2011

Forecasting with individual models and combinations

With practical applications to Value at Risk and options

Abstract

This paper studies the forecasting accuracy of individual models and combinations of those in different practical settings. Historical volatility models and models of the GARCH-family are used to perform this research. This research strives to answer the question whether combinations are superior to individual models. It seems this is only the case when looking at an error measure for over predictions and the mean absolute error. The individual models tend to outperform combinations in a VaR setting and with the error measure for under predictions. The performance of the new MAE-method provides evidence for the observation that the chosen individual models tend to over predict the actual volatility.

JEL classification: C22, C51, C52, C53, G17, G32

Keywords: volatility, forecast combinations, option context, value at risk. GARCH

Erasmus University Rotterdam
Erasmus School of Economics
MSc Economics & Business
Master Specialization Financial Economics
Author: J.H.M. Pijnenburg
Student number: 336201
Thesis supervisor: Dr. R.C.J. Zwinkels
Date this version: 17 May 2011

Preface and acknowledgements

This paper was written for the graduation process of the education Economics and Business (specialization Financial Economics) at the Erasmus University in Rotterdam, the Netherlands.

Ever since I followed the seminar Risk Management at the Erasmus University, I have been interested in forecasting the volatility. That is why I have chosen to research this topic for my master thesis. During my master thesis I considerably improved my knowledge of the subject. I especially enjoyed performing the several statistic tests and combining the different methods.

I would like to take this opportunity to thank Remco Zwinkels, my supervisor, for the guidance and advice he has given me during my research. My gratitude also goes to the co-reader of this report; thank you for taking the time to read this.

Zevenbergen, May 17nd 2011

John Pijnenburg

Table of contents

List of Figures

List of Tables

Chapter 1 Introduction

Forecasting of financial volatility is extensively researched in empirical literature. It is also used in practice for many financial activities, including option pricing and the calculation of the Value at Risk. When one compares the function for the volatility forecasting for option pricing and Value at Risk, a difference occurs. While for option pricing, one would be more interested in over predictions or under predictions or vice versa. For Value at Risk on the other hand, one would be interested in the tails of the return distribution. This example points out that forecasting financial volatility is not only a matter of minimizing the forecast error. They also have a different purpose for different financial activities. This article focuses on forecasting volatility in combination with the financial activities option pricing and Value at Risk.

As said before, there is an extensive amount of empirical literature on this topic. Firstly, this article discusses individual forecast models. Two sorts of individual models are used, namely the historical volatility models and models of the GARCH-family. In a review of Poon and Granger (2003) about 93 studies, it came forward that there is no superior sort of individual models. Historical volatility models were superior to models of the GARCH family in 56% in the cases. Brailsford and Faff (1996) back up the results of Poon and Granger (2003). They state that they could also not find a superior model using both sorts of models for the Australian Stock Exchange. Subsequently, Brailsford and Faff (1996) also say that the choice for error statistic is highly relevant for the ranking of models. The results of Brailsford and Faff (1996) and McMillan et al. (2000) (research on FTSE1000) indicate that the individual models seem to over predict the actual volatility more than they under predict.

Secondly, Timmermann (2005) states that combinations of individual models could lead to superior models. He states several reasons to work with combinations. To begin with, different individual models react differently to structural breaks, where in a combination different adaptabilities for those structural breaks could be combined. Secondly, individual models could be biased by misspecification, whereas the combination of models could be more robust versus that misspecification. When working with combinations, it is important to choose individual models that (1) are based on different methods and/or information and (2) make different assumptions of how the different variables are related to each other (Bates and Granger, 1969). There are many combination methods for combining individual forecasts. This article focuses on the regression based combination technique of Granger and Ramanathan (1984) and relative performance methods. Results of Shin and

Sohn (2007) and Becker and Clements (2008) indicate that combinations of individual models are superior to a single individual model.

Thirdly, a combination of models or an individual model could be superior in an empirical setting. However, it also needs to work in a practical setting like Value at Risk. Research on individual models in a Value at Risk setting is extensive. In researches of Angelidis and Degiannakis (2005) and Veiga et al. (2005), the exponential GARCH model seems to be superior over GARCH and EWMA in a Value at Risk setting. Combination of models seems to be superior to individual models in an empirical setting. This paper contributes to the existing literature by also examining combinations of models in a Value at Risk setting.

The main purpose of this study is to evaluate and compare the performances of single individual forecasting models and forecast combination of individual models. The main research question of this study is: are combinations of individual models superior to a single individual model? This paper contributes to the existing literature in several ways. Firstly, forecast combinations are evaluated in a Value at Risk setting. Secondly, a new forecasting combination model that incorporates the tendency of individual models to over predict rather than under predict is introduced. Thirdly, individual models and combinations are evaluated in different time periods, with different distributions and with different backtesting methods. The S&P 500 is the market that is researched in this paper.

The methodology used in this research contains five stages. The first stage describes the individual volatility models. These are moving average (20) and (60), EWMA, GARCH-N, GARCH-T, NAGARCH, NAGARCH-VIX and CGARCH. These models are used for forecasting the volatility. The encompassing tests are described in the second stage. These test tests which models could be combined to make the volatility forecast more accurate. The third stage lists several ways to combine the individual models into one combination model, namely the regression based combination (50 days and 250 days), the MSE performance method, the switching model and the new MAE model with a correction for over prediction. The backtesting methods are stated in the fourth stage. The last stage discusses a practical implementation of volatility models. For this, Value at Risk and the accompanying backtesting methods are used.

The main finding of this study is that there is no superior model between individual models and combinations for every context. Combinations seem to perform better in terms of over prediction, whereas individual models seem to do better in terms of under prediction. This is an important difference in the option context. When examining both groups of models in a VaR context, individual models seem to outperform the combinations, especially in high volatility periods. It is also worth mentioning that the assumption for a normal distribution does not seem to be accurate in high volatility periods.

This paper continues as following. In the next chapter, the methodology is described, which is divided into five sections. Firstly, the individual models are discussed. Secondly, the encompassing test is stated. Thirdly, the different combination techniques are described. Fourthly, the different backtesting methods are explained. Lastly, the practical application of Value at Risk is discussed. The third chapter elaborates on the data that is used for this study. In the fourth chapter, the results are presented and interpreted. The fifth section concludes.

Chapter 2 Methodology

This section describes the methodology that is used in this research. There are five stages. The first stage names the individual volatility models that are used for forecasting the volatility. In the second stage, the encompassing test that tests which models could be combined in order to improve the accuracy of the volatility forecast, are described. The third stage lists different ways of combining the individual models into one combination model. The fourth stage discusses the backtesting methods. Lastly, a practical implementation of volatility models is discussed using Value at Risk and the accompanying backtesting methods.

2.1 Individual models

The first stage of the research contains seven individual volatility models. It is important that each individual model can add something in a combination of models. Poon and Granger (2003) mention several findings about volatility in financial markets, namely that returns have fat tail distributions, that volatilities are likely to cluster, that volatilities show asymmetric reactions and that volatilities show mean reversion. These findings are further explained below with their accompanying models. The returns of all models are calculated in line with Hull (2008) using Equation (1):

$$u_t = ln\left(\frac{S_t}{S_{t-1}}\right) \tag{1}$$

Where u_t is the daily return at time t, S_t is the daily spot closing price at time t and S_{t-1} is the yesterday spot closing price. Christoffersen (2003) assumes that the mean value of the returns is zero because the standard deviation of the returns dominates the mean of returns at daily horizons. This assumption is adopted in this research.

2.1.1 Moving average

The first two individual models imply the simple moving average (hence forth MA). In this model (Christoffersen, 2003), the variance of tomorrow, σ_{t+1}^2, is the average of the past *m* observations. Equation (2) is used for this calculation.

$$\sigma_{t+1}^2 = \sum_{\tau=1}^{m} \frac{1}{m}\, u_{t+1-\tau}^2 \tag{2}$$

The choice for the number of past observations is arbitrary. According to Jorion (2001), the choice of *m* is a trade-off between stability in the case of m = 60 and the ability of capturing the volatility's variation in the case of m = 20. For this research, both choices are examined (m = 20 and m = 60). The disadvantage of this model is that it equally weights the past *m* returns. So the model assumes that older data is just as important as recent data. This assumption may not hold.

2.1.2 EWMA

The third model is the exponentially weighted moving average (hence forth EWMA). An advantage of EWMA in comparison to MA is that EWMA gives more weight to recent returns than to distant returns.

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)u_t^2 \tag{3}$$

Equation (3) states that tomorrow's forecasted variance is the weighted average of the variance and the squared return of today. Another feature of EWMA is that only one parameter needs to be estimated, namely lambda. In line with JP Morgan (1996), the value of lambda is set to 0.94, which is assumed to be appropriate for daily variance forecasting.

2.1.3 GARCH with normal distribution

The generalized autoregressive conditional heteroskedasticity (1,1) (hence forth GARCH) of Bollerslev (1986) is the fourth individual model that is researched. The model expands the ARCH-model of Engle (1982). This model incorporates one of the findings of Poon and Granger (2003), namely mean reversion. This is also the main difference between GARCH and EWMA. GARCH is estimated with Equation (4).

$$\sigma_{t+1}^2 = \omega + \alpha u_t^2 + \beta\sigma_t^2 \tag{4}$$

Where ω is equal to γV_L and where V_L is the long-run variance. As a result, the variance of tomorrow is calculated as a weighted average of the long-run variance (V_L), today's squared return and today's variance. GARCH is reduced to EWMA when $\gamma = 0$, $\alpha = 1 - \lambda$ and $\beta = \lambda$. That means that EWMA does not incorporate a long-run variance. GARCH also allows for mean reversion, which means that the model assumes that the future variance will return to the long-run variance. Mean reversion happens when $\alpha + \beta < 1$ and $\gamma V_L > 0$. On the other hand, when γV_L turns out to be negative, then GARCH is unstable.

A disadvantage of GARCH in comparison to EWMA is that one has to estimate three parameters instead of one. The quasi-maximum likelihood approach (Christoffersen, 2003) is used to estimate those parameters (5).

$$Max \ln L = Max \sum_{t=1}^{T}\left[-\frac{1}{2}(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{u_t^2}{\sigma_t^2}\right] \tag{5}$$

The approach is used to estimate the parameters that maximize the probability of the data appearing. Eviews is used to perform Equation (5). According to Angelidis et al. (2004), the estimated parameters include information about trading behavior that varies through time. Therefore, the parameters of this and the upcoming models from the GARCH family are re-estimated on a daily basis. Christoffersen (2003) states that the number of observations that is suited for estimating the parameters is arbitrary. It is a trade-off between a large number of observations with more accurate estimates (no structural breaks) and a small number of observations that reduces the risk of crossing a structural break. Taking this trade-off into account, a large number of observations (1500 observations) is used for estimating the parameters of a certain point in time for all the models of the GARCH family. For example, for 11 March 2008 an estimation period of 11 June 2002 until 10 March 2008 (1500 trading days) is used.

2.1.4 GARCH with student-t distribution

The fourth individual model, GARCH with student-t distribution, captures the finding that the returns have fat tail distributions. A student-t distribution is considered to better fit the fat tail distribution. Equation (4) remains the same with this model. Only the equation for the estimation of the parameters changes into the following maximum likelihood approach (6) (Angelidis et al., 2004).

$$Max \ln L = Max\, T \left[\ln\Gamma\left(\frac{v+1}{2}\right) - \ln\Gamma\left(\frac{v}{2}\right) - \frac{1}{2}\ln[\pi(v-2)]\right] - \frac{1}{2}\sum_{t=1}^{T}\left[\ln(\sigma_t^2) + (1+v)\ln\left(1+\frac{z_t^2}{v-2}\right)\right] \tag{6}$$

Where v stands for the degrees of freedom. Eviews is used to estimate Equation (6). The degrees of freedom are also estimated for each day. They are used for the practical application of value at risk.

2.1.5 NAGARCH

Non linear asymmetric GARCH (hence forth NAGARCH) of Engle and Ng (1993) captures the asymmetric reactions of volatility of returns. This phenomenon is also called the leverage effect by Black (1976). The leverage effect indicates that negative returns influence the volatility more than positive returns. Black (1976) states that the leverage effect exists because negative returns lead to a decrease in the equity of a company, while it is assumed that the debt remains constant. This means an increase in the debt/equity ratio of that company. An increase in the debt/equity ratio means that the firm is more leveraged and the company’s future becomes more uncertain. More uncertainty leads to more volatility in the equity price.

With the implication of NAGARCH, this leverage effect is captured. NAGARCH is expressed in Equation (7) (Christoffersen, 2003).

$$\sigma_{t+1}^2 = \omega + \alpha(u_t - \theta\sigma_t)^2 + \beta\sigma_t^2 \tag{7}$$

The parameters (α, β, ω and γ) are estimated with Equation (5). A positive value of θ signifies the existence of the leverage effect.

2.1.6 NAGARCH-VIX

Another way to expand a GARCH model is to add an explanatory variable to the equation. This is done in the seventh individual model, NAGARCH-VIX.

$$\sigma^2_{t+1} = \omega + \alpha(u_t - \theta\sigma_t)^2 + \beta\sigma_t^2 + \gamma VIX_t \tag{8}$$

NAGARCH is extended with the implied volatility index of the Chicago Board Options Exchange (hence forth VIX). VIX could play a significant role when forecasting the future volatility. The VIX variable in Equation (8) is calculated in line with Hao and Zhang (2010), as done in Equation (9).

$$VIX_t = \frac{1}{252}\left(\frac{V_t}{100}\right)^2 \tag{9}$$

VIX_t is calculated on a daily basis using Equation (9), where V_t is the value of the VIX-index on time *t*. The parameters are estimated with Equation (5) in Eviews.

2.1.7 CGARCH

The eight model is component GARCH (hence forth CGARCH) that is established by Engle and Lee (1999). CGARCH makes a distinction between a short-run component and a long-run component. In normal GARCH, the long-run component is ω and the conditional variance reverts to this constant long-run variance. In CGARCH, ω is replaced by ϖ_{t+1} that stands for a time-varying long-run variance. The tomorrow's volatility is given by Equation (10).

$$\sigma^2_{t+1} = \varpi_{t+1} + \alpha(u_t^2 - \varpi_t) + \beta(\sigma_t^2 - \varpi_t) \tag{10}$$

The time-varying long-run component is calculated using Equation (11).

$$\varpi_{t+1} = \omega + \rho(\varpi_t - \omega) + \phi(u_t^2 - \sigma_t^2) \tag{11}$$

Where ϖ_{t+1} stands for the long-run component and the short-run component could be calculated by $\sigma^2_{t+1} - \varpi_{t+1}$.

2.2 Encompassing tests

In the first stage, the eight individual models are described with their characteristics. Before combinations of those models can be tested, an encompassing test has to be performed. Encompassing tests examine whether a model or a combination of models incorporate(s) all the information in comparison to another model that was excluded from the combination. If it does not incorporate all the information, then the excluded model needs to be incorporated in the combination of models.

In order to run the upcoming test and to backtest the models in the fourth stage, the conditional variance of the returns is needed. The problem with the conditional variance is that it is unobservable. A simple way to overcome this problem is to use the squared returns as a proxy for the conditional variance. However, Parkinson (1980) states that a range-based variance is a better estimator of the actual variance than the squared returns. The range-based variance is also used by Bannouh et al. (2009) for calculating co variances between assets. They also find that the range-based variance is a better estimator than the squared returns. The range-based variance is calculated as in Equation (12) (Hung et al., 2009).

$$\sigma_t^2 = \frac{ln^2(h_t/l_t)}{4\ln 2} \tag{12}$$

Where h_t and l_t are the highest and the lowest price observed on day *t*. This range-based variance is used as a proxy of the unobserved conditional variance in the rest of the article.

The encompassing test is from Chong and Hendry (1986). It is a simple regression, as is presented in Equation (13).

$$y_{t+1} = \beta_0 + \beta_1\hat{y}^1_{t+1,t} + \beta_2\hat{y}^2_{t+1,t} + \varepsilon_{t+1,t} \tag{13}$$

Where y_{t+1} is the actual volatility at time t + 1 and $\beta_1\hat{y}^1_{t+1,t}$ and $\beta_1\hat{y}^2_{t+1,t}$ are the forecasted volatilities of models 1 and 2 made on time t for time t + 1. If model 1 incorporates all the relevant information of y_{t+1}, then $\beta_0 = \beta_2 = 0$ and $\beta_1 = 1$. This means that model 1 encompasses model 2 with forecasting power. But if both β_1 and β_2 are significantly different from zero, both models contain relevant forecasting information. SPSS' function "forward" is used to estimate Equation (13). Firstly, this function selects the model with the highest correlation with y_{t+1}. Secondly, the function selects the model that explains the most of the unexplained variation the first model could not explain. It is important to note that the added model is significantly different from zero. The other models are examined in the same way.

Multicollinearity

It is important to note that there is a danger for multicollinearity in performing regression (13). Multicollinearity happens when two or more independent variables in a regression have a high correlation with each other. This could lead to bias in the estimates of the standard errors of the regression. One way to detect multicollinearity is to look at the Variance Inflation Factor (hence forth VIF). Although there are some rules of thumb for the value of VIF that indicate multicollinearity (i.e. 10 or even lower 4), O'Brien (2007) argues that values for VIF higher than 10 or even 40 do not have to lead to a bias in the results of the regression. This report follows the findings of O'Brien and only mentions the VIF's and the possibility of a multicollinearity problem. Besides the VIF, one can look at the correlations between the models to detect a cause for multicollinearity. If the correlation is higher than 0.995 between two individual models, this report assumes that the last added model in the SPSS-procedure does not have added value to the combination of models.

2.3 Combination of models

After the encompassing test in the second stage, individual models for combinations are chosen. Subsequently, the third stage of this research examines several methods of combining individual forecasts into one combination model. In this research, two sorts of methods are explored, namely the regression based method and the relative performance methods. The first combination model that is researched, does not belong to the two sorts above. The combination model gives equal weight to each individual model.

2.3.1 Regression based methods

There is a large number of regression models to combine individual forecasts. A simple regression model of Granger and Ramanathan (1984) is used. This is shown in Equation (14).

$$\sigma_{t+1} = \omega_{0,t} + \omega_{i,t}\hat{\sigma}_{i,t+1,t} + \cdots + \omega_{n,t}^{'}\hat{\sigma}_{n,t+1,t} + \varepsilon_{t+1} \tag{14}$$

Where $\omega_{i,t}\hat{\sigma}_{i,t+1,t}$ is the weight of individual model *i* multiplied with the forecasts of individual model *i* and the same happens until model *n*. This regression has been chosen because this one does not need the assumption of unbiased individual models because of the intercept term, $\omega_{0,t}$. It is now important to choose an estimation window for this model. Of course, this choice is subjective. A short window can capture the variation in the variance and a long window follows a more stable path. Therefore, two estimation windows are selected, namely 50 days (short) and 250 days (long). For example, suppose one wants to estimate the individual weights for 19 August 2009. Then the short window would be from 10 June 2009 until 18 August 2009 and the long window would be from 3 September 2008 until 18 August 2009.

2.3.2 Relative performance methods

Unfortunately, regression based methods have disadvantages. For example, Bates and Granger (1969) discuss the probability that correlations between the individual forecasts are badly estimated. Therefore, regression based models could perform insufficiently. A method that pays no attention to correlations between individual forecasts is the Mean Squared Error (hence forth MSE) performance method of Stock and Watson (2001). They use Equations (15) and (16) to calculate the weights for the individual models.

$$SE_{t+1} = (\hat{\sigma}_t^2 - \sigma_t^2)^2 \tag{15}$$

$$\hat{\omega}_{t+1,t,i} = \frac{\left({}^{1}/_{SE_{t+1,t,i}}\right)}{\sum_{j=1}^{N}\left({}^{1}/_{SE_{t+1,t,j}}\right)} \tag{16}$$

$$\sigma_{t+1,c}^2 = \widehat{\omega}_{t+1,t,i}\hat{\sigma}_{i,t+1,t}^2 + \cdots + \widehat{\omega}_{t+1,t,n}\hat{\sigma}_{n,t+1,t}^2 \quad (17)$$

Where SE is the squared error. In Equation (16) the inverse of the SE of model *i* is divided by the sum of the inverse of SE of *n* models. The inverse of the SE is taken because that way the models that performed well according to SE receive higher weights than models that performed badly. Lastly, Equation (17) combines the individual forecasts and the estimated weights into one combination forecast, $\sigma_{t+1,c}^2$. Another difference between the relative performance methods and the regression based methods lies in the absence of negative individual weights.

The second relative performance method is derived from the switching mechanism of Frijns et al. (2008). The model is based on Equation (18) that describes the way that the individual weights are calculated with the switching mechanism.

$$\widehat{\omega}_{t+1,t,i} = \frac{\left(1+e^{\gamma\left(\left|\hat{\sigma}_{i,t}^2-\sigma_t^2\right|\right)}\right)^{-1}}{\sum_{j=1}^{N}\left(1+e^{\gamma\left(\left|\hat{\sigma}_{j,t}^2-\sigma_t^2\right|\right)}\right)^{-1}} \quad (18)$$

Where γ lies between zero and infinity. When $\gamma = 0$, all investors are equally distributed among the individual models. On the other hand, when γ goes to infinity, the investors are switching to the model that best performed in the past 50 days. γ is estimated by minimizing the mean absolute error of the volatility forecast of the combination over the past 50 days prior to time *t* with a value for γ. This estimation is done at a daily basis by solver in Excel. Again, the inverse is taken to give higher weights to the better performing models. The volatility of the total combination (Equation (17)) is again a combination of individual forecasts and the estimated weights of Equation (18). A difference between this model and the model of Frijns et al. (2008) is that Frijns et al. (2008) use the absolute percentage error to calculate the individual weights, whereas this model uses absolute error.

The last relative performance method uses an unexplored (as far as known) characteristic of the GARCH models which implies that GARCH models have a tendency to over predict the ‘actual’ volatility more than they under predict. This characteristic is retrieved from the results of Brailsford and Faff (1996) and McMillan et al. (2000) which indicate that the number of over predictions is significantly higher than the number of under predictions for GARCH models. In line with this characteristic, the new model starts with the volatility of the individual models. If the forecasted variances really over predict instead of under predict the actual volatility, then it is possible that the over predictions are clustered. To partly avoid the clustering of over predictions, Equation (19) is used.

$$\hat{\sigma}_{in,t+1}^2 = \hat{\sigma}_{i,t+1}^2 * (1 - O_t\eta) \quad (19)$$

Where $\hat{\sigma}_{in,t+1}^2$ is the new forecasted volatility of individual model *i* and O_t is a dummy variable that equals one if the individual model over predicts the actual volatility on time *t* and zero if this is not the case. η is the coefficient that is estimated with a regression based on Equation (19). 1500 past observations are used to estimate η for each individual model separately. This estimation is performed on a trimester basis. After that, the trimester estimations are turned into daily estimations using linear interpolation. Once η is estimated, the $\hat{\sigma}^2_{in,t+1}$ is calculated. These steps are done for each individual model separately.

After this first step, the individual weights are calculated using the absolute errors as shown in Equations (20) and (21). Subsequently, the volatility of the combination is calculated using Equation (22).

$$AE_{t+1} = |\hat{\sigma}^2_t - \sigma^2_t| \tag{20}$$

$$\widehat{\omega}_{t+1,t,i} = \frac{\left(1/_{AE_{t+1,t,i}}\right)}{\sum_{j=1}^{N}\left(1/_{AE_{t+1,t,j}}\right)} \tag{21}$$

$$\sigma^2_{t+1,c} = \widehat{\omega}_{t+1,t,i}\hat{\sigma}^2_{in,t+1,t} + \cdots + \widehat{\omega}_{t+1,t,n}\hat{\sigma}^2_{nn,t+1,t} \tag{22}$$

Equations (21) and (22) work in the same way as the equations of the MAE-model.

2.4 Backtesting methods

The fourth stage includes the backtesting of the individual models as well as the combination of models. The backtesting methods described in this section are divided into two parts, namely 'error' measures and regression based evaluation. The different models are ranked based on the results of these backtesting methods

2.4.1 Error measures

There is a large number of error measures to choose from. In line with Kovačić (2008), the following four error measures have been chosen. The following statement goes for all error measures: the higher the error, the worsen the model is.

Mean absolute error

The mean absolute error (hence forth MAE) is the average absolute difference between the forecasted volatility and the actual volatility, as shown in Equation (23) (Kovačić, 2008).

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{\sigma}^2_t - \sigma^2_t| \tag{23}$$

Where $\hat{\sigma}^2_t$ is the forecasted volatility and σ^2_t is the actual volatility (range based volatility).

Root mean squared error

The root mean squared error (hence forth RMSE) is calculated as in Equation (24) (Kovačić, 2008).

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \sigma_t^2)^2} \tag{24}$$

A reason why MAE and RMSE are both chosen, is the convenience that both can be used to analyze the variation in the errors. It is a rule that RMSE is equal to or larger than MAE. Also, the larger the difference between both measures, the larger the variance in the individual forecasting errors in the sample.

Mean mixed error (under and over)

Besides the two symmetric measures MAE and RMSE, it is also interesting to look at two asymmetric measures. Brailsford and Faff (1996) argue that the symmetric assumption of the previous two measures does not necessarily have to hold because it could be that investors do not value under- and over prediction equally. Therefore, Brailsford and Faff (1996) came up with the mean mixed error (under) (hence forth MME(U)) and the mean mixed error (over) (hence forth MME(O)). MME(U) and MME(O) are calculated with Equations (25) and (26) (Brailsford and Faff, 1996).

$$MME(U) = \frac{1}{T}\left[\sum_{t=1}^{O}|\hat{\sigma}_t^2 - \sigma_t^2| + \sum_{t=1}^{U}\sqrt{|\hat{\sigma}_t^2 - \sigma_t^2|}\right] \tag{25}$$

$$MME(O) = \frac{1}{T}\left[\sum_{t=1}^{U}|\hat{\sigma}_t^2 - \sigma_t^2| + \sum_{t=1}^{O}\sqrt{|\hat{\sigma}_t^2 - \sigma_t^2|}\right] \tag{26}$$

Where U is the total number of under predictions and O is the total number of over predictions. MME(U) gives more weight to the absolute errors of the under predictions than to the absolute errors of the over predictions. MME(O) does this vice versa. These asymmetric measures exist because investors do not equally value under- and over predictions.

2.4.2 Regression based evaluation

Another way to backtest and subsequently rank the different models is the regression based evaluation. The first regression is presented in Equation (27).

$$\sigma_t^2 = \alpha + \beta\hat{\sigma}_t^2 + \varepsilon \tag{27}$$

The R^2 of regression (27) measures the proportion of the actual volatility that is explained by the forecasted volatility. The higher R^2, the better the model. However, Engle and Patton (2000) argue that this regression model could be influenced by excessive values of the actual returns. Therefore, Engle and Patton (2000) came up with a log regression as a solution (Equation (28)).

$$\log(\sigma_t^2) = \alpha + \beta log(\hat{\sigma}_t^2) + \varepsilon \tag{28}$$

In this regression, the large values receive less weight than in the first regression. Again, the R^2-values are used for the ranking of the models.

2.5 Practical application (Value at risk)

Volatility forecasting is commonly used in practice. This research points out Value at risk (hence forth VaR). The 5% VaR is the amount of money that a company does not expect to lose in 95 percent of the cases. In this section, the calculation of the VaR and the accompanying backtesting methods are described.

2.5.1 VaR calculation

The VaR is a combination of an applied distribution and the volatility forecast, as can be seen in Equation (29).

$$VaR_{t+1} = -\sigma_{t+1} \, x \, \alpha_p \tag{29}$$

Where α_p is the probability density function of a normal distribution. The p stands for the quantile from the distribution that is set to 1% and 5% in this report. The (one-sided) z-values that correspond with an alpha of 1% and of 5% are 2.33 and 1.645. However, the normality assumption does not have to hold. Therefore, the alpha is replaced by the student-t distribution with v degrees of freedom. The degrees of freedom are collected while estimating the parameters of the GARCH with a student-t distribution. All the models are examined under two scenarios, namely the normal distribution and the student-t distribution.

2.5.2 Backtesting methods (Haas)

Once the VaR is calculated, backtesting methods are necessary in order to calculate the accuracy of the different models. In this research, the mixed Kupiec's test from Haas (2001) and Christoffersen (2003) is used. The mixed Kupiec's test can be divided into three different tests, namely the unconditional coverage test, the independence test and the conditional coverage test. This research speaks of a violation when the actual volatility exceeds the VaR.

Unconditional coverage test

Firstly, the unconditional coverage test of Christoffersen (2003) investigates whether the proportion of violations of a model(π) significantly differs from the expected probability (p). To test the null hypothesis of π = p, Christoffersen (2003) constructed the following likelihood ratio test (hence forth LC) (30).

$$LR_{UC} = -2 \ln \left[\frac{(1-p)^{T_o}}{\{(1-T_1/T)^{T_o}(T_1/T)^{T_1}\}} \right] \sim \chi_1^2 \tag{30}$$

Where T_1 is the number of violations and T_0 is the number of non-violations. The null hypothesis is rejected when the LR_{UC} is larger than the χ_1^2-statistic with a 10% significance level. The 10% significance level is chosen in line with Christoffersen (2003), who argues that a Type 2 error (accepting an incorrect model) is more damaging in risk management than a Type 1 error (rejecting a correct model).

Independence test

Secondly, the independence of the violations has to be tested. It is important that the violations are not clustered, because clustering could lead to a bankruptcy (among other things). The independence test from Christoffersen (2003) only looks at the clustering of two violations, whereas the independence test of Haas (2001) examines the independency of the violations. In a perfect scenario with 1,000 observations and 1% VaR, a violation occurs after 100 observations. The null hypothesis is that the violations occur independent of each other. The test statistic measures the time between two violations. This is calculated as in Equation (31) (Haas, 2001).

$$LR_{ind} = \sum_{i=2}^{n} \left[-2 \ln \left(\frac{p(1-p)^{v_i - 1}}{\left(\frac{1}{v_i}\right)\left(1 - \frac{1}{v_i}\right)^{v_i - 1}} \right) \right] - 2 \ln \left(\frac{p(1-p)^{v-1}}{\left(\frac{1}{v}\right)\left(1 - \frac{1}{v}\right)^{v-1}} \right) \sim \chi_n^2 \tag{31}$$

Where v is the time between two violations. The null hypothesis is rejected when the LR_{ind} exceeds the χ_n^2 with a 10% significance level.

Conditional coverage test

Lastly, the unconditional coverage test and the independence test are combined in the mixed Kupiec's test of Haas (2001). The null hypothesis of this test is that the violations occur independent of each other and the proportion of violations does not differ from the expected probability. Haas (2001) formulated Equation (32).

$$LR_{Mix} = LR_{UC} + LR_{ind} \sim \chi_{n+1}^2 \tag{32}$$

The null hypothesis is rejected when LR_{Mix} is larger than χ_{n+1}^2 with a 10% significance level.

Chapter 3 Data

Now that the methodology for this research is explained, the data is described in this chapter. Returns of the S&P 500 stock index are chosen to perform this research. The S&P 500 is selected because of the high liquidity of the index. The data for the S&P 500 and the VIX variable are collected from DataStream. The total sample period is from 5 April 1993 to 31 December 2010. The out of sample period that is researched is the period from 3 January 2005 to 31 December 2010. The in-sample period is defined as the period from 5 April 1993 to 31 December 2004.

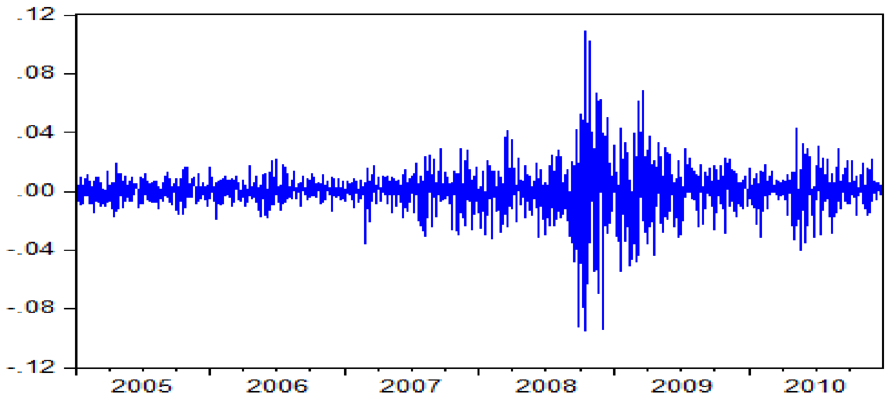
According to Brunnermeier (2009), the financial crisis started around July 2007. Therefore it is assumed that the period before 2007 (from 3 January 2005 to 29 June 2007) is a low volatility period and the period after 2007 is a high volatility period (from 2 July 2007 to 31 December 2010). This can also be seen in Table 1, in which the descriptive statistics are presented. For the encompassing test, an equally large period of 4 January 1999 until 31 December 2004 has been chosen. This period also captures periods of high (internet bubble burst) and low volatility and is assumed to be a good estimation of the out of sample period that is examined.

Table 1 Descriptive statistics

	Period	No. Of Obs.	Mean (x1/1,000)	Std. Dev. (x1/1,000)	Skewness	Kurtosis	Jarque Bera
Returns	2005-2007	650	0.332	6.428	-0.294	4.818	98.851
	2007-2010	915	-0.195	17.967	-0.175	9.310	1,522.91
	2005-2010	1565	0.024	14.348	-0.249	13.466	7,158.96
	1993-2010	4630	0.226	12.009	-0.205	12.059	15,864.8
Squared Returns	2005-2007	650	0.041	0.080	6.829	85.550	189,614
	2007-2010	915	0.323	0.931	7.247	70.127	179,802
	2005-2010	1565	0.206	0.727	9.355	116.085	856,728
	1993-2010	4630	0.144	0.479	12.229	219.690	9,173,741
Range based Variance	2005-2007	650	0.032	0.038	7.441	109.135	311,083
	2007-2010	915	0.211	0.458	5.476	38.491	52,595.9
	2005-2010	1565	0.137	0.362	7.056	63.233	249,562
	1993-2010	4630	0.102	0.242	9.219	117.554	2,597,129

(The null hypotheses of the Jarque Bera statistic is rejected in every case with a 5% significance level)

Figure 1 Graph of the volatility of the S&P500



From Table 1 and Figure 1, one can conclude that the first period (2005-2007) is a low volatility period and the second period (2007-2010) is a high volatility period. This conclusion is drawn because of the difference in the two standard deviations. Another reason for this conclusion is the average VIX-values of the two periods. The first and second period have an average VIX-value of 12.84 and 27.86. This indicates that the first period is less volatile than the second period. This conclusion has been given more strength by looking at Figure 3 of the appendix, where the average yearly VIX-values are drawn.

According to the standard deviations in Table 1, the range based variance is less volatile than the squared returns as a proxy for the conditional variance. Furthermore, returns of S&P500 do not seem to follow a normal distribution when looking at the values of skewness and kurtosis. Also, the null hypothesis of normality of the Jarque Bera test is rejected in every case. This leaves room to test with a student-t distribution.

Figure 1 shows the ARCH effect if one looks at the time-varying amplitude of the returns. The ARCH effect also comes back in the positive significant autocorrelations in Tables 13 and 14 in the appendix. Another interesting point that comes from Table 13 is that there is a difference between the two periods. The first period (2005-2007) shows less signs of autocorrelation (no significant results in the first nine lags excluding lag 5 and 6) than the second period (2007-2010), where every lag is positively significant.

Chapter 4 Results

In this section the results are discussed and analyzed. Firstly, the results for the individual models are interpreted. Secondly, a select number of individual models are chosen for the combinations of models with an encompassing test. Thirdly, the results for the combination of models are discussed. Lastly, the results of the practical application Value at Risk are analyzed.

4.1 Individual models

Firstly, the model coefficients of the GARCH models are given and explained. Secondly, the backtesting results for the individual models are shown and interpreted.

4.1.1 Model coefficients

The parameters of the GARCH models are estimated using the quasi maximum likelihood procedure in order to calculate the forecasted volatilities. The results of the estimations over the total period (1993-2010) are stated below in Table 2.

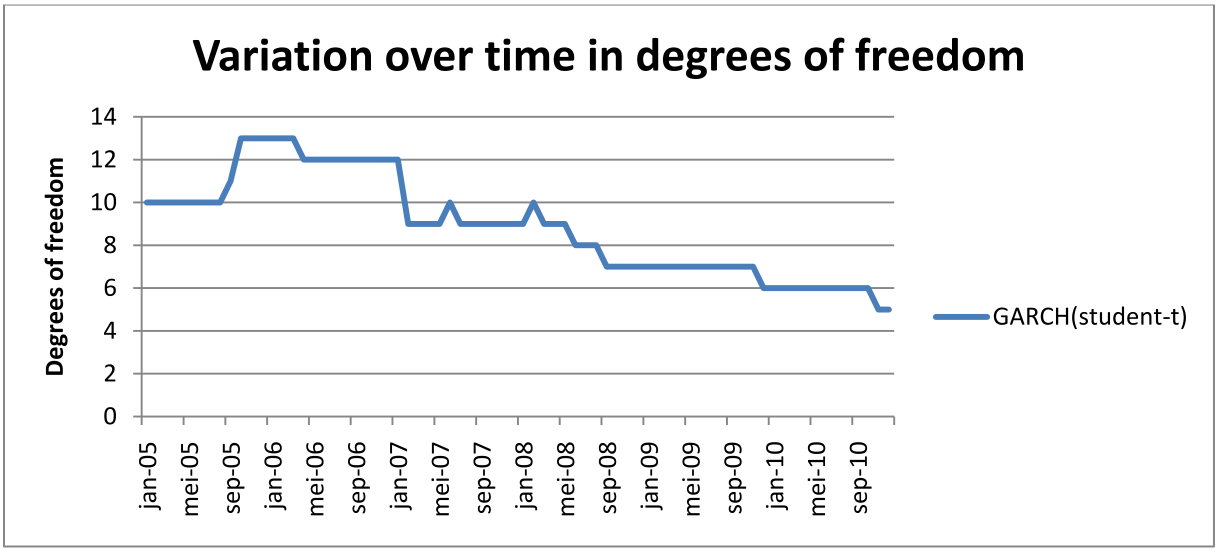
Table 2 Coefficients of the GARCH models

	GARCH-N	GARCH-T	NAGARCH	NAGARCH-VIX	CGARCH
ω	0.000 *** [0.000]	0.000 *** [0.000]	0.000 *** [0.000]	0.000 [0.000]	0.000 *** [0.000]
α	0.071 *** [0.005]	0.067 *** [0.007]	0.065 *** [0.005]	0.062 *** [0.010]	-0.045 *** [0.008]
β	0.923 *** [0.005]	0.932 *** [0.007]	0.851 *** [0.008]	0.518 *** [0.033]	-0.564 *** [0.142]
θ			1.099 *** [0.087]	2.008 *** [0.264]	
γ				0.113 *** [0.012]	
ρ					0.995 *** [0.002]
ϕ					0.074 *** [0.005]

*(Number in parenthesis indicates standard error. The symbols *,** and *** indicate statistical significance at the 0.10, 0.05 and 0.01 levels.)*

Table 2 states the results of the estimations in terms of coefficients and standard errors. Some interesting points arise from Table 2. Firstly, there is only a minor difference between the coefficients of GARCH with a normal distribution and GARCH with a student-t distribution. Only the standard errors are larger with GARCH-T than in GARCH-N. The assumption of another distribution does not make a difference in model coefficients, but could make a difference in the calculation of the VaR. As one can see in Figure 2, the analyzed out of sample period does not seem to follow a normal distribution. Figure 2 gives the number of degrees of freedom for the out of sample period.

Figure 2 Variation over time in degrees of freedom



After January 2007, the degrees of freedom decrease. This means that the tails of the student-t distribution are fatter. That is important to know for the application of value at risk.

Further on Table 2. The positive values for theta indicate the existence of a leverage effect. This means that negative returns influence the volatility more than positive returns, which is in line with Black (1976). Once VIX is added to NAGARCH, the average leverage effect increases from 1,099 to 2,008. This means that with the VIX-effect integrated into the model, negative returns receive even more weight compared to NAGARCH. Overall, the coefficients are all significant except for the ω of NAGARCH-VIX.

4.1.2 Backtesting

After the model coefficients are estimated, the forecasted volatilities are calculated. Subsequently, these forecasted volatilities are back tested and the results of those backtests are stated in Tables 3 and 4.

Table 3 Regression based evaluation of individual models

	Normal regression					Log regression				
Model	Alpha	Beta	R-square	Wald	Rank	Alpha	Beta	R-square	Wald	Rank
MA (20)	0.000*** [0.000]	0.546*** [0.017]	0.403	0.000 ***	7	-1.116*** [0.101]	0.801*** [0.025]	0.399	0.000 ***	7
MA (60)	0.000*** [0.000]	0.488*** [0.022]	0.236	0.000 ***	8	-1.196*** [0.112]	0.788*** [0.028]	0.339	0.000 ***	8
EWMA	0.000 [0.000]	0.612*** [0.019]	0.410	0.000 ***	5	-0.959*** [0.105]	0.844*** [0.026]	0.405	0.000 ***	6
GARCH-N	0.000 [0.000]	0.685*** [0.021]	0.413	0.000 ***	3	-0.523*** [0.183]	0.957*** [0.044]	0.406	0.000 ***	5
GARCH-T	0.000 [0.000]	0.622*** [0.019]	0.407	0.000 ***	6	-0.654*** [0.114]	0.928*** [0.028]	0.407	0.000 ***	4
NAGARCH	0.000 [0.000]	0.663*** [0.017]	0.502	0.000 ***	2	-0.663*** [0.106]	0.911*** [0.026]	0.439	0.000 ***	2
NAGARCH-VIX	**0.000*** [0.000]**	**0.946*** [0.021]**	**0.564**	**0.000** ***	**1**	**-0.488*** [0.106]**	**0.955*** [0.026]**	**0.463**	**0.000** ***	**1**
CGARCH	0.000* [0.000]	0.661*** [0.020]	0.412	0.000 ***	4	-0.493*** [0.118]	0.961*** [0.029]	0.407	0.000 ***	3

*(Number in parenthesis indicates standard error. The Wald test tests for the null hypothesis that α=0 and β=1. The p-value is given for each Wald test. The symbols *,** and *** indicate statistical significance at the 0.10, 0.05 and 0.01 levels.)*

Table 3 states the results of the regression based evaluation of the individual models. The ranking of the individual models is based on the R-squares. Several conclusions can be drawn from Table 3. Firstly, the capturing of the volatility's variation seems to dominate stability, because MA(20) seems to perform better than MA(60) in both evaluation methods. MA(20), EWMA, GARCH-N, GARCH-T and CGARCH are almost equal in their performance based on these two regression methods. This could mean that the long run variance adds little to the forecasted volatilities. Otherwise, the difference between MA(20) and EWMA in comparison to GARCH-N, GARCH-T and CGARCH would have been larger. It is also worth mentioning that the Wald test is rejected for each individual model. Subsequently, that means there is no proof for a perfect predictor among the individual models.

The addition of the leverage effect in NAGARCH seems to have the desired effect because the model is ranked second with a gap to the models that performed worse. When the VIX-variable is also added to the equation for forecasted volatility in NAGARCH-VIX, the results become even better

(0.564 and 0.463). So, the conclusion for the regression based evaluation is that NAGARCH-VIX is the best model.

It is important to rely on more backtesting methods instead of only one backtesting method because one method could give biased results. So, Table 4 gives the results for the error measures.

Table 4 Error measures of individual models

Model	MAE (/10,000)	Rank	RMSE (/10,000)	Rank	MME(U) (/1,000)	Rank	No. of under predictions	MME(O) (/1,000)	Rank	No. of over predictions	Bin. Prob.
MA (20)	1.358	7	3.460	7	2.135	6	409	6.712	5	1,156	0.000
MA (60)	1.516	8	3.722	8	2.286	8	378	7.173	8	1,187	0.000
EWMA	1.321	6	3.215	6	1.999	3	359	6.885	6	1,206	0.000
GARCH-N	1.224	4	3.016	3	2.051	5	337	6.702	4	1,228	0.000
GARCH-T	1.315	5	3.190	5	**1.912**	**1**	319	7.132	7	1,246	0.000
NAGARCH	1.162	2	2.920	2	2.013	4	377	6.279	2	1,188	0.000
NAGARCH-VIX	**1.018**	**1**	**2.418**	**1**	1.982	2	356	**6.061**	**1**	1,209	0.000
CGARCH	1.218	3	3.058	4	2.185	7	364	6.528	3	1,201	0.000

(The binominal probability tests whether the number of under predictions equals the number of over predictions.)

The results of the error measures give almost the same results as the regression based evaluations. MA(60) is also the worst performing model with all the error measures. The differences between MA(20), EWMA, GARCH-N, GARCH-T and CGARCH seem a bit larger than with the regression based evaluations. But NAGARCH-VIX still stands out as a winner by three out of four error measures. Also, NAGARCH-VIX shows the lowest variance in the individual forecasting errors because of the lowest difference between MAE and RMSE.

Only with MME(U), NAGARCH-VIX is the second best model behind GARCH-T. The reason why GARCH-T performs better with MME(U) lies in the number of under predictions, which is the lowest number of all models. On the other hand, GARCH-T performs second worst with MME(O) because of the high number of over predictions. This difference in results is exactly the reason why multiple backtesting methods have been chosen. For example, a seller of a call option is more worried about under prediction than a buyer of the same call option. Vice versa with a put option (Brailsford and Faff (1996)). The seller of a call option would prefer GARCH-T, whereas the buyer of the call option would prefer NAGARCH-VIX. So, there is not one best model for every context. An investor has to make his own choice regarding his own purposes with the forecasted volatilities.

The binominal probability of an equal distribution between over- and under predictions is lower than 0.05, which indicates inequality between the number of over- and under predictions. The conclusion can be drawn that all models give more over predictions than under predictions. In terms of Brailsford and Faff (1996), these individual models are 'biased' forecast models. This leaves room for the new model of this report which incorporates that characteristic.

4.2 Encompassing test

After the evaluation of the individual models, the results of the encompassing test are discussed in this subsection. The encompassing test is needed to evaluate which individual models will be combined in the third stage of this research. Table 5 states the SPSS-output as a result of the encompassing test.

Table 5 SPSS output of the encompassing test

		Coefficient	Std. Error	VIF	R-square
1	(Constant)	0.000	0.000		0.310
	NAGARCH-VIX	0.821***	0.031	1.000	
2	(Constant)	.000**	0.000		0.316
	NAGARCH-VIX	0.591***	0.071	5.361	
	NAGARCH	0.236***	0.066	5.361	
3	(Constant)	0.000**	0.000		0.317
	NAGARCH-VIX	0.577***	0.072	5.390	
	NAGARCH	0.189***	0.068	5.740	
	MA (20)	0.087***	0.033	1.595	
4	(Constant)	0.000	0.000		0.319
	NAGARCH-VIX	0.540***	0.073	5.709	
	NAGARCH	0.279***	0.080	7.901	
	MA (20)	0.244***	0.080	9.663	
	GARCH-N	-0.272**	0.127	13.291	

*(The symbols *,** and *** indicate statistical significance at the 0.10, 0.05 and 0.01 levels.)*

As expected from the evaluation of the individual models separately, NAGARCHVIX is the model that is first selected in the forward-procedure of SPSS. This is done because NAGARCH-VIX has the highest correlation with range based variance in the period where the encompassing test is performed. Table 15 in the appendix reports the correlations with the range based variance as well as the correlations between the individual models. NAGARCH-VIX, NAGARCH, MA(20) and GARCH-N were added to the combination of models in that order. It is important to note that the added value of a combination is only 0.9 when looking at the R square changes.

The other models were not added because they did not add value at a 5% significance level after GARCH-N was added. They could not add value because of the high correlations between several individual volatilities. When the other individual models would have been added, the VIF-values would also have been very high. The R square also did not increase significantly when the other models were added. So in the third stage, NAGARCH-VIX, NAGARCH, MA(20) and GARCH-N are used for the combinations of models.

4.3 Combination models

Now, the individual models for the combinations are chosen, the five different combinations of individual models are constructed. In order to construct the switching method and the MAE method, γ and η were estimated. Table 6 describes the results for these estimations. The estimation over the whole out of sample period is given for η.

Table 6 Coefficients of the switching model and the MAE-model

Switching model

	Mean	Std. Dev.	Min	Max
γ	368,229.2	465,592	0	1,500,000

MAE-model with over prediction correction

	Ma (20)	GARCH-N	NAGARCH	NAGARCH-VIX
η	0.502***	0.399***	0.387***	0.163***
	[0.016]	[0.021]	[0.017]	[0.024]

*((Number in parenthesis indicates standard error. The symbols *,** and *** indicate statistical significance at the 0.10, 0.05 and 0.01 levels. Due to limitations of the Excel solver, the maximum value for γ was set to 1,500,000.)*

When looking at the estimation results of the γ, one can notice that there are times that an equal distribution between the individual models seems to be the best method ($\gamma = 0$), but there are also times where γ seems to go to infinity whereby the best model of time *t* is chosen as model for time $t + 1$. The positive significant values of η indicate a possibility for improvement of the forecasted volatilities of the individual models that are used in the combinations. The estimation of η for NAGARCH-VIX shows the lowest value, which could indicate that NAGARCH-VIX is less vulnerable for clustering of over predictions.

After the estimations for those two models, the forecasted volatilities for all combinations were calculated. Tables 7 and 8 give the backtesting results for the combinations.

Table 7 Regression based evaluation of combinations

Model	Normal regression					Log regression				
	Alpha	Beta	R-square	Wald	Rank	Alpha	Beta	R-square	Wald	Rank
Equal combination	0.000 [0.000]	0.726*** [0.019]	0.489	0.000 ***	5	-0.552*** [0.108]	0.945*** [0.027]	0.446	0.000 ***	3
Regression (50 days)	0.000*** [0.000]	0.265*** [0.013]	0.213	0.000 ***	6	-1.430*** [0.100]	0.696*** [0.024]	0.357	0.000 ***	6
Regression (250 days)	**0.000*** [0.000]**	**0.741*** [0.016]**	**0.580**	**0.000 *** **	**1**	-0.876*** [0.103]	0.828*** [0.024]	0.427	0.000 ***	5
MSE performance method	0.000 [0.000]	0.849*** [0.020]	0.525	0.000 ***	3	**-0.533*** [0.108]**	**0.937*** [0.026]**	**0.448**	**0.000 *** **	**1**
Switching method	0.000 [0.000]	0.878*** [0.021]	0.529	0.000 ***	2	-0.547*** [0.108]	0.929*** [0.026]	0.447	0.000 ***	2
MAE method (o.c.)	0.000 [0.000]	0.977*** [0.024]	0.507	0.634	4	-0.527*** [0.110]	0.915*** [0.026]	0.439	0.000 ***	4

*(Number in parenthesis indicates standard error. The Wald test tests for the null hypothesis that α=0 and β=1. The p-value is given for each Wald test. The symbols *,** and *** indicate statistical significance at the 0.10, 0.05 and 0.01 levels.)*

The first thing worth mentioning from Table 7, is the bad performance of the regression combination (50 days). It performs even worse than the individual model MA(20). So it can be concluded that 50 days is a too small period for the regression combination. Regression combination (250 days) is the only combination model that performs better than every individual model with the normal regression based evaluation.

With the log regression based evaluation, the MSE performance method performs the best. It seems that the MSE performance method does have more outliers than the regression combination (250 days). This conclusion is drawn by looking at the difference in performance between the normal regression and the log regression based evaluation. The same holds for the switching method and the MAE method. Another interesting point from Table 7 is that the null hypothesis of the Wald test only holds for the MAE method in the normal regression. That is the only model in this research where the null hypothesis is not rejected.

Based on these regression based evaluations of the combinations, one could argue the relevance of combinations of forecasted volatilities. However, using more complex techniques seems to pay-off when comparing the equal combination with the other five combinations. In Table 8, the error measures for the combinations are stated.

Table 8 Error measures of combinations

Model	MAE (/10,000)	Rank	RMSE (/10,000)	Rank	MME(U) (/1,000)	Rank	No. of under predictions	MME(O) (/1,000)	Rank	No. of over predictions	Bin. Prob.
Equal combination	1.153	5	2.808	5	**1.902**	**1**	338	6.536	6	1.227	0.000
Regression (50 days)	1.473	6	5.661	6	3.050	6	615	4.961	3	950	0.000
Regression (250 days)	0.924	2	2.539	2	2.773	4	575	4.407	2	990	0.000
MSE performance method	1.025	4	2.555	4	2.207	2	405	5.710	5	1160	0.000
Switching method	0.993	3	**2.520**	**1**	2.350	3	438	5.434	4	1127	0.000
MAE method (o.c.)	**0.898**	**1**	2.540	3	2.948	5	578	**4.267**	**1**	987	0.000

(The binominal probability tests whether the number of under predictions equals the number of over predictions.)

In Table 8, the regression combination (50 days) is again the worst performing model except for the MME(O)-measure, where it is ranked as third combination model. The regression combination (250 days), the switching method and the MAE method perform better than the individual models based on the mean absolute error. The performance of the MAE method is especially worth noting because of the lowest MAE whereas the model was ranked fourth with both regression based evaluations. A reason for this observation could be that the MAE method uses the absolute error of individual models to minimize the forecasting error. The RMSE of the four best ranked models are almost equal but still higher than the NAGARCH-VIX. Also, the variation in the individual forecasting errors is the lowest for NAGARCH-VIX and not for the combinations.

The power of the combinations lies in the MME(O)-measure. The numbers of over predictions of the combinations are lower than the individual models. This leads to a better performance of the combinations based on MME(O). Again, the MAE method performs the best based on MME(O). To recall the option story, the MAE method is most preferred by buyers of call options and sellers of put options. The purpose of the MAE method is thereby partly accomplished by reducing the number of over predictions and increasing the number of under predictions. Based on MME(U), the MSE performance method is the best performing combination model. However, it is dominated by all individual models except for MA(60). So, the power of the combinations does not lie in the MME(U)-measure.

Besides the out of sample period, the in-sample period is also briefly researched. The results are stated in Tables 16 and 17 in the appendix. The results indicate that the MSE performance performs the best with all evaluation criteria except the MME(O). Another conclusion that can be drawn, is that the combinations seem to outperform the individual models in the in-sample period. Furthermore, the relative differences between the results of the models of the in-sample period and the out of sample period are small. NAGARCH-VIX still seems to be the best individual model whereas the other individual models do not seem to differ from each other, just like in the out of sample period.

4.4 Value at Risk

In the previous subsection, theoretical backtesting methods were used to evaluate the models, whereby the option context is used to make the connection with the practicalities of forecasted volatilities. In this subsection, another practical application of forecasted volatilities is used to evaluate the individual models as well as the combinations and in different time periods. Firstly, the numbers of violations are stated. Secondly, the results of the different backtesting methods for VaR are analyzed.

4.4.1 Number of violations

Before the results of the backtesting methods are analyzed, the number of violations for each model is stated. In Table 9, the number of violations under the assumption of a normal distribution is stated.

Table 9 Number of violations (normal distribution)

Period	Period 1 (2005-2007)		Period 2 (2007-2010)		Total period (2005-2010)	
VAR	1%	5%	1%	5%	1%	5%
Number of observations	650	650	915	915	1565	1565
Expected number of violations	7	33	10	46	16	79
Individual model:						
Moving Average (20)	12	38	30	67	42	105
Moving Average (60)	14	33	27	70	41	103
EWMA	14	35	28	62	42	97
GARCH-N	8	28	33	69	41	97
GARCH-T	8	27	29	62	37	89
NAGARCH	11	31	38	75	49	106
NAGARCH-VIX	12	37	31	75	43	112
CGARCH	8	28	30	70	38	98
Combination model:						
Equal combination	11	30	32	66	43	96
Regression combination (50 days)	23	59	61	96	84	155
Regression combination (250 days)	22	52	54	89	76	141
MSE performance method	12	34	37	81	49	115
Switching method	13	38	43	83	56	121
MAE method (over prediction correction)	20	52	52	100	72	152

From Table 9, it can be concluded that the number of violations is higher in the period of high volatility (period 2) than in the period of low volatility (period 1). Another observation is the difference in the number of violations of individual models and combinations. The number of violations is higher for the combinations than for the individual models. Therefore, the combinations seem to be more influenced by negative outliers than the individual models. A reason for this could lie in the observation that combinations have more under predictions than individual models. The chance for an extreme under prediction could be increased because of this.

During the first period, the individual models seem to perform accurately when looking at the small difference between the actual number of violations and the expected number of violations. However, in the second period and therefore also in the total period, the difference between the actual number of violations and the expected number of violations increases. This could be caused by the possibility that the range based variance is not an accurate estimator of the actual variance during periods of high volatility. Another reason for this difference could lie in the assumed normal distribution in the calculation of VaR. Therefore, Table 10 denotes the number of violations under the assumption of a student-t distribution.

Table 10 Number of violations (student-t distribution)

Period	Period 1 (2005-2007)		Period 2 (2007-2010)		Total period (2005-2010)	
VAR	1%	5%	1%	5%	1%	5%
Number of observations	650	650	915	915	1565	1565
Expected number of violations	7	33	10	46	16	79
Individual model:						
Moving Average (20)	10	27	10	50	20	77
Moving Average (60)	9	27	13	48	22	75
EWMA	6	27	7	53	13	80
GARCH-N	3	20	10	56	13	76
GARCH-T	4	20	7	54	11	74
NAGARCH	5	24	14	63	19	87
NAGARCH-VIX	6	30	8	56	14	86
CGARCH	4	21	10	60	14	81
Combination model:						
Equal combination	3	24	7	54	10	78
Regression combination (50 days)	15	50	32	80	47	130
Regression combination (250 days)	13	41	27	71	40	112
MSE performance method	5	27	13	63	18	90
Switching method	6	30	15	68	21	98
MAE method (over prediction correction)	10	36	29	81	39	117

The difference between the actual number of violations and the expected number of violations decreases when the student-t distribution is used in the calculation of VaR. In the first period, six individual models and two combinations even show a lower number of violations than expected. The majority of the number of violations of combinations is higher than those of the individual models, which is the same conclusion as when the normal distribution was used. However, it seems that the importance of the assumption for the distribution outweighs the importance of the forecasted volatility when comparing Tables 9 and 10.

4.4.2 VaR backtesting methods

The violations are used in the backtesting methods for VaR. The results for backtesting methods for VaR with the normal distribution are presented in Tables 11 (first period), 12 (second period) and 20 (appendix) (total period). The results of the backtesting with the student-t distribution are stated in Tables 18, 19 and 21 in the appendix.

Table 11 Results backtesting VaR, first period, normal distribution

Period	Period 1 (2005-2007)					
Alpha	1%			5%		
Test statistic	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**
Individual model:						
Moving Average (20)	3.762	10.930*	14.691*	0.931*	49.342*	50.273*
Moving Average (60)	6.571	22.165	28.736	0.008*	55.580	55.588
EWMA	6.571	21.403	27.974	0.198*	49.981	50.179
GARCH-N	0.326*	5.881*	6.207*	0.687*	39.874	40.561
GARCH-T	0.326*	5.881*	6.207*	1.037*	37.677	38.714
NAGARCH	2.606*	18.965	21.570	0.074*	36.446*	36.520*
NAGARCH-VIX	3.762	19.250	23.012	0.629*	36.347*	36.976*
CGARCH	0.326*	4.797*	5.123*	0.687*	35.310*	35.997*
Combination model:						
Equal combination	2.606*	16.158*	18.763	0.208*	36.275*	36.483*
Regression combination (50 days)	25.557	58.758	84.314	18.517	94.670	113.187
Regression combination (250 days)	23.023	46.459	69.482	10.503	74.376	84.879
MSE performance method	3.762	15.340*	19.102*	0.072*	36.554*	36.626*
Switching method	5.088	18.432*	23.520	0.931*	39.088*	40.019*
MAE method (over prediction correction)	18.242	31.284	49.527	10.503	75.421	85.924

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*

The backtesting results for the first period indicate that only CGARCH is accepted by each test with a 10% significance level. All other models are rejected either for lack of independence between the violations or the number of violations is too high. When the assumption of a normal distribution is replaced by the assumption of a student-t distribution in Table 18 in the appendix, more models are significant at a 10% level for all tests, namely MA(20), EWMA, GARCH-T, NAGARCH, NAGARC-VIX, the equal combination, the MSE performance method and the switching model.

It may seem strange that CGARCH is not significant anymore for all tests when the student-t distribution is applied, but it is not. The unconditional coverage test is not only rejected when the number of violations is too high, but also when it is too low in the case of CGARCH. A reason for this rejection lies in the overabundance of costs that is related to a too high VaR-value. For example, banks and financial institutions have to hold a capital reserve based on the VaR. When the calculated VaR is too high, it brings unnecessary cost with it. Therefore, a too low number of violations is also rejected by the unconditional coverage test.

Lastly, it seems that the regression combinations do seem to perform accurately for VaR purposes in a period of low volatility with both distributional assumptions, where as the regression combination (50 days) performed badly in the regression based evaluation and error measures. To come back at the bank-example, the VaR for banks in the Basel requirements are based on the 1% VaR. Subsequently, that means that all individual models seem to be an accurate model for banks to use.

Table 12 Results backtesting VaR, second period, normal distribution

Period	Period 2 (2007-2010)					
Alpha	1%			5%		
Test statistic	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
Individual model:						
Moving Average (20)	30.030	80.853	110.883	9.145	81.739*	90.884
Moving Average (60)	23.087	89.263	112.350	11.725	111.606	123.332
EWMA	25.328	85.413	110.742	5.495	72.711*	78.205
GARCH-N	37.595	87.115	124.711	10.834	81.678*	92.511
GARCH-T	27.644	77.209	104.853	5.495	71.941*	77.435*
NAGARCH	51.440	102.667	154.107	16.640	91.776	108.416
NAGARCH-VIX	32.486	78.265	110.751	16.640	80.978*	97.618
CGARCH	30.030	73.037	103.068	11.725	81.275*	93.001
Combination model:						
Equal combination	35.008	79.198	114.206	8.349	74.380*	82.729
Regression combination (50 days)	130.775	197.094	327.869	44.765	130.883	175.648
Regression combination (250 days)	104.283	151.341	255.624	34.138	111.602	145.740
MSE performance method	48.555	89.197	137.752	23.493	92.794*	116.287
Switching method	66.661	107.556	174.217	25.997	102.583	128.580
MAE method (over prediction correction)	97.059	151.311	248.369	51.354	135.517	186.872

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*

Table 12 describes the results of the backtesting for the second period. The first thing one can see, is the rejection of all unconditional coverage tests. Although, the individual models show some significance in the independence test with the 5% VaR, the assumption of normality does not seem to be accurate in a VaR-setting. Table 19 in the appendix seems to show more proof for that observation. The results for the backtesting with the student-t distribution give more significant results in Table 19. The unconditional coverage test is accepted for all individual models with the 1% VaR. However, the independence between the violations seems to be the problem in a high volatility period for some of the models. Only MA(20), EWMA, GARCH-T and NAGARCH-VIX show significance at a 10% level for all tests.

The change from a normal distribution to a student-t distribution did not help the combination much. Only the unconditional coverage test of MSE performance method and the equal combination in the 1% VaR is accepted but it seems that the violations do not happen independently of each other. So, the conclusion could be drawn that (some) individual models outperform the combinations in a VaR setting in a high volatility period.

The previously drawn conclusion also seems to hold when one looks at the total period results in Table 20 and Table 21 in the appendix. MA(20), EWMA, GARCH-T and NAGARCH-VIX are the only four models that passed each test with the assumption of student-t distribution. The importance of the choice for distribution seems to be large when comparing the results of both distributions. The assumption of the distribution seems to be more important than the forecasted volatility in the VaR-setting. It is also important to notice that the results of the backtesting of the practical application VaR give different results than the other backtesting methods. There does not seem to be a best model for every context. It depends on the context and purposes of the user of forecasted volatilities.

Chapter 5 Conclusion

This paper studies the forecasting accuracy of individual models and combinations. In particular, it studies whether combinations are superior to individual models in forecasting the volatility in an out of sample period. The results of the different backtesting methods show that there is not one conclusive answer to this question. Based on an evaluation by the mean absolute error, the MAE method (with a correction for over prediction) is superior to the other models. This could be caused by minimizing the forecast errors with the absolute errors of the individual models in the MAE method. The good performance of the MAE-measure provides evidence for the observation that the chosen individual models seem to over predict the actual volatility more than they under predict. This observation is captured in the new MAE-model of this paper.

The MAE method does also seem to be the best model with the MME(O)-measure. So, it is assumed that the MAE method is the model that is preferred by buyers of call options and sellers of put options. Sellers of call options and buyers of put options worry more about under predictions. Subsequently, looking at the MME(U)-measure, individual models are superior to combinations. So it is important to think about the purpose of the forecasted volatility before a model is chosen.

Besides the option context, this study focuses on another practical application, namely Value at Risk. In the VaR context, the choice for the distribution seems more important than the forecasted volatility. The student-t distribution outperforms the normal distribution, especially in periods with high volatility. Furthermore, the individual models seem to predict the tails of the return distribution more accurate than the combinations.

So based on this study, it seems that there is no superior model between individual models and combinations for every context when examining the forecasted volatilities. In further research, other individual models could be used to form combinations. Some of the individual model that have been used for this study, seem to have many similarities when examining at the correlation matrix. A way to extend this study in further research, is to take a look at emerging markets and see if the conclusions of this study still holds.

References

Angelidis, T., Benos, A. and Degiannakis, S. (2004). The use of GARCH models in VaR estimation, *Statistical Methodology* 1: 105-128.

Angelidis, T. and Degiannakis, S. (2005). Modeling risk for long and short trading positions, *Journal of Risk Finance* 6: 226-238.

Bannouh, K., van Dijk, D., and Martens, M. (2009). Range-based covariance estimation using high-frequency data: the realized co-range, *Journal of Financial Econometrics* 7: 341-372.

Bates, J.M. and Granger, C.W.J. (1969). The combination of Forecasts, *Operations Research Quarterly* 20: 451-468.

Becker, R. and Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior?, *International Journal of Forecasting* 24: 122–133.

Black, F. (1976). Studies in stock price volatility changes, *American Statistical Association*: 177-181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics 31*: 307-327.

Brailsford, T.J. and Faff, R.W. (1996). An evaluation of volatility forecasting techniques, *Journal of Banking & Finance* 20: 419-438.

Brunnermeier, M.K. (2009). Deciphering the Liquidity and Credit Crunch 2007-2008, *Journal of Economic Perspectives* 23: 77-100.

Christoffersen, P.F. (2003). Elements of Financial Risk Management, Elsevier.

Chong, Y.Y. and Hendry, D.F. (1986). Econometric Evaluation of Linear Macro-Economic Models, *The review of Economic Studies* 53: 671-690.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of

United Kingdom Inflation, *Econometrica* 50: 987-1007.

Engle, R.F. and Ng, V.K. (1993). Measuring and Testing of the Impact of News on Volatility,

Journal of Finance 48: 1749-1778.

Engle, R.F. and Lee, G.G.J. (1999). A long-run and short-run component model of stock return volatility, in co integration, causality and forecasting, *Oxford University Press*.

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

Frijns, B., Lehnert, T. and Zwinkels, R.C.J. (2008). Behavioral Heterogeneity in the Options Market, *Journal of Economic Dynamics and Control*, forthcoming.

Granger, C.W.J. and Ramanathan, R. (1984). Improved Methods of Combining Forecasts, Journal of Forecasting 3: 197-204

Haas, M. (2001). New Methods in Backtesting. *Center of Advanced European Studies and Research*.

Hao, J. and Zhang, J. E. (2010). GARCH Option Pricing Models, the CBOE VIX and Variance Risk Premium.

Hull, J.C. (2008). Options, Futures and Other Derivatives, Pearson Education, 7th Edition.

Hung, C., Tzang, S. and Hsyu, D.S. (2009). Forecasting Efficiency of Implied Volatility and the Intraday High-Low Price Range in Taiwan Stock Market, *International Research Journal of Finance and Economics* 27: 192-202.

Jorion, P. (2001). Value at Risk. The new benchmark for controlling derivatives risk, McGraw Hill.

JP Morgan (1996). RiskMetrics-technical document, 4th ed. New York.

Kovačić, Z.J. (2008). Forecasting Volatility: Evidence from the Macedonian Stock Exchange, *International Research Journal of Finance and Economics* 18: 182-212.

McMillan, D., Speight, A. and Apgwilym, O. (2000). Forecasting UK stock market volatility, *Applied Financial Economics* 10: 435-448.

O'Brien, R.M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors, *Quality & Quantity* 41: 673-690

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

Poon, S. and Granger, C.W.J. (2003). Forecasting Financial Market Volatility: A Review, Journal of Economic Literature 61: 478-539.

Shin, H.W. and Sohn, S.Y. (2007). Application of an EWMA combining technique to the prediction of currency exchange rates, *IIE transactions* 39: 639-644.

Stock, J.H. and Watson, M.W. (2001). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, *Cambridge University Press*: 1-44.

Timmermann, A. (2006). Forecast combinations, *Handbook of Economic Forecasting*, Elsevier.

Veiga B., Chan, F. and McAleer, M. (2005). It pays to violate: model choice and critical value assumption for forecasting value-at-risk thresholds, *Zerger A, Argent R. (Eds.), MODSIM 2005 International Congress on Modeling and Simulation*: 2304–2311.

Appendix

Table 13 Autocorrelation of the squared returns

	Total out of sample period			2005-2007			2007-2010		
Lag	AC	Q-Stat	Prob.	AC	Q-Stat	Prob	AC	Q-Stat	Prob
1	0.198	61.507	0.000	0.003	0.0056	0.940	0.169	26.096	0.000
2	0.382	290.46	0.000	0.002	0.0089	0.996	0.360	145.46	0.000
3	0.180	341.35	0.000	0.068	30.105	0.390	0.149	165.96	0.000
4	0.281	465.54	0.000	0.086	78.949	0.096	0.255	225.66	0.000
5	0.354	662.12	0.000	0.079	11.955	0.035	0.330	326.16	0.000
6	0.312	814.85	0.000	0.041	13.040	0.042	0.286	401.88	0.000
7	0.349	1006.3	0.000	0.009	13.088	0.070	0.325	499.69	0.000
8	0.260	1112.9	0.000	-0.017	13.284	0.102	0.233	549.91	0.000
9	0.325	1279.1	0.000	0.032	13.963	0.124	0.300	633.34	0.000
10	0.277	1400.2	0.000	0.235	50.655	0.000	0.249	690.98	0.000
11	0.384	1633.5	0.000	0.008	50.695	0.000	0.362	812.91	0.000
12	0.318	1793.3	0.000	0.025	51.096	0.000	0.293	892.73	0.000
13	0.261	1901.2	0.000	0.030	51.710	0.000	0.234	943.48	0.000
14	0.140	1932.1	0.000	0.017	51.903	0.000	0.107	954.10	0.000
15	0.219	2007.7	0.000	-0.011	51.980	0.000	0.189	987.50	0.000

Table 14 Autocorrelation of the range based variance

	Total out of sample period			2005-2007			2007-2010		
Lag	AC	Q-Stat	Prob.	AC	Q-Stat	Prob	AC	Q-Stat	Prob
1	0.637	635.30	0.000	0.126	10.326	0.001	0.616	347.89	0.000
2	0.552	1113.9	0.000	0.167	28.552	0.000	0.525	601.41	0.000
3	0.564	1613.8	0.000	0.122	38.235	0.000	0.538	867.76	0.000
4	0.578	2138.1	0.000	0.105	45.424	0.000	0.552	1148.8	0.000
5	0.563	2636.6	0.000	0.156	61.345	0.000	0.536	1414.1	0.000
6	0.484	3005.1	0.000	0.066	64.195	0.000	0.452	1602.9	0.000
7	0.513	3418.6	0.000	0.061	66.642	0.000	0.483	1818.1	0.000
8	0.495	3805.2	0.000	0.034	67.383	0.000	0.464	2017.6	0.000
9	0.522	4234.5	0.000	0.074	70.973	0.000	0.492	2242.1	0.000
10	0.423	4516.3	0.000	0.204	98.563	0.000	0.386	2380.0	0.000
11	0.429	4807.0	0.000	0.073	102.05	0.000	0.393	2523.4	0.000
12	0.540	5266.8	0.000	0.078	106.12	0.000	0.511	2765.9	0.000
13	0.413	5535.9	0.000	0.055	108.16	0.000	0.375	2896.9	0.000
14	0.335	5713.0	0.000	0.069	111.29	0.000	0.292	2976.1	0.000
15	0.376	5936.6	0.000	0.042	112.44	0.000	0.336	3081.2	0.000

Figure 3 Average yearly VIX-values

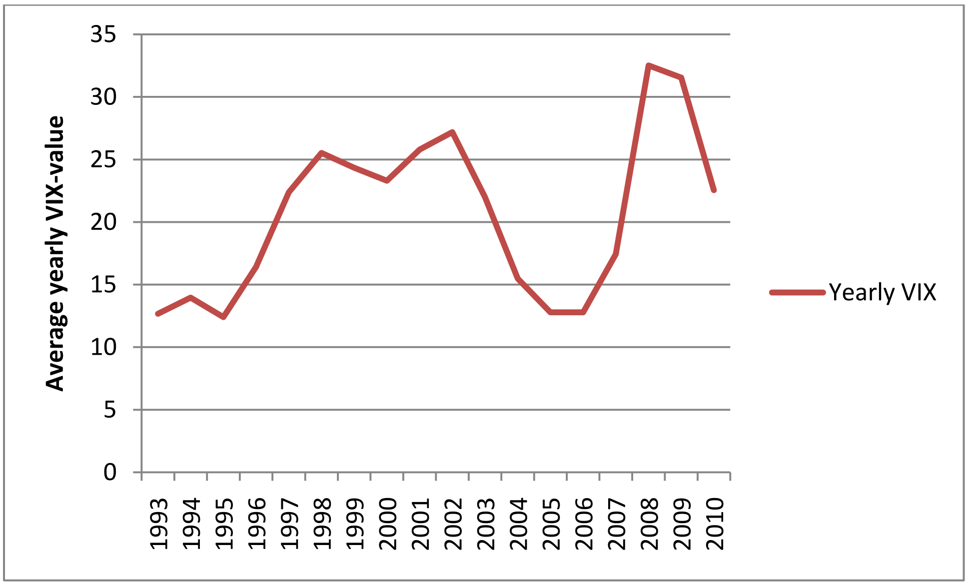


Table 15 Correlation of the individual models

	RBV	MA(20)	MA(60)	EWMA	GARCH-N	GARCH-T	NAGARCH	NAGARCH-VIX	CGARCH
RBV	1.000	0.379	0.338	0.411	0.413	0.408	0.534	0.557	0.411
MA(20)	0.379	1.000	0.772	0.956	0.936	0.944	0.608	0.574	0.924
MA(60)	0.338	0.772	1.000	0.878	0.779	0.818	0.519	0.593	0.753
EWMA	0.411	0.956	0.878	1.000	0.967	0.980	0.678	0.665	0.946
GARCHN	0.413	0.936	0.779	0.967	1.000	0.996	0.737	0.656	0.982
GARCHT	0.408	0.944	0.818	0.980	0.996	1.000	0.712	0.646	0.974
NAGARCH	0.534	0.608	0.519	0.678	0.737	0.712	1.000	0.902	0.728
NAGARCHVIX	0.557	0.574	0.593	0.665	0.656	0.646	0.902	1.000	0.645
CGARCH	0.411	0.924	0.753	0.946	0.982	0.974	0.728	0.645	1.000

Table 16 Regression based evaluation of the in sample period

Model	Normal regression					Log regression				
	Alpha	Beta	R-square	Wald	Rank	Alpha	Beta	R-square	Wald	Rank
MA (20)	0.000*** [0.000]	0.495*** [0.020]	0.173	0.000	12	-1.268*** [0.091]	0.759*** [0.022]	0.282	0.000	12
MA (60)	0.000*** [0.000]	0.561*** [0.025]	0.145	0.000	13	-1.138*** [0.100]	0.797*** [0.024]	0.260	0.000	13
EWMA	0.000*** [0.000]	0.619*** [0.022]	0.207	0.000	9	-1.009*** [0.095]	0.826*** [0.023]	0.298	0.000	9
GARCH-N	0.000*** [0.000]	0.648*** [0.022]	0.215	0.000	8	-0.488*** [0.109]	0.962*** [0.027]	0.300	0.000	8
GARCH-T	0.000*** [0.000]	0.639*** [0.023]	0.204	0.000	10	-0.668*** [0.104]	0.918*** [0.026]	0.297	0.000	10
NAGARCH	0.000 [0.000]	0.732*** [0.020]	0.310	0.000	4	-0.507*** [0.103]	0.953*** [0.025]	0.321	0.000	6
NAGARCH-VIX	0.000*** [0.000]	0.958*** [0.023]	0.356	0.000	3	-0.128 [0.105]	1.039*** [0.025]	0.353	0.000	2
CGARCH	0.000*** [0.000]	0.612*** [0.022]	0.197	0.000	11	-0.661*** [0.106]	0.920*** [0.026]	0.292	0.000	11
Equal combination	0.000 [0.000]	0.764*** [0.022]	0.282	0.000	5	-0.364*** [0.103]	0.988*** [0.025]	0.336	0.000	4
Regression (total period)	0.000 [0.000]	1.000*** [0.024]	0.358	0.000	2	-1.005*** [0.087]	0.798*** [0.020]	0.337	0.000	3
MSE performance method	**0.000*** [0.000]**	**0.989*** [0.022]**	**0.388**	**0.000**	**1**	**-0.097 [0.098]**	**1.041*** [0.024]**	**0.390**	**0.000**	**1**
Switching method	0.000* [0.000]	0.810*** [0.023]	0.281	0.000	6	-0.553*** [0.100]	0.926*** [0.024]	0.326	0.000	5
MAE method (o.c.)	0.000*** [0.000]	0.911*** [0.027]	0.274	0.000	7	-0.746*** [0.100]	0.862*** [0.023]	0.307	0.000	7

Table 17 Error measures of the in sample period

Model	MAE (/10,000)	Rank	RMSE (/10,000)	Rank	MME(U) (/1,000)	Rank	No. of under predictions	MME(O) (/1,000)	Rank	No. of over predictions	Bin. Prob.
MA (20)	0.762	12	1.486	13	2.047	10	876	5.300	6	2186	0.000
MA (60)	0.769	13	1.440	12	1.997	9	809	5.538	10	2253	0.000
EWMA	0.730	8	1.384	9	1.921	8	785	5.392	7	2277	0.000
GARCH-N	0.739	9	1.373	8	1.779	6	688	5.714	11	2374	0.000
GARCH-T	0.753	10	1.386	10	1.760	4	679	5.791	13	2383	0.000
NAGARCH	0.684	7	1.276	6	1.734	2	719	5.458	9	2343	0.000
NAGARCH-VIX	0.611	4	1.181	3	1.860	7	756	4.947	5	2306	0.000
CGARCH	0.757	11	1.402	11	1.775	5	688	5.783	12	2374	0.000
Equal combination	0.677	6	1.282	7	1.744	3	713	5.398	8	2.349	0.000
Regression (total period)	0.540	2	1.164	2	2.572	12	1182	**3.567**	**1**	1880	0.000
MSE performance method	**0.514**	**1**	**1.148**	**1**	**1.515**	**1**	816	4.319	3	2246	0.000
Switching method	0.615	5	1.252	5	2.145	11	909	4.564	4	2153	0.000
MAE method (o.c.)	0.575	3	1.240	4	2.718	13	1175	3.645	2	1886	0.000

Table 18 Results backtesting VaR, first period, student-t distribution

Period	Period 1 (2005-2007)					
Alpha	1%			5%		
Test statistic	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**
Individual model:						
Moving Average (20)	1.635*	5.040*	6.675*	1.037*	27.957*	28.994*
Moving Average (60)	0.867*	11.854*	12.722*	1.037*	49.469	50.506
EWMA	0.040*	3.717*	3.757*	1.037*	33.112*	34.149*
GARCH-N	2.380*	2.482*	4.861*	5.831	19.866*	25.697*
GARCH-T	1,126*	2,580*	3,705*	5,831*	19,866*	25,697*
NAGARCH	0.380*	4.486*	4.865*	2.564*	20.934*	23.497*
NAGARCH-VIX	0.040*	6.415*	6.455*	0.208*	24.614*	24.821*
CGARCH	1.126*	2.580*	3.705*	4.871	29.019*	33.890
Combination model:						
Equal combination	2.380*	2.482*	4.861*	2.564*	17.591*	20.155*
Regression combination (50 days)	8.200	29.861	38.061	8.579	63.345	71.924
Regression combination (250 days)	5.088	24.400	29.488	2.169*	52.849*	55.018
MSE performance method	0.380*	1.594*	1.973*	1.037*	17.305*	18.342*
Switching method	0.040*	2.428*	2.468*	0.208*	22.166*	22.374*
MAE method (over prediction correction)	1.635*	18.141	19.776	0.384*	34.253*	34.637*

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*

Table 19 Results backtesting VaR, second period, student-t distribution

Period	Period 2 (2007-2010)					
Alpha	1%			5%		
Test statistic	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**
Individual model:						
Moving Average (20)	0.077*	10.782*	10.859*	0.404*	54.865*	55.269*
Moving Average (60)	1.447*	40.490	41.937	0.115*	77.268	77.383
EWMA	0.555*	9.130*	9.686*	1.153*	57.350*	58.503*
GARCH-N	0.077*	23.517	23.594	2.263*	63.631*	65.894*
GARCH-T	0,555*	9,130*	9,686*	1,484*	61,088*	62,572*
NAGARCH	2.235*	26.472	28.707	6.157	70.811*	76.968*
NAGARCH-VIX	0.152*	9.729*	9.881*	2.263*	50.661*	52.924*
CGARCH	0.077*	24.143	24.221	4.273	65.579*	69.852*
Combination model:						
Equal combination	0.555*	14.806	15.362	1.484*	55.382*	56.866*
Regression combination (50 days)	35.008	77.900	112.908	22.281	106.574	128.855
Regression combination (250 days)	23.087	59.314	82.401	12.648	76.721*	89.369
MSE performance method	1.447*	28.627	30.074	6.157	63.485*	69.642*
Switching method	3.167	28.046	31.213	9.973	62.069*	72.042*
MAE method (over prediction correction)	27.644	67.228	94.872	23.493	90.183*	113.676

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*

Table 20 Results backtesting VaR, total period, normal distribution

Period	Total period (2005-2010)					
Alpha	1%			5%		
Test statistic	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**
Individual model:						
Moving Average (20)	30.675	91.783	122.458	8.735	131.081	139.816
Moving Average (60)	28.691	111.428	140.120	7.527	167.186	174.713
EWMA	30.675	106.816	137.491	4.409	122.692	127.101
GARCH-N	28.691	92.997	121.688	4.409	121.552	125.961
GARCH-T	21.269	83.091	104.359	1.491*	109.617	111.109
NAGARCH	45.875	121.632	167.507	9.370	128.222	137.591
NAGARCH-VIX	32.708	97.515	130.224	13.596	117.325*	130.921*
CGARCH	23.045	77.835	100.879	4.875	116.585	121.460
Combination model:						
Equal combination	32.708	95.356	128.064	3.965	110.656*	114.621*
Regression combination (50 days)	148.659	255.852	404.511	62.422	225.553	287.975
Regression combination (250 days)	121.882	197.800	319.682	43.243	185.978	229.221
MSE performance method	45.875	104.537	150.413	15.971	129.349*	145.320
Switching method	63.147	125.988	189.135	21.225	141.671	162.895
MAE method (over prediction correction)	109.147	182.595	291.742	58.068	210.939	269.007

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*

Table 21 Results backtesting VaR, total period, student-t distribution

Period	Total period (2005-2010)					
Alpha	1%			5%		
Test statistic	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**	**LR_{uc}**	**LR_{ind}**	**LR_{cc}**
Individual model:						
Moving Average (20)	1.123*	15.822*	16.944*	0.021*	82.822*	82.843*
Moving Average (60)	2.311*	52.344	54.655	0.144*	126.737	126.881
EWMA	0.481*	12.848*	13.329*	0.041*	90.462*	90.503*
GARCH-N	0.481*	25.998	26.479	0.069*	83.497*	83.565*
GARCH-T	1,557*	11,710*	13,267*	0,247*	80,953*	81,201*
NAGARCH	0.678*	30.958	31.636	0.995*	91.745*	92.740*
NAGARCH-VIX	0.182*	16.144*	16.326*	0.784*	75.274*	76.058*
CGARCH	0.182*	26.723	26.905	0.101*	94.597*	94.698*
Combination model:						
Equal combination	2.363*	17.288	19.651	0.001*	72.973*	72.974*
Regression combination (50 days)	41.308	107.761	149.070	30.305	169.919	200.224
Regression combination (250 days)	26.757	83.715	110.472	13.596	129.570*	143.166
MSE performance method	0.340*	30.220	30.560	1.775*	80.789*	82.565*
Switching method	1.669*	30.474	32.143	4.875	84.235*	89.110*
MAE method (over prediction correction)	24.875	85.369	110.244	17.649	124.436*	142.085

*(The symbol * means that the model is accepted by that backtesting method with a 10% significance level)*