Contents

List of Figures

List of Tables

1 Introduction

Due to the large number of assets being traded in the financial markets, the measure of risk associated with the movements of market prices is crucial in financial applications. Market risk is measured by volatility which usually is defined as the conditional variance of returns (relative price changes). However since volatility of an asset is not observable, the true level of volatility has to be estimated. The construction of a model for measuring and predicting future volatility is crucial in areas such as option pricing and risk management. In risk management, future volatility is needed for estimating potential future losses of a portfolio. It is well-known that financial returns satisfies some general characteristics. Mandelbrot(1963) first noted that volatility is time varying with periods of either large or small movements in prices referred to as volatility clustering. In order to capture volatility clusters in financial returns, Engle(1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model. Bollerslev(1986) extended the ARCH model to the Generelized Autoregressive Conditional Heteroscedasticity (GARCH) model which is able to capture volatility clusters in a larger degree using fewer parameters for estimation. Mandelbrot(1963) also noted that the tail distribution of returns often is thick, resulting in a relative high probability for extreme values. For GARCH models, even if the conditional distribution of returns is assumed as normally distributed, the unconditional distribution has thicker tails compared to the normal distribution. However, GARCH model are usually not able to capture all level of tail thickness in returns suggesting that the conditional distribution is non-normal. Bollerslev(1987) suggested the use of student-t distribution for the conditional distribution. Black(1976) noted that positive and negative price changes tend to have an asymmetric impact of volatility where a large negative price decrease tend to increase future volatility. This is referred to as the leverage effect. Nelson(1991) and Glosten et al.(1993) proposed their extensions known as Exponential GARCH (EGARCH) and GJR-GARCH in order to capture the leverage effect evident in financial returns.
The goal of this thesis is to compare the forecasting performance of several GARCH volatility models. Evaluating the forecasting accuracy of volatility models is challenging. A proxy for the unobservable true volatility is used combined with a loss function in order to measure the accuracy of forecasts. The range-based proxies used include opening, high, low and closing prices. Bennett,Gil(2012) argues that the Yang,Zhang proxy, defined

in Yang,Zhang(2000), is theoretically more efficient as a range-based proxy compared to the Parkinson proxy, defined in Parkinson(1980), which works better for empirical data according to some studies. Patton(2006) showed that the MSE and QLIKE loss functions are robust to the choice of an unbiased proxy in the sense that the ranking of volatility forecasting models is true even if using an unbiased volatility proxy. GARCH models are also evaluated in terms of value-at-risk forecasting performance. The forecasting performance comparison between GARCH models are in focus rather than their individual forecasting performance.
The rest of the thesis is structured as follows: theory in section 2 defines all statistics, hypothesis tests and models in order to fully understand the results. Section 3 describes the methods used for analysis of the data set. Section 4 describes the results in detail where tables and figures are presented. Section 5 summarizes the main conclusions of the results and section 6 discusses the results.
The overall conclusions for volatility forecasts using GARCH models suggests that asymmetric GARCH models are able to capture the leverage effect in the data set in order to generate more accurate forecasts. The scheme used for selecting data for estimation clearly affects the accuracy of forecasts. The forecasting performance of volatility and value-at-risk commbined suggest that GARCH models which tend to overestimate volatility forecasts also tend to overestimate value-at-risk forecasts.

2 Theory

Hypothesis tests and statistics are defined in order to describe return data. GARCH models are defined in terms of the conditional variance, error distributions and methods used for estimation. Proxies for the unobservable true volatility and robust loss functions are defined in order to evaluate one-day ahead volatility forecasts. The concept of value-at-risk is introduced and the loss function used for evaluation of value-at-risk forecasts is defined.

2.1 Empirical properties of returns

Volatility models are constructed in order to capture typical patterns of asset returns. Engle,Patton(2001) summarized the empirical properties of asset returns as:
a) Volatility fluctuates in a continuous manner over time around its unconditional mean and does not diverge to infinity. Volatility models are constructed such that long-run forecasts of volatility converges to the unconditional volatility.
b) Periods of either large or small changes in returns tend to come in clusters which is referred to volatility clustering.
c) Volatility tend to increase after a price decrease compared to a price increase of same size. It is referred to as leverage effect.
d) Return data tend to include relatively high frequencies of large price changes. The tail distribution of returns is usually thick.

2.2 Hypothesis tests

2.2.1 Skewness

Skewness is defined as a normalized form of the third central moment and measures the symmetry of a distribution around its mean. Skewness is estimated by:

$$S = \sqrt{T}\frac{\sum_{i=1}^{T}(x_i - \bar{x})^3}{(\sum_{i=1}^{T}(x_i - \bar{x})^{\frac{3}{2}})^2} \sim N(0, 6/T). \qquad (1)$$

Under the assuption of normality, S is asymptotically distributed as normal with zero mean and variance $6/T$. To test for negative skewness of a distribution, consider the null hypothesis $H_0 : S = 0$ versus $H_a : S < 0$. The t-ratio statistic of the sample skewness is

$$z = \frac{S}{\sqrt{6/T}} \sim N(0,1), \tag{2}$$

where z is asymptotically standard normal. H_0 is rejected at significance level α if $z < Z_\alpha$ where Z_α is the lower 100α quantile of the standard normal distribution. The null hypothesis is also rejected if the p-value of the test statistic is less than α.

2.2.2 Kurtosis

Kurtosis is defined as a normalized form of the fourth central moment and measures the peakedness and tail behaviour of a distribution. Excess kurtosis is defined as $K - 3$ since a normal distribution has zero excess kurtosis. Kurtosis is estimated by:

$$K = T\frac{\sum_{i=1}^{T}(x_i - \bar{x})^4}{(\sum_{i=1}^{T}(x_i - \bar{x})^2)^2} \sim N(3, 24/T). \tag{3}$$

Under the normality assumption, $K - 3$ is asymptotically distributed as normal with zero mean and variance $24/T$. A distribution with positive excess kurtosis has a higher peak and a heavier tail compared to the normal distribution and is called leptokurtic. To test for positive excess kurtosis of a distribution, consider the null hypothesis $H_0 : K - 3 = 0$ versus $H_a : K - 3 > 0$. The t-ratio statistic of the sample excess kurtosis is:

$$z = \frac{K-3}{\sqrt{24/T}} \sim N(0,1), \tag{4}$$

where z is asymptotically standard normal. H_0 is rejected at significance level α if $z < Z_{1-\alpha}$ or if the p-value of the test statistic is less than α.

2.2.3 Jarque-Bera test

Jarque, Bera(1987) combines the skewness and kurtosis to test for normality of a distribution and uses the test statistic:

$$JB = \frac{S^2}{6/T} + \frac{(K-3)^2}{24/T} \sim \chi^2_2, \tag{5}$$

where JB is asymptotically chi-squared distributed with 2 degrees of freedom under the null hypothesis that data are i.i.d. normal. The normality assumption in H_0 is rejected if the p-value of the JB-statistic is less than α.

2.2.4 One-sample t test

The one-sample t test is a parametric test for the location parameter μ when σ is unknown. The test-statistic is defined as:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{T}} \sim t_{T-1}, \tag{6}$$

where s is the sample standard deviation of the sample data. Under the assumption of the null hypothesis, t is student-t distributed with T-1 degrees of freedom, $t \sim t_{T-1}$. The null hypothesis, $\bar{x} = \mu$, is rejected at significance level α if the p-value of the test is less than α.

2.2.5 Chi-Square Variance Test

The chi-square variance test tests if the sample variance is constant. The test-statistic is defined as:

$$V = (T-1)\frac{s^2}{\sigma^2} \sim \chi^2_{T-1}. \tag{7}$$

Under the assumption of the null hypothesis, V is chi-squared distributed with T-1 degrees of freedom, $V \sim \chi^2_{T-1}$. The null hypothesis of constant variance is rejected at significance level α if the p-value of the test is less than α.

2.2.6 Modified Q-test

The modified Q-statistic measures the autocorrelation in squared returns and is defined as:

$$MQ(l) = T(T+2)\sum_{j=1}^{l}\frac{\rho_j^2}{T-j} \sim \chi^2_l, \tag{8}$$

where ρ_j is the j-lag sample autocorrelation of squared returns. If data is i.i.d. then MQ(l) is asymptotically chi-square distribution with l degrees of freedom. Tsay(2002) argues that $log(T)$ is a suitable value for l. To test for autocorrelation in raw returns given present ARCH effects, Diebold and Lopez(1996) suggests using the heteroskedasticity robust version of MQ(p):

$$MQ^{HC}(l) = T(T+2)\sum_{j=1}^{l}\frac{\rho_j^2}{T-j}\left(\frac{\sigma^4}{\sigma^4+\gamma_j}\right) \sim \chi^2_l, \tag{9}$$

where σ^4 is the squared sample variance of returns, γ_j is the sample autocovariance of squared returns and ρ_j is the j-lag sample autocorrelation of the returns.

2.3 Volatility Models

Let P_t be the closing price at day t and define $r_t = log(P_{t-1}) - log(P_t)$ as the continuously compounded return, or logarithmic return, over the period t-1 to t. GARCH models are defined as:

$$r_t = \mu_t + \epsilon_t, \quad \mu_t = E[r_t|F_{t-1}], \quad \epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = Var(r_t|F_{t-1}), \tag{10}$$

where μ_t is the conditional mean and σ_t^2 is conditional variance given the information set F_{t-1} at time $t-1$. z_t is defined as i.i.d. with zero mean and unit variance and ϵ_t is serially uncorrelated but serially dependent. If z_t is standard normal, the conditional distribution of ϵ_t is normal with zero mean and time-varying conditional variance σ_t^2. However, GARCH models are constructed in such a way that the unconditional distribution of z_t is non-normal.

2.4 Conditional Mean

For significant serial correlation in returns, let r_t follow a stationary ARMA(p,q) model with lags p and q:

$$\mu_t = E[r_t|F_{t-1}] = \phi_0 + \sum_i^p \phi_i r_{t-i} - \sum_i^q \theta_i \epsilon_{t-i}, \tag{11}$$

where p, q are non-negative integers defining the AR and MA orders. For daily returns, the conditional mean is typically specified as constant or an autoregressive model with $p = 1$.

2.5 Conditional Variance

The volatility process, measured by the conditional variance, has strong serial correlation meaning that future level of variance tomorrow is affected by the current level of today. Let the conditional variance of today be the average of the m most recent conditional variances:

$$\sigma_t^2 = \frac{1}{m}\sum_{i=1}^m \sigma_{t-i}^2 \approx \frac{1}{m}\sum_{i=1}^m r_{t-i}^2. \tag{12}$$

Note that the forecast for the variance of tomorrow is a proper forecast since it is available at the end of today. The one-day ahead forecast is defined by replacing t with $t+1|t$. However since it puts equal weights on each observations, an extreme observation affects future variances for m periods. The choice of m greatly affects the pattern of the volatility process. Empirical data suggests a more gradual decline for the effect of past returns on future conditional variance. The RiskMetric^{TM} model, where the weights of past squared returns decline exponentially, is defined as:

$$\sigma_t^2 = (1-\lambda)\sum_{i=2}^{\infty}\lambda^{i-1}r_{t-i}^2, 0 < \lambda < 1. \tag{13}$$

The level of variance today can be written as:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2, \tag{14}$$

where $\lambda = 0.94$ often is used. The volatility level of today is as a weighted average of the conditional variance and squared return from yesterday. The construction of estimation of variance only uses 100 observations (99.8% of the weights included), which is a huge advantage for large portfolios. The one-day ahead forecast is defined by replacing t with $t+1|t$. However only one-day ahead forecasts are available using the RiskMetric^{TM} model. In risk management and option pricing, multiperiod volatility forecasts are usually needed.

2.5.1 ARCH

Engle(1982) suggested that the serial correlation in squared returns can be modelled using a linear function of past squared innovations. An autoregressive conditional heteroskedasticity (ARCH) model for the conditional variance is defined as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i\epsilon_{t-i}^2, \tag{15}$$

where the ARCH(q) process is well-defined and the conditional variance σ_t^2 is positive if $a_0 > 0, a_i \geq 0, i = 1, .., q$. If $\sum_{i=1}^{q} a_i < 1$ then ϵ_t is covariance stationary, and the unconditional variance of ϵ_t is $a_0/(1-\sum_{i=1}^{q} a_i)$. The one-day ahead forecast is defined by replacing t with $t+1|t$. Also k-ahead forecasts can be constructed for ARCH models.

2.5.2 GARCH

In empirical applications, a relatively long lag in the ARCH(q) model is needed to describe the heavy tails and volatility clustering in the volatility process of returns. Bollerslev(1986) proposed a generalization of the ARCH(q) model to allow for past squared innovations but also past conditional variances in the current conditional variance equation. The generalized ARCH (GARCH) model is defined as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_i \sigma_{t-i}^2, \tag{16}$$

where the GARCH(p,q) process is well-defined and the conditional variance σ_t^2 is positive if $a_0 > 0, a_i \geq 0, i = 1, .., q$ and $b_0 > 0, b_j \geq 0, j = 1, .., p$. If $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1$ then ϵ_t is covariance stationary, and the unconditional variance of ϵ_t is $a_0/(1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j)$. It can be shown that the GARCH(p,q) model is equivalent to an ARCH(∞) model. The one-day ahead forecast is defined by replacing t with $t + 1|t$. Also k-ahead forecasts can be constructed for GARCH models.

2.6 Asymmetric Conditional Variance

Standard GARCH models are able to capture thick tails and volatility clusters in returns, however these models are not able to capture the leverage effect often evident in returns. Empirical data for stock returns suggests that there is a negative correlation between the return of today and future volatility. Asymmetry is built in to the conditional variance functions reflecting that a decrease in stock prices tends to larger increase in volatility compared to the effect of an increase of same size in stock prices.

2.6.1 EGARCH

Nelson(1991) introduced the exponential GARCH (EGARCH) model to allow for asymmetric effects in its conditional variance function. The EGARCH model is defined as:

$$log(\sigma_t^2) = a_0 + \sum_{i=1}^{q} a_i \frac{|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} \sum_{j=1}^{p} b_j log(\sigma_{t-j}^2). \tag{17}$$

If γ_i is negative then bad news will have a larger impact on volatility since the total effect of ϵ_{t-i} is $(1-\gamma_i)|\epsilon_{t-i}|$ for bad news and $(1+\gamma_i)|\epsilon_{t-i}|$ for good news. σ_t^2 is always guaranteed to be positive because the logarithm of σ_t^2 is modelled. EGARCH is covariance stationary if $\sum_{j=1}^{p} b_i < 1$. The one-day ahead forecast is defined by replacing t with $t+1|t$. Also k-ahead forecasts can be constructed for EGARCH.

2.6.2 GJR

Glosten et al.(1993) presented the GJR(p,q) model to allow for asymmetric effects in its conditional variance function with the use of an indicator function. GJR(p,q) is defined as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i + \gamma_i S_{t-i})\epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2, \tag{18}$$

where $S_{t-i} = 1$ if $\epsilon_{t-i} < 0$ and $S_{t-i} = 0$ if $\epsilon_{t-i} \geq 0$. When ϵ_t is positive then the total effects are $a_i\epsilon_{t-i}^2$ and $(a_i+\gamma_i)\epsilon_{t-i}^2$ if ϵ_t is negative. If γ_i is positive, bad news will have a larger impact on σ_t^2. The one-day ahead forecast is defined by replacing t with $t+1|t$. Also k-ahead forecasts can be constructed for GJR.

2.7 Distributions

For GARCH models, even if the conditional distribution of innovations is normally distributed, the unconditional distribution of innovations has thicker tails compared to the normal distribution.

2.7.1 Normal Distribution

The density function of a normal distribution is defined as:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, -\infty < z < \infty, \tag{19}$$

where z has standard normal distribution for $\mu = 0, \sigma = 1$.

2.7.2 Student's t-Distribution

Bollerslev(1987) suggested the use of a standardized Student's t-distribution in GARCH models. For a random variable u_t with t distribution and v degrees of freedom and scale parameter s_t, the density function of u_t is:

$$f(u_t) = \frac{\Gamma[(v+1)/2)]}{(v\pi)^{1/2}\Gamma(v/2)} \frac{s_t^{-1/2}}{[1+u_t^2/(s_t v)]^{(v+1)/2}}, \tag{20}$$

where Γ is the gamma function, $\Gamma(x) = \int_0^\infty y^{x-1}\mathrm{e}^{-y}\,\mathrm{d}y$. The t-distribution is standardized for $s_t = \sigma_t^2(v-2)/v$.

2.8 Maximum Likelihood Estimation

In ARCH models, the most commonly used method for estimating model parameters is the method of maximum likelihood estimation. The likelihood function of a GARCH(p,q) model, used for estimation, is defined as:

$$f(\epsilon_1, ..., \epsilon_T|\theta) = \prod_{t=p+q+1}^{T} f(\epsilon_t|F_{t-1})f(\epsilon_1, ..., \epsilon_{p+q}|\theta), \tag{21}$$

where θ is the set of parameters to be estimated in the assumed distribution, conditional mean and conditional variance. For sufficiently large samples, $f(\epsilon_1, ..., \epsilon_{p+q}|\theta)$ is usually dropped since its exact form is complicated. Maximizing the likelihood function is equivalent to maximizing the log-likelihood function. The conditional log-likelihood function for the tth observation is:

$$l_t(\theta) = log(f(z_t)) - \frac{1}{2}log(\sigma_t^2(\theta), \tag{22}$$

where the total conditional log-likelihood is the sum of the conditional log-likelihoods. Under the normality assumption of z_t, the resulting conditional log-likelihood is:

$$l(z_{p+q+1}, ..., z_T|\theta, z_1, ..., z_{p+q}) = -\frac{1}{2}\sum_{t=p+q+1}^{T}(log(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2}), \tag{23}$$

where $log(2\pi)$ is dropped and σ_t^2 is evaluated recursively. For GARCH models an iterative method is needed to obtain a solution of θ since there is no closed-form solution. For initial values of $\theta_0, \epsilon_0, \sigma_0$, the log-likelihood can be maximized using a numerical optimization method. Newton-Raphson is often used, where an iteration is defined as:

$$\theta_{n+1} = \theta_n - \lambda_n H(\theta_n)^{-1} s(\theta_n), \tag{24}$$

where θ_n is the vector of estimated parameters at iteration n, $s(\theta_n)$ and $H(\theta_n)$ are the gradient vector and Hessian matrix of the log-likelihood at iteration n. A scalar λ_n is chosen such that $log(\theta_{n+1}) \geq log(\theta_n)$. The BHHH algorithm is often used for GARCH models which uses first order derivatives as an approximation for the Hessian matrix.

$$H(\theta) = -\sum_{t=1}^{T} \frac{\delta l_t}{\delta \theta} \frac{\delta l_t}{\delta \theta}. \tag{25}$$

Under suitable regularity conditions, even for conditional non-normality these estimates are consistent and asymptotically normally distributed. An asymptotic covariance matrix of the estimates is constructed using the final Hessian matrix. In practice, the necessary regularity conditions are usually satisfied.

2.9 Volatility Proxy

Since volatility σ_t^2 is unobservable, a proxy is needed to estimate the true volatility. The true volatility is defined as the integrated volatility (IV),

$$\sigma_t^{2(IV)} = \int_{t-1}^{t} \sigma^2(x) dx, \tag{26}$$

over time interval $t-1$ to t. The daily squared innovations ϵ_t^2, using close-to-close prices is an unbiased but noisy proxy of the true volatility. Its variance can be reduced by using more daily squared innovations for estimation. Time series are however rarely stationary for longer periods meaning that volatility could change slowly over time.
Andersen,Bollerslev(1998) introduced the realized volatility (RV) which uses high-frequency data in order to measure daily volatility more accurate. RV is the sum of squared intra-day returns observed over small non-overlapping intervals. Under weak regularity conditions, RV converges in probability to IV as the sampling frequency increases to infinity.

2.9.1 Range-based proxies

If high-frequency data is not available, the accuracy of estimated volatility still can be improved using range-based proxies consisting of opening, high, low and closing prices. Let C_t be the closing price at trading day t, O_t be the opening at trading day t, H_t the highest price at trading day t and L_t the lowest price at trading day t. Define the normalized open as $o_t = log(O_t) - log(C_{t-1})$, the normalized high as $u_t = log(H_t) - log(O_t)$, the normalized low as $d_t = log(H_t) - log(O_t)$ and the normalized close as $c_t = log(C_t) - log(O_t)$. Assume that $log(P_t)$ follows a continuous random walk $\mathrm{d}log(P(t)) = \sigma(t)\mathrm{d}N(t)$ where the expected squared range value is $E[d_t^2] = 4log(2)\sigma_t^2$. Parkinson's(1980) range-based estimate of the daily volatility is defined as:

$$\hat{\sigma}_t^2 = \frac{d_t^2}{4log(2)}. \tag{27}$$

The variance of the daily squared innovation estimate is 5.2 larger than the variance of the Parkinson estimate. Since the closing price of yesterday is assumed to equal the opening price of today, the volatility estimate of parkinson is usually underestimated.
Yang,Zhang(2000) introduces their range-based proxy which is unbiased, independent of the conditional mean and independent of the difference between yesterday's closing price and today's opening price. The variance of the daily squared innovation estimate is 14 larger than the variance of the daily volatility estimate using the Yang,Zhang proxy. Assume that volatility is constant over n $(n > 1)$ trading days. The Yang,Zhang estimate is defined as:

$$\hat{\sigma}_{yz}^2 = \hat{\sigma}_o^2 + k\hat{\sigma}_c^2 + (1-k)\hat{\sigma}_{rs}^2, \tag{28}$$

where $\hat{\sigma}_o^2$ is the sample variance of $\{o_t\}_{t=1}^n$, $\hat{\sigma}_c^2$ is the sample variance of $\{c_t\}_{t=1}^n$, $\hat{\sigma}_{rs}^2$ is the mean of $\{u_t(u_t - c_t) + d_t(d_t - c_t)\}_{t=1}^n$ and $k = \frac{0.34}{1.34+(n+1)/(n-1)}$. The Yang,Zhang volatility estimate is the sum of the over-night volatility, the open to close volatility and the weighted average of the Rogers-Satchell. The estimate of the daily volatility is defined using $n = 2$.

2.9.2 Downward Bias of Volatility Estimators

Volatility estimators, which uses the daily high and low prices, are unbiased while assuming continuous trading. Garman and Klass(1980) and Rogers and Satchell(1991) showed that these estimators will have a downward bias since prices are only observed at discrete time points. The observed daily high (low) is likely higher (lower) than the theoretical high (low). Yang, Zhang(2000) showed that for a correction of the downward bias, the volatility estimate is dependent on the conditional mean and the variance of the volatility estimate is too large. Even though the Yang,Zhang proxy is more accurate than the Parkinson estimate for simulated data, Bennett,Gil(2012) argues that some studies have shown that the Parkinson estimate is a more accurate proxy for empirical data.

2.10 Statistical Loss Functions

In order to measure the performance of GARCH models, statistical loss functions are constructed. The Schwarz's(1978) Bayesian Criterion (SBC) is a measure for the in-sample performance in order to decide which model that fits the data set best. SBC is defined as:

$$SBC = -2 * log(L(\theta)) + n * log(T), \tag{29}$$

where $L(\theta)$ is the optimal log-likelihood value, T is the number of observations and n is the number of parameters needed for estimation. The model with the lowest SBC value fits the data set best.
In order to evaluate the forecasting performance of GARCH models, loss functions are constructed to measure the distance between the one-day forecasting variance $\hat{\sigma}_t^2$ and the actual, but unobservable variance σ_t^2. The mean squared error (MSE) measures the squared difference between the forecast and its proxy of the true volatility,

$$(\hat{\sigma}_t^2 - \sigma_t^2)^2. \tag{30}$$

But underestimating the true volatility could be costly for a conservative risk manager compared to overestimation of the same amount. The QLIKE loss function allows for an asymmetric loss in order to evaluate volatility forecasts,

$$\frac{\sigma_t^2}{\hat{\sigma}_t^2} - log(\frac{\sigma_t^2}{\hat{\sigma}_t^2}) - 1. \tag{31}$$

Patton(2006) showed that the MSE and QLIKE loss functions belong to a family of robust loss functions. Loss functions are robust in the sense that the rankings of volatility forecasting models with use of an unbiased volatility proxy is the same as if using the unobservable true volatility.

2.11 Value-At-Risk

In the beginning of the 1970s, financial institutions faced an increase of instability in the financial markets. Measures were created in order to calculate capital charges given the risk financial institutions were facing. Value-at-Risk (VaR) measures the worst likely outcome for a specific period and confidence level. For a long trading position, VaR with probability level $1-p$ satisfies

$$p = P(\epsilon_t \leq VaR_t^{(1-p)}) \Rightarrow VaR_t^{1-p} = \mu_t + z_p\sigma_t, \tag{32}$$

where z_p is the 100pth percentile of the assumed distribution of z_t. For $p = 0.05$ and r_t distributed as standard normal, $VaR_t^{0.95} = -1.645$ which refers to the risk at a 95% confidence level. For a capital of 10 million, the 95% VaR equals 164 500. Combining VaR with GARCH modelling, the one-step ahead VaR forecast is given by:

$$VaR_{t+1|t}^{(1-p)} = \mu_{t+1|t} + z_p\sigma_{t+1|t}, \tag{33}$$

where $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ are one-step ahead conditional mean and conditional variance forecasts. Lopez(1999) proposed a loss function for measuring the accuracy of VaR forecasts as the squared distance between the forecasted VaR and the actual return given that VaR is less than the return. Sarma et al.(2003) extended the loss function using a penalty for overestimation of VaR. With inspiration of Sarma et al.(2003), the following loss function is used for measuring the accuracy of the VaR forecasts on the basis of the distance between the observed returns and the forecasted VaR values:

$$L_{t+1} = \begin{cases} (r_{t+1} - VaR_{t+1|t}^{(1-p)})^2 & r_{t+1} < VaR_{t+1|t}^{(1-p)} \\ \alpha(r_{t+1} - VaR_{t+1|t}^{(1-p)})^2 & r_{t+1} \geq VaR_{t+1|t}^{(1-p)}, \end{cases}$$

where α is the size of the penalty.

3 Method

The data set used for volatility analysis is the stock index S&P500 and is collected from Yahoo! Finance[32]. S&P500 is an American stock market index based on 500 large companies which are all listed on NYSE or NASDAQ. The data set starts at 2003-11-21 and ends at 2015-11-27 where stock indices are converted into logarithmic returns.

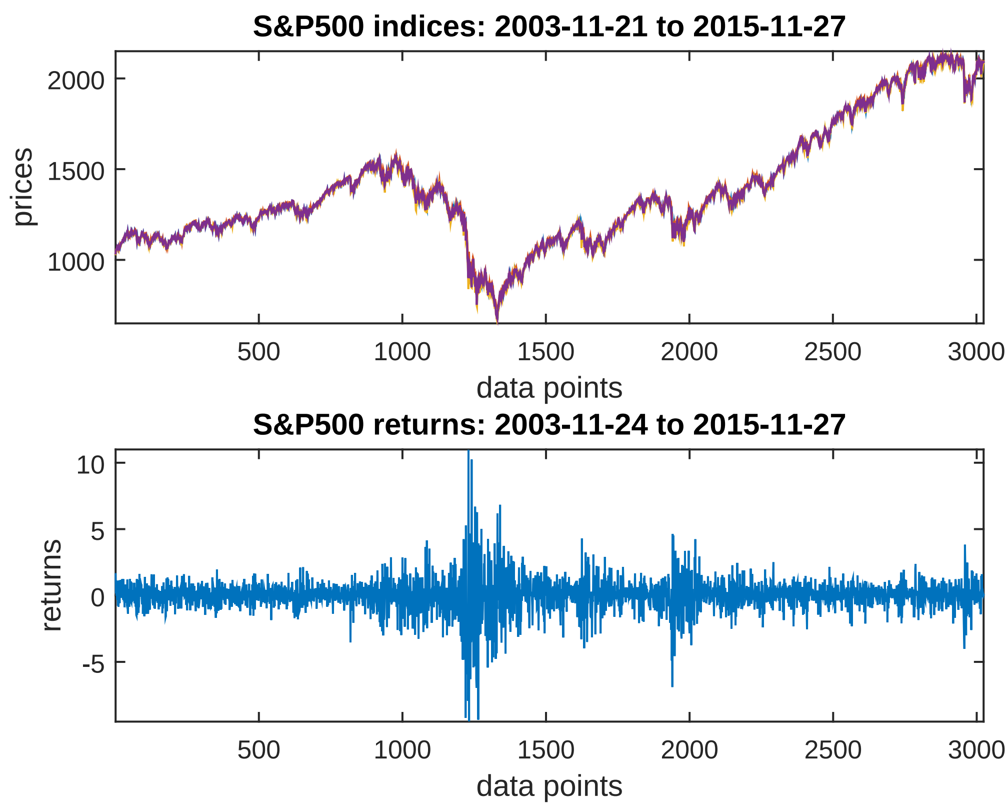

Figure 1: Plot of a) daily closing indices and b) logarithmic returns for S&P500.

The stylized facts of return data are checked including: a) volatility clusters (returns are serially dependent), b) leverage effect (returns are negatively skewed), c) heavy-tailed distribution (returns have positive excess kurtosis), d) weak or no serial correlation (if returns are serially correlated then returns are modelled as an AR(1) model in order to remove significant autocorrela-

tion). Three different series from the data set are used in order to check if properties of returns are changing over time.
Statistical packages are used in MATLAB®for estimation of model parameters and generating forecasting of GARCH models. The in-sample performance of GARCH models are compared using the SBC values. The model with lowest SBC values is regarded as the model which fits the data set best. An estimated GARCH model is correctly specified if the estimated standardized residuals are i.i.d. and distributed according to the assumed distribution. The autocorrelation of residuals and squared residuals are plotted for up to 8 lags in order to verify that residuals are serially uncorrelated and independent. A quantile-quantile plot compares the empirical distribution of estimated standardized residuals with the theoretical assumed distribution.
One-day ahead volatility forecasts are computed and evaluated as follows:
a) An initial sample using data from $t = 1, ..., T$ is used to estimate models and generate the one-day ahead forecasts $\sigma^2_{t+1|t}$. For a rolling scheme, data is moved ahead one period $t = 2, ..., T + 1$ and for a recursive scheme, data is increased by one period $t = 1, ..., T + 1$. The rolling window length T is set to 1260 (5 years of data) and 1764 forecasts (7 years) are generated for each model and scheme.
b) Forecasts are evaluated using a proxy of the true volatility (Parkinson, Yang,Zhang) where the distance between the forecast and proxy is measured using a loss function (MSE,QLIKE). Models which minimizes the loss functions are most accurate in generating accurate one-day ahead forecasts of volatility.
c) The forecasting performance of value-at-risk is evaluated using a loss function which measures the squared distance between the VaR forecast and the return value if value-at-risk is underestimated. A penalty is used for overestimation of value-at-risk where the squared distance is multiplied by a scale factor α. The use of $\alpha = 0.001$ and $\alpha = 0.1$ compares the effect of overestimation compared to underestimation of value-at-risk.

4 Results

GARCH models are constructed in order to capture the empirical properties of returns such as volatility clusters and leverage effect. GARCH models should be able to generate accurate forecasts of future volatility. One-day ahead volatility forecasts are evaluated using robust loss function which measures the distance between forecasts and an unbiased proxy of true volatility. Forecasts are also evaluated in terms of the value-at-risk performance.

4.1 Data Analysis

The empirical properties of returns are measured using statistics and tests in order to specify the GARCH model correctly.

4.1.1 Conditional Variance

Conditional variance functions are constructed such that volatility forecasts are a combination of past conditional variances and squared returns meaning that future volatility is highly correlated with the current level of volatility. Using the absolute values of returns $|r_t|$ as a approximation for volatility σ_t, figure 1b suggests that the volatility process σ_t^2 is varying over time with periods of either low or high variance. Daily volatility could also be approximated by the sample variance of the last 21 days. For a two-sided 95% chi-squared variance test, figure 2 confirms that the estimated volatility process σ_t^2 is varying over time. Daily volatility σ_t^2 could also be approximated by the daily squared return r_t^2. Volatility clusters could be detected by measuring the autocorrelation of squared returns. The p-values of the modified Q-statistic of squared returns with 20 lags, in table 1, confirms that the volatility process is serially correlated or equivalent that returns are serially dependent.

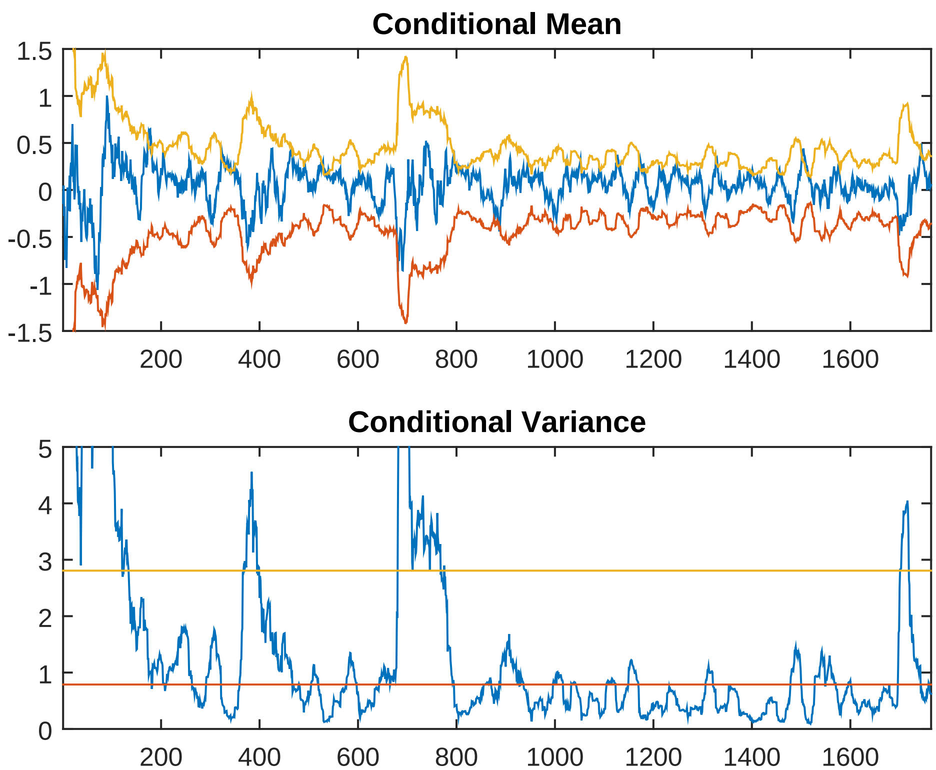

Figure 2: Plot of the a) estimated conditional mean and b) conditional variance. Estimations are calculated using 21 past observations and are compared to the two-sided 95% confidence bound using the one-sample t-test and chi-squared variance test.

Standard GARCH models are able to capture volatility clusters in financial returns, however these models are not able to capture the leverage effect usually evident in financial data. The price plot and its returns combined, in figure 1, suggests that a price drop tend to be followed by an increase in the absolute return r_t. The distribution of returns around its mean is measured using the skewness statistic. The one-sided 95% test of skewness, in table 1, confirms that the distribution of returns is negatively skewed. Also by comparing the mean and the median of returns, in table 1, large negative return values affects the mean. Asymmetric GARCH models such as EGARCH and GJR are therefore likely to capture the leverage effect in returns.

Period (years)	4-8	1-9	7-11
Sample size	1260	2268	1260
Min	-9.5	-9.5	-6.9
Mean (%)	-1.5	1.3	5.0
Median (%)	8.3	7.4	7.4
Max	11	11	4.6
Std	1.67	1.34	1.01
Skewness	-0.26 (1)	-0.31 (1)	-0.50 (1)
Ex. Kurtosis	6.7 (1)	10 (1)	4.7 (1)
$MQ^{HC}(r, 8/20)$	p=0/(3%)	p=0/(0.1%)	p=0.1%/(17%)
$MQ(r^2, 20)$	p=0	p=0	p=0
JB	1	1	1

Table 1: Measurements of statistics for three different series of the returns. For example the serie 1-9 uses data from the 9 first years of the used data. The minimum, mean, median, maximum, standard deviation, the value of the skewness and the excess kurtosis and their outcome given a 95% one-sided test, p-values of the modified Q-statistic for the returns and squared returns and the outcome of the JB-statistic are calculated. H_0 is rejected for an outcome of 1.

4.1.2 Conditional Mean

If returns r_t are weakly serially correlated, returns r_t could be modelled as an AR(1) in order to remove significant autocorrelation in returns. If returns are serially uncorrelated but has a non-zero mean, a constant is set for the conditional mean μ_t. The plot of returns r_t, in figure 1b, suggests that returns

r_t varies around a constant zero mean. The conditional mean μ_t could be approximated by the sample mean of the last 21 days. The two-sided 95% one sample t-test for the estimated conditional mean, in figure 2a, confirms that the conditional mean μ_t could be approximated as zero constant since these are dominated by the conditional variance σ_t^2. Serial correlation in returns could be detected by measuring the autocorrelation of returns. The p-values of the heteroskedasticity robust version of the modified Q-statistic of returns with either 8 or 20 lags, in table 1, confirms that the conditional mean is serially correlated, however the amount of serial correlation is lower than the serial correlation in squared returns.

4.1.3 Data Distribution

The normality distribution of returns is usually questioned. The return series, in figure 1b, contains a lot of extreme values (specially during the financial crisis in 2008). These extreme values suggests that the distribution of returns has thick tails. The one-sided 95% test of the excess kurtosis, in table 1, confirms that returns are non-normal distributed with heavy tails. But GARCH models are constructed in such a way that even though the conditional distribution of innovations ϵ_t is normally distributed, the volatility process σ_t^2 is heavy-tailed with positive excess kurtosis. (explain more?)

4.2 Estimation Evaluation

The estimated standardized residuals $\hat{z}_t = \hat{\epsilon}_t/\hat{\sigma}_t$ of a correctly specified GARCH model is independent and identically distributed (i.i.d.) according to the assumed distribution for z_t. The in-sample performance of GARCH models are evaluated using SBC and the model with lowest SBC fits data best. The specified GARCH models are: a) ARCH(5), b) GARCH(1,1), c) GARCH(1,2), d) EGARCH(1,1), e) GJR(1,1) where z_t is set to the standard normal distribution or the standardized t-distribution and μ_t is zero or the expected value of a AR(1) model for r_t. The methods used for selecting data for estimation are the rolling scheme and the recursive scheme.

4.2.1 In-sample Performance

The in-sample performance of GARCH models is compared using SBC. Table 2, confirms that EGARCH and GJR models fit the data set more accurate than standard GARCH models. It is of no surprise since data is negatively skewed. Table 2, also confirms that the use of a t-distribution fits the data set more accurate compared to the use of a normal distribution. Data has high positive excess kurtosis meaning that the student-t distribution is able to capture the thick tails in data in a larger degree. Table 2, also confirms that the extension of an AR(1) model does not improve the fit of the data. Neither that is surprising since the conditional mean could be approximated as zero. It remains to be seen if the use of asymmetric GARCH models, student-t distribution and AR(1) model is useful for prediction of future volatility.

SBC/model	ARCH(5)	GARCH(1,1)	GARCH(1,2)	EGARCH(1,1)	GJR(1,1)
4-8,N (AR)	3.42(3.42)	3.36(3.36)	3.34(3.34)	3.33(3.33)	3.32(3.33)
4-8,t (AR)	3.37(3.36)	3.33(3.32)	3.31(3.30)	3.30(3.30)	3.29(3.29)
1-9,N (AR)	2.90(2.90)	2.83(2.83)	2.82(2.81)	2.80(2.80)	2.78(2.79)
1-9,t (AR)	2.85(2.85)	2.80(2.80)	2.79(2.78)	2.77(2.77)	2.76(2.76)
7-11,N (AR)	2.62(2.62)	2.58(2.58)	2.57(2.57)	2.52(2.53)	2.52(2.53)
7-11,t (AR)	2.58(2.57)	2.55(2.54)	2.55(2.54)	2.49(2.49)	2.50(2.50)

Table 2: The SBC of GARCH models for choices of distribution (N or t) where the use of an AR(1) is presented in brackets. A model with lowest SBC value fits the data sets best.

4.2.2 Estimation Validation

GARCH models are correctly specified if the estimated standardized residuals $\hat{z}_t$ are i.i.d. according to the assumed distribution for z_t. The autocorrelation of estimated standardized residuals and squared standardized residuals are plotted, in figure 3, for up to 8 lags using the GJR model with normal distribution. For 95% two-sided tests for the autocorrelation of estimated standardized residuals and squared standardized residuals, in figure 3, the estimated standardized residuals are clearly serially independent but weakly serially correlated for some GARCH models. EGARCH will generate similar results.

The distribution of estimated standardized residuals should also match the initial distribution assumed for z_t. The distribution of estimated standardized residuals is compared to the theoretical one using a quantile-quantile plot. The qq-plot of estimated standardized residuals using GJR, in figure 4, suggests that neither the normal distribution or the student-t distribution fits data accurate in the tails. The use of normal distribution seems to fit the data set well in the right tails however poorly in the left tails. The student-t distribution seems to fit the data set more accurate in the left tail however with the loss of fitting the right tail accurate. The question remains which distribution is most useful for prediction of future volatility.

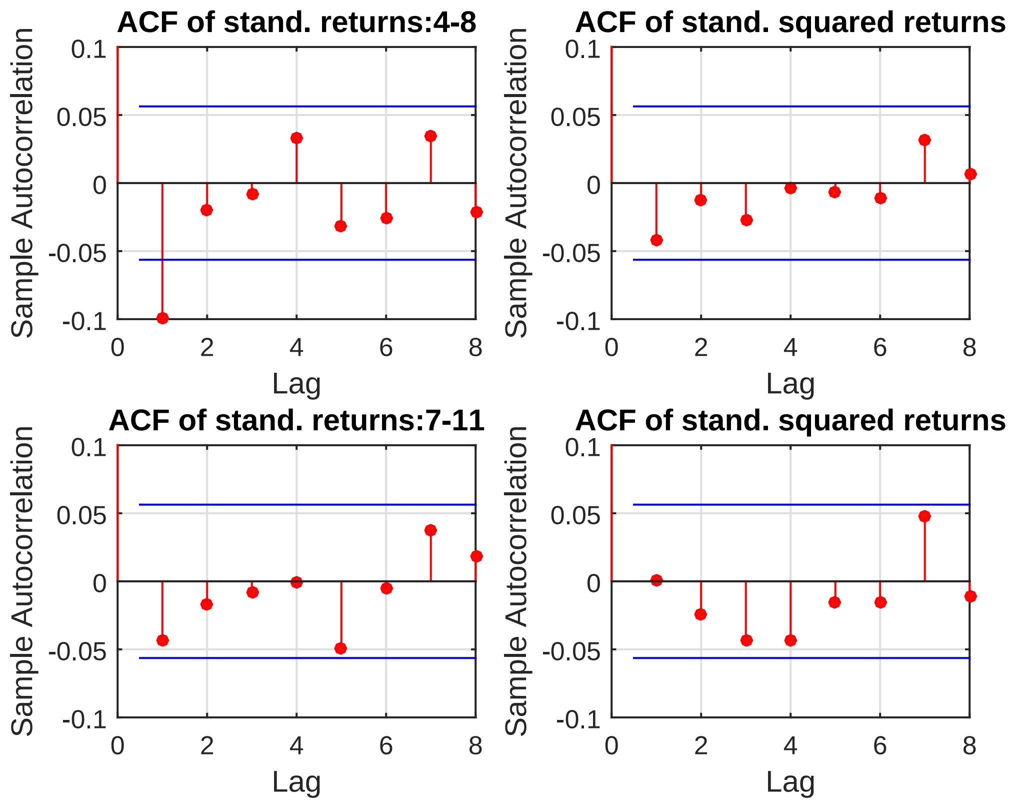

Figure 3: The autocorrelation of estimated standardized residuals and squared standardized residuals using GJR with normal distribution are plotted for up to 8 lags for two different series and compared to the 95% confidence bound $\pm 1.96/\sqrt{(T)}$.

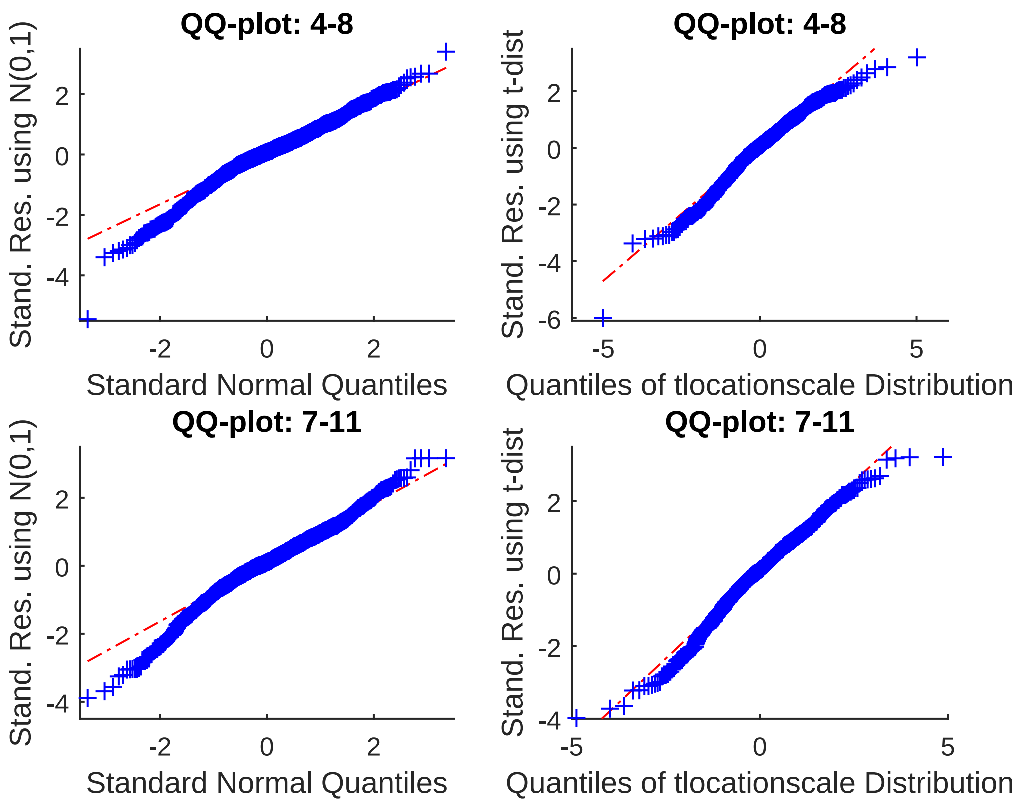

Figure 4: QQ-plot which compares the estimated standardized residuals of GJR assuming standard normal distribution or student-t distribution for two different series, with the theoretical standardized student-t distribution where the degrees of freedom is chosen as 6.1 for the first serie and 6.6 for the second one.

4.3 Forecasting Performance

GARCH models should be able to generate accurate forecasts of future volatility. One-day ahead volatility forecasts are evaluated using robust loss functions which measure the distance between forecasts and an unbiased proxy of true volatility. Loss functions are robust in the sense that the rankings of volatility models in forecasting are true even though unbiased proxies are used as the true volatility. The proxies used for true volatility are the Parkinson proxy and the Yang,Zhang proxy. The loss functions used for evaluating the accuracy of volatility forecasts are the mean squared error (MSE) and the QLIKE function. The mean squared error (MSE) is defined in terms of squared difference compared to the QLIKE which is defined in terms of the quote. The definition of MSE suggests that MSE reflects the accuracy of forecasts in a larger degree in periods of high volatility. The definition of QLIKE suggests that QLIKE reflects the accuracy of forecasts in a larger degree in periods of low volatility. MSE is symmetric around the true volatility which means that underestimation and overestimation of volatility forecasts are equally weighted in terms of forecasting erorrs. However QLIKE is defined in such a way that the forecasting erorrs in terms of underestimation is considered worse than overestimation.

4.3.1 Volatility Forecasting Performance

Forecasting performance is evaluated in terms of estimation methods, distributions, conditional means and conditional variances. It is clear that the asymmetric models EGARCH and GJR outperform standard GARCH models in terms of one-day ahead volatility forecasts, seen in table 3. This make sense since there is a significant level of leverage effect in data (verified by the skewness statistic in table 1). Surprisingly the RiskMetric model performs better than asymmetric GARCH models during periods of low volatility in period 5 and 6. However the RiskMetric only uses 100 observations for estimation comparing to GARCH models which uses at least 1260 observations. This suggest that the estimation of GARCH model parameters are affected by the extremes during the financial crisis in 2008, seen in figure 1b. Figure 5, verifies that EGARCH overestimates volatility forecasts during period 5 and 6.

The combination of asymmetric GARCH models, error distribution and estimation scheme are kept for further analysis in detail. The conditional mean is set to constant zero since it has little impact in forecasting volatility (see Appendix, table 10-13). Since the conditional mean could be approximated as zero, it is unsurprising that the adding AR(1) model only has limited effect on improving forecasting performance. MSE and QLIKE will be used as loss functions and Parkinson and Yang,Zhang will be used as proxies for estimation of true volatility. Two observations are considered outliers, plotted in figure 6, since the values of proxies for observations 364 and 365 are too large compared to the forecasting values. The reason for the extreme values is the extreme daily low in observation 364. Seen in table 3, the removal of outliers clearly decreases the forecasting errors in period 2.

MSE	1	2	3	4	5	6	7
RiskMetric	22	0.77* (5.2)	2.5	0.59	0.12	0.16	0.76
ARCH(5)	9.5	1.1* (4.9)	3.5	0.68	0.31	0.30	0.70
GARCH(1,1)	14	0.71* (5.0)	2.2	0.47	0.16	0.19	0.65
GARCH(1,2)	13	0.74* (5.1)	2.6	0.46	0.18	0.20	0.68
EGARCH	4.8	0.76* (4.8)	1.7	0.44	0.15	0.18	0.62
GJR	13	0.69* (4.6)	2.9	0.32	0.15	0.20	0.71
QLIKE	1	2	3	4	5	6	7
RiskMetric	0.33	0.29* (0.37)	0.32	0.32	0.35	0.36	0.45
ARCH(5)	0.26	0.31* (0.34)	0.33	0.36	0.45	0.52	0.43
GARCH(1,1)	0.27	0.27* (0.33)	0.32	0.31	0.40	0.42	0.42
GARCH(1,2)	0.25	0.27* (0.34)	0.32	0.31	0.40	0.43	0.42
EGARCH	0.17	0.26* (0.30)	0.28	0.27	0.33	0.38	0.42
GJR	0.22	0.25* (0.29)	0.30	0.24	0.34	0.41	0.38

Table 3: The combination of conditional mean, distribution and data estimation method which generates the lowest one-day ahead forecasting errors using the Yang,Zhang proxy and the MSE and QLIKE robust loss functions are compared for each GARCH model. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each. GARCH models are compared to the performance of RiskMetric. The effect of outliers is shown in period 2.

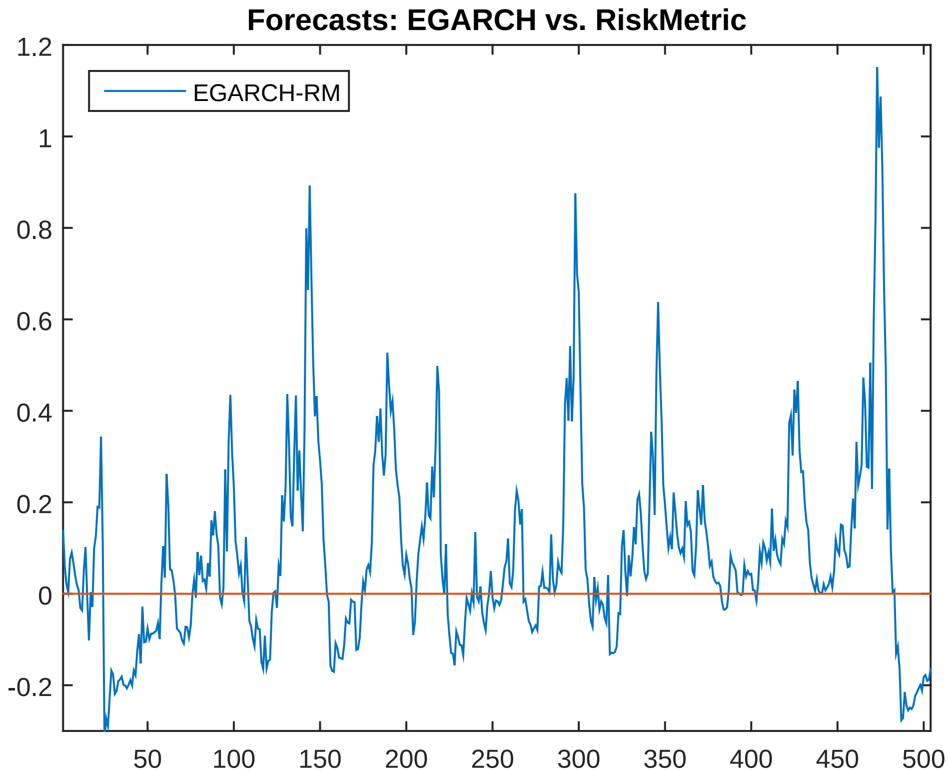

Figure 5: The difference between the forecasts of EGARCH (using normal distribution and recursive scheme) and the RiskMetric is plotted for period 5 and 6.

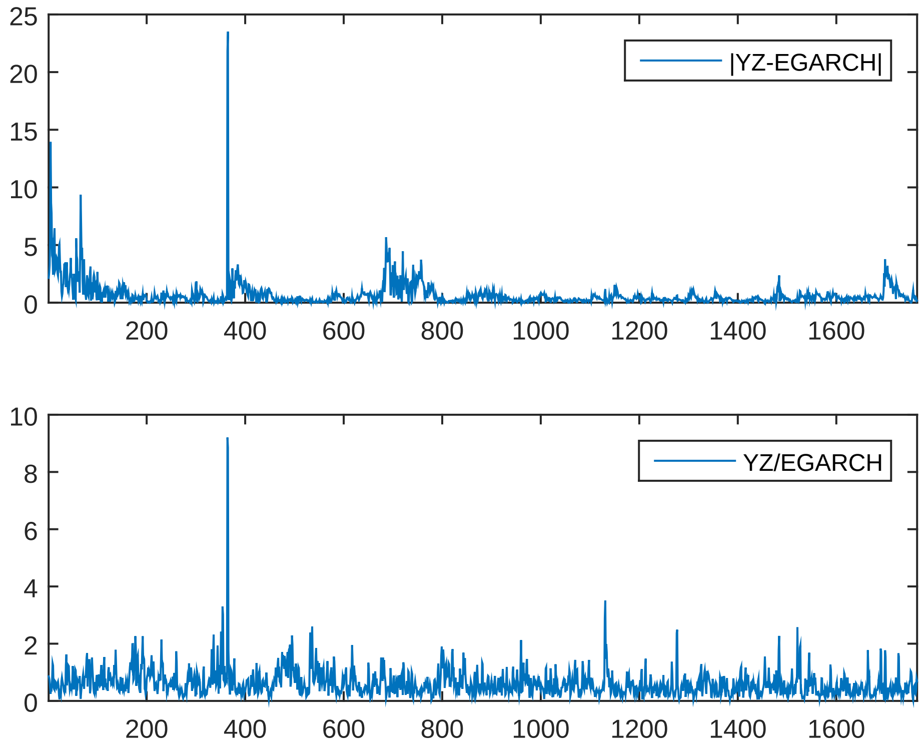

Figure 6: The forecasts of EGARCH with normal distribution and recursive scheme are plotted against the Yang,Zhang proxy. It compares a) the absolute difference between the Yang,Zhang proxy and EGARCH forecasts and b) the quote between the Yang,Zhang proxy and EGARCH forecasts.

4.3.2 Estimation Scheme

GARCH models requires historical data in order to estimate model parameters. The scheme used for selecting data for estimation clearly affects the one-day ahead forecasting accuracy of volatility, seen in table 5. The recursive scheme outperforms the rolling scheme in period 2 to 6 but the rolling scheme outperforms the recursive scheme in period 7. Obviously, the rolling scheme is more sensitive to the choice of historical data for estimation than the recursive scheme. Seen in figure 1b, data contains the extreme returns during the financial crisis in 2008. For period 1 to 6, both estimation schemes include these extremes. However the recursive scheme still contains the calm period before the crisis which suggests that the rolling scheme is highly dominated by the extremes which leads to overestimation of volatility forecasts, seen in figure 7. In period 7, however the data for the rolling scheme does not include these extremes which then results in more accurate volatility forecasts. Figure 7, suggests that the use of recursive scheme overestimate volatility forecasts for the last 50 observations.

	1	2	3	4	5	6	7
MSE,YZ,rek	4.9	0.69	1.7	0.32	0.15	0.19	0.68
MSE,P,rek	6.0	0.71	2.4	0.41	0.17	0.16	0.58
MSE,YZ,roll	4.8	0.79	2.3	0.39	0.20	0.25	0.62
MSE,P,roll	5.9	0.80	2.5	0.46	0.22	0.20	0.39
QLIKE,YZ,rek	0.17	0.25	0.28	0.24	0.33	0.38	0.40
QLIKE,P,rek	0.25	0.34	0.34	0.34	0.41	0.47	0.44
QLIKE,YZ,roll	0.17	0.26	0.33	0.29	0.39	0.45	0.38
QLIKE,P,roll	0.25	0.35	0.38	0.38	0.46	0.53	0.40

Table 4: The combination of conditional variance and distribution which generates the lowest one-day ahead forecasting errors, in terms of the MSE and QLIKE loss functions using the Yang,Zhang and Parkinson proxies, are compared for the use of recursive and rolling scheme. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

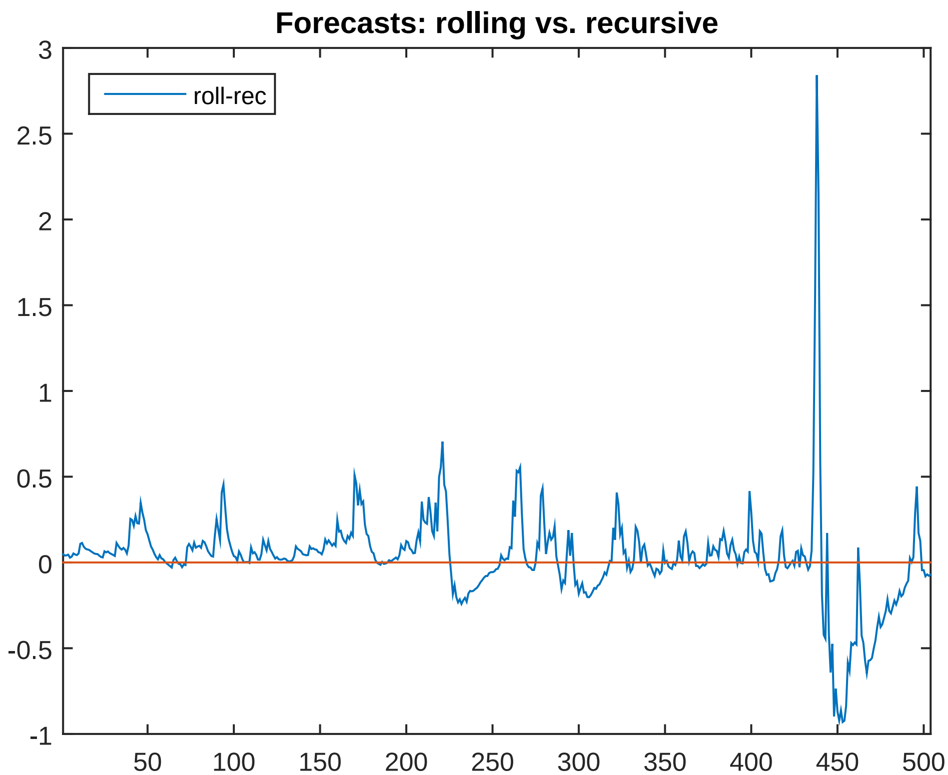

Figure 7: The difference between the forecasts of rolling and recursive schemes for EGARCH with normal distribution are plotted for period 6 and 7.

4.3.3 Distribution

GARCH models requires historical data in order to model the assumed distribution. The in-sample performance, in table 2, suggests that the use of t-distribution fits the data set better. However, seen in figure 4, the tail distribution of the empirical estimated standardized residuals were not convincing for either the normal distribution or the student-t distribution. The results in table 5, suggest that the use of normal distribution generates more accurate volatility forecasts during periods of high volatility. Seen in figure 8, the use of t-distribution overestimates volatility forecasts during periods of high volatility such as period 7.

	1	2	3	4	5	6	7
MSE,YZ,N	4.8	0.69	1.7	0.32	0.15	0.19	0.62
MSE,P,N	5.9	0.71	2.4	0.41	0.17	0.16	0.39
MSE,YZ,t	6.9	0.77	2.1	0.37	0.15	0.19	0.75
MSE,P,t	7.5	0.80	2.7	0.44	0.18	0.17	0.47
QLIKE,YZ,N	0.17	0.25	0.28	0.24	0.34	0.39	0.38
QLIKE,P,N	0.25	0.34	0.34	0.34	0.42	0.48	0.40
QLIKE,YZ,t	0.19	0.26	0.29	0.25	0.33	0.38	0.40
QLIKE,P,t	0.26	0.35	0.35	0.34	0.41	0.47	0.41

Table 5: The combination of conditional variance and estimation scheme which generates the lowest one-day ahead forecasting errors, in terms of the MSE and QLIKE loss functions using the Yang,Zhang and Parkinson proxies, are compared for the use of normal and student-t distribution. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

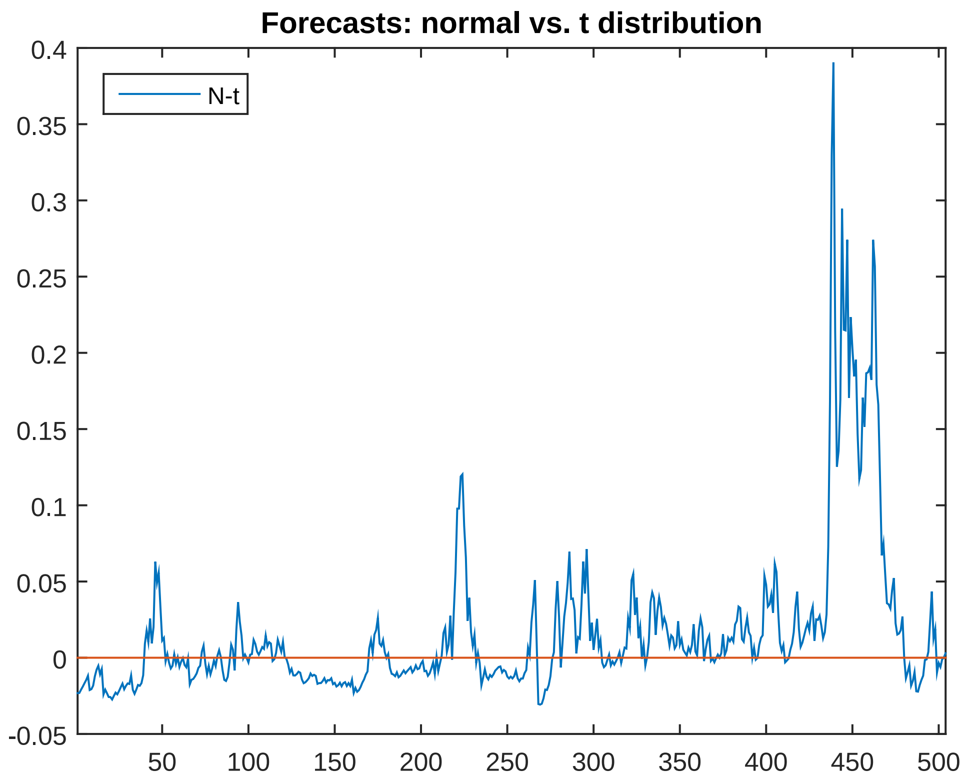

Figure 8: The difference between the forecasts of normal and t-distribution for EGARCH with recursive scheme are plotted for period 6 and 7.

4.3.4 Volatility Model

It is clear that the choice of estimation scheme and distribution assumption affects the accuracy of forecasts. However it is less clear which of the asymmetric GARCH models perform best in terms of accurate volatility forecasts. During periods of high volatility, EGARCH generate more accurate forecasts. GJR clearly overestimates volatility during periods of high volatility, seen in figure 9 for period 1 and 3.

	1	2	3	4	5	6	7
MSE,YZ,EGARCH	4.8	0.76	1.7	0.32	0.15	0.19	0.62
MSE,P,EGARCH	5.9	0.80	2.4	0.51	0.18	0.16	0.39
MSE,YZ,GJR	13	0.69	2.1	0.37	0.15	0.19	0.75
MSE,P,GJR	13	0.71	3.2	0.41	0.17	0.17	0.51
QLIKE,YZ,EGARCH	0.17	0.26	0.28	0.27	0.33	0.38	0.42
QLIKE,P,EGARCH	0.25	0.35	0.34	0.37	0.41	0.47	0.45
QLIKE,YZ,GJR	0.22	0.25	0.30	0.24	0.34	0.41	0.38
QLIKE,P,GJR	0.29	0.34	0.35	0.34	0.41	0.50	0.40

Table 6: The combination of distribution and estimation scheme which generates the lowest one-day ahead forecasting errors, in terms of the MSE and QLIKE loss functions using the Yang,Zhang and Parkinson proxies, are compared for EGARCH and GJR. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

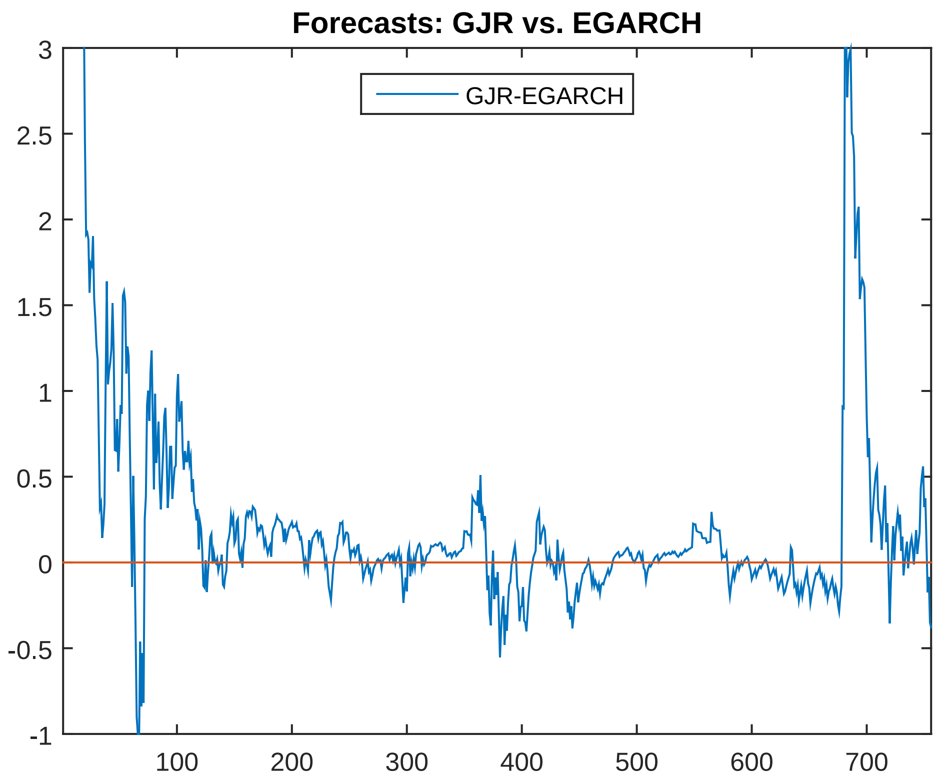

Figure 9: The difference between the forecasts of GJR and EGARCH with use of normal distribution and recursive scheme are plotted for periods 1 to 3

4.4 Value-at-Risk

Volatility forecasts are also evaluated in terms of value-at-risk performance. Value-at-risk performance is evaluated using a loss function which measures the squared distance of underestimation of value-at-risk. Underestimation of value-at-risk refers to the scenario that the actual return is less than the value-at-risk measure. A penalty for overestimation of value-at-risk is used in order to compare the effect of underestimation and overestimation of value-at-risk forecasts using $\alpha = 0.001$ and $\alpha = 0.1$. The value-at-risk measure is defined by using p-value of 5%.

4.4.1 Value-at-Risk performance

Comparing the value-at-risk measure with the actual return is equivalent to comparing the estimated standardized residuals $\hat{z}_t = \hat{\epsilon}_t / \hat{\sigma}_t$ with the 5th percentile of the assumed distribution z_p for $p = 0.05$. An increase in volatility forecasts $\hat{\sigma}_t$ results in a decrease in absolute value of the estimated standardized residuals $|\hat{z}_t|$. A decrease of $|\hat{z}_t|$ results in less risk of underestimation of the value-at-risk measure. Now assume that volatility models which fail to generate accurate volatility forecasts overestimates volatility forecasts. Given the assumption, a larger value of penalty for overestimation of value-at-risk results in that volatility models, which generates accurate volatility forecasts, also generate accurate value-at-risk forecasts. For $\alpha = 0.001$, the value-at-risk loss function prefers the use of rolling scheme (seen in table 7), t-distribution (seen in table 8) and GJR for periods of high volatility (seen in table 9). However for a larger penalty of $\alpha = 0.1$, the value-at-risk loss function prefers the use of recursive scheme (seen in table 7), normal distribution (seen in table 8) and EGARCH for periods of high volatility (seen in table 9). These results suggest that volatility models which generate less accurate volatility forecasts overestimates one-day ahead volatility forecasts. But there is one exception: in period 7, the use of a small penalty prefers the use of rolling scheme, seen in table 7. As seen in figure 7, the high peak for the rolling scheme results in an accurate volatility forecast but inaccurate value-at-risk forecast, seen in figure 10.

L($\alpha = 0.001$)	1	2	3	4	5	6	7
rec	0.096	0.051	0.103	0.015	0.016	0.017	0.028
roll	0.094	0.045	0.083	0.012	0.013	0.011	0.014
L($\alpha = 0.1$)	1	2	3	4	5	6	7
rec	1.49	0.54	0.79	0.35	0.25	0.24	0.38
roll	1.49	0.55	0.82	0.38	0.27	0.25	0.38

Table 7: Value-at-risk performance in terms of estimation scheme for penalty $\alpha = 0.001$, $\alpha = 0.1$ and p-value $p = 0.05$. The degrees of freedom for student-t distribution is 7.5.

L($\alpha = 0.001$)	1	2	3	4	5	6	7
N	0.112	0.049	0.095	0.012	0.013	0.012	0.019
t	0.094	0.045	0.083	0.012	0.013	0.011	0.014
L($\alpha = 0.1$)	1	2	3	4	5	6	7
N	1.49	0.54	0.79	0.35	0.25	0.24	0.38
t	1.57	0.55	0.82	0.36	0.26	0.24	0.40

Table 8: Value-at-risk performance in terms of distribution for penalty $\alpha = 0.001$, $\alpha = 0.1$ and p-value $p = 0.05$. The degrees of freedom for student-t distribution is 7.5.

L($\alpha = 0.001$)	1	2	3	4	5	6	7
EGARCH	0.138	0.045	0.086	0.012	0.013	0.011	0.014
GJR	0.094	0.045	0.083	0.014	0.014	0.013	0.021
L($\alpha = 0.1$)	1	2	3	4	5	6	7
EGARCH	1.49	0.54	0.79	0.38	0.25	0.24	0.38
GJR	1.66	0.54	0.83	0.36	0.25	0.24	0.38

Table 9: Value-at-risk performance in terms of volatility model for penalty $\alpha = 0.001$, $\alpha = 0.1$ and p-value $p = 0.05$. The degrees of freedom for student-t distribution is 7.5.

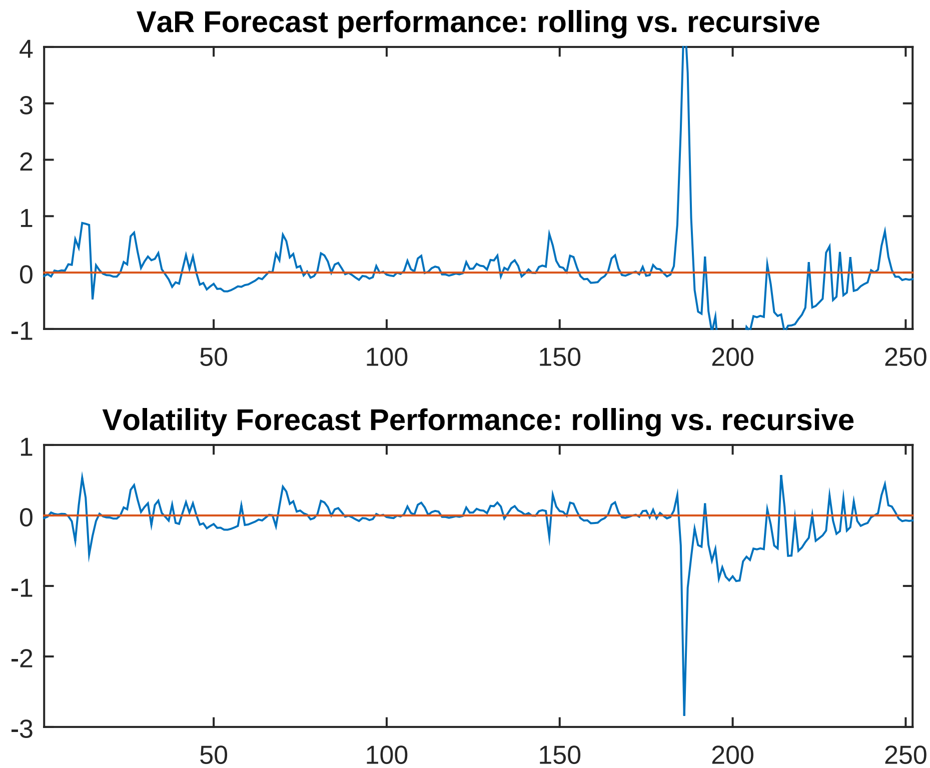

Figure 10: Compares the forecasting accuracy between the use of rolling and recursive scheme, using EGARCH and normal distribution, for a) between the value-at-risk forecasts with the actual return in period 7 and b) between the volatility forecasts and the Parkinson proxy. The distance is defined in terms of $|z - x| - |y - x|$, where z is the rolling forecasts and y is the recursive forecasts.

5 Conclusions

The results suggests that asymmetric GARCH models fits the data set better but also generates more accurate forecasts compared to standard GARCH models. It seems that the EGARCH model generates more accurate volatility forecasts in periods of large volatility. The forecasting performance given the choice of proxy yields the same rankings for volatility forecasting models in almost all cases. Also the use of different loss functions yields the same outcome in most cases. For the choice of error distribution, the use of t-distribution fits data better however the use of normal distribution generates more accurate forecasts in periods of high volatility. The use of an AR(1) model for the conditional mean does not improve either the data fit or the forecasting accuracy. The estimation scheme clearly affects the forecasting performance. In most cases, the recursive scheme generates more accurate forecasts suggesting that GARCH models require a large amount of historical data in order to improve forecasting of volatility. But in period 5 and 6, GARCH models generate less accurate volatilitiy forecasts compared to the RiskMetric model suggesting that the extremes during the financial crisis in 2008 results in overestimation of forecasts using GARCH models. The need for selecting useful historical data for estimation of GARCH models is crucial in order to generate accurate volatility forecasts. The results of the value-vt-risk performance suggests that there is a strong connection between overestimation of volatility forecasts and overestimation of value-at-risk forecasts. The use of a large value for the penalty of overestimation suggests that the volatility models which generates more accurate volatility forecasts also generates more accurate value-at-risk forecasts.

6 Discussion

The focus of this thesis has been on the one-day ahead volatility forecasting performance of GARCH models using robust loss functions and unbiased range-based proxies for the true volatility. Andersen,Bollerslev(1998) introduced the realized volatility (RV) which uses high-frequency data in order to estimate daily volatility more accurate. RV is defined as the sum of squared intra-day returns over a specific interval length. For small time intervals, the RV estimate tend to be affected by potential bias due to market micro-structure frictions such as the bid-ask spread (transactions at discrete time points). To avoid potential bias, in studies such as Andersen,Bollerslev(1998), 5 minute intervals were used as sampling frequency. If forecasts of RV are useful then the log(RV) could often be approximated as a Gaussian ARIMA(0,1,q) model according to Tsay(2002). Martens,van Dijk(2007) introduced the realized range volatility (RRV) by replacing each squared intra-day returns by the high-low range. RRV is a more efficient estimator in theory compared to RV but is highly affected by micro-structure frictions. To account for potential bias, RRV is multiplied by the ratio of the average daily range and the average daily realized range. Using Monte-Carlo simulations, the scaled RRV outperformed several RV estimators.
Correct specification of GARCH models is needed in order to capture the empirical properties of returns such as the amount of thick tails and asymmetry in data. GARCH models are usually not able to capture all of the excess kurtosis in daily returns. It seems that the conditional distribution of innovations is non-normal. Distributions such the GED (see Nelson(1991)), the skewed student-t (see Fernandez, Steel(1998)) and the skewed GED (see Theodossiou(2002)) could be used to capture thick tails and asymmetry in daily returns. For correct specified GARCH models, historical data used for estimation should be able to model the assumed distribution correctly in order to estimate accurate model parameters. Diebold(1986) suggested that structural changes affected the level of the unconditional variance. The structural change during the financial crisis in 2008 clearly affects the volatility forecasts of GARCH models. In order to detect structural changes, regime-switching GARCH models can be constructed where different states, such as high versus low volatility, is determined by either an observable variable, such as the sign of innovations (see Fornari,Mele(1995,1996)), or an unobservable Markov process (see Gray(1996)). Also the idea of a smooth transition in order to gradually change values of parameters were introduced

by Hagerud(1996) and Gonzalez Rivera(1996).
In this thesis only one-day ahead volatility forecasts have been used which requires daily re-estimation. In financial applications, the use of multiperiod volatility forecasts are of interest.
For GARCH models using daily returns, Zivot(2008) stated that the use of trading volume, macroeconomic news announcements, implied volatility from option prices (such as VIX) and realized volatility and overnight volatility are variables which help to predict volatility. The construction of GARCH models should be specified such that these include variables which all improve the accuracy of volatility forecasts.

7 References

[1] Andersen, T. and Bollerslev, T. Answering the skeptics:Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39, 885−905, 1998.
[2] Bennett C, Gil M, Measuring historical volatility, Equity Derivatives, 2012.
[3] Black, F. Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177−181, 1976.
[4] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307−327, 1986.
[5] Bollerslev, T. Aconditional heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics, 69, 542−547, 1987.
[6] Christoffersen, Peter F., Elements of Financial Risk Managment, 2nd edition, 2012.
[7] Diebold, F.X. Modelling the persistence of conditional variances: A comment. Econometric Reviews, 5, 51−56, 1986.
[8] Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica, 50, 987−1008, 1982.
[9] Engle, R.F., Patton, A.J. What good is a volatility model?. Quantitative Finance, 1: 237−245, 2001.
[10] Fernandez, C. and Steel, M. On Bayesian modeling of fat tails and skewness. Journal of the American Statistical Association, 93, 359−371, 1998.
[11] Fornari, F. and Mele, A. Sign- and volatility-switching ARCH models: Theory and applications to international stock markets. Banca dItalia Discussion Paper 251, 1995.
[12] Fornari, F. and Mele, A. Modeling the changing asymmetry of conditional variances. Economics Letters, 50, 197−203, 1996.
[13] Garman, M.B. and Klass, M.J. On the estimation of security price volatilities from historical data. Journal of Business, 53(1), 67−78, 1980.
[14] Glosten, L., Jagannathan, R. and Runkle, D. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48, 1779−1801, 1993.
[15] Gonzalez-Rivera, G. Smooth transition GARCH models. Working paper, Department of Economics, University of California, Riverside, 1996.
[16] Gray, S.F, Modelling the conditional distribution of interest rates as a

regime-switching process, Journal of Financial Economics, 42, 27−62, 1996.
[17] Hagerud, G.E. A smooth transition Arch model for asset returns. Working Paper Series in Economics and Finance 162, Department of Finance, Stockholm School of Economics, 1996.
[18] Lopez, J.A. Methods for evaluating value-at-risk estimates. Economic Policy Review, Federal Reserve Bank of New York, 2, 3−17, 1999.
[19] Mandelbrot, B. The variation of certain speculative prices. Journal of Business, 36, 394−419, 1963.
[20] Martens, M. and van Dijk, D. measuring volatility with the realized range. Journal of Econometrics, 138, 181−207, 2007.
[21] Nelson, D. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347−370, 1991.
[22] Parkinson, M. The extreme value method for estimating the variance of the rate of return. Journal of Business, 53(1), 61−65, 1980.
[23] Patton, A.J. Volatility forecast comparison using imperfect volatility proxies. Working paper, London School of Economics, 2006.
[24] Rogers, L.C.G. and Satchell, S.E. Estimating variance from high, low and closing prices. Annals of Applied Probability, 1, 504−512, 1991.
[25] Sarma, M., Thomas, S. and Shah, A. Selection of VaR models. Journal of Forecasting, 22(4), 337−358, 2003.
[26] Schwarz, G. Estimating the dimension of a model. Annals of Statistics, 6, 461−464, 1978.
[27] Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. Management Science, 44, 1650−1661.
[28] Tsay, R. Analysis of Financial Time Series. John Wiley & Sons, 2002.
[29] Xekalaki E., Degiannakis S. ARCH Models for Financial Applications.
[30] Yang D., Zhang Q., Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices, Journal of Business, vol. 73, no. 3, 2000.
[31] Zivot, E. Practical Issues in the Analysis of Univariate GARCH Models, 2008.
[32] http://finance.yahoo.com/

8 Appendix

MSE, ARCH(5)	1	2	3	4	5	6	7
roll, N	9.71	1.34*	5.47	1.00	0.460	0.325	0.702
roll, N, AR	9.93	1.34*	4.46	1.00	0.450	0.316	0.732
rek, N	9.46	1.07*	4.17	0.693	0.319	0.309	0.938
rek, N, AR	9.73	1.06*	3.50	0.682	0.312	0.299	1.01
roll, t	10.9	1.73*	7.99	1.26	0.539	0.358	0.773
roll, t, AR	11.5	1.84*	7.08	1.26	0.548	0.368	0.869
rek, t	10.5	1.20*	4.99	0.805	0.359	0.337	1.03
rek, t,AR	11.1	1.27*	4.74	0.828	0.365	0.339	1.17
MSE, GARCH(1,1)	1	2	3	4	5	6	7
roll, N	14.6	0.801*	2.96	0.577	0.213	0.207	0.652
roll, N, AR	14.0	0.796*	2.64	0.574	0.206	0.203	0.672
rek, N	14.3	0.712*	2.43	0.477	0.166	0.191	0.733
rek, N, AR	13.8	0.705*	2.23	0.471	0.163	0.189	0.757
roll, t	19.1	0.908*	3.36	0.653	0.217	0.213	0.697
roll, t, AR	18.6	0.903*	3.11	0.647	0.212	0.214	0.769
rek, t	18.7	0.799*	2.76	0.540	0.168	0.192	0.775
rek, t,AR	18.1	0.808*	2.67	0.546	0.166	0.191	0.824
MSE, GARCH(1,2)	1	2	3	4	5	6	7
roll, N	13.6	0.868*	3.60	0.598	0.242	0.223	0.679
roll, N, AR	12.8	0.849*	3.09	0.590	0.232	0.217	0.693
rek, N	13.4	0.758*	2.89	0.470	0.183	0.201	0.796
rek, N, AR	12.5	0.737*	2.55	0.461	0.178	0.197	0.820
roll, t	17.4	1.02*	4.33	0.697	0.251	0.232	0.751
roll, t, AR	16.9	1.02*	3.89	0.676	0.245	0.233	0.817
rek, t	16.9	0.854*	3.34	0.526	0.188	0.204	0.869
rek, t,AR	16.3	0.860*	3.15	0.534	0.186	0.202	0.931

Table 10: The combination of conditional mean, distribution and data estimation method, for each volatility model, are compared in terms of MSE. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

MSE, EGARCH	1	2	3	4	5	6	7
roll, N	4.83	0.933*	2.33	0.623	0.241	0.268	0.621
roll, N, AR	4.82	0.922*	2.24	0.646	0.235	0.264	0.617
rek, N	4.93	0.763*	1.74	0.439	0.151	0.185	0.677
rek, N, AR	4.93	0.758*	1.71	0.457	0.147	0.184	0.696
roll, t	6.85	1.17*	3.57	0.739	0.252	0.323	0.757
roll, t, AR	7.34	1.18*	3.86	0.819	0.244	0.317	0.764
rek, t	7.08	0.869*	2.13	0.521	0.149	0.186	0.752
rek, t,AR	7.57	0.870*	2.15	0.553	0.145	0.186	0.788
MSE, GJR	1	2	3	4	5	6	7
roll, N	13.0	0.788*	3.99	0.390	0.203	0.246	0.864
roll, N, AR	12.8	0.812*	3.58	0.401	0.204	0.244	0.857
rek, N	13.0	0.699*	3.12	0.321	0.145	0.201	0.714
rek, N, AR	12.7	0.692*	2.86	0.324	0.145	0.201	0.756
roll, t	16.9	0.973*	5.77	0.453	0.210	0.277	1.08
roll, t, AR	16.1	0.999*	5.41	0.498	0.218	0.282	1.10
rek, t	16.7	0.774*	3.70	0.365	0.150	0.207	0.797
rek, t,AR	15.8	0.786*	3.34	0.378	0.152	0.208	0.848

Table 11: The combination of conditional mean, distribution and data estimation method, for each volatility model, are compared in terms of MSE. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

QLIKE, ARCH(5)	1	2	3	4	5	6	7
roll, N	0.265	0.328*	0.398	0.462	0.544	0.559	0.435
roll, N, AR	0.267	0.323*	0.389	0.459	0.534	0.546	0.430
rek, N	0.256	0.306*	0.341	0.361	0.460	0.528	0.466
rek, N, AR	0.259	0.302*	0.333	0.357	0.449	0.516	0.462
roll, t	0.281	0.356*	0.439	0.486	0.575	0.577	0.451
roll, t, AR	0.286	0.353*	0.424	0.484	0.562	0.564	0.450
rek, t	0.266	0.320*	0.361	0.379	0.481	0.548	0.483
rek, t,AR	0.271	0.318*	0.359	0.380	0.470	0.536	0.484
QLIKE, GARCH(1,1)	1	2	3	4	5	6	7
roll, N	0.276	0.284*	0.350	0.358	0.444	0.451	0.422
roll, N, AR	0.271	0.281*	0.345	0.356	0.432	0.440	0.417
rek, N	0.272	0.275*	0.327	0.313	0.407	0.431	0.450
rek, N, AR	0.266	0.271*	0.321	0.310	0.399	0.423	0.446
roll, t	0.309	0.297*	0.345	0.365	0.452	0.463	0.435
roll, t, AR	0.304	0.292*	0.346	0.359	0.437	0.451	0.435
rek, t	0.301	0.286*	0.336	0.321	0.408	0.427	0.458
rek, t,AR	0.296	0.283*	0.333	0.320	0.400	0.419	0.458
QLIKE, GARCH(1,2)	1	2	3	4	5	6	7
roll, N	0.263	0.281*	0.353	0.355	0.449	0.466	0.422
roll, N, AR	0.255	0.276*	0.345	0.352	0.434	0.453	0.415
rek, N	0.257	0.272*	0.330	0.310	0.411	0.441	0.447
rek, N, AR	0.248	0.266*	0.323	0.307	0.401	0.432	0.443
roll, t	0.289	0.295*	0.363	0.360	0.456	0.477	0.435
roll, t, AR	0.283	0.289*	0.351	0.355	0.438	0.464	0.431
rek, t	0.280	0.280*	0.340	0.316	0.412	0.437	0.455
rek, t,AR	0.273	0.276*	0.336	0.315	0.402	0.427	0.453

Table 12: The combination of conditional mean, distribution and data estimation method, for each volatility model, are compared in terms of QLIKE. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

QLIKE, EGARCH	1	2	3	4	5	6	7
roll, N	0.174	0.265*	0.325	0.334	0.412	0.447	0.422
roll, N, AR	0.172	0.266*	0.326	0.340	0.409	0.446	0.421
rek, N	0.170	0.257*	0.282	0.272	0.343	0.391	0.450
rek, N, AR	0.169	0.258*	0.284	0.278	0.342	0.391	0.451
roll, t	0.192	0.284*	0.350	0.338	0.402	0.457	0.446
roll, t, AR	0.192	0.287*	0.358	0.350	0.395	0.458	0.448
rek, t	0.188	0.270*	0.292	0.278	0.330	0.380	0.460
rek, t, AR	0.188	0.271*	0.294	0.286	0.328	0.383	0.466
QLIKE, GJR	1	2	3	4	5	6	7
roll, N	0.225	0.257*	0.328	0.288	0.397	0.445	0.383
roll, N, AR	0.220	0.261*	0.325	0.293	0.397	0.444	0.383
rek, N	0.223	0.252*	0.303	0.244	0.346	0.420	0.402
rek, N, AR	0.217	0.254*	0.299	0.247	0.345	0.420	0.406
roll, t	0.250	0.276*	0.350	0.288	0.392	0.447	0.395
roll, t, AR	0.244	0.281*	0.356	0.298	0.395	0.450	0.402
rek, t	0.246	0.263*	0.314	0.249	0.340	0.411	0.406
rek, t, AR	0.239	0.266*	0.311	0.254	0.342	0.413	0.414

Table 13: The combination of conditional mean, distribution and data estimation method, for each volatility model, are compared in terms of the QLIKE. Forecasts are divided into 7 (years) boxes consiting of 252 forecasts each.

MSE, EGARCH, YZ	1	2	3	4	5	6	7
rek, N	4.9	0.76	1.7	0.44	0.15	0.19	0.68(0.62)
roll, N	4.8	0.93	2.3	0.62	0.24	0.27	0.62
rek, t	7.1	0.87	2.1	0.52	0.15	0.19	0.75(0.69)
roll, t	6.9	1.2	3.6	0.74	0.25	0.32	0.76
MSE, GJR, YZ	1	2	3	4	5	6	7
rek, N	13	0.69	3.1	0.32	0.15	0.20	0.71(0.69)
roll, N	13	0.79	4.0	0.39	0.20	0.25	0.86 (0.77*)
rek, t	17	0.77	3.7	0.37	0.15	0.21	0.80(0.77)
roll, t	17	0.97	5.8	0.45	0.21	0.28	1.1 (0.97*)
MSE, EGARCH, P	1	2	3	4	5	6	7
rek, N	6.0	0.80	2.4	0.51	0.18	0.16	0.61(0.57)
roll, N	5.9	0.95	2.5	0.66	0.26	0.23	0.39
rek, t	7.7	0.91	2.7	0.58	0.18	0.17	0.66(0.61)
roll, t	7.5	1.2	3.3	0.75	0.27	0.28	0.47
MSE, GJR, P	1	2	3	4	5	6	7
rek, N	13	0.71	3.2	0.41	0.17	0.17	0.58(0.58)
roll, N	13	0.80	3.6	0.46	0.22	0.20	0.51
rek, t	16	0.80	3.6	0.44	0.18	0.17	0.63(0.62)
roll, t	16	0.97	5.0	0.50	0.22	0.22	0.65

Table 14: Combinations of asymmetric volatility model, distributions and estimation scheme are compared in terms of MSE using both volatility proxies.

QLIKE, EGARCH, YZ	1	2	3	4	5	6	7
rek, N	0.17	0.26	0.28	0.27	0.34	0.39	0.45(0.44)
roll, N	0.17	0.27	0.33	0.33	0.41	0.45	0.42
rek, t	0.19	0.27	0.29	0.28	0.33	0.38	0.46(0.45)
roll, t	0.19	0.28	0.35	0.34	0.40	0.46	0.45
QLIKE, GJR, YZ	1	2	3	4	5	6	7
rek, N	0.22	0.25	0.30	0.24	0.35	0.42	0.40(0.40)
roll, N	0.23	0.26	0.33	0.29	0.40	0.45	0.38
rek, t	0.25	0.26	0.31	0.25	0.34	0.41	0.41(0.41)
roll, t	0.25	0.28	0.35	0.29	0.39	0.45	0.40
QLIKE, EGARCH, P	1	2	3	4	5	6	7
rek, N	0.25	0.35	0.34	0.37	0.42	0.48	0.49(0.48)
roll, N	0.25	0.36	0.38	0.43	0.48	0.53	0.45
rek, t	0.26	0.36	0.35	0.38	0.41	0.47	0.50(0.49)
roll, t	0.27	0.37	0.40	0.43	0.47	0.54	0.48
QLIKE, GJR, P	1	2	3	4	5	6	7
rek, N	0.29	0.34	0.35	0.34	0.42	0.51	0.44(0.44)
roll, N	0.29	0.35	0.38	0.38	0.47	0.53	0.40
rek, t	0.32	0.35	0.36	0.34	0.41	0.50	0.44(0.44)
roll, t	0.32	0.36	0.39	0.38	0.46	0.53	0.41

Table 15: Combinations of asymmetric volatility model, distributions and estimation scheme are compared in terms of QLIKE using both volatility proxies.

L($\alpha = 0.001$) EGARCH	1	2	3	4	5	6	7
N,rek	0.167	0.055	0.122	0.015	0.016	0.017	0.029
t,rek	0.138	0.054	0.114	0.015	0.017	0.017	0.026
N,roll	0.169	0.049	0.101	0.012	0.013	0.012	0.019
t,roll	0.140	0.045	0.086	0.012	0.013	0.011	0.014
L($\alpha = 0.001$) GJR	1	2	3	4	5	6	7
N,rek	0.112	0.054	0.110	0.018	0.017	0.017	0.030
t,rek	0.096	0.051	0.103	0.017	0.017	0.017	0.028
N,roll	0.112	0.051	0.095	0.015	0.014	0.014	0.023
t,roll	0.094	0.045	0.083	0.014	0.014	0.013	0.021
L($\alpha = 0.1$) EGARCH	1	2	3	4	5	6	7
N,rek	1.49	0.542	0.787	0.376	0.254	0.236	0.399
t,rek	1.57	0.558	0.819	0.392	0.256	0.240	0.414
N,roll	1.49	0.557	0.821	0.411	0.280	0.252	0.384
t,roll	1.57	0.588	0.897	0.427	0.287	0.268	0.409
L($\alpha = 0.1$) GJR	1	2	3	4	5	6	7
N,rek	1.66	0.536	0.830	0.354	0.253	0.237	0.384
t,rek	1.75	0.550	0.862	0.364	0.256	0.241	0.397
N,roll	1.66	0.548	0.853	0.375	0.271	0.245	0.383
t,roll	1.76	0.576	0.930	0.387	0.276	0.255	0.406

Table 16: The ValueAtRisk loss function values, using $p = 5\%$, for $\alpha = 0.001$ and $\alpha = 0.1$ are compared for each combination of volatility model, distribution for residuals and the choice of recursive and rolling method.