td>(0.314)</td>
<td>(0.573)</td>
<td></td>
</tr>
<tr>
<td>GARCH-SSTD</td>
<td>4.6%</td>
<td>0.346</td>
<td>0.707</td>
<td>0.6%</td>
<td>1.886</td>
<td>1.959</td>
</tr>
<tr>
<td></td>
<td>(0.557)</td>
<td>(0.702)</td>
<td></td>
<td>(0.170)</td>
<td>(0.376)</td>
<td></td>
</tr>
<tr>
<td>GARCH-NIG</td>
<td>4.4%</td>
<td>0.788</td>
<td>1.342</td>
<td>0.6%</td>
<td>1.886</td>
<td>1.959</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.511)</td>
<td></td>
<td>(0.170)</td>
<td>(0.376)</td>
<td></td>
</tr>
<tr>
<td>EGARCH-N</td>
<td>4.1%</td>
<td>1.812</td>
<td>4.389</td>
<td>1.1%</td>
<td>0.998</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>(0.178)</td>
<td>(0.111)</td>
<td></td>
<td>(0.754)</td>
<td>(0.842)</td>
<td></td>
</tr>
<tr>
<td>EGARCH-STD</td>
<td>4.2%</td>
<td>1.421</td>
<td>3.753</td>
<td>0.9%</td>
<td>0.105</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>(0.233)</td>
<td>(0.153)</td>
<td></td>
<td>(0.746)</td>
<td>(0.875)</td>
<td></td>
</tr>
<tr>
<td>EGARCH-SSTD</td>
<td>3.6%</td>
<td>4.553</td>
<td>6.351</td>
<td>0.4%</td>
<td>4.706</td>
<td>4.738</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.042)</td>
<td></td>
<td>(0.03)</td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>EGARCH-NIG</td>
<td>3.5%</td>
<td>5.268</td>
<td>7.682</td>
<td>0.4%</td>
<td>4.706</td>
<td>4.738</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.026)</td>
<td></td>
<td>(0.03)</td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>APARCH-N</td>
<td>4.4%</td>
<td>0.788</td>
<td>2.672</td>
<td>1.2%</td>
<td>0.380</td>
<td>0.672</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.263)</td>
<td></td>
<td>(0.538)</td>
<td>(0.715)</td>
<td></td>
</tr>
<tr>
<td>APARCH-STD</td>
<td>4.7%</td>
<td>0.193</td>
<td>1.506</td>
<td>0.8%</td>
<td>0.434</td>
<td>0.563</td>
</tr>
<tr>
<td></td>
<td>(0.660)</td>
<td>(0.471)</td>
<td></td>
<td>(0.510)</td>
<td>(0.755)</td>
<td></td>
</tr>
<tr>
<td>APARCH-SSTD</td>
<td>3.9%</td>
<td>2.747</td>
<td>3.982</td>
<td>0.5%</td>
<td>3.094</td>
<td>3.144</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.137)</td>
<td></td>
<td>(0.079)</td>
<td>(0.208)</td>
<td></td>
</tr>
<tr>
<td>APARCH-NIG</td>
<td>3.9%</td>
<td>2.747</td>
<td>3.982</td>
<td>0.5%</td>
<td>3.094</td>
<td>3.144</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.137)</td>
<td></td>
<td>(0.079)</td>
<td>(0.208)</td>
<td></td>
</tr>
</tbody>
</table>

Note: For the tests, UC denotes unconditional coverage and CC conditional coverage. For each test, the test-statistic is displayed as well as the corresponding p-value within parenthesis. Tests which led to a significant result on $\alpha = 5\%$ have been highlighted.
6.1.2 DAX

The results for the DAX index returns are displayed in Table 3. As can be seen in the table, the models perform well, and the vast majority pass the tests. Only the GARCH-N model gets rejected for the 99% VaR. In this case, the model underestimates the risk.

**Table 3. Results from the backtesting procedure with DAX.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Hit ratio</th>
<th>LR\textsubscript{UC}</th>
<th>LR\textsubscript{CC}</th>
<th>Hit ratio</th>
<th>LR\textsubscript{UC}</th>
<th>LR\textsubscript{CC}</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH-N</td>
<td>5.0%</td>
<td>0</td>
<td>4.204</td>
<td>1.7%</td>
<td>4.091</td>
<td>5.304</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(0.122)</td>
<td>(0.074)</td>
<td></td>
<td>(0.043)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>GARCH-STD</td>
<td>5.5%</td>
<td>0.510</td>
<td>5.110</td>
<td>0.9%</td>
<td>1.050</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td>(0.475)</td>
<td>(0.078)</td>
<td>(0.746)</td>
<td></td>
<td>(0.883)</td>
<td>(0.0883)</td>
</tr>
<tr>
<td>GARCH-SSTD</td>
<td>4.9%</td>
<td>0.021</td>
<td>2.579</td>
<td>0.6%</td>
<td>1.886</td>
<td>1.947</td>
</tr>
<tr>
<td></td>
<td>(0.884)</td>
<td>(0.275)</td>
<td>(0.170)</td>
<td></td>
<td>(0.378)</td>
<td>(0.378)</td>
</tr>
<tr>
<td>GARCH-NIG</td>
<td>4.5%</td>
<td>0.544</td>
<td>4.210</td>
<td>0.5%</td>
<td>3.094</td>
<td>3.134</td>
</tr>
<tr>
<td></td>
<td>(0.461)</td>
<td>(0.122)</td>
<td>(0.079)</td>
<td></td>
<td>(0.209)</td>
<td>(0.209)</td>
</tr>
<tr>
<td>EGARCH-N</td>
<td>4.3%</td>
<td>1.081</td>
<td>1.807</td>
<td>1.2%</td>
<td>0.380</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>(0.299)</td>
<td>(0.405)</td>
<td>(0.538)</td>
<td></td>
<td>(0.724)</td>
<td>(0.724)</td>
</tr>
<tr>
<td>EGARCH-STD</td>
<td>4.4%</td>
<td>0.788</td>
<td>1.397</td>
<td>1.1%</td>
<td>0.098</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.497)</td>
<td>(0.754)</td>
<td></td>
<td>(0.852)</td>
<td>(0.852)</td>
</tr>
<tr>
<td>EGARCH-SSTD</td>
<td>3.9%</td>
<td>2.747</td>
<td>2.924</td>
<td>0.8%</td>
<td>0.434</td>
<td>0.547</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.232)</td>
<td>(0.510)</td>
<td></td>
<td>(0.761)</td>
<td>(0.761)</td>
</tr>
<tr>
<td>EGARCH-NIG</td>
<td>3.8%</td>
<td>3.294</td>
<td>3.534</td>
<td>0.8%</td>
<td>0.434</td>
<td>0.547</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.171)</td>
<td>(0.510)</td>
<td></td>
<td>(0.761)</td>
<td>(0.761)</td>
</tr>
<tr>
<td>APARCH-N</td>
<td>4.1%</td>
<td>1.812</td>
<td>2.811</td>
<td>1.2%</td>
<td>0.380</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>(0.178)</td>
<td>(0.245)</td>
<td>(0.538)</td>
<td></td>
<td>(0.724)</td>
<td>(0.724)</td>
</tr>
<tr>
<td>APARCH-STD</td>
<td>4.6%</td>
<td>0.346</td>
<td>0.751</td>
<td>0.8%</td>
<td>0.434</td>
<td>0.547</td>
</tr>
<tr>
<td></td>
<td>(0.557)</td>
<td>(0.687)</td>
<td>(0.510)</td>
<td></td>
<td>(0.761)</td>
<td>(0.761)</td>
</tr>
<tr>
<td>APARCH-SSTD</td>
<td>3.9%</td>
<td>2.747</td>
<td>4.068</td>
<td>0.7%</td>
<td>1.016</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.131)</td>
<td>(0.314)</td>
<td></td>
<td>(0.577)</td>
<td>(0.577)</td>
</tr>
<tr>
<td>APARCH-NIG</td>
<td>3.9%</td>
<td>2.747</td>
<td>4.068</td>
<td>0.7%</td>
<td>1.016</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.131)</td>
<td>(0.314)</td>
<td></td>
<td>(0.577)</td>
<td>(0.577)</td>
</tr>
</tbody>
</table>

Note: For the tests, UC denotes unconditional coverage and CC conditional coverage. For each test, the test-statistic is displayed as well as the corresponding p-value within parenthesis. Tests which led to a significant result on \( \alpha = 5\% \) have been highlighted.
6.1.3 FTSE100

The results for the FTSE100 index returns are displayed in Table 4. Similar to the previous results, the model generally performs well, and the majority avoids rejection. None of the models fail at the 95% level. However, GARCH-N, GARCH-STD, and EGARCH-N fail to estimate the 99% VaR accurately. In all these cases, the risk is underestimated.

Table 4. Results from the backtesting procedure with FTSE100.

<table>
<thead>
<tr>
<th>Model</th>
<th>95% VaR</th>
<th>99% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit ratio</td>
<td>LR_{UC}</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>4.9%</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.884)</td>
</tr>
<tr>
<td>GARCH-STD</td>
<td>5.1%</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.885)</td>
</tr>
<tr>
<td>GARCH-SSTD</td>
<td>4.4%</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.375)</td>
</tr>
<tr>
<td>GARCH-NIG</td>
<td>4.4%</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.375)</td>
</tr>
<tr>
<td>EGARCH-N</td>
<td>5.1%</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.855)</td>
</tr>
<tr>
<td>EGARCH-STD</td>
<td>5.3%</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.666)</td>
</tr>
<tr>
<td>EGARCH-SSTD</td>
<td>5.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>EGARCH-NIG</td>
<td>4.6%</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.557)</td>
</tr>
<tr>
<td>APARCH-N</td>
<td>5.2%</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.773)</td>
</tr>
<tr>
<td>APARCH-STD</td>
<td>5.5%</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.475)</td>
</tr>
<tr>
<td>APARCH-SSTD</td>
<td>4.8%</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.770)</td>
</tr>
<tr>
<td>APARCH-NIG</td>
<td>4.7%</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.660)</td>
</tr>
</tbody>
</table>

Note: For the tests, UC denotes unconditional coverage and CC conditional coverage. For each test, the test-statistic is displayed as well as the corresponding p-value within parenthesis. Tests which led to a significant result on α = 5% have been highlighted.
6.2 Evaluation Discussion

The results from Table 2-4 show that the models perform well. The majority passed the backtesting procedure for each index. However, some failed more than others, and some patterns were recognizable based on the rejections. Both the ARCH model specification and the distributional assumption seems to play an important role in a model’s ability to estimate VaR accurately. Of the three different ARCH specifications, the Asymmetric Power ARCH seems to be the most robust model. It was able to avoid getting rejected in a single backtest for both 95% and 99% VaR. For the other two specifications, this was not the case. The EGARCH model and GARCH model both were rejected in five respectively three instances. However, it is possible that these two models work differently well for specific conditional distributions. Given the results, it cannot be said that any one of these two is better to the other. Only, that these two provide accurate estimates in general for the considered data.

An expected result is that the assumption of conditional error distribution affects the conservativeness of the predictions. The models with a skewed t-distribution or NIG-distribution tend to produce a lower number of hits than the corresponding models with a normal or symmetric t-distribution. Based on the number of rejections, it appears as it is more difficult to estimate the 99% VaR than the 95% VaR correctly. In these cases, the normality assumption of the conditional distribution lead to the most rejections. As expected, the risk in these instances was significantly underestimated which likely is due to the distributions inability to account for thick tails. When the models with skewed t- and NIG-distributions were rejected, the risk was on the other hand significantly overestimated.

It might be the case that leptokurtosis is the most important factor to account for in the conditional distribution. Since introducing skewness in the distribution did not lead to fewer rejections, it is hard to say that this aspect brings further predictive power. However, it should be acknowledged that in general terms, all distributions seemed to work relatively well (apart from the normal distribution for the 99% level), and it is hard to argue for either superiority or inferiority of any one of these. Even though the t-distribution marginally lead to the fewest rejections, more data would have to be tested to draw strict conclusions.
7. Conclusions

The purpose of this study was to find out which of the selected models accurately could estimate Value at Risk for the three stock indices. Three different ARCH-models were considered together with four different conditional distributions leading to a total of 12 models. The results show that these models tend to be accurate. However, since the APARCH model avoided to get rejected in a single test, we argue that this model did the best job generally. When it comes to the conditional distribution, the results indicate that the normal distribution might be a bad choice due to inability to provide accurate 99% VaR predictions. Hence, a distribution which can account for fat tails is likely a better choice overall.
8. Suggestions for Further Research

Interestingly, the results of this study seem to indicate that the considered models perform well. It would be interesting to see if this is also the case for different time periods and assets, perhaps under more volatile periods. It could be meaningful to include a comparison of leptokurtic distributions with and without skewness under such periods as well. Finally, only one day ahead forecasts were considered in this study. Therefore, a possible extension could explore whether the results of this study can be generalized over longer forecast horizons and possibly see how the models compare with long-memory volatility models such as FIGARCH or FIEGARCH.
9. References


Jánský, I., Rippel, M., n.d. “Value at Risk forecasting with the ARMA-GARCH family of models in times of increased volatility 20.”


10. Appendix

A. Non-key Results

Figure 8. PACF for the returns of the indices.

Figure 9. Partial autocorrelation function (PACF) for absolute returns of the indices.

Figure 10. Partial autocorrelation function (PACF) for squared returns of the indices.
Table 5. Ljung-Box test and ARCH-LM test for ARCH-effects in returns.

<table>
<thead>
<tr>
<th>Test</th>
<th>OMXS30</th>
<th>DAX</th>
<th>FTSE100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ljung-Box test $r_t^2$ [20 lags]</td>
<td>1145.45</td>
<td>1193.80</td>
<td>1112.70</td>
</tr>
<tr>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td>Ljung-Box test $</td>
<td>r_t</td>
<td>$ [20 lags]</td>
<td>1796.00</td>
</tr>
<tr>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td>(p&lt;0.01)</td>
<td></td>
</tr>
</tbody>
</table>

Note: P-values are displayed within parentheses.

B. Parameter Estimates from the Rolling Window Estimation

In theory, there is no specific reason to assume that the underlying process we model is constant across time. On the contrary, it might be more reasonable to assume that the process is subject to structural changes over time due to the nature of financial markets. This is one of the reasons why a relatively small moving estimation-window is used in this study. More specifically, the models are re-estimated after every forecast-step. For each specific model, this has yielded a total of 1000 estimates of every parameter. These estimates provide information about the underlying process and whether it is constant over time. Based on these rolling coefficients, it does not appear as this necessarily always is the case.

Figure 11. GARCH-NIG coefficients with robust standard errors for OMXS30.

Figure 11 and 12 show the rolling coefficients for the GARCH-NIG model estimated with the OMXS30 and the DAX-index respectively. The first figure clearly shows that the estimates are not constant over the estimation period. In the second figure, the estimates look more stable. They tend to move somewhat but stays roughly around the same value.
This tendency is also found for the parameters which relates to the shape of the conditional distribution. The rolling coefficients for the GARCH-SSTD estimated on the FTSE-data are displayed in Figure 13. The skewness parameter seems to be relatively stable while the shape parameter (degrees of freedom) trends downward as the standard errors clearly decreases.

To summarize, it appears as some parameter estimates are stable over time, and others possibly not. This might imply that there are changes in the underlying process. Consequently, it is appropriate to use of a moving estimation window for the backtesting procedure. A recursive estimation window, or a single estimation procedure would not be able to capture time variation of the parameters and might lead to bad estimates.