Volatility Forecasting Performance: Evaluation of GARCH type volatility models on Nordic equity indices

Amadeus Wennström

Master of Science Thesis, Spring 2014

Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden
Supervisor: Associate Professor Jimmy Olsson (KTH)
Examiner: Professor Boualem Djehiche (KTH)

Submission Date: June 11, 2014

Abstract

This thesis examines the volatility forecasting performance of six commonly used forecasting models; the simple moving average, the exponentially weighted moving average, the ARCH model, the GARCH model, the EGARCH model and the GJR-GARCH model. The dataset used in this report are three different Nordic equity indices, OMXS30, OMXC20 and OMXH25. The objective of this paper is to compare the volatility models in terms of the in-sample and out-of-sample fit. The results were very mixed. In terms of the in-sample fit, the result was clear and unequivocally implied that assuming a heavier tailed error distribution than the normal distribution and modeling the conditional mean significantly improves the fit. Moreover a main conclusion is that yes, the more complex models do provide a better in-sample fit than the more parsimonious models. However in terms of the out-of-sample forecasting performance the result was inconclusive. There is not a single volatility model that is preferred based on all the loss functions. An important finding is however not only that the ranking differs when using different loss functions but how dramatically it can differ. This illuminates the importance of choosing an adequate loss function for the intended purpose of the forecast. Moreover it is not necessarily the model with the best in-sample fit that produces the best out-of-sample forecast. Since the out-of-sample forecast performance is so vital to the objective of the analysis one can question whether the in-sample fit should even be used at all to support the choice of a specific volatility model.

Keywords: Conditional Variance, ARCH, GARCH, EGARCH, GJR-GARCH, volatility forecasting, Parkinson's estimator

Acknowledgements

I would like to thank my thesis supervisor, Jimmy Olsson, Associate Professor at the Department of Mathematics at KTH Royal Institute of Technology for his encouragement throughout the intense work of finishing this paper.

Stockholm, June 2014
Amadeus Wennström

Contents

1 Introduction

The study of finance is to a large extent a study of volatility[1] and volatility really permeates finance. In the not too distant past, several theoretical models assumed constant volatility, see Merton (1969) and Black and Scholes (1973). Today it is a well-known fact that volatility of asset returns is time-varying and predictable, see Andersen and Bollerslev (1997). Volatility also has some commonly seen characteristics. Three important stylized facts about volatility is that volatility exhibits persistence, volatility is mean reverting and innovations have an asymmetric impact on volatility, see Engle and Patton (2001).

Measuring and forecasting volatility of asset returns is vital for risk management, asset allocation, and option pricing. Risk management is to a large extent about measuring potential future losses of a portfolio, and to be able to estimate the potential losses, the future volatility must be estimated. The same holds for pricing options; when determining the price of an option the future volatility of the underlying asset is a very important parameter. In asset allocation the future volatility of different asset classes is an input in quadratic investment and hedging principles. Due to the high demand for accurate volatility forecasts there has been an immense interest amongst both practitioners and researchers to model the time varying volatility.

For volatility forecasts there are two major sources, volatility models based on time series and the volatility implied from option prices. From a theoretical point of view the implied volatilities from options prices should contain all relevant, available information and thus be a good estimate of the future realized volatility. The evidence supporting that so is the case is however mixed. In option prices from which the volatility is implied there is, in general, a risk premium due to the fact that the volatility risk cannot be perfectly hedged, see Bollerslev and Zhou (2005). Moreover, one of the most quoted phenomenons illuminating the limitations of the classic Black-Scholes model from which the volatility is implied is the so-called *smile effect*. The smile effect is the effect when calculating the implied volatility for options with different strikes on the same underlying with the same time to maturity one does not necessarily get the same implied volatility. In general the implied volatility will be a u-shaped curve with a minimum implied volatility for at-the-money options. Thus, if the implied volatility would be used as a forecast of volatility, the same market is giving multiple forecasts for the future volatility of the same underlying during the same time period. Furthermore, implied volatilities are only available for specific time horizons for a very limited set of assets. Based on this the only objective source for volatility forecasts, available for all financial assets, are time series models which will be the object of study in this report.

There are two major types of conditional heteroscedastic models. In the first type of volatility models the evolution of the variance is determined using an exact function and in volatility models of the second type the evolution of the variance is governed by a stochastic equation. This report will focus on volatility models of the first type, in which the volatility forecasts are determined using an exact function. Due to the enormous interest in forecasting volatility there is today a large number of different models that try to mimic the evolution and

[1] In financial jargon volatility usually refers to the standard deviation which is the square root of the variance. In this paper the standard deviation σ and the variance σ^2 will interchangably be referred to as volatility to simplify the terminology. Since the two measures are linked by a simple monotonic transformation this should cause no conceptual confusion.

characteristics of financial asset volatility. The *Autoregressive Conditional Heteroscedasticity* (ARCH) model introduced by Engle (1982) was one of the first models that provided a way to model conditional heteroscedasticity in volatility. The model was simple and intuitive but required usually many parameters to describe adequately the volatility process. Bollerslev (1986) extended the ARCH model to the *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) which had the same key properties as the ARCH but required far less parameters to adequately model the volatility process. Both the ARCH and the GARCH model are able to model the persistence of volatility, the so-called volatility clustering but the models both assume that positive and negative shocks have the same impact on volatility. It is well known that for financial asset volatility the innovations have an asymmetric impact. To be able to model this behavior and overcome the weaknesses of the GARCH model Nelson (1991) proposed an extension to the GARCH model called the *Exponential GARCH* (EGARCH) which is able to allow for asymmetric effects of positive and negative asset returns. Another widely used extension of the GARCH model is the *GJR-GARCH* proposed by Glosten, Jagannathan and Runkle (1993).

Forecasting the future level of volatility is far from trivial and evaluating the forecasting performance presents even further challenge. Even if a model has been chosen and fitted to the data and the forecasts have been calculated, evaluating the performance of that forecast is troubling due to the latent nature of realized conditional volatility. A proxy for the realized volatility is therefore needed and moreover the choice of statistical measure is, as pointed out by Bollerslev, Engle and Nelson (1994), far from clear.

In this paper the focus will be restricted to examining the forecasting performance of six commonly used forecasting models; the simple moving average, the exponentially weighted moving average, the ARCH model, the GARCH model, the EGARCH model and the GJR-GARCH model. The dataset used in this report are three different Nordic equity indices; OMXS30, OMXC20 and OMXH25. OMX Stockholm 30 (OMXS30), OMX Copenhagen 20 (OMXC20) and OMX Helsinki 25 (OMXH25) are the leading stock exchange indices at their respective markets and consist of the 30, 20 and 25 most actively traded shares, respectively. The data was provided by Nasdaq OMX®.

The objective of this paper is to compare the volatility models in terms of the in-sample and out-of-sample fit. Three main themes are studied. First, the basic structure of the modeling framework will be investigated with respect to the error distribution and the conditional mean. The purpose is to gain insight in how the assumed error distribution and different models for the conditional mean impacts the in-sample and out-of-sample fit regardless of which specific conditional variance model is used. The second theme is to analyze whether more complex models, which are able to exhibit more of the stylized facts and characteristics of asset price volatility, provide a better in-sample fit and/or out-of-sample-fit than more parsimonious ones. The third and final theme is whether the model with the best in-sample fit also produces the best out-of-sample volatility forecast. The aim of the analysis is thus not evaluating whether the GARCH type volatility models do provide accurate forecasts but if the more complex models outperforms the more parsimonious ones. That GARCH type volatility models do provide strikingly accurate volatility forecasts was shown by Andersen and Bollerslev (1998) and is not the object of study in this paper.

This paper adds to the existing literature in primarily three important directions. First, the specific time period investigated is, until today, unique. Secondly, the sole focus on volatility modeling on Nordic equity indices provides an interesting insight to that specific market. Finally, some previous papers have tried to find the model with a superior predictive ability based on several different asset types such as foreign exchange, commodities and equities, see for example Hansen and Lunde (2001). By limiting the data used to be based on a single asset type, in this case equity indices, one increases the chance of distinguishing the superior model. This is due to the fact that there might be different models that are best at forecasting the volatility of the different asset types.

The results were very mixed. There is not a single volatility model that is preferred based on all the loss functions used in this paper. An important finding is however not only that the ranking differs when using different loss functions but how dramatically it can differ. This illuminates the importance of choosing an adequate loss function for the intended purpose of the forecast. Moreover it is not necessarily the model with the best in-sample fit that produces the best out-of-sample forecast. Since the out-of-sample forecast performance is so vital to the objective of the analysis one can question whether the in-sample fit should even be used at all to support the choice of a specific volatility model.

The remainder of this paper is organized as follows. In section 2 the methodology used and the data employed in this paper are described. Section 3 starts by presenting the key characteristics of volatility and its implications on modeling it. Section 3 continues by first presenting the basic structure of a volatility model and then the details of the specific models studied in this paper. Each volatility model is defined, their respective properties and weaknesses are discussed and an explicit expression of the forecasts of each model is presented. Section 4 explores the parameter calibration for the studied volatility models and presents ways to compare the in-sample model fit. Section 5 discusses the problems with evaluating the out-of sample forecasting performance of the different volatility models. It presents various models to use as a proxy for the latent daily realized volatility and also presents the loss functions used in the evaluation of the forecasting performance. In section 6 the empirical results are presented and analyzed. Section 7 summarizes the report and presents the main conclusions and findings from the empirical results. Section 8 concludes the report and suggests topics for further research.

2 Data and methodology

2.1 Methodology

The objective of this paper is to compare the six volatility models presented in the introduction in terms of their in-sample fit and their out-of-sample forecasting performance. The meaning of in-sample and out-of-sample will be explained later in this section. The forecasting ability of these models will be compared using three financial time series, more specifically three Nordic equity indices. The Nordic equity indices used are; OMXS30, OMXC20 and OMXH25. OMX Stockholm 30 (OMXS30), OMX Copenhagen 20 (OMXC20) and OMX Helsinki 25 (OMXH25) are the leading stock exchange indices at their respective market and consist of the 30, 20 and 25 most actively traded shares respectively. To only include such few of the most actively traded shares guarantees that all the underlying shares have excellent liquidity. The excellent liquidity reduces the market microstructure effects such as the bid-ask bounce. The data was provided by Nasdaq OMX®.

The reasoning for looking at the out-of-sample forecasting performance in addition to the in-sample fit comes from the objective of the analysis. In forecasting it is not necessarily the model that provides the best in-sample fit that produces the best out-of-sample volatility forecast, which is the main objective of the analysis; see Shamiri and Isa (2009). Hence it is common to use the out-of-sample forecast to aid the selection of which model is best suited for the series under study; see Andersen and Bollerslev (1998), Hansen and Lunde (2001) and Brandt and Jones (2006). The so-called out-of-sample forecast refers to that the data used in the fitting of the model is different from the data used for evaluating the forecasts. Typically one divides the data into two subsets, one in which the model parameters are fitted (estimation subsample) and another subset used to evaluate the forecasting performance of the models (forecasting subsample). The basic structure is as follows: if the complete set consists of T number of data points, $p_1, p_2, \dots, p_T$, the data is divided into the subset $\{p_1, p_2, \dots, p_n\}$ and $\{p_{n+1}, p_{n+1}, \dots, p_T\}$ where n is the initial forecast origin. A reasonably choice is $n = \frac{2T}{3}$, see Tsay (2008). Let h denote the maximum forecast horizon of interest, that is, one is interested in the 1-step up until the h-step ahead forecasts. Then the out-of-sample forecasting evaluation process, using a so-called recursive scheme, works as follows:

1. Set $m = n$ to be the initial forecast origin. Then fit each of the models using the data $\{p_1, p_2, \dots, p_m\}$. Now calculate the 1-step to h-step ahead forecasts from the forecast origin m using the fitted models.

2. Compute the forecasts errors, for each of the 1- to h-step ahead forecast, for each model, as the difference between the forecasted volatility and the actual volatility.

$$e(i, m) = \sigma^2_{actual,1} - \sigma^2_{Forecaste,1}, i = 1, 2.., h$$

3. Advance the origin by 1, that is $m = m + 1$ and start over from step 1. Repeat this process until the forecast origin m is equal to the last data point T.

Once all the forecasts and respective forecasts errors for each of the models have been computed the only thing left is to use a loss function to evaluate the 1-step to h-step ahead forecasts for each of the models.

The forecasting scheme described above where the estimation sample grows as the forecast origin is increased, has the advantage that at each step all forecasts is based on all past information. However, in this study due to the computational burden, each model will only be fitted once, a so called fixed scheme. Each model will be fitted to the data until the initial forecast origin from which the forecasts can be computed. To compute the forecasts from the next data point, the same estimates for the parameters are used but when computing the forecasts from that origin the observations until that day are available, i.e. at each step only the data is updated with new information. Furthermore the forecast horizon will be set to one, thus the forecasts computed at each data point will only be the 1-day ahead forecasts. The focus of this report is to compare the different volatility models not to compare different forecasting horizons and due to space limitations the focus is restricted to 1-day ahead forecasts. The forecast performance for different forecast horizons based on the same model exceeds this papers scope.

The rest of the paper will deal with the process described above in a chronological manner. In summary: each of the models, which are described further in section 3, is fitted to the data in the estimation subsample for each of the three datasets[2]. The fitting process for these models are described and discussed in section 4. Once the models have been fitted, the out-of-sample evaluation of the volatility forecasts is described and discussed in section 5. All of the empirical results, model fitting and out-of-sample forecast performance is then presented in section 6.

The complete set of data used in this study is the daily price data for the three equity indices from 2002-01-02 until 2014-04-15. The data is divided into a nine year in-sample period and a two year, approximately, out-of-sample period. The in-sample period is 2002-01-02 until 2010-12-30 and the out-of-sample period is 2011-01-03 until 2014-04-15. For the in-sample period the data are the daily closing prices for the three indices. In the out-of-sample period daily high and low prices will be used together with the daily closing prices when computing the proxy for the latent daily volatility. The reasoning for this will be fully explained in section 5.

When studying financial time series most researchers study the return time series rather than the raw price data. In 1997 Campbell, Lo and MacKinlay gave two main reasons for this. First, the return of an asset is a complete, scale free summary of that particular investment opportunity. Secondly, the return series are much easier to handle than the raw price series since it has more attractive statistical properties. There are several different definitions of returns. In this paper the returns of study will be the log returns. The variables of interest are daily log returns, r_t, defined by the interdaily difference of the natural logarithm of the daily asset prices, p_t. The daily returns are thus defined by

$$r_t = \log(p_t) - \log(p_{t-1}) = \log\left(\frac{p_t}{p_{t-1}}\right) \tag{1}$$

and will be the time series of study in this paper. The rest of this section presents the three datasets and provides some descriptive statistics.

[2] The fitting process only apply to the GARCH type volatility models in the study. The Simple Moving Average and the Exponentially Weighted Moving Average forecast methods are non-parametric and thus do not require any parameter fitting.

2.2 Descriptive statistics

2.2.1 OMXS30

For OMXS30 there were 3,089 daily data points during the entire period 2002-01-02 until 2014-04-15. The in-sample period from 2002-01-02 until 2010-12-30 consisted of 2,263 daily data points and the out-of-sample period from 2011-01-03 until 2014-04-15 consisted of 826 daily data points. In Figure 1 the daily closing price during the entire period is plotted. The in-sample period is plotted in blue and the horizontal red line indicates where the out-of-sample period starts which is then indicated by the black line.

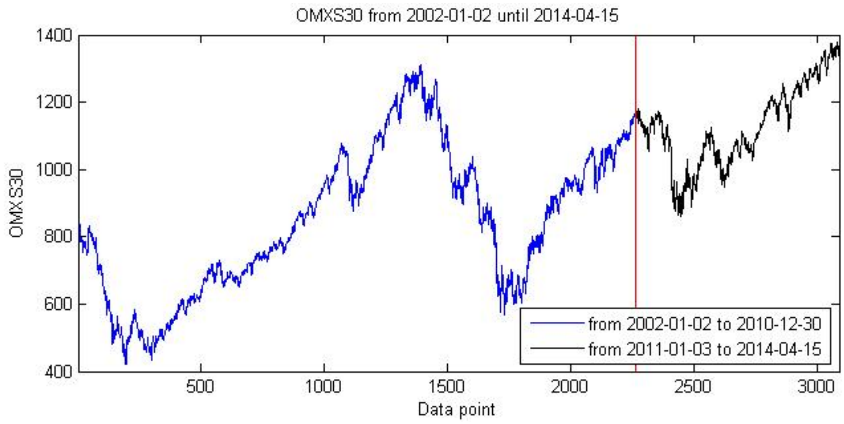

Figure 1: OMXS30 daily index level from 2002-01-12 until 2014-04-15. In total there are 3,089 observations. The blue line represents the index level during the in-sample period and the vertical red line indicates where the out-of-sample period starts which is then represented by the black line.

The main variable of study is as mentioned not the price process but the daily log return defined in equation (1). The left plot of Figure 2 shows the daily return for the in-sample period. The daily return series seems to be a stationary process with a mean close to zero but with volatility exhibiting relatively calm periods followed by more turbulent periods. This is one of the key characteristics mentioned in the introduction of asset return volatility and is referred to as volatility clustering. The right plot in Figure 2 shows the Sample Autocorrelation Function for the daily returns of lags 0 to 20. The Sample Autocorrelation function is a very useful tool and can be used for checking for serial correlation in the return data. Based on ocular inspection of the Sample Autocorrelation Function plot it is not completely clear whether the data is serially correlated or not, even though it has minor significant serial correlation at lag 3.

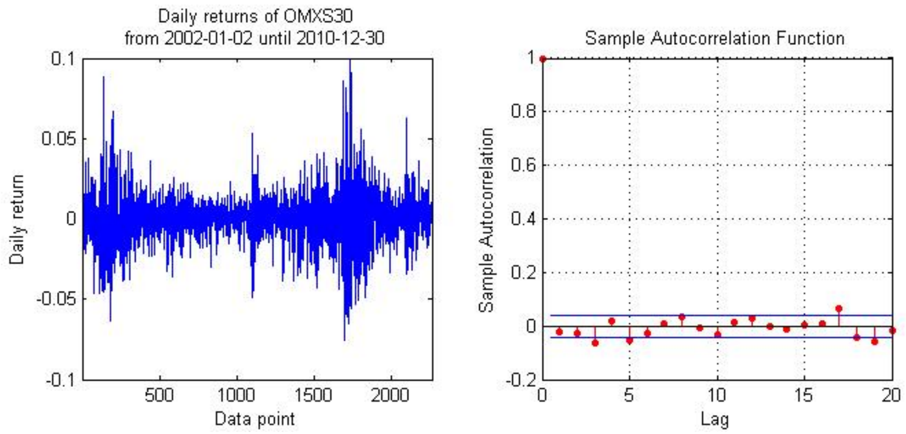

Figure 2: Left plot: Daily returns of OMXS30 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,262 observations. Right plot: Sample Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

However the Ljung-Box Q-test null hypothesis that all autocorrelations up to the tested lags are zero is rejected for lags 5, 10 and 15 at a 5% significance level, see appendix A for explanation of the Ljung-Box Q-test. This suggests that a conditional mean model is required for this return series. The focus of this paper is the conditional variance and not the conditional mean but for the conditional variance model to work properly the conditional mean needs to be modeled as well. In section 3, three different conditional mean models will be presented and in the empirical study each of the conditional variance models will be used together with each of the conditional mean models to be able to see the effect of the conditional mean model on the forecasting performance.

The next descriptive statistic is that of the squared returns. The left plot of Figure 3 shows the daily squared returns for the in-sample period where the volatility clustering is even more evident than in Figure 2. The right plot of Figure 3 shows the Sample Partial Autocorrelation Function which shows clearly significant autocorrelation. Engle's ARCH test rejects the null hypothesis, that there is no autocorrelation, for lags 6 and even 14 at a 5% significance level and thus confirms that the squared returns are serially correlated, see appendix A for description of Engle's ARCH test.

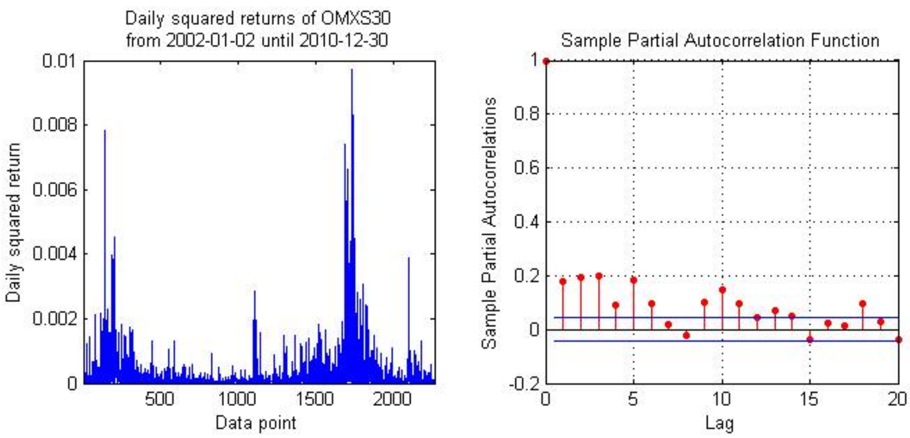

Figure 3: Left plot: Daily squared returns of OMXS30 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,262 observations. Right plot: Sample Partial Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

Next the empirical distribution of the return series is examined. In Figure 4 two q-q plots are presented. The left plot is a q-q plot of the empirical distribution of the daily returns (y-axis) against the best fitted normal distribution (x-axis). The plot to the right is a q-q plot of the empirical distribution (y-axis) against the best fitted t location-scale distribution (x-axis).

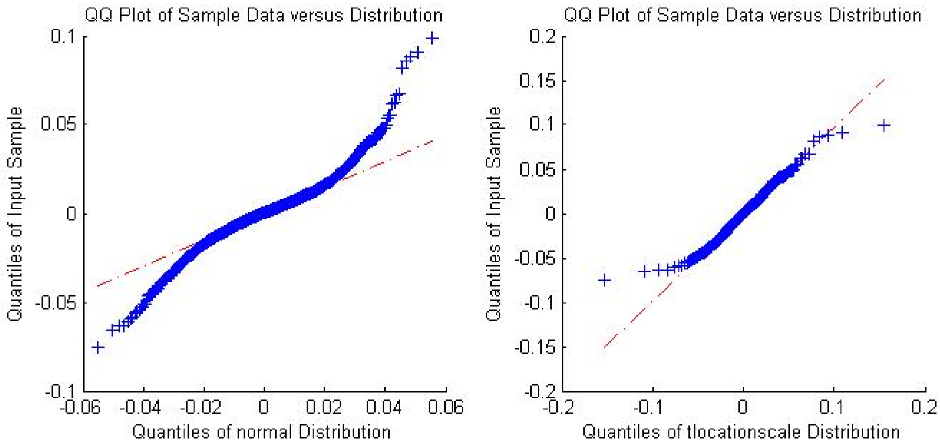

Figure 4: Left plot: q-q plot of the empirical distribution (y-axis) against best fitted normal distribution (x-axis). Right plot: q-q plot of the empirical distribution (y-axis) against best fitted t location-scale distribution (x-axis).

The left plot in Figure 4 shows clearly that even the best fitted normal distribution does not provide a particularly good fit as a reference distribution. The empirical distribution of the daily returns exhibits significantly heavier tails than the reference distribution which implies that another choice of parametric family should be considered. From the right plot of Figure 4 it is evident that the t location-scale distribution[3] is a much better fit and actually shows the opposite behavior that the empirical distribution has slightly lighter tails than the reference distribution. Though in the case of the right plot there are only a very limited number of points that are off the linear red line. In Table 1 some summary statistics of the return series is presented.

Table 1: Summary statistics for OMXS30 daily returns in the in-sample period.

	Sample size	**Mean**	**Variance**	**Skewness**	**Excess kurtosis**	**Jarque-Bera**
OMXS30	2,262	0.00014502	0.00024859	0.203321332	6.982098267	1[4]

In Table 1 the sample size, unconditional mean, unconditional variance, skewness, excess kurtosis and result of Jargue-Bera test is presented. The mean and the variance will be conditionally modeled so the more interesting statistics are the skewness and the excess kurtosis which can be used to test whether the empirical distribution have kurtosis and skewness similar to a normal distribution. This was done with the Jargue-Bera test which rejected the null hypothesis that sample distribution comes from a normal distribution at the 5% significance level, see appendix A for short description of Jargue-Bera test. This was in line with expectations from the ocular inspection of the q-q plots in Figure 4 which implied that the empirical distribution of the daily returns exhibit significantly heavier tails than the normal distribution.

[3] The t location-scale distribution is a non-standardized Student's t-distribution, meaning with a modified mean and variance.

[4] A value of 1 implies that the null hypothesis was rejected.

The same analysis was done for the other two indices, OMXC20 and OMXH25, but the full analysis of those is available in appendix B to facilitate reading of the report. The main conclusions were the same and are presented in the following summary.

2.2.2 Summary

The analysis of the data suggests some common characteristics of the three datasets. Firstly, all of the equity indices do exhibit some weak serial correlation of lower order, in their daily returns and thus need a model for the conditional mean. Secondly, all the squared return series exhibit ARCH effects which motivate the use of GARCH type volatility models. Finally, the empirical distribution of the return series display significantly heavier tails than the normal distribution which implies that another choice of parametric family should be considered.

Moreover some of the well-known key characteristics of asset return volatility are evident in all the covered datasets. The volatility is exhibiting periods of relative calm followed by more turbulent periods, referred to as volatility clustering and the volatility seems to be mean reverting to some extent. This shows that volatility is not diverging to infinity but is moving within some range.

Based on these characteristics the basic structure for the modeling framework is presented in section 3. The section then continues by specifying the conditional mean models, conditional variance models and the error distributions that will be used in the forecasting models.

3 Forecasting models

As mentioned in the introduction, it is a well-known fact that volatility of asset returns is time-varying and predictable, see Andersen and Bollerslev (1997). Volatility also has some commonly seen characteristics. Three of the most prominent stylized facts about volatility are that volatility exhibits persistence, volatility is mean reverting and innovations have an asymmetric impact on volatility, see Engle and Patton (2001). The volatility persistence based on empirical data suggests that volatility exhibits periods of relative calm followed by more turbulent periods and is also referred to as volatility clustering. Empirical data also suggests that volatility is mean reverting to some extent. Hence, volatility is not diverging to infinity but is moving within some range. The volatility's asymmetric dependency of positive and negative shocks is also referred to as the leverage effect, first noted by Black (1976), which suggests that negative shocks have a larger impact on the volatility than an equal size positive shock. Any model that tries to forecast volatility should be able to incorporate as many of these characteristics as possible to accurately describe the conditional variance.

The basic structure of volatility modeling is presented next which is followed by a more specific theory about the conditional mean and more importantly the different conditional variance models and their properties.

3.1 Basic structure

Let r_t denote the daily log return defined in equation (1) section 2. The general idea of GARCH type volatility modeling is that r_t is serially uncorrelated or exhibits some minor lower order serial correlation and thus need a model for the conditional mean. The aim of volatility models is to be able to capture this dependency in the return series. Consider the conditional mean μ_t and the conditional variance h_t^2 defined as

$$\mu_t = E(r_t|F_{t-1}), \qquad h_t^2 = Var(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}] \tag{2}$$

where F_{t-1} denotes the information set available at time t-1. As was evident in the descriptive statistics of the datasets in section 2 the serial dependence of the return series was quite weak and therefore suggests that the conditional mean should be able to be modeled by a relatively simple model, see Tsay (2002). If the conditional mean is assumed to follow a stationary ARMA(p, q) model, the model framework is described by

$$r_t = \mu_t + Z_t, \quad \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i Z_{t-i}$$

for r_t and where p and q are the nonnegative parameters of the ARMA(p, q) model.

3.2 Conditional mean

The main focus of this paper is the conditional variance and not the conditional mean but for the conditional variance model to work properly the conditional mean needs to be modeled and understood as well. In this paper three different models for the conditional mean will be used. The zero mean model advocated by Figlewski (1997) where the conditional mean is

$$\mu_t = 0,$$

the constant mean model with the conditional mean given by

$$\mu_t = \phi_0$$

where ϕ_0 is the unconditional mean of the in-sample period and the AR(1) mean model defined by

$$\mu_t = \phi_0 + \phi_1 r_{t-1}.$$

With the AR(1) conditional mean model the parameters are first calibrated to the in-sample period then the conditional means are computed and subtracted from the return series to get the mean adjusted return series Z_t.

3.3 Conditional variance

The main object of study in this paper is the conditional variance which, using the same notation as in equation (2), is defined by

$$h_t^2 = Var(r_t|F_{t-1}) = Var(Z_t|F_{t-1})$$

where Z_t

$$Z_t = r_t - \mu_t.$$

In this paper the focus will be restricted to examining the forecasting performance of six commonly used forecasting models; the simple moving average, the exponentially weighted moving average, the ARCH model, the GARCH model, the EGARCH model and the GJR-GARCH model. Next, each model is presented and discussed. First the two non GARCH models included as benchmarks are presented. Even though they are in no sense GARCH the above model with the conditional mean and conditional variance holds.

3.3.1 Simple Moving Average

One of the simplest and most straightforward forecast models for conditional variance from period $k + 1$ through $k + h$ is the simple moving average of the squared return series. The *n* day simple moving average at time k is defined by

$$\mathrm{MA_k(n)} = \frac{1}{n}\sum_{j=0}^{n-1} r_{k-j}^2.$$

This metric is the same as the historical variance over the *n* day historical period. Each of the squared observations is given an equal weight of $\frac{1}{n}$ up until the k+1-*n* day and all observations before that day get a zero weight. One obvious question is what *n* should be for a *h* day ahead forecast. A common convention is to use $n = h$. That means that the h day ahead forecast is given by the n day historical variance. Figlewski (1997) argued that calculating the historical variance over a significantly longer period generally yields lower forecast errors. Therefore, in this paper the 1 day ahead forecast, denoted $\mathrm{h}_{\mathrm{k}}^2(1)$, using this model will be calculated as the simple moving average of the 10 last squared observations (n=10) and is given by the equation

$$\mathrm{h}_{\mathrm{k}}^2(1) = \frac{1}{10}\sum_{j=0}^{9} r_{k-j}^2.$$

3.3.2 Exponential Weighted Moving Average

Another simple way to forecast the conditional variance is the so called Exponentially Weighted Moving Average which is quite similar to the Simple Moving Average but with a different weighting scheme. The exponentially weighted moving average at time k is defined

$$\mathrm{EWMA}_{\mathrm{k}} = \frac{1}{\Gamma}\sum_{j=0}^{J} \beta^j r_{k-j}^2$$

where

$$\Gamma = \sum_{j=0}^{J} \beta^j$$

and J is the total number of data points available prior to k. As for the parameter n in the Simple Moving Average model here β has to be defined. Riskmetrics™ use $\beta = 0.94$ which is the value used in this paper. The one day ahead volatility forecast at time k is given by

$$h_k^2(1) = \frac{1}{\Gamma}\sum_{j=0}^{J} \beta^j r_{k-j}^2$$

with the same notation as above. With this model it can easily be shown, by recursion, that the h day ahead forecast is the same as the one day ahead forecast

$$h_k^2(\mathrm{h}) = h_k^2(1) = \frac{1}{\Gamma}\sum_{j=0}^{J} \beta^j r_{k-j}^2.$$

Both the Moving Average and the Exponential Weighted Moving Average are non-parametric since they do not require any model calibration; they are the same independent of the in-sample data. These very blunt measures will be used as a benchmark for the four parametric models presented next.

3.3.3 The ARCH Model

The Autoregressive Conditional Heteroscedasticity (ARCH) model introduced by Engle (1982) was one of the first models that provided a way to model conditional heteroscedasticity in volatility. The model was simple and intuitive but usually required many parameters to adequately describe the volatility process. In the ARCH model Z_t is assumed to have the following representation

$$Z_t = h_t e_t, \{e_t\} \sim IID(0,1)$$

where h_t^2 is the conditional variance and is a function of $\{Z_s, s < t\}$, defined by

$$h_t^2 = \propto_0 + \sum_{i=1}^{p} \propto_i Z_{t-i}^2$$

where $\propto_0 > 0$ and $\propto_j \geq 0, j = 1,2 \dots p$ to guarantee that the conditional variance is positive. One parameter that has to be specified before fitting the model to the in-sample data is the order p. The order of the model needs to be specified for each of the parametric models. For the ARCH model the order p can be suggested by the Sample Partial Autocorrelation function of the squared returns were the Sample PACF should be insignificantly different from zero for lags greater than p. One weakness of the ARCH model is that it usually required quite a high order to accurately be able to model the conditional variance. For example, by looking at the Sample PACF for OMXS30 in Figure 3 the order p is suggested to be as high as 14 which is infeasible, or at least unpractical. For the GARCH, EGARCH and the GJR-GARCH the order cannot easily be estimated using the PACF or ACF. There are several studies showing that higher order versions of these models rarely perform better in terms of their out-of-sample volatility forecast, see Hansen and Lunde (2005) and Bollerslev et al. (1992). Based on this and due to the computational burden of incorporating higher order versions of all the models the analysis in this study will be restricted to lower order version. The models studied are the ARCH(3), GARCH(1,1), GARCH(1,2), GARCH(2,1), EGARCH(1,1) and the GJR-GARCH(1,1). That the ARCH(3) is unable to fully model the volatility clustering is well known but is included as it has as many parameters as the GARCH(1,2), EGARCH(1,1) and the GJR-GARCH(1,1) and will be used as a benchmark that is expected to be clearly outperformed.

3.3.3.1 Properties

The strength of the ARCH model is that it manages to model the volatility clustering and the mean reverting characteristics. The ability to model volatility clustering can be seen in the definition of the conditional variance where it is evident that a large Z_{t-i}^2 will give rise to a large h_t^2. In other words, large and small chocks tend to be followed by large and small chocks respectively of either sign. To further increase the understanding of the ARCH model's dynamics it is worth noting that the ARCH(1) model can be rewritten as an AR(1) model on the squared residuals Z_t^2. The ARCH model however suffers from some major drawbacks. Firstly the ARCH model generally requires many parameters to correctly describe the volatility process. Secondly the ARCH model models the conditional variance with only the squared shocks as a variable and is thus not able to model the asymmetric effects of positive and negative shocks. Furthermore the ARCH model imposes restrictive intervals for the parameters if it should have finite fourth moments and is likely to over predict volatility since it responds slowly to large, isolated shocks.

3.3.3.2 Forecasting

The forecasts of the ARCH model are obtained recursively as the forecasts of an AR model. If we consider an ARCH(p) model at the forecast origin k, the 1-step ahead forecast of h_{k+1}^2 is

$$h_k^2(1) = \alpha_0 + \alpha_1 Z_k^2 + \cdots + \alpha_p Z_{k+1-p}^2.$$

The 2-step ahead forecast is then given by

$$h_k^2(2) = \alpha_0 + \alpha_1 h_k^2(1) + \alpha_2 Z_k^2 + \cdots + \alpha_p Z_{k+2-p}^2.$$

Repeating this procedure yields the j-step ahead forecast for h_{k+j}^2 is

$$h_k^2(j) = \propto_0 + \sum_{i=1}^{p} \propto_i h_k^2(j-i).$$

Where $h_k^2(j-i) = \mathrm{Z}_{k+j-i}^2$ if $j - i \leq 0$. Thus if we consider an ARCH(3) model, which is the model used in this report, at the forecast origin k, the 1-step ahead forecast is

$$h_k^2(1) = \propto_0 + \propto_1 h_k^2(0) + \propto_2 h_k^2(-1) + \propto_3 h_k^2(-2) = \propto_0 + \propto_1 \mathrm{Z}_k^2 + \propto_2 \mathrm{Z}_{k-1}^2 + \propto_3 \mathrm{Z}_{k-2}^2.$$

3.3.4 The GARCH Model

Bollerslev (1986) extended the ARCH model to the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) which had the same key properties as the ARCH but required far less parameters to adequately model the volatility process. In the GARCH model Z_t is assumed to have the same representation as in the ARCH model

$$Z_t = h_t e_t, \{e_t\} \sim IID(0,1) \qquad (3)$$

but with a different model for h_t defined by

$$h_t^2 = \propto_0 + \sum_{i=1}^{p} \propto_i Z_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2$$

where $\propto_0 > 0, \propto_i \geq 0, \beta_j \geq 0 \; and \; \sum_{i=1}^{\max(p,q)} (\propto_i + \beta_i) < 1$. Where $\alpha_i \equiv 0 \; for \; i > p \; and \; \beta_j \equiv 0 \; for \; j > q$.

3.3.4.1 Properties

The properties of the GARCH model is quite similar to that of the ARCH but requires far less parameters to adequately model the volatility process. The GARCH model is able to model the volatility clustering but does not address the problem with the ARCH models lack of ability to model the asymmetric effect of positive and negative returns. The GARCH model also imposes restrictions on the parameters to have a finite fourth moment as was the case for the ARCH model. The GARCH model is similar to the ARCH model but with the addition of lagged conditional variances,h_{t-j}^2, as well as the lagged squared returns, Z_{t-i}^2. The addition of the lagged conditional variances avoids the need for adding many lagged squared returns as was the case for the ARCH model to be able to appropriately model the volatility. This greatly reduces the number of parameters that need to be estimated. In fact, considering the GARCH(1,1) the conditional variance can be rewritten as

$$h_t^2 = \propto_0 + \propto_1 Z_{t-1}^2 + \beta_1(\propto_0 + \propto_1 Z_{t-2}^2 + \beta_1 h_{t-2}^2)$$

Or, by continuing the recursive substitution, as

$$h_t^2 = \frac{\propto_0}{1-\beta_i} + \propto_1 \sum_{i=0}^{\infty} Z_{t-1-i}^2 \beta_1^i$$

which shows that the GARCH(1,1) model corresponds to an ARCH(∞) model with a certain structure for the value of the parameters of the lagged returns Z_{t-i}^2. Furthermore, just as the ARCH(1) could be seen as an AR(1) model of the squared returns the GARCH(1,1) model can in a similar way be rewritten as an ARMA(1,1) model on the squared returns.

3.3.4.2 Forecasting

The forecasts of the GARCH model are obtained recursively as the forecasts of an ARMA model. If we consider an GARCH(1,1) model which is one of the GARCH models under study at the forecast origin k, the 1-step ahead forecast of h_{k+1}^2 is

$$h_k^2(1) = \alpha_0 + \alpha_1 Z_k^2 + \beta_1 h_k^2$$

When calculating multistep ahead forecasts the volatility equation (3) can be rewritten as $Z_t^2 = h_t^2 e_t^2$, which gives

$$h_{t+1}^2 = \propto_0 + (\propto_1 + \beta_1)h_t^2 + \alpha_1 h_t^2(e_t^2 - 1).$$

If $t = k + 1$ the equation yields

$$h_{k+2}^2 = \propto_0 + (\propto_1 + \beta_1)h_{k+1}^2 + \alpha_1 h_{k+1}^2(e_{k+1}^2 - 1)$$

with $E(e_{k+1}^2 - 1|F_h) = 0$, the 2-step volatility forecast is

$$h_k^2(2) = \propto_0 + (\propto_1 + \beta_1)h_k^2(1).$$

The general j-step ahead forecast for h_{k+j}^2, at the forecast origin k, is

$$h_k^2(j) = \propto_0 + (\propto_1 + \beta_1)h_k^2(j-1), j > 1.$$

Repeating the substitutions for $h_k^2(j-1)$ until the j-step forecast can be written as a function of $h_k^2(1)$ gives the explicit expression for the j-step ahead forecast

$$h_k^2(j) = \frac{\propto_0 [1 - (\propto_1 + \beta_1)^{j-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{j-1} h_k^2(1).$$

The derivation of the forecasts for the other two GARCH models in this study, GARCH(1,2) and GARCH(2,1) is similar but is omitted in the report.

3.3.5 The EGARCH Model

Both the ARCH and the GARCH model are able to model the persistence of volatility, the so-called volatility clustering but the models both assume that positive and negative shocks have the same impact on volatility. It is well known that for financial asset volatility the innovations have an asymmetric impact. To be able to model this behavior and overcome the weaknesses of the GARCH model Nelson (1991) proposed the first extension to the GARCH model, called the Exponential GARCH (EGARCH), which was able to allow for asymmetric effects of positive and negative asset returns. In the EGARCH(p,q) model Z_t is assumed to have the same representation as before, see equation (3) with the conditional variance now given by

$$\log(h_t^2) = \propto_0 + \sum_{i=1}^{p} [\propto_i Z_{t-i} + \gamma_i(|Z_{t-i}| - E(|Z_{t-i}|))] + \sum_{j=1}^{q} \beta_j \log(h_{t-j}^2).$$

Here no restrictions are imposed on the parameters to guarantee a nonnegative conditional variance. The EGARCH(1,1) is thus given by

$$\log(h_t^2) = \propto_0 + \propto_1 Z_{t-1} + \gamma_1(|Z_{t-1}| - E(|Z_{t-1}|)) + \beta_1 \log(h_{t-1}^2).$$

To illustrate the ability to model for asymmetrical effects of positive and negative asset returns consider the function g defined by

$$g(Z_t) \equiv \propto_1 Z_{t-1} + \gamma_1\big(|Z_{t-1}| - E(|Z_{t-1}|)\big).$$

By the assumed properties of Z_t, $g(Z_t)$ has zero mean and is uncorrelated. The function can be rewritten as

$$g(Z_t) = (\propto_1 + \gamma_1) Z_t I(Z_t > 0) + (\propto_1 - \gamma_1) Z_t I(Z_t < 0) - \gamma_1 E(|Z_t|)$$

where the asymmetrical effect of positive and negative asset returns is evident. Positive shocks have an impact $(\propto_1 + \gamma_1)$ on the logarithm of the conditional variance while negative shocks have an impact $(\propto_1 - \gamma_1)$. Typically $\propto_1 < 0, 0 \leq \gamma_1 < 0$ and $\beta_1 + \gamma_1 < 1$. With this configuration negative shocks have a larger impact than positive shocks which is in line with empirical evidence by the so called leverage effect.

3.3.5.1 Properties

The EGARCH model requires no restrictions on the parameters to assure that the conditional variance is nonnegative. The EGARCH model is able to model volatility persistence, mean reversion as well as the asymmetrical effect. To allow for positive and negative shocks to have different impact on the volatility is the main advantage of the EGARCH model compared to the GARCH model.

3.3.5.2 Forecasting

If we consider an EGARCH(1,1) model, which is the model used in this report, at the forecast origin k, the 1-step ahead forecast of h_{k+1}^2 is

$$\log(h_k^2(1)) = \propto_0 + \propto_1 Z_k + \gamma_1(|Z_k| - E(|Z_k|)) + \beta_1 \log(h_k^2).$$

Since all of the parameters at the right hand side are known at time k the 1 day ahead volatility forecast is simply $h_k^2(1) = h_{k+1}^2$. The multi day ahead volatility forecast of the EGARCH(1,1) model is not as trivial as for the other models used in this paper and since the forecast evaluation will be based only on their 1 day ahead forecast the general expression for the multi day ahead volatility forecast of the EGARCH(1,1) model is omitted.

3.3.6 The GJR-GARCH Model

An alternative way of modeling the asymmetric effects of positive and negative asset returns was presented by Glosten, Jagannathan and Runkle (1993) and resulted in the so called GJR-GARCH model. In the GJR-GARCH(p,q) model Z_t is assumed to have the same representation as before, see equation (3) with the conditional variance now given by

$$h_t^2 = \propto_0 + \sum_{i=1}^{p} (\alpha_i Z_{t-i}^2 (1 - I[Z_{t-i} > 0]) + \gamma_i Z_{t-i}^2 I[Z_{t-i} > 0]) + \sum_{j=1}^{q} \beta_j h_{t-j}^2$$

where $\propto_0 > 0, \propto_i \geq 0, \beta_j \geq 0 \; and \; \gamma_i \geq 0$ to guarantee that the conditional variance is nonnegative. The GJR-GARCH(1,1) is thus given by

$$h_t^2 = \propto_0 + \alpha_1 Z_{t-1}^2 (1 - I[Z_{t-1} > 0]) + \gamma_1 Z_{t-1}^2 I[Z_{t-1} > 0] + \beta_1 h_{t-1}^2.$$

As in the case of the EGARCH the asymmetrical effect of shocks can be seen by considering the function

$$g(Z_t) \equiv \alpha_1 Z_{t-1}^2 (1 - I[Z_{t-1} > 0]) + \gamma_1 Z_{t-1}^2 I[Z_{t-1} > 0].$$

Positive shocks thus have an impact γ_1 on the logarithm of the conditional variance while negative shocks have an impact $\propto_1$. Typically $\propto_1 > \gamma_1$ which imposes a larger weight for negative shocks than for positive shocks in line with the leverage effect.

3.3.6.1 Properties

The properties of the GJR-GARCH model are very similar to the EGARCH model which both are able to capture the asymmetric effect of positive and negative shocks. The GJR-GARCH and the EGARCH may both be considered for the same series and it is hard to distinguish a criterion for choosing either one of the two models.

3.3.6.2 Forecasting

If we consider an GJR-GARCH(1,1) model, which is the model used in this report, at the forecast origin k, the 1-step ahead forecast of h_{k+1}^2 is

$$h_k^2(1) = \propto_0 + \alpha_1 Z_k^2 (1 - I[Z_k > 0]) + \gamma_1 Z_k^2 I[Z_k > 0] + \beta_1 h_k^2.$$

When calculating multistep ahead forecasts the volatility equation (3) can be rewritten using $Z_t^2 = h_t^2 e_t^2$ which gives

$$h_k^2(2) = \mathrm{E}[\propto_0 + \alpha_1 Z_{k+1}^2 (1 - I[Z_{k+1} > 0]) + \gamma_1 Z_{k+1}^2 I[Z_{k+1} > 0] + \beta_1 h_{k+1}^2 | F_k].$$

With $E(e_{k+1}^2 - 1 | F_h) = 0$, the 2-step volatility forecast is

$$h_k^2(2) = \propto_0 + \left(\frac{\propto_1 + \gamma_1}{2} + \beta_1\right) h_k^2(1).$$

The general j-step ahead forecast for h_{k+j}^2, at the forecast origin k, is

$$h_k^2(\mathrm{j}) = \propto_0 + \left(\frac{\propto_1 + \gamma_1}{2} + \beta_1\right) h_k^2(j-1).$$

Repeating the substitutions for $h_k^2(j-1)$ until the j-step forecast can be written as a function of $h_k^2(1)$ gives the explicit expression for the j-step ahead forecast

$$h_k^2(\mathrm{j}) = \propto_0 \sum_{i=0}^{j-2} \left(\frac{\propto_1 + \gamma_1}{2} + \beta_1\right)^i + \left(\frac{\propto_1 + \gamma_1}{2} + \beta_1\right)^{j-1} h_k^2(1).$$

3.4 The distribution of $\boldsymbol{e_t}$

For all of the GARCH type volatility models besides determining their order one must also assume a distribution for the error, e_t, to fully specify each model. In the parametric models Z_t was assumed to have the representation

$$Z_t = h_t e_t, \{e_t\} \sim IID(0,1).$$

The error terms e_t should be identically distributed and independent with zero mean and unit variance but what distribution it should follow must be specified. In this paper two different types of error distributions are considered, the standard normal distribution
$e_t \sim N(0,1)$ and the heavier tailed student's t-distribution $\sqrt{v/(v-2)}\, e_t \sim t_v$. The scale factor of e_t when the error distribution is assumed to be a student-t distribution is introduced to make the variance of e_t equal to 1. As was evident in the descriptive statistics in section 2 the return series is typically not normally distributed but display significantly heavier tails than the normal distribution which implies that another choice of parametric family should be considered which in this paper is done with the student t-distribution. Next the two distributions are defined in terms of their density function.

3.4.1 Normal distribution

The density function of the normal distribution is defined as

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, \qquad -\infty < z < \infty.$$

3.4.2 Student t-Distribution

The density function of the Student t-distribution is defined by

$$f(z) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{z^2}{v}\right)^{-(\frac{v+1}{2})}, \qquad -\infty < z < \infty$$

where v denotes the number of degrees of freedom and Γ denotes the gamma function, $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$. To get the standardized distribution the density function is scaled with $\sqrt{\frac{v-2}{v}}$.

4 In-sample model fitting and evaluation

Once the conditional variance models and their respective order have been specified the next step is to estimate the parameters in each of the models. Each of the GARCH type models is fitted to the in-sample data using the Maximum Likelihood Estimation method which given the dataset and a statistical model provides estimates for the parameters. Once the density function of e_t is provided the estimation is easy to implement.

4.1 Maximum Likelihood estimation

The likelihood function is essentially a joint probability density function but instead of regarding it as a function of the data with the parameters given, $g(z_1, z_2, \dots, z_n|\Theta)$, it is viewed as a function of the parameters with the data given, $L(\Theta|z_1, z_2, \dots, z_n)$ where Θ is the set of parameters that are to be estimated. If the returns were independent of each other the joint density function would simply be the product of the marginal densities. In the GARCH model returns are of course not independent however, the joint density function can still be written as the product of the conditional density functions as

$$f(z_1, z_2, \dots, z_n|\Theta) = f(z_n|F_{n-1})f(z_{n-1}|F_{n-2}) \dots f(z_1).$$

The Likelihood function is to be maximized with respect to the unknown parameters and is defined as

$$L(\Theta|F_{n-1}) = \prod_{t=1}^{n} \varphi(Z_t|F_{t-1})$$

where F_t denotes the information available at time t and φ is the density function of e_t. Hence, the exact form of the likelihood function depends on the parametric form of the distribution of the innovations. Maximizing the likelihood function is equivalent to maximizing the logarithm of the likelihood function which is usually easier to handle and therefore the log likelihood functions will be the function of focus. With the innovations, z_k, assumed to be realizations from a normal distribution, the log likelihood function takes the specific form

$$\log[L(\Theta|F_{n-1})] = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log(h_i^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{z_i^2}{h_i^2}$$

where the variance h_i^2 is substituted recursively with the specified conditional variance model. If the innovations are assumed to be realizations from the student t distribution instead, with v degrees of freedom, the log likelihood function is

$$\log[L(\Theta|F_{n-1})] = n\,log\left[\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma\left(\frac{v}{2}\right)}\right] - \frac{1}{2}\sum_{i=1}^{n}\log(h_i^2) - \frac{v+1}{2}\sum_{i=1}^{n}\log[1 + \frac{z_i^2}{h_i^2(v-2)}]$$

where again the variance h_i^2 is substituted recursively with the specified conditional variance model. The maximization of the respective log-likelihood function was done with Matlab's® function *estimate*. The maximization with respect to the parameters yields the optimal in-sample fit for each model. These parameters are then used for calculating the volatility forecasts for the respective model further explained in section 3.

4.2 Evaluation of in-sample fit

When the model fitting is completed there is a need to compare the goodness of fit for the different models. A comparison of how well each of the models fit the in-sample data can be done using different metrics depending on whether the models are nested or not. If the models are nested, meaning if the more complex model can be transformed into the simpler model by setting constraints on some of the parameters, the in-sample fit can be evaluated by using a likelihood ratio. A more general evaluation method is to use an information criterion that can compare any model's fit to the same data. The basic idea of information criteria tests are that the maximum likelihood for each of the model is subjected to a penalty for its complexity, usually the numbers of parameters. How the penalty is calculated differs from different information criteria tests. In this paper two of the most well used criteria will be used, Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The Akaike information criterion is defined by

$$AIC = -2logL(\theta) + 2k$$

where $logL(\theta)$ is the maximized log likelihood function for a model with k numbers of parameters. The Bayesian information criterion is defined by

$$BIC = -2logL(\theta) + klog(N)$$

with the same notation as for the AIC but with the additional parameter N which denotes the number of data points in the in-sample period. When comparing the in-sample fit of different models using the AIC and BIC information criteria tests the smaller value of the criterion the better. A model with a smaller AIC or BIC thus provides the best in-sample fit taking into account the numbers of parameters needed and for the BIC the number of data points used as well.

5 Out-of-sample evaluation of volatility forecast

When the conditional variance models being studied have been fully specified with their respective order and error distribution, have been fitted to the in-sample data and the 1 day ahead forecasts for each of the days in the out-of-sample period has been computed, the out-of-sample evaluation is the final step in the study performed in this paper. However, even with the future level of volatility forecasts computed, it is far from trivial to evaluate the forecasting performance of the different models. Even if a model has been chosen and fitted to the data and the forecasts have been calculated, evaluating the performance of that forecast is difficult due to the latent nature of conditional volatility, it is unobservable. A proxy for the realized volatility is thus needed and moreover the choice of statistical measure, as pointed out by Bollerslev, Engle and Nelson (1994), is far from clear. This section starts by revealing the problems with the inherently latent nature of volatility and advocates the use of a different volatility proxy than the squared observations, more specifically the High-Low Range. The second part of this section discusses the issues of what evaluation criteria to use when comparing the volatility forecasts with the proxy of the "true" volatility. Several different loss functions are presented and discussed. The section ends by specifying the loss functions used in this paper.

5.1 Volatility proxies

To understand the difficulties with estimating the "true" latent volatility consider the univariate random variable X following a zero mean normal distribution. The variance is defined as

$$Var(X) = E[(X-\mu)^2] = E[X^2] - E[X]^2 \; where \; \mu \text{ denotes the mean.}$$

With a zero mean $\mu = E[X] = 0$ the variance is thus $Var(X) = E[X^2] - E[X]^2 = E[X^2]$. The empirical distributions of the squared realizations X^2 from N(0,var) where var=0.5, 1 and 2 respectively are shown in Figure 5. The empirical distributions are based on 100,000 observations from the underlying process.

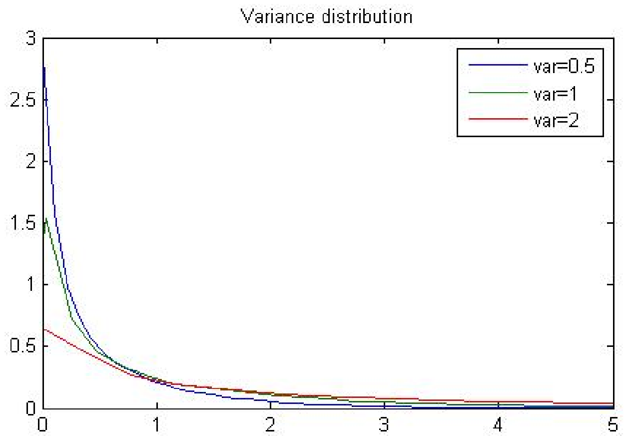

Figure 5: The empirical distributions for squared realizations from N(0, var) where var= 0.5, 1 and 2 respectively. The empirical distribution is based on 100,000 observations from the underlying process.

In Figure 5 the empirical density exhibits less kurtosis with a higher level of variance. In addition the empirical density of squared innovations from a distribution with higher variance has a heavier right tail than the empirical density of observations from a distribution with lower variance which is expected since large squared innovations is more likely with a high variance. Using the daily squared innovation as a proxy for the underlying volatility is thus to use a single squared innovation to try to distinguish from which of the distributions in figure 5 that specific innovation is a sample from. That is evidently a very blunt proxy and is likely to be extremely noisy.

The variance can be estimated using a simple moving average which models the variance as the average of a fixed number of squared observations. In Table 2 the standard error, SE, for the moving average with a 100 day window is presented for six different levels of the underlying variance.

Table 2: Standard error of variance estimations using MA(100) for 6 different values of the latent, unconditional variance. The standard error is calculated based on 100,000 observations from the underlying process of N(0,var).

VAR	**0.5**	**1**	**2**	**4**	**8**	**16**
SE	0.072042	0.143672	0.284956	0.566776	1.119889	2.286552

In Table 2 it is evident that the moving average proxy of the volatility has a higher standard error the higher the underlying variance is. Hence, it is harder to be accurate, in an absolute sense, for the moving average model when estimating high variances. Another interesting aspect is how the length of the moving average window affects the reliability of the variance measure. In Figure 6 the empirical estimation error distribution for moving average models using 10, 30 and 100 day windows is presented. The underlying process was an N(0, 2) distribution and the error was computed as, $\in$=MA(d)-2. The empirical distribution is based on 100,000 observations from the underlying process.

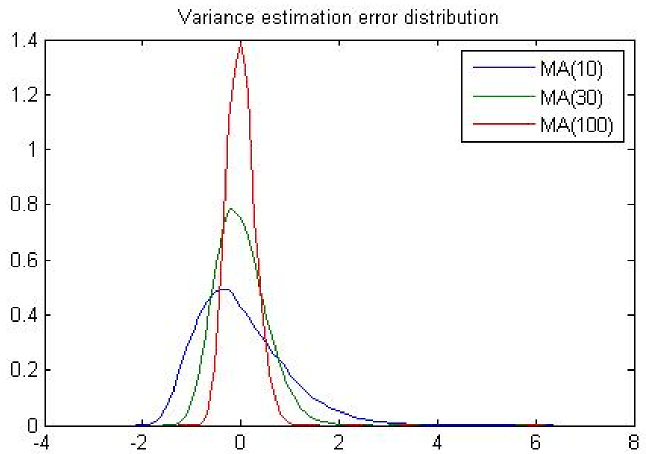

Figure 6: The empirical estimation error distribution for Moving Average models using 10, 30 and 100 day windows. The underlying process was an N(0, 2) distribution and thus the error was computed as, $\in = \mathbf{MA}(\mathbf{d}) - \mathbf{2}$. The empirical distribution is based on 100,000 observations from the underlying process.

In Figure 6 it is evident that with a constant variance the moving average estimate will be more accurate the more observations that are included, that is the longer the window used by the moving average. This can be seen as the variance estimation error distribution gets centered at zero and is more and more leptokurtic with an increasing window length. This is expected and in line with the law of large numbers. Another interesting fact evident in Figure 14 is that with a shorter window, the variance is in general underestimated which is seen by the left skew which is most prominent for the MA(10).

To further underline the difficulties with estimating variance even when it is constant, that is not heteroscedastic, the ARCH(1) variance measurement error is investigated for different levels of variance. In Figure 7 the standard error of the ARCH(1) variance estimation is plotted for six different levels of the underlying variance. Since parameter calibration will be different depending on the specific observations the standard error was computed four times with a completely new set of observations for each of the variance levels. Therefore, four different values of the standard error will be presented for each variance level. A linear regression was then performed based on the mean of these four measurements at each variance level.

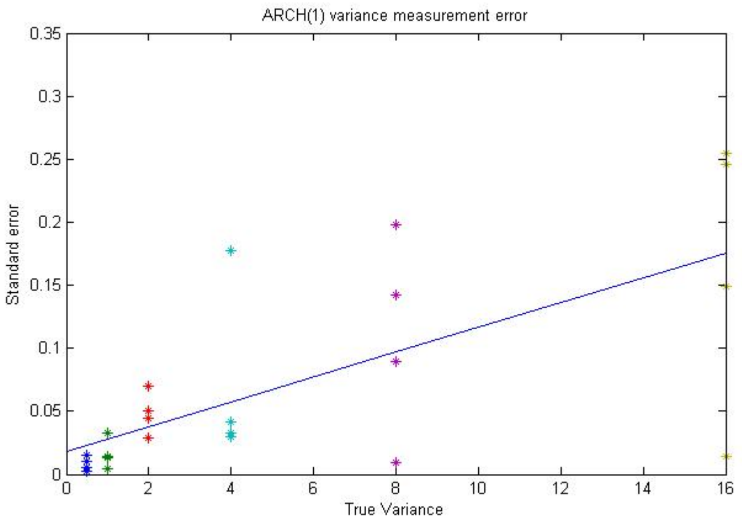

Figure 7: The standard error of the ARCH(1) estimated variance for 6 different values of the latent, unconditional variance. The standard error is calculated based on 10,000 observations from the underlying process of N(0,var) where var=0.5, 1, 2, 4, 8 and 16.

As was the case for the moving average variance estimation, evident in Table 2, the ARCH(1) proxy of the volatility has a higher standard error the higher the underlying variance is. It is therefore harder to be accurate, in an absolute sense, for the ARCH(1) model when estimating high variances.

Based on the discussion in this section so far it is more difficult to be accurate the higher the underlying variance is both for the blunt moving average model and the ARCH(1) model. Moreover it is advantageous to include as many previous observations as possible when estimating a constant variance with the moving average model. In this paper the variance is assumed to be time varying, that is not constant, which even further complicates matters. When the variance is time varying there will be a trade-off between getting an accurate measure of the variance and getting a measure for the variance that truly reflects the variance at that particular time. The ability to model changes in time can be seen in appendix C where it is evident that the MA(50) is able to react to changes much faster than the MA(100), however the MA(100) has a lower variance estimation error which was exhibited in Figure 6.

In this paper the conditional variance is the object of study and the variance is considered completely time varying and thus should be able to take different values at each time point. If only daily closing prices are available the only volatility proxy that fully emphasizes the time varying property is the daily squared returns. The daily squared returns have for a long time and in a vast amount of papers been used as the proxy for the unobserved variance, see Cumby et al (1993), Figlewski (1997) and Jorion (1995, 1996). Squared returns are however an extremely noisy proxy of the latent realized variance which will be demonstrated. In Figure 8 the squared returns from an N(0,16) distribution is plotted in blue. The unconditional variance of 16 is plotted in red. It is strikingly evident that the squared returns provides a highly variable proxy of the, in this case, constant volatility.

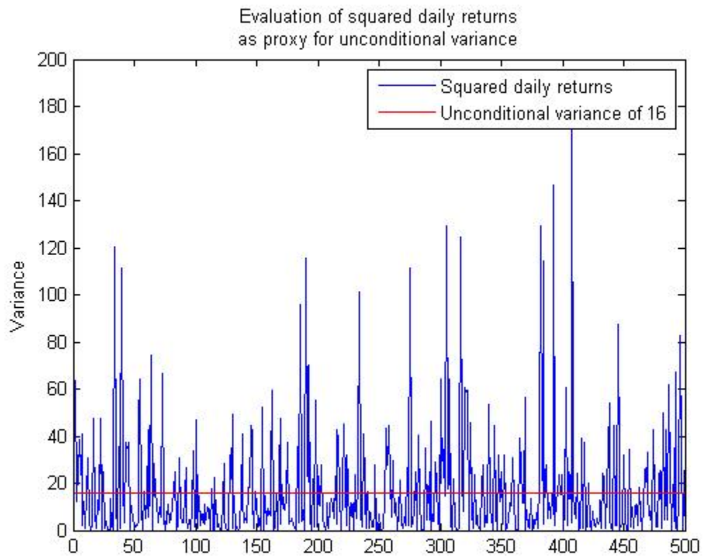

Figure 8: Plot of the squared innovations from an N(0, 16) distribution in blue and the unconditional variance of 16 in red. The plot is based on 500 observations from the underlying process.

In Figure 9 the estimated variance of an ARCH(1) model and a MA(100) is plotted. The plot is based on 10,000 observations from an N(0, 16) distribution. In the figure it is evident that the ARCH(1) model is quite close to the constant unconditional variance of 16 whereas the MA(100) exhibits highly variable estimations of the variance. The motive for including Figure 9 is however that even if the ARCH(1) volatility estimates are used, which evidently are very close to the actual volatility of 16, evaluating those estimates using squared returns would still lead to the conclusion that the estimates are far from accurate since the squared returns are such a noisy proxy.

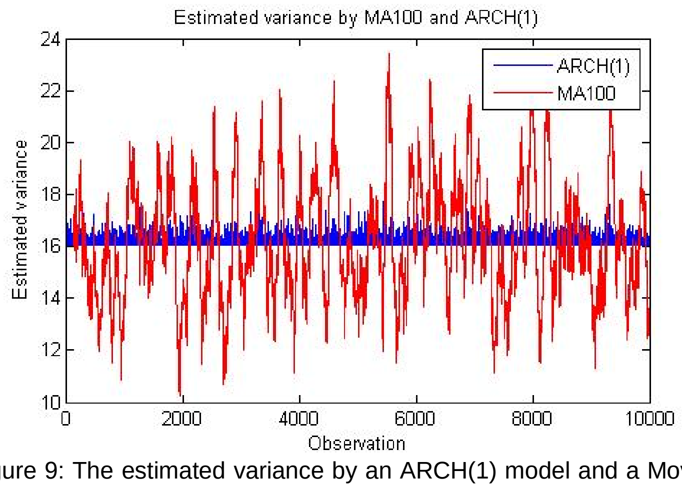

Figure 9: The estimated variance by an ARCH(1) model and a Moving Average with a 100 day window. The plot is based on 10,000 observations from an N(0, 16) distribution.

The use of daily squared returns as a proxy led to very poor out-of-sample performance in spite of highly significant in-sample fitting, see Andersen and Bollerslev (1998). This led to the conclusion that the volatility models explained very little of the time variability of the volatility and thus had limited practical value. However Andersen and Bollerslev (1998) answered the skeptics and showed that the GARCH type volatility in fact provides strikingly accurate volatility forecasts when the out-of-sample forecasting performance was evaluated with a more suited proxy. They argued that there is in fact no contradiction between a good volatility forecast estimate and a poor predictive power for the daily squared returns. The advocated proxy was to estimate the daily volatility using cumulative squared intra-day returns. The intra-day returns are used as a proxy for the volatility in the following manner. Assume there are m equally spaced observations per trade day and denote the ith intra-day return during day t by $r_{i,m,t}$. The cumulative squared intra-day returns are then computed as

$$\widehat{\sigma^2_{RV,m,t}} = \sum_{i=1}^{m} r^2_{i,m,t}.$$

If m=1 the cumulative squared intra-day return proxy is equal to the daily squared return proxy. With m=1 the proxy is unbiased but very noisy. As $m \to \infty$ $\widehat{\sigma^2_{RV,m,t}} \to^p \sigma^2_t$ where σ^2_t denotes the true latent volatility. However, using the cumulative squared intra-day returns as a proxy for the realized daily volatility requires high frequency data which in many cases aren't available or is only available for shorter time periods. Furthermore, such data is costly to obtain and time consuming to process. In addition, as noted by Dacorogna et al. (2001), the estimation of volatility from high frequency data is a complex issue due to market microstructure effect. Reliable open and close prices and intraday high and low prices are often available for most financial assets over long time horizons. There are volatility proxies that use such data instead of the high frequency data to estimate the volatility. In this paper only daily data was available resulting in that another proxy than the cumulative squared intra-day return proxy had to be used. A simplified proxy will be used, first introduced by Parkinson (1980) usually referred to as the High-Low range proxy. The high low range at day t denoted RG_t is defined as

$$RG_t = \max_{\tau}(\log(p_\tau)) - \min_{\tau}(\log(p_\tau)), \quad t-1 < \tau \leq t$$

where p_τ is the price level at time τ during the day. The log range is thus the difference of the logarithm of the highest price level during the given day and the logarithm of the lowest price during the same day. This range contains more information than the simple daily return based on the closing price since it incorporates how the price has fluctuated throughout the day. For example, on a day when the price fluctuates substantially during the day but the closing price still is close to the opening price, the daily return would suggest a day of low volatility while the log range would reflect the intraday price movements and thus imply correctly that the volatility was high. Assuming a geometric Brownian motion with zero drift and with a constant volatility σ the expected value of the squared log range is directly related to the volatility of the process by the following expression

$$E_{t-1}[RG^2_t] = 4\log(2)\,\sigma^2_t.$$

Hence, the Parkinson estimator for the volatility denoted $\widehat{\sigma_{P,t}^2}$ is defined by

$$\widehat{\sigma_{P,t}^2} = \frac{(\log(H_t) - \log(L_t))^2}{4\log(2)}$$

where H_t, L_t denote the daily high and low price respectively.
An extension to Parkinson's volatility estimator was provided by Garman and Klass (1980) which in addition to the daily high and low prices also utilized the information in the opening and closing prices. The estimator is defined as

$$\widehat{\sigma_{GK,t}^2} = 0.5\log\left(\frac{H_t}{L_t}\right)^2 - (2\log(2) - 1)\log\left(\frac{C_t}{O_t}\right)^2$$

where H_t, L_t denote the daily high and low price respectively and C_t, O_t denote the closing and opening price respectively. Under idealized conditions the Garman-Klass volatility estimator is a less noisy proxy than the Parkinson estimator. However empirical studies have shown that the Parkinson's estimator performs well with real data, see Chou et al. (2010) and Alizadeh, Brandt and Diebold (2002). In addition the Parkinson's estimator appears to be very robust to market microstructure effects, see Brandt and Diebold (2006). For these reasons the Parkinson's estimator will be used as the proxy for the volatility in this paper.

5.2 Forecast evaluation

With the volatility proxy, $\widehat{\sigma_{P,t}^2}$, and the 1 day ahead volatility forecasts, $h_{t-1}^2(1)$, computed for each of the days in the out-of-sample period it is far from trivial to evaluate the performance of the respective volatility models which was pointed out by Bollerslev, Engle and Nelson (1994) and Diebold and Lopes (1996). There is not a unique, universal criterion for selecting the "best" model but will depend on the preferences and the intended use of the forecasts. The preferences are usually expressed through a loss function. A well-used way to evaluate the out-of-sample fit is through the R^2 of the Mincer-Zarnowitz (MZ) regression defined by

$$\widehat{\sigma_{P,t}^2} = a + bh_{t-1}^2(1) + u_t \qquad (4)$$

where $\widehat{\sigma_{P,t}^2}$ is the volatility proxy, $h_{t-1}^2(1)$ is the one day ahead volatility forecast at time t-1, u_t is the conditional mean and a and b are the parameters to be estimated in the regression. An optimal, unbiased forecast would satisfy a=0, b=1 with an R^2 close to 1. The MZ regression, equation 4, is very sensitive to extreme values of $\widehat{\sigma_{P,t}^2}$ and thus the parameters will primarily be determined by these extreme values. This was illuminated by Engle and Patton (2000) who advocated the use of the log regression

$$\log(\widehat{\sigma_{P,t}^2}) = a + b\log(h_{t-1}^2(1)) + u_t \qquad (5)$$

which is less sensitive to outliers. However the R^2 of the regressions is not an optimal criteria for the evaluation since it does not penalize a biased forecast, that is with $a \neq 0$ or $b \neq 1$. So purely based on the R^2 a poor biased forecast can be preferred to a good unbiased forecast due to the parameters a and b not fulfilling the unbiased requirements. Thus in this paper, instead of the R^2 of the regressions defined in equation (4) and (5) the following four loss functions will be used in the models out-of-sample forecast performance evaluation

$$MSE = n^{-1} \sum_{t=1}^{n} \left(\widehat{\sigma_{P,t}^2} - h_{t-1}^2(1)\right)^2 \tag{6}$$

$$R^2LOG = n^{-1} \sum_{t=1}^{n} \left(\log(\frac{\widehat{\sigma_{P,t}^2}}{h_{t-1}^2(1)})\right)^2 \tag{7}$$

$$PSE = n^{-1} \sum_{t=1}^{n} \left(\widehat{\sigma_{P,t}^2} - h_{t-1}^2(1)\right)^2 h_{t-1}^{-4}(1) \tag{8}$$

$$MAD = n^{-1} \sum_{t=1}^{n} |\widehat{\sigma_{P,t}^2} - h_{t-1}^2(1)| \tag{9}$$

where *n* is the number of days in the out-of-sample period, $\widehat{\sigma_{P,t}^2}$ is the volatility proxy at day t and $h_{t-1}^2(1)$ is the one day ahead volatility forecast at time t-1. Before comparing these loss functions it is important to understand their economic interpretation and what they actually measure to be able to draw the right conclusions. The criteria MSE (6), R^2LOG (7) and PSE (8) were all suggested by Bollerslev, Engle and Nelson (1994) and the MAD criteria (9) were suggested by Hansen and Lunde (2001). The criteria MSE (6) and R^2LOG (7) is similar to the R^2 of the regressions defined in equation (4) and (5) respectively except for the constant term a. The Mean Square Error (MSE) is the average of the squared deviations of the forecast from the volatility proxy. Hence, one large deviation is given a much higher weighting than a sum of small deviations even if the sum of the deviations is equal to the one time large deviation. When evaluating volatility forecasting performance this can seem quite illogical since in general one large deviation is not more troublesome than a sum of small deviations that sum up to the size of the large deviation since the returns accumulate over time. Moreover single outliers will have a significant impact on the MSE criteria. The R^2LOG still assigns higher weighting to large deviations but they are not as penalized as in the case of MSE. The Percentage Squared Errors (PSE) measures the average of the squared percentage deviation. The deviations are expressed as a percentage of the forecasted volatility. Hence the PSE takes into account the fact that it is harder to be accurate, in an absolute sense, when estimating high variances and thus measures the relative error as a percentage to account for this. The fact that it is harder to measure higher variances was evident in the previous sub section *Volatility proxies* where the standard error for the moving average and the ARCH model was investigated for different levels of volatility. The Mean Absolute Deviation (MAD) is interesting since it is very robust to outliers and this criteria actually gives equal weighting to a large deviation of size z as to a sum of several deviations accumulating to z.

6 Empirical Results

In this section all of the empirical results are presented and discussed. As mentioned in the introduction the objective of this paper is to compare the volatility models in terms of the in-sample and out-of-sample fit. There will be three main themes being studied, first the basic structure of the modeling framework will be investigated with respect to the error distribution and the conditional mean to gain insight in how the assumed error distribution and different models for the conditional mean impacts the in-sample and out-of-sample fit, regardless of which specific conditional variance model is used. The second theme is whether the more complex models which are able to exhibit more of the stylized facts and characteristics of asset price volatility provide a better in-sample fit and/or out-of-sample-fit than the more parsimonious models. The third and final theme is whether the model with the best in-sample fit also produces the best out-of-sample volatility forecast. This question will be investigated when considering the conditional mean and the error distribution as well as when considering the different conditional variance models. Hence the third theme will be investigated parallel with the two others. The aim of the analysis is to evaluate whether the more complex models, in terms of the conditional mean, error distribution and conditional variance, outperforms the more parsimonious ones not if the GARCH type volatility models do provide accurate forecasts. That GARCH type volatility models do provide strikingly accurate volatility forecasts was shown by Andersen and Bollerslev (1998) and is not the object of study in this paper. However, based on the R^2 of the regressions defined in equation (4) and (5) the result available in appendix D suggest that the models provide good volatility forecasts. The R^2, using the Parkinson's proxy, are as high as 0.51 which is even higher than those found by Andersen and Bollerslev (1998).

Since there are three different conditional mean processes compared and two different error distributions there will be six sets of results for each of the data sets OMXS30, OMXC20 and OMXH25. Hence there will be a total of 18 sets of results, moreover in each of the result set, data for all of the compared volatility models must be included both in terms of in-sample and out-of-sample performance. Due to this large quantity of data, the complete results are available in appendix D to facilitate reading and only essential parts are presented here.

6.1 Impact of error distribution and conditional mean

The first variable investigated is how the assumed error distribution affects the in-sample and out-of-sample fit. As discussed in section 3 one must assume a distribution for the error e_t to fully specify each GARCH type volatility model. In this paper two different types of error distributions are considered, the standard normal distribution and the heavier tailed student's t-distribution. In Table 3 the average AIC and BIC for all of the compared volatility models is presented. The average AIC and BIC is presented both assuming a normal error distribution and a student t distribution for each respective dataset. The AIC and BIC for each specific model and error distribution is available for all datasets in appendix D.

Table 3: Average AIC and BIC for the compared parametric volatility models assuming a normal and a student t error distribution. In each row the result based on the respective data series is presented.

	AIC		BIC	
	Normal Distribution	**Student t Distribution**	**Normal Distribution**	**Student t Distribution**
OMXS30	-13128	-13174	-13108	-13148
OMXC20	-13656	-13725	-13635	-13698
OMXH25	-13476	-13510	-13455	-13483

Keeping in mind that a smaller value of AIC and BIC is preferred it is clearly evident in Table 3 that assuming a student t distribution provides a better in-sample fit across all the datasets both in terms of the AIC and the BIC. The result is unequivocal and is not too surprisingly considering the q-q plots combined with the jarque-bera test in section 2 which implied that the empirical distribution of the return series displayed significantly heavier tails than the normal distribution. The big question is however not the in-sample fit but the out-of-sample fit which is used to compare the forecasting performance.

In Table 4 the average MSE, R2LOG, MAD and PSE for the compared volatility models is presented assuming a normal error distribution and a student t distribution for each respective dataset.

Table 4: Average MSE, R2LOG, MAD and PSE for the compared parametric volatility models assuming a normal and a student t error distribution. In each row the result based on the respective data series is presented.

	MSE		R2LOG		MAD		PSE	
	Normal Dist	Student t Dist	Normal Dist	Student t Dist	Normal Dist	Student t Dist	Normal Dist	Student t Dist
OMXS30	3,2828E-08	3,44604E-08	1,56255879	1,568586045	0,00010925	0,000111465	0,60737559	0,605401525
OMXC20	1,9084E-08	1,94771E-08	1,5365959	1,547049561	8,4711E-05	8,58898E-05	0,6893038	0,680849797
OMXH25	4,234E-08	4,35193E-08	1,37610757	1,396047009	0,00013147	0,000133771	0,57914054	0,576024709

For all the four loss functions a smaller value is preferred. After the results in Table 3 which unequivocally suggested that assuming a student t distribution provides a better in-sample fit across all the datasets it is quite surprising that the MSE, R2LOG and MAD in Table 4 suggest the opposite; that assuming a normal error distribution unequivocally provides a better out-of-sample fit in terms of the MSE, R2LOG and MAD loss functions. The only loss function that suggests assuming a student t distribution is preferred compared to assuming a normal distribution, is the PSE. Quite a lot can be understood from these results. First of all, it is not necessarily the model with the best in-sample fit that provides the best out-of-sample fit. Moreover what is even more striking is that using different loss functions not only gives different ranks for the models but can even imply completely opposite conclusions. This is a quite major finding.

The next variable being studied is the conditional mean and how the model for the conditional mean affects the in-sample and out-of-sample performance. In Table 5 the average AIC for the compared volatility models with three different models for the conditional mean: Zero Mean, Constant Mean and AR(1) Mean. In each row the result based on the respective data series is presented.

Table 5: Average AIC for the compared volatility models with three different models for the conditional mean: Zero Mean, Constant Mean and AR(1) Mean. In each row the result based on the respective data series is presented.

	AIC		
	Zero Mean	Constant Mean	AR(1) Mean
OMXS30	-13147.05	-13151.95	-13152.73
OMXC20	-13686.07	-13692.85	-13692.40
OMXH25	-13487.46	-13495.35	-13496.56

In Table 5 it is evident that a constant mean process outperforms a zero mean process for all the three datasets in terms of the in-sample measure AIC. Furthermore, for the OMXS30 and OMXH25 the AR(1) mean process improves the in-sample fit even further. For OMXC20 the AR(1) mean process does not improve the in-sample fit compared to the simpler constant mean process but is only marginally worse. As well as a better in-sample fit the conditional mean must also provide a better out-of-sample fit to add any value to the forecasting framework.

In Table 6 the average MSE, R2LOG, MAD and PSE for the compared volatility models is presented with three different models for the conditional mean: Zero Mean, Constant Mean and AR(1) Mean. For all the four loss functions a smaller value is preferred and the smallest value for each loss function and data series is highlighted in green.

Table 6: Average MSE, R2LOG, MAD and PSE for the compared volatility models presented with three different models for the conditional mean: Zero Mean, Constant Mean and AR(1) Mean.

	MSE			R2LOG		
	Zero Mean	Constant Mean	AR(1) Mean	Zero Mean	Constant Mean	AR(1) Mean
OMXS30	3.3672E-08	3.35133E-08	3.3747E-08	1.57097953	1.559021659	1.56671607
OMXC20	1.923E-08	1.91972E-08	1.9415E-08	1.54577534	1.538384115	1.54130874
OMXH25	4.3052E-08	4.26177E-08	4.3118E-08	1.3907512	1.379129764	1.3883509
	MAD			PSE		
	Zero Mean	Constant Mean	AR(1) Mean	Zero Mean	Constant Mean	AR(1) Mean
OMXS30	0.00011054	0.000109972	0.00011056	0.60713397	0.606701664	0.60533004
OMXC20	8.5376E-05	8.51487E-05	8.5376E-05	0.6850329	0.683475906	0.6867216
OMXH25	0.00013293	0.000132048	0.00013288	0.57935484	0.576144115	0.57724891

From Table 6 it is clear that the constant mean model provides the best out-of-sample fit in terms of the MSE, R2LOG and MAD for the three data series. The only loss function suggesting another mean model is the PSE based on the OMXS30 data. Apart from that, the conclusion is that the constant mean model provides the best out-of-sample fit for all the three datasets. Again, it was not the model that provided the best in-sample fit that had the best out-of-sample performance. However, here the AR(1) Mean model was only marginally better in-sample than the constant mean model. An interesting point is that the zero mean model that was clearly outperformed in-sample is actually the second best model in terms of out-of-sample performance when measured by the MSE.

6.2 Conditional variance models

The previous subsection considered the impact of the conditional mean and the error distribution. Next the six different volatility models will be compared in terms of their in-sample fit and out-of-sample forecasting performance. The variable under study here is the different models for the conditional variance in each of the volatility models. Each of the volatility model was presented and explained in section 3. The main theme is whether the more complex models which are able to exhibit more of the stylized facts and characteristics of asset price volatility provide a better in-sample fit and/or out-of-sample-fit than the more parsimonious models. Based on previous discussion in terms of in-sample fit assuming a student t error distribution and an AR(1) conditional mean was the preferred choice. Thus, when comparing the different volatility models in terms of their in-sample fit that will be the scenario presented here.

In Table 7 the AIC for the different parametric volatility models is presented for each of the three data sets. The basic structure is assuming a student t error distribution and an AR(1) mean model. The smallest value in each column is highlighted in green and the largest value is highlighted in red.

Table 7: AIC for the different volatility models for each of the three data sets OMXS30, OMXC20 and OMXH25. The smallest value in each column is highlighted in green and the largest value is highlighted in red. The basic structure is assuming a student t error distribution and an AR(1) mean model.

	OMXS30	**OMXC20**	**OMXH25**
MA10			
Exponential MA			
ARCH(3)	-12989	-13596	-13327
GARCH(1,1)	-13186	-13745	-13529
GARCH(1,2)	-13180	-13743	-13540
GARCH(2,1)	-13186	-13742	-13516
EGARCH(1,1)	-13261	-13776	-13585
GJR(1,1)	-13251	-13763	-13578

A smaller value of AIC is preferred and across the three data sets it is evident that the ARCH(3) has the worst in-sample fit and the EGARCH(1,1) has the best in-sample fit. The in-sample fit of the different volatility models for each data set assuming a normal error distribution and a different mean model is available in appendix D and the same pattern is evident there. The main conclusion here is that yes, the more complex models do provide a better in-sample fit than the more parsimonious models. The GARCH(1,1) clearly outperforms ARCH(3) and the EGARCH(1,1) and the GJR(1,1) clearly outperforms the GARCH(1,1). This is quite expected especially since, for example the ARCH model is nested in the GARCH model. So if the GARCH(1,1) model did not provide a better in-sample fit there would be no point in setting the extra parameter to anything other than zero which would then reduce the GARCH model to the simple ARCH model. Another interesting detail in Table 7 is that the higher order GARCH does not necessarily provide a better fit than the GARCH(1,1) which is in line with previous studies and briefly discussed in section 3.

As emphasized throughout this paper in addition to the in-sample fit it is vital to look at the out-of-sample forecasting performance. The reasoning for this comes from the objective of the analysis which is to forecast volatility, not to model its behavior in the past. Based on the result comparing the error distribution and the conditional mean in terms of out-of-sample fit assuming a normal error distribution and a constant conditional mean was the preferred choice. Thus, when comparing the different volatility models in terms of their out-of-sample performance that will be the setting presented here.

In Table 8 MSE, R2LOG, MAD and PSE for the compared volatility models is presented based on OMXS30. For all the four loss functions a smaller value is preferred and the smallest value for each loss function is highlighted in green and the largest value is highlighted in red.

Table 8: MSE, R2LOG, MAD and PSE for the compared volatility models based on OMXS30. The smallest value for each loss function is highlighted in green and the largest value is highlighted in red. The basic structure is assuming a normal error distribution and a constant mean model.

OMXS30				
	MSE	**R2LOG**	**MAD**	**PSE**
MA10	3.6637E-08	1.26282	0.00010369	0.90665
Exponential MA	3.3486E-08	1.34067	0.00010289	0.59486
ARCH(3)	3.7806E-08	2.29117	0.00013449	0.68066
GARCH(1,1)	3.1706E-08	1.53339	0.00010593	0.55379
GARCH(1,2)	3.2369E-08	1.55659	0.00010744	0.57186
GARCH(2,1)	3.1701E-08	1.53315	0.00010592	0.55383
EGARCH(1,1)	2.6092E-08	1.46512	0.00010277	0.48574
GJR(1,1)	3.3084E-08	1.49757	0.00011073	0.50954

The exhibit in Table 8 is quite inconclusive. There is not a single volatility model that is preferred based on all the four loss functions. When focusing on the MSE however, the result is that the best model is the EGARCH and the ARCH(3) is the worst model. Furthermore the GARCH model is preferred to the ARCH model and the EGARCH is preferred to the GARCH which is in line with the finding in terms of the in-sample fit. This suggests that the more complex models do provide a better out-of-sample fit than the more parsimonious models in line with the findings for the in-sample fit. For this data set, the same conclusion is implied by the more robust measure MAD and a similar conclusion is also drawn when comparing the PSE with the exception of MA10 now being the worst model. So using MSE, MAD and PSE as loss functions the more complex models do provide a better out-of-sample fit than the more parsimonious models. An important finding is however how the ranking differs when using the R2LOG loss function. Based on this loss function ARCH is still considered the worst model but more strikingly the blunt non-parametric Moving Average with a 10 day window is the overall preferred model.

Table 9 presents the MSE, R2LOG, MAD and PSE for the compared volatility models based on OMXC20. The MSE and R2LOG gives the same ranking as was the case based on OMXS30 but the MAD and the PSE loss functions give significantly different ranking than when based on the OMXS30.

Table 9: MSE, R2LOG, MAD and PSE for the compared volatility models based on OMXC20. The smallest value for each loss function is highlighted in green and the largest value is highlighted in red. The basic structure is assuming a normal error distribution and a constant mean model.

OMXC20				
	MSE	R2LOG	MAD	PSE
MA10	2.0758E-08	1.32376	8.1899E-05	0.98464
Exponential MA	1.8772E-08	1.38078	7.9222E-05	0.84234
ARCH(3)	2.2695E-08	1.95662	0.00010074	0.55190
GARCH(1,1)	1.8463E-08	1.53928	8.3542E-05	0.64894
GARCH(1,2)	1.847E-08	1.53922	8.3558E-05	0.64924
GARCH(2,1)	1.8473E-08	1.53836	8.352E-05	0.65082
EGARCH(1,1)	1.6393E-08	1.46709	8.0701E-05	0.60166
GJR(1,1)	1.8411E-08	1.52869	8.4132E-05	0.56151

Table 9 makes the findings in Table 8 even more inconclusive. Now the three loss functions R2Log, MAD and PSE all suggest that the best model is the non-parametric MA10, Exponential MA and based on PSE the ARCH model is the overall preferred model. It is quite a contrast that one loss function suggests that a particular model is the worst and another loss function can suggest that same model to be the best. This illuminates the importance of choosing an adequate loss function for the intended purpose of the forecast.

Finally in Table 10 the MSE, R2LOG, MAD and PSE for the compared volatility models based on OMXH25 is presented. Table 10 presents the most coherent ranking across the different loss functions yet. For all the loss functions the more complex models do provide a better out-of-sample fit than the more parsimonious models. The GARCH is preferred over the ARCH and the GJR and the EGARCH is preferred over the GARCH. Furthermore the worst model suggested by all of the four loss functions is either the non-parametric MA10 and the Exponential MA or the ARCH(3).

Table 10: MSE, R2LOG, MAD and PSE for the compared volatility models based on OMXH25. The smallest value for each loss function is highlighted in green and the largest value is highlighted in red. The basic structure is assuming a normal error distribution and a constant mean model.

OMXH25				
	MSE	R2LOG	MAD	PSE
MA10	4.8103E-08	1.29239	0.00013331	0.85867
Exponential MA	4.5607E-08	1.34304	0.00013347	0.56969
ARCH(3)	4.7754E-08	1.66242	0.00014415	0.67791
GARCH(1,1)	4.1802E-08	1.38439	0.00013168	0.52741
GARCH(1,2)	4.0251E-08	1.35188	0.00012774	0.54164
GARCH(2,1)	4.1746E-08	1.38313	0.00013155	0.52757
EGARCH(1,1)	3.3664E-08	1.30790	0.00012231	0.44834
GJR(1,1)	3.8114E-08	1.25662	0.00012525	0.46686

The out-of-sample performance of the compared volatility models in terms of the different loss functions based on the three data sets presents a bit of a conundrum. It is far from evident which of the specific conditional volatility models that outperforms the other. First, the ranking of models based on a specific loss function differs for the three data sets. Secondly, for the OMXC20 and OMXS30 the best and worst model respectively depends heavily on which loss function is used. So to answer which model has the best out-of-sample performance one must first consider the specific data set used and then which loss function to use as the criteria. Moreover it is not necessarily the model with the best in-sample fit that produces the best out-of-sample forecast. Since the out-of-sample forecast performance is so vital to the objective of the analysis one can question whether the in-sample fit should even be used at all to support the choice of a specific volatility model.

The out-of-sample fit of the different volatility models for each data set assuming a student t error distribution and a different mean model is available in appendix D and exhibits similar inconsistency in ranking the volatility models.

7 Summary and Conclusions

In this section a short summary of the result is presented and the main conclusions are stated.

In the Result section there were three main themes being studied, first the basic structure of the modeling framework was investigated, with respect to the error distribution and the conditional mean to gain insight in how the assumed error distribution and different models for the conditional mean and impacts the in-sample and out-of-sample fit regardless of which specific conditional variance model is used. The second theme was whether the more complex models which are able to exhibit more of the stylized facts and characteristics of asset price volatility provide a better in-sample fit and/or out-of-sample-fit than the more parsimonious models. The third and final theme was whether the model with the best in-sample fit also produces the best out-of-sample volatility forecast.

The impact of the error distribution and the conditional mean was quite clear when looking at the in-sample fit. It was clearly evident that assuming a student's t-distribution provides a better in-sample fit than assuming a normal distribution across all the datasets both in terms of the AIC and the BIC. The result is unequivocal and is expected considering the q-q plots combined with the jarque-bera test in section 2 which implied that the empirical distribution of the return series displayed significantly heavier tails than the normal distribution. Considering the conditional mean it was evident that a constant mean process outperforms a zero mean process for all the three datasets in terms of the in-sample measure AIC. Furthermore, for the OMXS30 and OMXH25 the AR(1) mean process improves the in-sample fit even further. For OMXC20 the AR(1) mean process does not improve the in-sample fit compared to the simpler constant mean process but is only marginally worse. So in terms of the in-sample fit assuming a student t distribution and modeling the mean improves the forecast model. However, when looking at the out-of-sample performance the result was quite strikingly different. With the in-sample performance in mind it was quite surprising that the out-of-sample performance measures MSE, R2LOG and MAD suggested that assuming a normal error distribution unequivocally provides a better out-of-sample fit. The only loss function that suggested assuming a student t distribution is preferred compared to assuming a normal distribution is the PSE. Looking at the out-of-sample performance of the conditional mean model it was clear that the constant mean model generally provided the best out-of-sample fit in terms of all the four loss functions. The conclusion is that the constant mean model provides the best out-of-sample fit for all the three datasets. Again it was not the model that provided the best in-sample fit that had the best out-of-sample performance. However, here the AR(1) Mean model was only marginally better in-sample than the constant mean model. An interesting point though is that the zero mean model that was clearly outperformed in-sample is actually the second best model in terms of out-of-sample performance when measured by the MSE.

Quite a lot can be understood from this. First of all, it is not necessarily the model with the best in-sample fit that provides the best out-of-sample fit. Moreover, considering the case of the PSE comparing the error distributions, what is even more striking is that using different loss functions not only gives different ranks for the models but can even imply completely opposite conclusions. These are quite major findings and give rise to yet more questions. How is it possible that the model that is preferred based on the in-sample performance at the same time can be the worst performing model considering the out-of-sample metrics? Maybe even more interesting is what would impose this: what does it mean that a model is worse in terms of the out-of-sample fit compared to the in-sample fit? One possible reason is the fact that the dynamics of the volatility might have changed during the quite long time horizon of the data. Not only does the volatility change over time but also the dynamics of the volatility, how the volatility evolves over time, might have shifted. The dynamics of volatility is not stationary and especially not over a time horizon from 2002 until 2014, just over 13 years. In addition, during that time period the world has witnessed one of the greatest financial crises of all times which quite likely have changed the dynamics of the market. Possible remedies are to reduce both the in-sample time horizon and maybe even more importantly reducing the out-of-sample time horizon. Another reason for the discrepancy between the in-sample and the out-of-sample performance might lie in the nature of model fitting. That a model that is backtested to perfection and has an extremely good in-sample fit can become sluggish and reacts slowly to changes in the volatility and sudden chocks while a model with a not so good in-sample fit might be more flexible and is able to more quickly allow for changes in volatility dynamics and chocks. There might also be a trade-off between fitting the model to the in-sample data and the models alertness to new inputs.

Next the six different volatility models were compared in terms of their in-sample fit and out-of-sample forecasting performance. The main theme is whether the more complex models which are able to exhibit more of the stylized facts and characteristics of asset price volatility provide a better in-sample fit and/or out-of-sample-fit than the more parsimonious models. Again, the in-sample results was very coherent and it was evident that the ARCH(3) has the worst in-sample fit and the EGARCH(1,1) has the best in-sample fit. The main conclusion here is that yes, the more complex models do provide a better in-sample fit than the more parsimonious models. The GARCH(1,1) clearly outperforms ARCH(3) and the EGARCH(1,1) and the GJR(1,1) clearly outperforms the GARCH(1,1). This is quite expected especially since, for example the ARCH model is nested in the GARCH model. If the GARCH(1,1) model did not provide a better in-sample fit there would be no point in setting the extra parameter to anything other than zero which would then reduce the GARCH model to the simple ARCH model. Another interesting result was that the higher order GARCH does not necessarily provide a better fit than the GARCH(1,1) which is in line with previous studies and is briefly discussed in section 3.

In terms of the out-of-sample forecasting performance the results was very inconclusive. There is not a single volatility model that is preferred based on all the four loss functions. An important finding is however not just that the ranking differs when using different loss functions, but how dramatically it can differ. It is quite a contrast that one loss function suggests that a particular model is the worst and another loss function can suggest that same model to be the best. This highlights the importance of choosing an adequate loss function for the intended purpose of the forecast.

The out-of-sample performance of the compared volatility models in terms of the different loss functions based on the three data sets thus suggests a bit of a challenge. It is far from evident which of the specific conditional volatility models that outperforms the other. First, the ranking of models based on a specific loss function differs for the three data sets. Secondly, for the OMXC20 and OMXS30 the best and worst model respectively depends heavily on which loss function is used. To answer which model has the best out-of-sample performance one must first consider the specific data set used and then which loss function to use as the criteria. Moreover, it is not necessarily the model with the best in-sample fit that produces the best out-of-sample forecast. Since the out-of-sample forecast performance is so vital to the objective of the analysis, one can question whether the in-sample fit should even be used at all to support the choice of a specific volatility model. Obviously one does only have information up until the current day and not in the future. However, as done in this paper, one can divide the historical data into an in-sample period and an out-of-sample period. The result of this paper then suggests that one should put more emphasis on the out-of-sample performance due to the nature of the analysis: to model volatility in the future. This can be argued by the highly inconsistency of the in-sample performance compared to the out-of-sample performance. Choosing the model with the best in-sample fit would not necessarily be the best choice of model in terms of the out-of-sample performance.

8 Suggestions for further research

To further increase the understanding of volatility forecasting and its performance one could consider different forecasting schemes, look at different forecast horizons, consider more complex conditional mean models, investigate the use of other error distributions and extend the number of conditional variance models being studied. Furthermore, another proxy for the latent volatility could be used and one could focus on other asset classes than equity indices such as commodities or foreign exchange. However, based on the result in this paper the next step forward that is most likely to be the most rewarding is to investigate which loss functions that provide the most appropriate out-of-sample performance measure for the intended purpose of the forecast. The objective of the forecast, the intended purpose, must be considered if the result is to be of any use for portfolio managers, risk managers and other practitioners. It is not until an appropriate out-of-sample metric has been established that one can truly find the models with superior predictive ability.

9 Bibliography

Alizadeh, S., Brandt, M., Diebold, F. Range-based estimation of stochastic volatility models. Journal of finance, 57: 1047 – 1091, 2002.

Andersen, T. G., Bollerslev, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39:885–907, 1998.

Andersen, T.G., Bollerslev, T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. Journal of Finance, 52: 975 – 1005, 1997.

Black, F., Scholes, M. The pricing of options and corporate liabilities. Journal of political Economy, 81: 637 – 654, 1973.

BLACK, F., Studies in stock price volatility changes. Proceedings of the 1976 business meeting of the business and economics section, American Statistical Association, 177 – 181, 1976.

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

Bollerslev, T., Chou, R.Y., Kroner, K.F. ARCH modeling in finance: a review of theory and empirical evidence. Journal of Econometrics 52:5 – 59, 1992.

BOLLERSLEV, T., ENGLE, R.F., NELSON, D. ARCH models. *Handbook of Econometrics*, ed. by Engle, R.F and McFadden D.L, vol. IV, pp. 2961–3038, 1994.

Bollerslev, T., Zhou, H. Volatility Puzzles: A Simple Framework for Gauging Return - Volatility Regressions. *Journal of Econometrics*, 2005.

Brandt, M., Diebold, F. A no-arbitrage approach to range-based estimation of return covariances and correlations. Journal of Business, 79: 61 – 74, 2006.

Brandt, M.W., Jones, C.S. Volatility Forecasting With Range-Based EGARCH Models. American Statistical Association, Journal of Business & Economic Statistics, 2006.

Campbell, J. Y., Lo, A. W., MacKinlay, A.C. The econometrics of financial Markets. Princeton University Press: New Jersey, 1997.

Chou, R. Y., Chou, H., Liu, N. Range Volatility Models and Their Applications in Finance. in Handbook of Quantitative Finance and Risk Management, ed. by C.-F. Lee and J. Lee, Springer, chap. 83, 2010.

Cumby, R., Figlewski, S., Hasbrouck J. Forecasting Volatility and Correlations with EGARCH Models. Journal of Derivatives, 51 – 63, 1993.

Dacorogna, M.M., Gencay, R., Muller, U., Olsen, R.B., Olsen, O.V. An introduction to high frequency finance, Academic Press, New York, 2001.

Diebold, F.X., Lopez, J.A. Forecast Evaluation and Combination. Handbook of Statistics, ed. By Maddala, G.s., Rao, C.R. vol 14, pp 241 – 268, 1996.

Engle, R. F. Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom ination. Econometrica, 50:987–1007, 1982.

Engle, R.F., Patton, A.J. What good is a volatility model?. Quantitative Finance, 1: 237 – 245, 2001.

Figlewski, S. *Forecasting volatility*, Financial Markets, Institutions, and Instruments 6, Blackwell Publishers: Boston, 1997.

Garman, M. B., Klass, M.J. On the Estimation of Security Price Volatilities from Historical Data. The Journal of Business, 53:67–78, 1980.

Glosten, L. R., Jagannathan, R., Runkle, D. E. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, 48:1779–1801, 1993.

Hansen, P. R., Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH(1,1)?. Journal of Applied Econometrics, 20:873–889, 2005.

Jorion, P. Predicting Volatility in the Foreign Exchange Market. Journal of Finance, 50:507 – 528, 1995.

Jorion, P. Risk and Turnover in the Foreign Exchange Market. In Frankel, J.A., Galli, G., Gioavnni, A. eds The Microstructure of Foreign Exchange Markets, 1996.

Merton, R.C. Lifetime portfolio selection under uncertainty: The continuous-time case. Review of Economics and Statistics, 51: 247 – 257, 1969.

Nelson, D. B. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59:347–370, 1991.

Parkinson, M. The Extreme Value Method for Estimating the Variance of the Rate of Return. The Journal of Business, 53:61–65, 1980.

Shamiri, A., Isa, Z. Modelling and forecasting volatility of Malaysian stock markets. Journal of Mathematics and Statistics, 3: 234 – 240, 2009.

Tsay, R. Analysis of Financial Time Series. John Wiley & Sons, 2002.

Tsay, R. Out-of-Sample Forecasts. Lecture Note of Bus 41402 Univariate Volatility Models, 2008.

Appendix A: Significance tests

Ljung-Box Q-test

The Ljung-Box Q test is a test for checking whether a specific time series are observed values from iid random variables. As opposed to the sample autocorrelation function, which can be used to check whether each sample autocorrelation $\hat{p}(j)$ falls inside the bounds $\pm\frac{1.96}{\sqrt{n}}$, the Ljung-Box Q test is a single statistic

$$Q = n(n+2)\sum_{j=1}^{h}\frac{\hat{p}^2(j)}{n-j}.$$

If *n* is large and $Y_1, \dots, Y_n$ is an iid sequence, the distribution of Q can be approximated as the sum of squares of the independent N(0,1) random variables, $\sqrt{n}\hat{p}(j), j = 1, \dots, h$. That is the distribution of Q can be approximated as chi-squared with *h* degrees of freedom. A too large value of Q would suggest that the sample autocorrelations are too high for the data to be observations from an iid sequence. The iid hypothesis is rejected with $\propto$ significance level if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the chi-squared distribution with *h* degrees of freedom.

Engle's ARCH test

Engle´s ARCH test is very similar to the Ljung-Box Q-test but tests whether the squared data is an iid sequence. The test statistics is the same as for the Ljung-Box Q-test but with the data autocorrelation replaced with the autocorrelation for the squared data, $\hat{p}_{ww}$.

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{p}^2_{ww}(k)}{n-k}.$$

The iid hypothesis is rejected on the same basis that is with $\propto$ significance level if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the chi-squared distribution with *h* degrees of freedom.. This test is performed on data to see whether the squared data is serially correlated which justifies modeling the conditional volatility using a GARCH type model.

Jarque-Bera Test

The Jarque-Bera test is a goodness-of-fit test which examines if the sample data have kurtosis and skewness similar to a normal distribution. The test statistics is JB which is defined by

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

where *n* is the sample size, S is the sample skewness and K is the sample kurtosis. If the sample data comes from a normal distribution JB should, asymptotically, have a chi-squared distribution with two degrees of freedom. The null hypothesis is that the sample data have a skewness of zero and an excess kurtosis of 3 which is what the normal distribution has. The Jarque-Bera test for the descriptive statistics was determined with Matlab's™ built in function jbtest(x). Jbtest(x) test if the sample distribution comes from a normal distribution at the 5% significance level.

Appendix B: Descriptive statistics

OMXC20

For OMXC20 there were 3,079 daily data points during the entire period 2002-01-02 until 2014-04-15. The in-sample period from 2002-01-02 until 2010-12-30 consisted of 2,256 daily data points and the out-of-sample period from 2011-01-03 until 2014-04-15 consisted of 823 daily data points. In Figure B1 the daily closing price during the entire period is plotted. The in-sample period is plotted in blue and the horizontal red line indicates where the out-of-sample period starts which is then indicated by the black line.

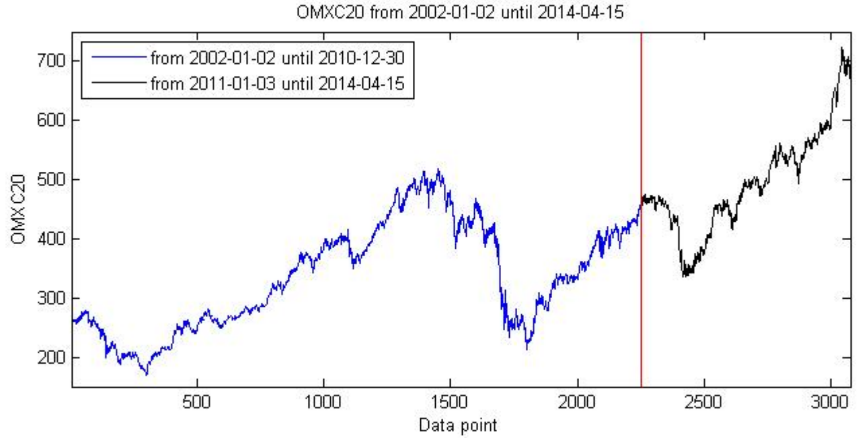

Figure B 1: OMXC20 daily index level from 2002-01-12 until 2014-04-15. In total there are 3,079 observations. The blue line represents the index level during the in-sample period and the vertical red line indicates where the out-of-sample period starts which is then represented by the black line.

The left plot of Figure B2 shows the daily return for the in-sample period. The right plot in Figure B2 shows the Sample Autocorrelation Function for the daily returns of lags 0 to 20. The Sample From the Sample Autocorrelation Function plot it is not completely clear based on ocular inspection whether the data is serially correlated or not, even though it has minor significant serial correlation at lag 3. However the Ljung-Box Q-test null hypothesis is rejected for lags 5 and 10 at a 5% significance level as for the OMXS30 however the null hypothesis is not rejected for 15 lags. This still suggests that a conditional mean model is required for this return series as well.

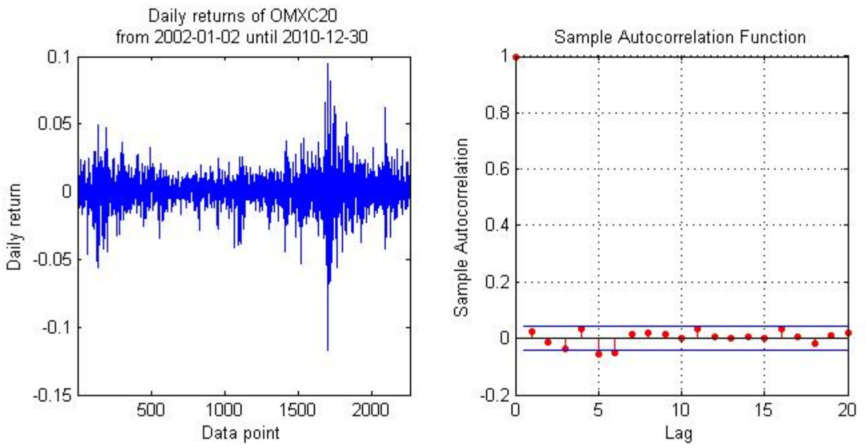

Figure B 2: Left plot: Daily returns of OMXC20 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,255 observations. Right plot: Sample Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

The left plot of Figure B3 shows the daily squared returns for the in-sample period which exhibits volatility clustering as was the case for OMXS30. The right plot of Figure B3 shows the Sample Partial Autocorrelation Function which, by ocular inspection, clearly shows significant autocorrelation. Engle’s ARCH test rejects the null hypothesis for lags 5 and 12 at a 5% significance level and thus confirms that the squared returns are serially correlated.

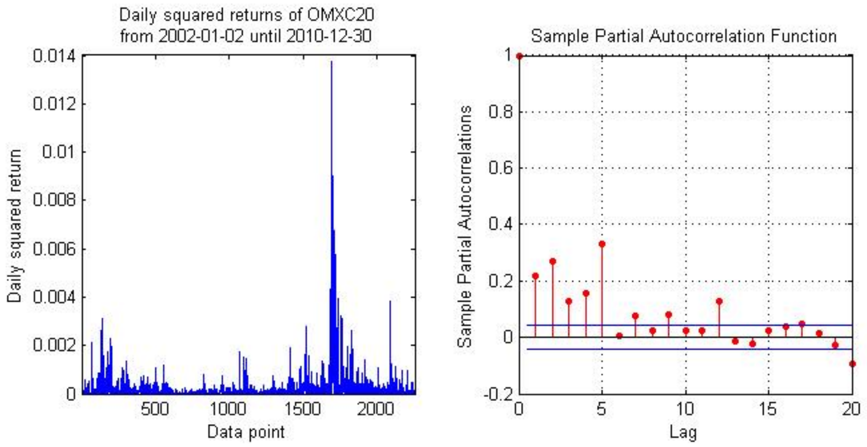

Figure B 3: Left plot: Daily squared returns of OMXC20 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,255 observations. Right plot: Sample Partial Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

In Figure B4 two q-q plots are presented. The left plot is a q-q plot of the empirical distribution of the daily returns (y-axis) against the best fitted normal distribution (x-axis). The plot to the right is a q-q plot of the empirical distribution (y-axis) against the best fitted t location-scale distribution (x-axis).

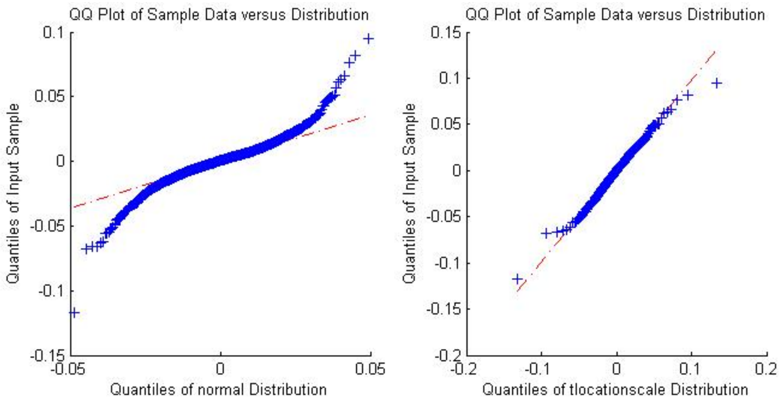

Figure B 4: Left plot: q-q plot of the empirical distribution (y-axis) against best fitted normal distribution (x-axis). Right plot: q-q plot of the empirical distribution (y-axis) against best fitted t location-scale distribution (x-axis).

The two q-q plots basically have the same shape as the q-q plots for OMXS30 and thus the same conclusions can be made. The empirical distribution of the daily returns exhibit significantly heavier tails than the normal distribution but is sufficiently close to the t location-scale distribution. In Table B1 some summary statistics of the return series is presented.

Table B 1: Summary statistics for OMXC20 daily returns in the in-sample period.

	Sample size	Mean	Variance	Skewness	Excess kurtosis	Jarque-Bera
OMXC20	2,255	0.00023297	0.00019459	-0.23732466	9.453134027	1

The Jargue-Bera test rejects the null hypothesis that sample distribution comes from a normal distribution at the 5% significance level which was in line with expectations from the ocular inspection of the q-q plots in Figure 8 which implied that the empirical distribution of the daily returns exhibit significantly heavier tails than the normal distribution.

OMXH25

For OMXH25 there were 3,089 daily data points during the entire period 2002-01-02 until 2014-04-15. The in-sample period from 2002-01-02 until 2010-12-30 consisted of 2,263 daily points and the out-of-sample period from 2011-01-03 until 2014-04-15 consisted of 826 daily points. In Figure B5 the daily closing price during the entire period is plotted. The in-sample period is plotted in blue and the horizontal red line indicates where the out-of-sample period starts which is then indicated by the black line.

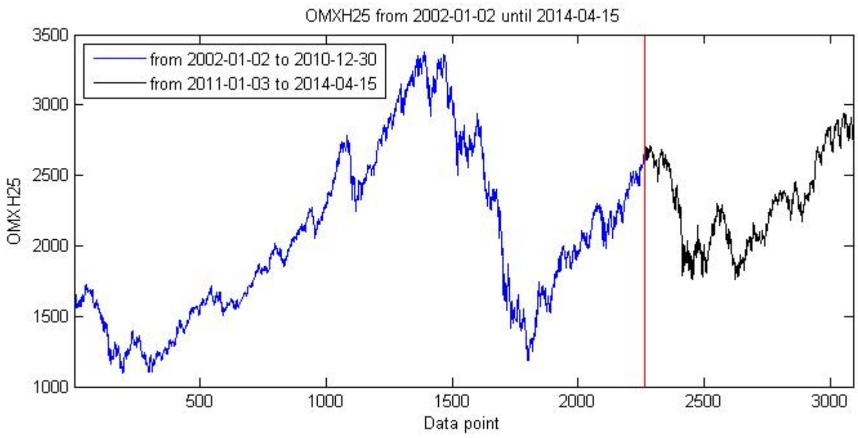

Figure B 5: OMXH25 daily index level from 2002-01-12 until 2014-04-15. In total there are 3,089 observations. The blue line represents the index level during the in-sample period and the vertical red line indicates where the out-of-sample period starts which is then represented by the black line.

In Figure B6 the daily return for the in-sample period and the respective Sample Autocorrelation Function of lags 0 to 20 is presented. The Ljung-Box Q-test null hypothesis is rejected for lags 5 and 10 at a 5% significance level however the null hypothesis is not rejected for 15 lags. This still suggests that a conditional mean model is required for this return series as was required for the two former series.

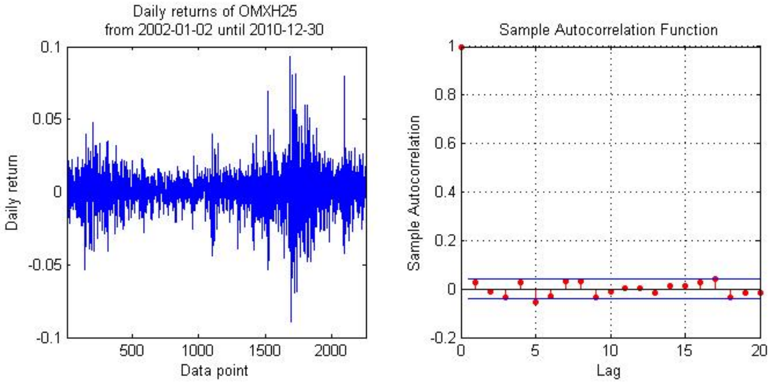

Figure B 6: Left plot: Daily returns of OMXH25 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,262 observations. Right plot: Sample Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

Figure B7 shows the daily squared returns for the in-sample period the respective Sample Partial Autocorrelation Function which, by ocular inspection, clearly shows significant autocorrelation. Engle's ARCH test rejects the null hypothesis for lags 7 and 13 at a 5% significance level and thus confirms that the squared returns are serially correlated.

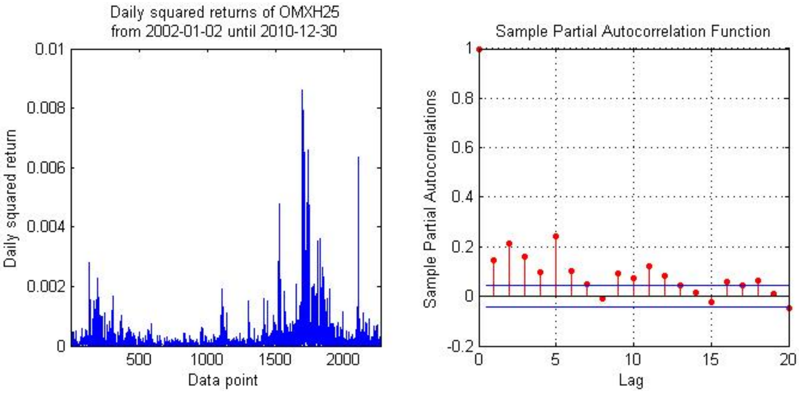

Figure B 7: Left plot: Daily squared returns of OMXH25 from 2002-01-02 until 2010-12-30, which is the in-sample period consisting of 2,262 observations. Right plot: Sample Partial Autocorrelation Function for the daily returns of lags 0 to 20 and the 5% confidence level in blue.

In Figure B8 two q-q plots are presented. The left plot is a q-q plot of the empirical distribution of the daily returns (y-axis) against the best fitted normal distribution (x-axis). The plot to the right is a q-q plot of the empirical distribution (y-axis) against the best fitted t location-scale distribution (x-axis).

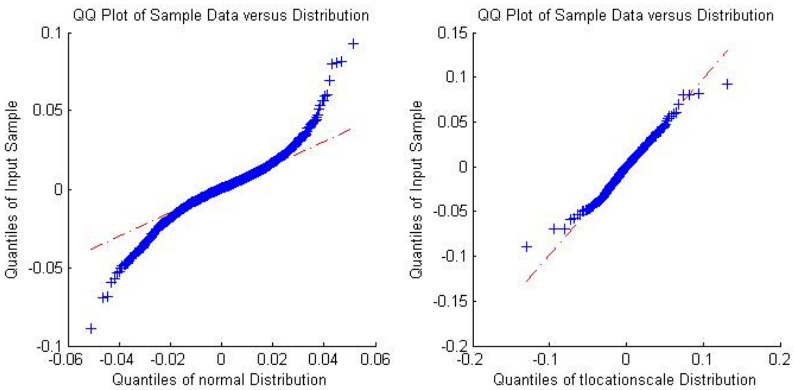

Figure B 8: Left plot: q-q plot of the empirical distribution (y-axis) against best fitted normal distribution (x-axis). Right plot: q-q plot of the empirical distribution (y-axis) against best fitted t location-scale distribution (x-axis).

The two q-q plots imply that the empirical distribution of the daily returns exhibit significantly heavier tails than the normal distribution which motivates the use of a more heavy-tailed distribution such as the t location-scale distribution. In Table B2 some summary statistics of the return series is presented.

Table B 2: Summary statistics for OMXH25 daily returns in the in-sample period.

	Sample size	**Mean**	**Variance**	**Skewness**	**Excess kurtosis**	**Jarque-Bera**
OMXH25	2,262	0.00021855	0.00021217	0.00812616	7.242869425	1

The Jargue-Bera test rejects the null hypothesis at the 5% significance level which implies, in line with the two other datasets, that the empirical distribution of the daily returns has significantly heavier tails than the normal distribution.

Appendix C: Moving Average

In Figure C1 9 different simulated processes and their respective Moving Average with a 100 step window is plotted. The simulated process is plotted in blue and the Moving Average is plotted in green. The ability to model changes in time can be seen by comparing the simulated process with its respective Moving Average. It is evident that when the simulated process is constantly increasing the moving average is underestimating the process and when simulated process is constantly decreasing the moving average is overestimating the simulated process. Furthermore the moving average has a similar pattern as the simulated process but is lagging.

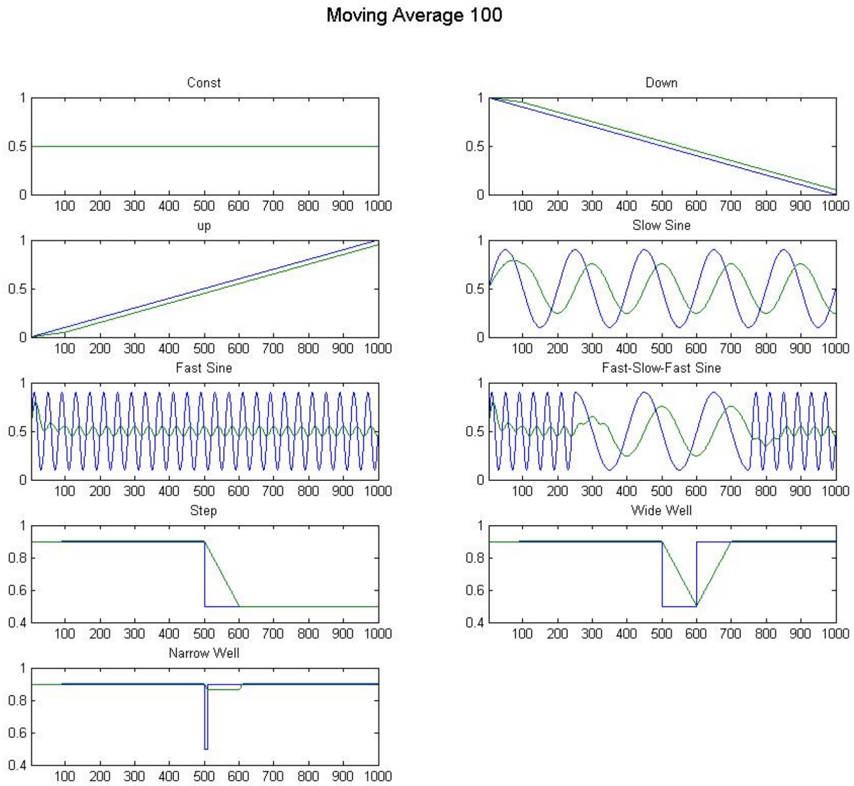

Figure C 1: Shows the Moving Average with a 100 step window for 9 different simulated processes. The simulated process is plotted in blue and the Moving Average is plotted in green.

In Figure C2 the same simulated processes as in Figure C1 are plotted but now with their respective Moving Average with a 50 step window. The Moving Average with a 50 step window shows the same characteristic as was the case with a 100 step window, although there are a few differences. With a shorter window the moving average is not underestimating as much when the simulated process is constantly increasing and is not overestimating as much when the simulated process is constantly decreasing as was the case with a 100 step window. Furthermore the moving average is still lagging but not as much and reacts quicker to changes in the simulated process than was the case with a 100 step window.

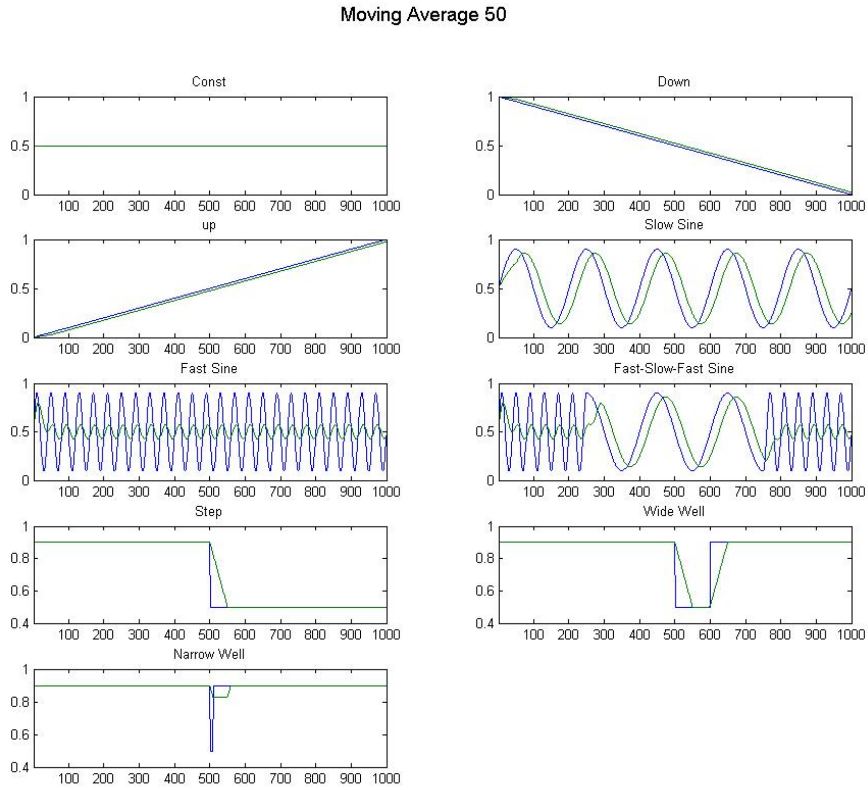

Figure C 2: Shows the Moving Average with a 50 step window for 9 different simulated processes. The simulated process is plotted in blue and the Moving Average is plotted in green.

Appendix D: Empirical results

Table D 1: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXS30 as the underlying process and assuming a normal error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Normal distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13445	0,30686	0,05912	0,40483	3,67027E-08	1,26849	0,000103818	0,91504
Exponential MA			0,15703	0,27613	0,07194	0,44368	3,35312E-08	1,34524	0,000103	0,59675
ARCH(3)	-12846	-12823	0,10373	0,29609	0,05061	0,30022	3,71517E-08	2,32476	0,000134724	0,67630
GARCH(1,1)	-13144	-13127	0,15747	0,30335	0,07240	0,45067	3,17774E-08	1,53891	0,0001061	0,55561
GARCH(1,2)	-13148	-13125	0,16557	0,29753	0,07307	0,44035	3,24724E-08	1,56204	0,000107633	0,57433
GARCH(2,1)	-13144	-13127	0,15750	0,30319	0,07241	0,45067	3,17569E-08	1,53797	0,000106036	0,55576
EGARCH(1,1)	-13237	-13214	0,18180	0,42296	0,09510	0,51391	2,61046E-08	1,46731	0,00010287	0,48594
GJR(1,1)	-13225	-13207	0,17658	0,37714	0,09596	0,50595	3,31735E-08	1,49873	0,000110804	0,50989
Average	-13124	-13104	0,15427	0,32291	0,07383	0,43878	3,28338E-08	1,56793	0,000109373	0,60870
	Normal distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13432	0,30845	0,05808	0,40700	3,66374E-08	1,26282	0,000103694	0,90665
Exponential MA			0,15684	0,27728	0,07300	0,44533	3,34856E-08	1,34067	0,000102887	0,59486
ARCH(3)	-12859	-12836	0,10315	0,29841	0,05208	0,30267	3,78058E-08	2,29117	0,000134492	0,68066
GARCH(1,1)	-13148	-13131	0,15717	0,30555	0,07328	0,45258	3,17059E-08	1,53339	0,000105932	0,55379
GARCH(1,2)	-13152	-13129	0,16505	0,29992	0,07334	0,44224	3,23691E-08	1,55659	0,000107438	0,57186
GARCH(2,1)	-13148	-13131	0,15718	0,30551	0,07328	0,45258	3,17012E-08	1,53315	0,000105916	0,55383
EGARCH(1,1)	-13238	-13215	0,18077	0,42278	0,09497	0,51400	2,60916E-08	1,46512	0,000102773	0,48574
GJR(1,1)	-13226	-13209	0,17582	0,37768	0,09661	0,50627	3,30841E-08	1,49757	0,000110729	0,50954
Average	-13129	-13109	0,15379	0,32445	0,07433	0,44033	3,28601E-08	1,56006	0,000109233	0,60712
	Normal distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13257	0,31035	0,06419	0,40663	3,64862E-08	1,26240	0,000103552	0,91332
Exponential MA			0,15561	0,27853	0,08135	0,44506	3,34083E-08	1,34006	0,000102819	0,59222
ARCH(3)	-12860	-12837	0,10245	0,29835	0,05830	0,30178	3,78107E-08	2,29078	0,000134365	0,67825
GARCH(1,1)	-13150	-13132	0,15583	0,30675	0,08166	0,45237	3,15949E-08	1,53212	0,000105778	0,55176
GARCH(1,2)	-13154	-13131	0,16376	0,30160	0,08080	0,44252	3,22211E-08	1,55502	0,000107266	0,56858
GARCH(2,1)	-13150	-13132	0,15583	0,30677	0,08166	0,45236	3,1619E-08	1,53251	0,000105829	0,55168
EGARCH(1,1)	-13240	-13217	0,17934	0,42325	0,10644	0,51289	2,61353E-08	1,46641	0,000102896	0,48590
GJR(1,1)	-13227	-13210	0,17478	0,37876	0,10807	0,50567	3,30379E-08	1,49819	0,000110657	0,50876
Average	-13130	-13110	0,15252	0,32555	0,08281	0,43991	3,27892E-08	1,55969	0,000109145	0,60631

Table D 2: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXS30 as the underlying process and assuming a student t error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Student distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13445	0,30686	0,05912	0,40483	3,67027E-08	1,26849	0,000103818	0,91504
Exponential MA			0,15703	0,27613	0,07194	0,44368	3,35312E-08	1,34524	0,000103	0,59675
ARCH(3)	-12983	-12955	0,10919	0,27827	0,05099	0,29616	4,30615E-08	2,33255	0,000140099	0,67003
GARCH(1,1)	-13180	-13157	0,15828	0,29406	0,07287	0,45016	3,29753E-08	1,54057	0,000107945	0,55553
GARCH(1,2)	-13172	-13144	0,15896	0,31659	0,07095	0,44412	3,60593E-08	1,64396	0,000116052	0,55210
GARCH(2,1)	-13181	-13158	0,15829	0,29364	0,07285	0,45007	3,26424E-08	1,51919	0,000106786	0,55805
EGARCH(1,1)	-13257	-13229	0,18346	0,42023	0,09528	0,51313	2,67821E-08	1,45802	0,000103926	0,48613
GJR(1,1)	-13247	-13225	0,17792	0,37274	0,09591	0,50567	3,43202E-08	1,48421	0,000112012	0,51088
Average	-13170	-13144	0,15470	0,31981	0,07374	0,43848	3,45093E-08	1,57403	0,000111705	0,60557
	Student distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13432	0,30845	0,05808	0,40700	3,66374E-08	1,26282	0,000103694	0,90665
Exponential MA			0,15684	0,27728	0,07300	0,44533	3,34856E-08	1,34067	0,000102887	0,59486
ARCH(3)	-12987	-12959	0,10868	0,28057	0,05233	0,29869	4,31735E-08	2,32335	0,000139995	0,66978
GARCH(1,1)	-13184	-13161	0,15802	0,29703	0,07374	0,45211	3,30459E-08	1,52915	0,000107852	0,55349
GARCH(1,2)	-13186	-13157	0,16595	0,29270	0,07393	0,44302	3,33922E-08	1,55704	0,000108904	0,57301
GARCH(2,1)	-13185	-13162	0,15809	0,29530	0,07385	0,45194	3,2623E-08	1,51348	0,000106693	0,55614
EGARCH(1,1)	-13259	-13231	0,18277	0,41931	0,09517	0,51320	2,68054E-08	1,45569	0,000103854	0,48577
GJR(1,1)	-13250	-13227	0,17739	0,37248	0,09659	0,50591	3,41684E-08	1,48168	0,000111809	0,51060
Average	-13175	-13149	0,15526	0,31789	0,07459	0,43965	3,41664E-08	1,55798	0,000110711	0,60629
	Student distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,13257	0,31035	0,06419	0,40663	3,64862E-08	1,26240	0,000103552	0,91332
Exponential MA			0,15561	0,27853	0,08135	0,44506	3,34083E-08	1,34006	0,000102819	0,59222
ARCH(3)	-12989	-12960	0,10791	0,28076	0,05816	0,29773	4,32463E-08	2,32310	0,000139968	0,66730
GARCH(1,1)	-13186	-13163	0,15674	0,29738	0,08230	0,45185	3,28176E-08	1,53491	0,000107701	0,55138
GARCH(1,2)	-13180	-13151	0,16418	0,30192	0,07904	0,43493	3,81588E-08	1,67812	0,00011951	0,56067
GARCH(2,1)	-13186	-13163	0,15678	0,29629	0,08231	0,45166	3,25508E-08	1,51233	0,000106601	0,55403
EGARCH(1,1)	-13261	-13232	0,18140	0,41951	0,10672	0,51208	2,68533E-08	1,45726	0,00010396	0,48600
GJR(1,1)	-13251	-13228	0,17630	0,37368	0,10797	0,50530	3,41226E-08	1,48177	0,000111717	0,50992
Average	-13175	-13150	0,15394	0,31980	0,08276	0,43815	3,47055E-08	1,57374	0,000111978	0,60435

Table D 3: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXC20 as the underlying process and assuming a normal error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Normal distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,04673	0,19514		0,19438	2,08543E-08	1,33156	8,22024E-05	0,98934
Exponential MA			0,05669	0,17661		0,24662	1,88456E-08	1,38807	7,94807E-05	0,84764
ARCH(3)	-13435	-13412	0,03446	0,19921		0,18061	2,23317E-08	1,96944	0,00010066	0,55116
GARCH(1,1)	-13679	-13662	0,05774	0,21088		0,24782	1,85662E-08	1,54830	8,38908E-05	0,65442
GARCH(1,2)	-13679	-13662	0,05775	0,21066		0,24786	1,85604E-08	1,54771	8,38522E-05	0,65502
GARCH(2,1)	-13677	-13654	0,05733	0,20988		0,24833	1,86065E-08	1,54832	8,39701E-05	0,65745
EGARCH(1,1)	-13724	-13701	0,07142	0,30571		0,28545	1,64343E-08	1,47138	8,08781E-05	0,60030
GJR(1,1)	-13716	-13694	0,05906	0,26379		0,26075	1,81845E-08	1,51931	8,32887E-05	0,57202
Average	-13652	-13631	0,05515	0,22149	#DIV/0!	0,23898	1,90479E-08	1,54051	8,47779E-05	0,69092
	Normal distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,04718	0,19877	0,02995	0,19688	2,07579E-08	1,32376	8,18987E-05	0,98464
Exponential MA			0,05695	0,17934	0,03692	0,24857	1,87715E-08	1,38078	7,92215E-05	0,84234
ARCH(3)	-13441	-13418	0,03272	0,19910	0,02607	0,18199	2,26954E-08	1,95662	0,000100741	0,55190
GARCH(1,1)	-13686	-13669	0,05803	0,21481	0,03685	0,25018	1,84626E-08	1,53928	8,35415E-05	0,64894
GARCH(1,2)	-13686	-13669	0,05804	0,21469	0,03686	0,25018	1,84698E-08	1,53922	8,35579E-05	0,64924
GARCH(2,1)	-13684	-13661	0,05755	0,21443	0,03664	0,25063	1,84729E-08	1,53836	8,35202E-05	0,65082
EGARCH(1,1)	-13728	-13706	0,07120	0,30644	0,04686	0,28507	1,63926E-08	1,46709	8,07008E-05	0,60166
GJR(1,1)	-13720	-13697	0,05751	0,27286	0,04481	0,25991	1,84115E-08	1,52869	8,41321E-05	0,56151
Average	-13658	-13637	0,05490	0,22506	0,03687	0,24043	1,90543E-08	1,53423	8,46643E-05	0,68638
	Normal distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,04586	0,19375	0,02565	0,19627	2,08251E-08	1,32331	8,18009E-05	0,98645
Exponential MA			0,05568	0,17576	0,03190	0,24664	1,88105E-08	1,38163	7,91389E-05	0,85082
ARCH(3)	-13439	-13416	0,03135	0,19314	0,02232	0,18119	2,28576E-08	1,95805	0,000100876	0,55266
GARCH(1,1)	-13686	-13669	0,05656	0,21020	0,03146	0,24844	1,8545E-08	1,53999	8,35454E-05	0,65338
GARCH(1,2)	-13686	-13669	0,05655	0,21024	0,03146	0,24844	1,85413E-08	1,53967	8,35281E-05	0,65337
GARCH(2,1)	-13684	-13661	0,05605	0,20758	0,03131	0,24907	1,86846E-08	1,53916	8,38389E-05	0,66119
EGARCH(1,1)	-13728	-13705	0,06946	0,30282	0,04051	0,28274	1,64369E-08	1,47014	8,07093E-05	0,60218
GJR(1,1)	-13721	-13698	0,05596	0,26531	0,03879	0,25845	1,85022E-08	1,52845	8,40917E-05	0,56485
Average	-13657	-13636	0,05343	0,21985	0,03167	0,23890	1,91504E-08	1,53505	8,46911E-05	0,69061

In Table D3 the R squared for the log regression based on the daily squared returns as a volatility proxy was zero when assuming a normal distribution and a zero mean. Thus the regression analysis failed to find any correlation between the volatility forecast and the volatility proxy.

Table D 4: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXC20 as the underlying process and assuming a student t error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Student distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,04673	0,19514		0,19438	2,08543E-08	1,33156	8,22024E-05	0,98934
Exponential MA			0,05669	0,17661		0,24662	1,88456E-08	1,38807	7,94807E-05	0,84764
ARCH(3)	-13589	-13560	0,03851	0,20386		0,18147	2,30607E-08	1,97752	0,000101987	0,54385
GARCH(1,1)	-13736	-13713	0,05748	0,21433		0,24670	1,88775E-08	1,55591	8,5085E-05	0,64723
GARCH(1,2)	-13734	-13706	0,05780	0,21435		0,24648	1,87517E-08	1,54721	8,45247E-05	0,64851
GARCH(2,1)	-13728	-13699	0,05392	0,23230		0,24240	1,92738E-08	1,59684	8,7491E-05	0,60580
EGARCH(1,1)	-13770	-13741	0,06844	0,31034		0,27962	1,66263E-08	1,47561	8,16938E-05	0,59324
GJR(1,1)	-13766	-13737	0,05527	0,27664		0,26112	1,90047E-08	1,53559	8,53313E-05	0,55757
Average	-13720	-13693	0,05436	0,22795	#DIV/0!	0,23735	1,94118E-08	1,55104	8,59745E-05	0,67915
	Student distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,04718	0,19877	0,02995	0,19688	2,07579E-08	1,32376	8,18987E-05	0,98464
Exponential MA			0,05695	0,17934	0,03692	0,24857	1,87715E-08	1,38078	7,92215E-05	0,84234
ARCH(3)	-13597	-13568	0,03849	0,20694	0,02899	0,18407	2,30933E-08	1,96645	0,000101725	0,54357
GARCH(1,1)	-13745	-13722	0,05752	0,22088	0,03625	0,24864	1,86753E-08	1,53535	8,42363E-05	0,64033
GARCH(1,2)	-13743	-13715	0,05800	0,21894	0,03652	0,24871	1,86757E-08	1,53536	8,42108E-05	0,64339
GARCH(2,1)	-13745	-13722	0,05778	0,21811	0,03643	0,24906	1,86605E-08	1,53179	8,40434E-05	0,64660
EGARCH(1,1)	-13776	-13747	0,06832	0,31059	0,04546	0,27953	1,65921E-08	1,47000	8,15043E-05	0,59518
GJR(1,1)	-13762	-13733	0,05468	0,27804	0,04303	0,26251	1,94944E-08	1,59686	8,82253E-05	0,54852
Average	-13728	-13701	0,05486	0,22895	0,03669	0,23975	1,93401E-08	1,54254	8,56332E-05	0,68057
	Student distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,04586	0,19375	0,02565	0,19627	2,08251E-08	1,32331	8,18009E-05	0,98645
Exponential MA			0,05568	0,17576	0,03190	0,24664	1,88105E-08	1,38163	7,91389E-05	0,85082
ARCH(3)	-13596	-13567	0,03712	0,20093	0,02478	0,18337	2,32475E-08	1,96886	0,000101857	0,54426
GARCH(1,1)	-13745	-13722	0,05603	0,21591	0,03094	0,24700	1,87408E-08	1,53392	8,4112E-05	0,64481
GARCH(1,2)	-13743	-13715	0,05660	0,21395	0,03119	0,24703	1,87268E-08	1,53325	8,4029E-05	0,64830
GARCH(2,1)	-13742	-13713	0,05581	0,21438	0,03101	0,24782	1,85652E-08	1,53139	8,34921E-05	0,64788
EGARCH(1,1)	-13776	-13747	0,06655	0,30621	0,03923	0,27732	1,66413E-08	1,47288	8,15014E-05	0,59520
GJR(1,1)	-13763	-13735	0,05160	0,27731	0,03834	0,25572	2,18783E-08	1,63529	9,25612E-05	0,54492
Average	-13727	-13700	0,05316	0,22478	0,03163	0,23765	1,96794E-08	1,54757	8,60616E-05	0,68283

In Table D4 the R squared for the log regression based on the daily squared returns as a volatility proxy was zero when assuming a student t distribution and a zero mean. Thus the regression analysis failed to find any correlation between the volatility forecast and the volatility proxy.

Table D 5: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXH25 as the underlying process and assuming a normal error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Normal distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,11618	0,34037	0,06819	0,36982	4,82714E-08	1,29835	0,000133526	0,87053
Exponential MA			0,14700	0,30075	0,08520	0,42852	4,57096E-08	1,34826	0,000133665	0,57368
ARCH(3)	-13213	-13191	0,09784	0,31161	0,05070	0,29162	4,76448E-08	1,66900	0,000144138	0,67907
GARCH(1,1)	-13502	-13485	0,14419	0,33676	0,08495	0,43506	4,19656E-08	1,39030	0,000131939	0,53114
GARCH(1,2)	-13497	-13474	0,14991	0,34725	0,07938	0,40836	4,15559E-08	1,37769	0,000130175	0,54438
GARCH(2,1)	-13502	-13485	0,14420	0,33673	0,08496	0,43507	4,19608E-08	1,39055	0,00013194	0,53109
EGARCH(1,1)	-13560	-13537	0,17460	0,43500	0,10776	0,49626	3,35959E-08	1,30746	0,000122195	0,44892
GJR(1,1)	-13554	-13531	0,16666	0,39848	0,10463	0,49054	3,83234E-08	1,25664	0,0001255	0,46720
Average	-13471	-13450	0,14257	0,35087	0,08322	0,41941	4,23784E-08	1,37978	0,000131635	0,58075
	Normal distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,11667	0,34309	0,07113	0,37284	4,81034E-08	1,29239	0,000133307	0,85867
Exponential MA			0,14731	0,30267	0,08989	0,43136	4,56066E-08	1,34304	0,000133471	0,56969
ARCH(3)	-13220	-13197	0,09803	0,31495	0,05220	0,29483	4,77539E-08	1,66242	0,00014415	0,67791
GARCH(1,1)	-13509	-13492	0,14438	0,33983	0,08958	0,43814	4,18021E-08	1,38439	0,000131684	0,52741
GARCH(1,2)	-13502	-13479	0,14901	0,35426	0,08326	0,41404	4,02508E-08	1,35188	0,000127743	0,54164
GARCH(2,1)	-13509	-13492	0,14437	0,33985	0,08958	0,43815	4,17458E-08	1,38313	0,000131553	0,52757
EGARCH(1,1)	-13563	-13540	0,17397	0,43413	0,11065	0,49651	3,36643E-08	1,30790	0,000122309	0,44834
GJR(1,1)	-13558	-13535	0,16606	0,39911	0,10785	0,49081	3,81141E-08	1,25662	0,00012525	0,46686
Average	-13477	-13456	0,14247	0,35349	0,08677	0,42208	4,21301E-08	1,37272	0,000131183	0,57726
	Normal distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	**BIC**	**r^2 proxy**	**Parkinson's proxy**	**r^2 proxy**	**Parkinson's proxy**	**MSE**	**R2LOG**	**MAD**	**PSE**
MA10			0,11685	0,33809	0,05860	0,37175	4,83029E-08	1,29153	0,000133273	0,85875
Exponential MA			0,14710	0,29969	0,07360	0,43079	4,56476E-08	1,34066	0,00013326	0,57616
ARCH(3)	-13221	-13198	0,09670	0,31105	0,03979	0,29519	4,78125E-08	1,66043	0,000143989	0,67648
GARCH(1,1)	-13511	-13493	0,14407	0,33641	0,07274	0,43744	4,1909E-08	1,38228	0,000131526	0,53277
GARCH(1,2)	-13516	-13493	0,15236	0,33015	0,07173	0,42397	4,25075E-08	1,38976	0,000131697	0,54069
GARCH(2,1)	-13511	-13493	0,14405	0,33652	0,07274	0,43744	4,18842E-08	1,38188	0,000131479	0,53278
EGARCH(1,1)	-13566	-13543	0,17414	0,43127	0,09086	0,49678	3,37244E-08	1,30498	0,000122115	0,44878
GJR(1,1)	-13561	-13538	0,16565	0,39525	0,08822	0,49053	3,82944E-08	1,25503	0,000125298	0,46883
Average	-13481	-13460	0,14261	0,34730	0,07103	0,42299	4,25103E-08	1,37582	0,000131579	0,57941

Table D 6: Presents the in-sample and out-of-sample fit of the six compared volatility forecast models with OMXH25 as the underlying process and assuming a student t error distribution. Each of the three sub tables presents the results with different models for the conditional mean, more specifically zero mean, constant mean an AR(1) mean respectively(top to bottom). The first two columns in each of the sub tables AIC and BIC shows the Akaike and the Bayesian information criteria and measures the in-sample fit. The next 8 columns are different measures of the out-of-sample fit. The two columns under MZ regressions, r^2 proxy and Parkinson's proxy shows the R squared of the Mincer-Zarnowitz regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The r^2 proxy and Parkinson's proxy under log regression similarly shows the R squared of the log regression using daily squared returns and Parkinson's proxy respectively as the volatility proxy. The last four columns presents the values of the different loss functions used for each of the compared volatility models. In each column the "best" value is highlighted in green and the "worst" value is highlighted in red. For the AIC and BIC columns the lowest value corresponds to the best in-sample fit. For the four columns representing the R squared from a performed regression analysis the highest value corresponds to the best out-of-sample fit. For the last four columns representing the compared loss functions the lowest value corresponds to the best out-of-sample fit. The last row in each of the three sub tables is simply the average value of the respective column.

	Student distribution | Zero Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,11618	0,34037	0,06819	0,36982	4,82714E-08	1,29835	0,000133526	0,87053
Exponential MA			0,14700	0,30075	0,08520	0,42852	4,57096E-08	1,34826	0,000133665	0,57368
ARCH(3)	-13317	-13288	0,10150	0,30541	0,05047	0,28749	4,95848E-08	1,70276	0,000147409	0,66573
GARCH(1,1)	-13519	-13496	0,14533	0,33038	0,08517	0,43389	4,32315E-08	1,37630	0,000132922	0,53947
GARCH(1,2)	-13529	-13500	0,15347	0,32721	0,08310	0,42096	4,30038E-08	1,40103	0,000132688	0,54088
GARCH(2,1)	-13511	-13488	0,14534	0,33031	0,08528	0,43481	4,75598E-08	1,52257	0,000144318	0,51874
EGARCH(1,1)	-13577	-13548	0,17607	0,43733	0,10890	0,49662	3,4055E-08	1,30874	0,000123255	0,44743
GJR(1,1)	-13569	-13541	0,16536	0,40398	0,10426	0,48985	3,8395E-08	1,25576	0,000125959	0,46719
Average	-13504	-13477	0,14378	0,34697	0,08382	0,42025	4,37264E-08	1,40172	0,000134218	0,57796
	Student distribution | Constant Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,11667	0,34309	0,07113	0,37284	4,81034E-08	1,29239	0,000133307	0,85867
Exponential MA			0,14731	0,30267	0,08989	0,43136	4,56066E-08	1,34304	0,000133471	0,56969
ARCH(3)	-13325	-13296	0,10159	0,30894	0,05187	0,29082	4,99009E-08	1,69747	0,000147714	0,66471
GARCH(1,1)	-13532	-13509	0,14491	0,33704	0,08975	0,43806	4,24074E-08	1,39478	0,000132827	0,52717
GARCH(1,2)	-13536	-13508	0,15348	0,33051	0,08741	0,42426	4,29265E-08	1,39610	0,000132617	0,53669
GARCH(2,1)	-13532	-13509	0,14531	0,33475	0,08986	0,43791	4,2479E-08	1,38900	0,000132594	0,52930
EGARCH(1,1)	-13582	-13554	0,17551	0,43659	0,11176	0,49691	3,4208E-08	1,30937	0,000123537	0,44675
GJR(1,1)	-13575	-13547	0,16660	0,39926	0,10825	0,49051	3,92108E-08	1,26216	0,000127233	0,46723
Average	-13514	-13487	0,14392	0,34911	0,08749	0,42283	4,31053E-08	1,38554	0,000132912	0,57503
	Student distribution | AR(1) Mean									
			MZ regression		Log regression					
	AIC	BIC	r^2 proxy	Parkinson's proxy	r^2 proxy	Parkinson's proxy	MSE	R2LOG	MAD	PSE
MA10			0,11685	0,33809	0,05860	0,37175	4,83029E-08	1,29153	0,000133273	0,85875
Exponential MA			0,14710	0,29969	0,07360	0,43079	4,56476E-08	1,34066	0,00013326	0,57616
ARCH(3)	-13327	-13298	0,10037	0,30428	0,03966	0,29085	5,00164E-08	1,69646	0,000147492	0,66199
GARCH(1,1)	-13529	-13506	0,14226	0,34406	0,07220	0,43796	4,17566E-08	1,42464	0,000133259	0,52264
GARCH(1,2)	-13540	-13511	0,15327	0,32596	0,07179	0,42235	4,31795E-08	1,39510	0,000132603	0,54325
GARCH(2,1)	-13516	-13487	0,13697	0,32966	0,07157	0,43387	4,73992E-08	1,49599	0,000143357	0,52144
EGARCH(1,1)	-13585	-13557	0,17577	0,43371	0,09189	0,49719	3,42653E-08	1,30633	0,00012335	0,44702
GJR(1,1)	-13578	-13549	0,16635	0,39474	0,08859	0,49032	3,92417E-08	1,25636	0,000126861	0,46947
Average	-13512	-13485	0,14237	0,34627	0,07099	0,42189	4,37262E-08	1,40088	0,000134182	0,57509