



UNIVERSITY OF AGDER

A Comparative Study of Volatility Forecasting Models

June 1, 2014

Rune Haddeland Einarsen

Supervisor

Valeriy I Zakamulin

This master's thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education. However, this does not imply that the University answers for the methods that are used or the conclusions that are drawn.

University of Agder, 2014

School of Business and Law

Abstract

The purpose of this thesis is to investigate which of the selected models that forecasts the out-of-sample volatility most accurate and to see if the regression based models can outperform the historical volatility models. Using the data from the S&P500, NASDAQ Composite, DJIA, CBOE Interest Rate, LBMA Gold and USD/GBP return series. The data is forecasted under different distribution assumptions and then evaluated against each other. Trough this thesis, it can be concluded that the asymmetric GJR-GARCH under Student-t distribution most accurately describes the S&P500 and DJIA while GJR-GARCH under the normal distribution provides the most accurate forecast for NASDAQ Composite. The asymmetric EGARCH under Student-t distribution and under normal distribution most accurately describes the CBOE Interest Rate return series and the LBMA Gold return series respectively. When it comes to the USD/GBP return series the EWMA model provided the best forecast. Among the models classified as historical volatility models in this thesis only the EWMA model could compete with the asymmetric GARCH models by being the preferred model for one of the series and close in terms of MSE for the other series.

Contents

1	Introduction	5
2	Literature review	6
3	Models	8
3.1	SMA	8
3.2	EWMA	8
3.3	ARMA	9
3.4	ARCH	10
3.5	GARCH	10
3.6	EGARCH	11
3.7	GJR-GARCH	11
3.8	TGARCH	11
4	Statistical Tests	12
4.1	Jarque-Bera test	12
4.2	ADF	12
4.3	ARCH LM test	13
4.4	Test for asymmetries in volatility	13
5	Data	14
6	Forecast Evaluation	16
6.1	Data and sampling procedure for the model based forecast	16
6.2	Forecasting procedure	17
6.3	Model specifications and tests	18
6.3.1	Naïve models	18
6.3.2	In-sample tests	18
6.3.3	Regression based forecast models	19
6.4	Problems with ARCH-family models	20
6.5	Evaluation with loss functions	20

7	Empirical results and analysis	21
7.1	In-sample test results	21
7.2	In-sample parameter estimates	23
7.2.1	S&P500 GJR-GARCH(1,1)-skew.Student-t	23
7.2.2	NASDAQ Composite GJR-GARCH(1,1)-norm	23
7.2.3	DJIA GJR-GARCH(1,1)-skew.Student-t	24
7.2.4	CBOE Interest Rate EGARCH(1,1)-skew.Student-t	24
7.2.5	LBMA Gold EGARCH(1,1)-norm	25
7.3	Out-of-sample comparison	26
7.3.1	Evaluation models for S&P500 index	26
7.3.2	Evaluation models for NASDAQ Composite index	29
7.3.3	Evaluation models for DJIA index	32
7.3.4	Evaluation models for CBOE Interest Rate	35
7.3.5	Evaluation models for USD/GBP	38
7.3.6	Evaluation models for LBMA Gold	41
8	Discussion	44
9	Conclusion	45

1 Introduction

Volatility forecasting is an important component in asset allocation, risk management and option pricing. High volatility means large deviations from the mean and deviation implies risk (see Figlewski (1997)[1]). For instance in the Black and Scholes (1973)[2] options pricing model volatility is the only unknown variable so the one with the most accurate volatility estimate is able to pinpoint the theoretical price most accurately and this has a lot of potential value. In modern portfolio theory when minimizing risk for a given level of expected return using Markowitz (1952)[3] optimization the return volatility is an important part of the optimization calculations. Similarly in the other fields obviously the better the prediction the better you will perform. This has pushed practitioners and researchers to develop a plethora of forecasting models as new features common to financial data has been discovered and technology have improved. It has still proven hard to beat the naive models of predicting next months volatility using the previous months volatility.

Previous research is ambiguous in its conclusions regarding what model predicts the volatility most accurately as shown in Poon and Granger (2003)[4] where the findings of 93 papers on the subject were summarized. This makes it meaningful to test several models against the naive models on a set of different financial time-series. There are several factors that are important when modeling volatility that might be the cause of the ambiguity in previous research results: different forecast horizons, data, sample periods and proxy for realized volatility are used to mention some.

The purpose of this thesis is to investigate which of the fitted models forecast the out-of-sample volatility most accurate and to see if the regression based models can outperform the historical volatility models. The models used in in this thesis is RW (random walk), SMA (simple moving average), EWMA (exponentially weighted moving average), ARMA (auto-regressive moving average), GARCH (generalized auto regressive conditional heteroskedasticity), exponential GARCH, Threshold GARCH and Glosten-Jagannathan-Runkle GARCH. These are some of the models that have been recommended in previous papers. There is a lot of models not tested in this thesis, so there might still be superior models available for forecasting the out-of-sample one month ahead volatility of the return data. It is important that when evaluating the models to have an accurate estimate of the true variance in the out-of-sample period and this is computed on the base of daily returns of each month. To investigate the phenomena of non-normal distribution in financial time-series data, the ARCH-family models are

also analyzed under different distributions. The distributions include the normal distribution, and distributions that allow for more skewness and kurtosis the skewed Student-t, and skewed general error distribution. As shown in Section 5.1 non of the return series follow the normal distribution manly due to large excess kurtosis and some skewness.

Following the introduction part is Section 2 is the literature review where I will present some previous research on volatility forecasting, some stylized facts of financial time-series, models and the tests used in this thesis. In Section 3 and 4 the models and tests used is presented. Section 5 presents the data and the descriptive statistics of the daily return series. In Section 6 I will describe the forecast evaluation method used to generate the results. Section 7 is a presentation of the results starting with the in-sample tests and parameter estimation followed by they out-of-sample comparison of the models. Section 8 is a discussion of the findings contrasted to findings of others, and pointing out new questions that could be explored. In Section 9 you find the conclusion with a short summary of the approach and the main findings. Then the acknowledgments are given followed by the reference list and an Appendix where the R code used to generate the results and the full URLs to the data sources can be found.

2 Literature review

Modeling volatility have for a long time been of grate interest among researchers. It was discovered that the volatility in financial where clustering meaning that for instance low volatility where more likely to be followed by low volatility than high and so on. Volatility clustering - a phenomenon in financial time-series modeling is that one turbulent trading day tends to be followed by another and vice versa concerning tranquil periods Poon (2005)[5]. In 1971 Box and Jenkins[6] popularized the ARMA model which could be used to identify a linear process that could have generated a given time series provided it is stationary. The ARMA model gained attention as a method for capturing the volatility movement. This volatility clustering can in principle be captured by ARMA models but it could breach the non-negativity cosnstrant and it had problems besting the naive benchmark model (RW). This lead to the development of new volatility forecasting models. The ARCH model by Engle (1982)[7] was the first one, and it was groundbreaking due to its ability to explain the non-linear dynamics of financial data in a better way. There are limitations to the ARCH-model due to the

possibility of it breaking the non-negativity constraints and need for many variables to capture the dynamics. A more parsimonious model that avoids over-fitting and is less likely to breach the non-negativity constraints than the ARCH-model was presented by Bollerslev and Taylor (1986)[8] the GARCH-model.

One weakness with the GARCH-model is that it is symmetric, meaning if shocks with different signs and the same magnitude have different effects on the volatility it can't capture it. This is called the leverage effect in financial data. Leverage effect - negative news leading to a fall in stock price shifts a firm's debt to equity ratio upwards. This means that the firm have increased leverage and thus higher risk. Christie (1982)[9] acknowledged this effect as the "leverage effect", and reference the observation that stock price volatility increases more from a negative shock than a positive shock of the same magnitude. This effect gave rise to the asymmetric GARCH-models. One such model is the exponential GARCH presented by Nelson (1991)[10] the model works on logarithmic volatility values and can not break the non-negativity constraints. The Threshold-GARCH by Zakoian 1991[11] and the GJR-GARCH by Glosten, Jagannathan, and Runkle (1993)[12] use an indicator function I to model the positive and negative shocks on the conditional variance asymmetrically. Many other asymmetric models have since been produced.

Most financial return series suffer from excess kurtosis and skewness. Kurtosis - also known as "the fourth moment", measures how fat the tails of the distribution are (see Brooks (2008)[13]). Skewness - also known as "the third moment", measures how the distribution deviates from its mean value (see Dowd (2005)[14]). This leads to data that does not follow the normal distribution and justify testing models using different distributions.

Poon and Granger (2003)[4] summarized the findings of 93 papers on volatility forecasting and found that GARCH-models perform better than ARCH-models, while asymmetric GARCH-models outperform symmetric GARCH-models. They did however not find the results homogenous. Villhelmsson (2006)[15] suggests that one reason for this is that researchers use different data, sampling periods, sample frequency and forecast horizon. He also states the proxy used for ex-post variance, the loss function and distribution used can influence the result. The results from Poon and Granger (2003)[4] show that the historical volatility models often outperform the more complex regression based models. For instance in the case of the studies where historical volatility models were compared to

the GARCH models the historical volatility models where preferred in almost 50% of the cases

3 Models

Here I present the models used with an addition of the original ARCH-model. The ARCH-family and ARMA formulas presented are in line with the formulas used by the rugarch package in R[16].

3.1 SMA

Simple moving average model takes the equally weighted average of the returns in a given historical period and presents this average as the prediction for the next periods volatility. The model can be presented as:

$$r_t = \mu + \varepsilon_t \tag{1}$$

$$\sigma_t^2 = \frac{\sum_{l=1}^T r_{t-l}^2}{T}, \tag{2}$$

where r is the daily returns consist of a mean μ and ε which account for the volatility. $t = 1, \dots, T$ denotes the time at which each return is measured. For daily returns $\mu \approx 0$ which is why the model is estimated only using r_{t-l}^2 and not $(r_{t-l} - E[r])^2$. The square root of time rule can be used to extend the the forecast horizon.

This models main weakness is that it dose not react well to large jumps in financial asset prices. When a long averaging period is used, the importance of a single extreme event is averaged out within a large sample of returns (see Frank J. Fabozzi (2008)[17]).

The random walk model is estimated using SMA with one month look back. The model assumes that the best forecast of next period's volatility is the volatility of the period prior $\sigma_{t+1}^2 = \sigma_t^2$.

3.2 EWMA

The EWMA model is similar to the SMA model the only difference lies in the weighting of the observations. The model tries to address the unresponsiveness that the SMA model have to extreme

events by implementing exponential weighting such that the data points gets exponentially more weight the more recent they are. The model can be presented as:

$$\sigma_t^2 = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{t-i}^2, \quad (3)$$

or in the form of recursions:

$$\sigma_t^2 = (1 - \lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2, \quad (4)$$

Where r is the daily returns assumed to be normally distributed with a zero mean and $t = 1, \dots, \infty$ denotes the time at which each return is measured. The λ is a constant, $0 < \lambda < 1$, called the smoothing or the decay constant. The term $(1 - \lambda)r_{t-1}^2$ determines the intensity of reaction of volatility to market events. The smaller the λ the more the volatility responds to the market information in yesterday's return. $\lambda\sigma_{t-1}^2$ is the persistence term that determines how much of yesterdays volatility will carry over to today irrespective of what happens in the market. The closer λ is to 1, the more prescient is volatility following a market shock (see Frank J. Fabozzi (2008)[17]). Similar to the SMA model a h-day ahead forecast can be obtained by using the square root of time rule.

The shortcomings of this model lies in the assumptions used as they contradict the features presented in the stylized facts of Section 2.2 that return series hardly ever is i.i.d. When it comes to the decay constant, one can use the λ presented in the RiskMetrics™ or try to estimate it.

3.3 ARMA

The ARMA-model popularized by Box and Jenkins (1971)[6] has been suggested to provide a good forecast of volatility in a lot of academic literature. It is less complex than many other models and relatively simple to implement. The ARMA(p,q)-model can be written as:

$$\sigma_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i \sigma_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (5)$$

where $|\phi| < 1$ and $|\theta| < 1$ for a stationary process. The innovations ε_t are assumed to be i.i.d. The variable σ_t depends on its own past which is the moving average part and previous values of linear combinations of the white noise process of the error term Verbeek (2008)[18].

3.4 ARCH

The auto-regressive conditional heteroskedasticity model by Engle in 1982[7] modeled the variance of a regression model's disturbances as a linear function of the lagged values of the squared regression disturbances. An ARCH(q) model can be written as:

$$r_t = \mu + \varepsilon_t, \quad (6)$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2, \quad (7)$$

where equation (6) is the conditional mean and equation (7) is the conditional variance, with the intercept ω , and where the innovations ε_t that are normally distributed with mean zero and standard deviation σ . α_i is the ARCH parameters. The model is useful when the data is of a non-linear character. In financial time-series heteroskedasticity might be present which means that the variance is not constant over time, also volatility of similar magnitude might appear in clusters. ARCH-family models have the ability to capture these effects.

This model have a few problems for instance it might need a large q to capture all of the dependence and as the q gets large the non-negatively constrain might get breached. It can also not handle asymmetric effects. I will not use this model, but it is important as it is the first of the ARCH-family models.

3.5 GARCH

The GARCH-model was developed by Bollerslev and Taylor in 1986[8] and was designed to get around some of the drawbacks in the ARCH-model. The GARCH-model is more parsimonious as in generally a lower order is required to capture the dynamics and its is thus less likely to breach the non-negatively constraints. The past squared residuals capture high frequency effects, while the lagged variance captures long term influences. The GARCH(q,p) model can be written as:

$$\sigma_t^2 = \omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (8)$$

where ω , α , β need to be non-negative to ensure the non-negativity constraint is not breached and with $\alpha + \beta < 1$ but should be close to unity for an accurate model specification.

3.6 EGARCH

The exponential GARCH model presented by Nelson 1991[10] is one of the models used in this thesis able to capture the asymmetric effect referred to as the levered effect in financial time series. The model has the ability to account for negative shocks and positive shocks of the same magnitude having an unequal destabilizing effect. The model also can not breach the non-negativity constraint. The EGARCH(q,p) model can be written as:

$$\log(\sigma_t^2) = \omega + \sum_{j=1}^q (\alpha_j \varepsilon_{t-j} + \gamma_j (|\varepsilon_{t-j}| - E|\varepsilon_{t-j}|)) + \sum_{j=1}^p \beta_j \log(\sigma_{t-j}^2), \quad (9)$$

where in the conditional variance equation (9) the coefficient α_j captures the sign effect and γ_j the size effect of the asymmetry. Positive estimates of the volatility is guaranteed due to working on the log variance. There are no restrictions on ω , α , and γ , but to maintain stability β must be positive and less than one.

3.7 GJR-GARCH

The GJR-GARCH model following the work of Glosten, Jagannathan, and Runkle (1993)[12] uses an indicator function I to model the positive and negative shocks on the conditional variance asymmetrically. The GJR-GARCH(q,p) model can be written as:

$$\sigma_t^2 = \omega + \sum_{j=1}^q (\alpha_j \varepsilon_{t-j}^2 + \gamma_j I_{t-j} \varepsilon_{t-j}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (10)$$

where γ_j is the leverage term. The indicator function I takes the value 1 for $\varepsilon < 0$ and 0 otherwise. The parameters have the same restrictions as for the GARCH model with addition of $\gamma > 0$. It is easy to recognize that the GARCH model is in fact a restricted version of the GJR-GARCH, with $\gamma = 0$.

3.8 TGARCH

The threshold GARCH model proposed by Zakoian 1991[11] is similar to the GJR-GARCH with the only noticeable difference being that it works on the conditional standard deviation as opposed

to the conditional variance and thus on the innovations directly not the squared innovations. The TGARCH(q,p) model can be written as:

$$\sigma_t = \omega + \sum_{j=1}^q (\alpha_j \varepsilon_{t-j} + \gamma_j I_{t-j} \varepsilon_{t-j}) + \sum_{j=1}^p \beta_j \sigma_{t-j}, \quad (11)$$

where γ_j is the leverage term. The indicator function I takes the value 1 for $\varepsilon < 0$ and 0 otherwise. The parameters have the same restrictions as for the GARCH model with addition of $\gamma > 0$.

4 Statistical Tests

In this Section I present the different test used on the data. The tests consist of the Jarque-Bera test for non-normality, augmented Dickey-Fuller (1979)[19] test for unit roots, Engle's (1982)[7] ARCH LM test for ARCH effects and the sign bias test by Engle and NG (1993)[20] for leverage effects.

4.1 Jarque-Bera test

The Jarque-Bera (1987)[21] test is used to test if a sample follows a normal distribution. The following formula is used:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right), \quad (12)$$

where n is the sample size, S is the sample skewness and K the sample kurtosis. The test is χ^2 distributed and have a joint null of $S = 0$ and $(K - 3) = 0$. Rejecting the null suggests that the sample does not follow a normal distribution.

4.2 ADF

The augmented Dickey-Fuller (1979)[19] test is used to determine if a time-series is stationary by checking for unit roots. This is important when applying an ARMA model to determine if the data needs to be differentiated. The formula is:

$$\Delta r_t = \alpha + \beta t + \theta r_{t-1} + \delta_1 \Delta r_{t-1} + \dots + \delta_{k-1} \Delta r_{t-k} + \varepsilon_t, \quad (13)$$

where α is a constant, β the coefficient on a time trend t and k the lag order of the auto-regressive process. Imposing constrains on the parameters α and β can be used to model random walk ($\alpha =$

$0, \beta = 0$) and random walk with a drift ($\beta = 0$). The null for the test is $\theta = 0$ suggests non-stationary data versus the alternative $\theta < 0$ for stationary data.

4.3 ARCH LM test

Financial time-series data are often assumed to be non-linear and a test to confirm this and thus warrant the use of non-linear models should be conducted. For this purpose I use Engle's ARCH test which is a Lagrange multiplier to assess the significance of ARCH effects Engle 1982[7]. The conditional heteroskedasticity in a variance process is equal to the autocorrelation in the squared innovation process. The residual series is given by:

$$e_t = r_t - \hat{\mu}_t, \quad (14)$$

where $\hat{\mu}_t$ is the conditional mean of the process and e_t the innovation. The alternative hypothesis for autocorrelation in the squared residuals is given by the regression

$$H_a : e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2 + u_t,$$

where u_t is a white noise error process. The null suggesting no autocorrelation in the squared residuals is

$$H_0 : \alpha_0 = \alpha_1 = \dots = \alpha_m = 0$$

4.4 Test for asymmetries in volatility

To test for asymmetry in the data the sign bias test by Engle and Ng's (1993) [20] is used. The test indicates if the residuals in the GARCH-model are sign biased suggesting that the leverage effect is present in the time-series or not. The joint test can be stated as

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- \varepsilon_{t-1} + \phi_3 S_{t-1}^+ \varepsilon_{t-1} + \nu_t, \quad (15)$$

where the residuals from the symmetric GARCH take value 1 if $\varepsilon_{t-1} < 0$ and gives the slope dummy value S_{t-1}^- . S_{t-1}^+ is defined as $1 - S_{t-1}^-$. If ϕ_1 is significant it suggests sign bias and that negative and positive shocks have different destabilizing effect. If ϕ_2 is significant it indicates negative size bias. A

significant ϕ_3 indicates the presence of positive size bias. If positive or negative size bias is significant it suggests that the size of a shock will have an asymmetric impact on volatility. The null for the joint test is $\phi_1 = \phi_2 = \phi_3 = 0$ for no significant asymmetry in the squares residuals. The alternative hypothesis that at least one ϕ_i is significant for $i = 1, 2, 3$ which suggests that the use of asymmetric models are reasonable.

5 Data

The closing price data form S&P500, NASDAQ Composite, DJIA and CBOE Interest Rate(TNX) have been gathered from Yahoo Finance. LBMA Gold prices can be found at Open Financial Data Project. The currency rate series for USD/GBP is from oanda.com. The period used is from 1990-01-01 to 2014-02-07. The returns are calculated according to,

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \quad (16)$$

where r_t is the return at time t and p_t is the daily closing price of the index at time t .

As described in Section 4.1, the Jarque-Bera test is χ^2 distributed with a joint null of skewness and kurtosis equal to zero and three respectively. At a one percent significant level, the normality are rejected for all the series. This is not surprising given the high kurtosis in the return series. The heavy tails indicate that the returns suffer from extreme events that fall outside the normal distribution assumption.

Table 3.1: Descriptive statistics of the daily return series including the Jarque-Bera test

	S&P-500	NASDAQ Composite	DJIA
Mean	0.0001959	0.000244	0.0003426
Median	0.0005437	0.001105	0.0004450
Maximum	0.1037823	0.124138	0.1108000
Minimum	-0.0993237	-0.107034	-0.0787300
Std.dev	0.01157099	0.0151518	0.01095204
Skewness	-0.4190841	-0.2621505	0.013709
Kurtosis	8.731976	5.91227	8.512445
JB p-value	0.00000	0.00000	0.00000
	CBOE Interest Rate	LBMA Gold	USD/GBP
Mean	-0.0003003	0.0001503	-0.0000011
Median	0.0000000	0.0002556	0.0000000
Maximum	0.0875332	0.0676617	0.047700
Minimum	-0.1857708	-0.1007168	-0.031900
Std.dev	0.01562485	0.01030453	0.004453823
Skewness	-0.2533819	-0.4081641	0.3459726
Kurtosis	7.237176	7.407597	6.609938
JB p-value	0.00000	0.00000	0.00000

In Table 3.1, the descriptive statistics of the daily return series are presented. The results for the S&P500 and NASDAQ Composite series is consistent with Glosten et al. (1993)[12], who found that stock index returns often exhibits negative skewness. The DJIA series does not share this feature as the skewness is low and positive. I have not removed any outliers in the raw data so the excess kurtosis found in all the series is not unexpected. Figlewski (1997)[1] stated that equities and many other securities suffer from fatter tails since the log normal diffusion model is inconsistent with large price changes. This implies that using different distributions that might capture the extreme events in the tails is justified.

It is worth to note that the statistics between the series is relatively similar except for the USD/GDP series where the min-max range is a lot closer and the skewness is relatively large and positive. The standard deviation is also less than half of all of the other series.

6 Forecast Evaluation

6.1 Data and sampling procedure for the model based forecast

I am going to investigate which of the forecasting models earlier presented in the theory (Section 3) that produce the most accurate forecast for the six return series earlier presented (Section 5), and see whether the regression based models outperform the naive models. The return series used are chosen to reflect the variety among financial time series. The tree stock exchange indexes represent different types of stocks, S&P500 where the largest stocks are listed DJIA represent the industrialized sector and NASDAQ Composite the technology sector in USA. I have also included the LBMA Gold returns, the exchange rate series USD/GBP and options index on 10 year US treasury notes the CBOE Interest Rate. I find it interesting to see how the different forecasting model perform on a varied selection of series. Most of the series are collected from Yahoo Finance the full URLs will be provided in the Appendix.

Market micro-structure problems, also referred to as noise in the data, is frequently found in real markets due to bid-ask spreads, non-trading, and serial correlation. This makes most intraday data unusable for calculations (see Figlewski (1997)[1]). I use the daily closing prices to produce the return series as earlier described (Section 5.1). However Figlewski (1997)[1] points out that positive serial correlation is often found in daily closing prices for equities and other securities. A way to limit the effect of serial dependence at high frequencies is to use longer sampling intervals, but this means reducing the amount of data points assuming one does not extend the time period in question to compensate for this. Having fewer data points leads to increased sampling error. The best choice of sampling frequency must depend on the statistical properties of the particular price series under consideration according to Figlewski (1997)[1]. The length of the forecasting horizon should also be taken into consideration when deciding which historical data to elaborate on. Having a large data sample does not guarantee an accurate model that provides unbiased volatility, because volatility tends to change over time. There is a trade-off between trying to collect as much data as possible and trying to eliminate data that is obsolete. When the forecasting horizon is short, it is more appropriate to choose a short sample of the latest observations, which captures volatility clustering, thereby, capturing the abilities/phenomenon of the current market conditions (see Figlewski (1997)[1]).

One can reduce the sampling error by using a large number of daily observations. Figlewski

(1997)[1] shows that the choice of frequency at which the data is collected can have a large effect on volatility. Practitioners and researchers tend to use the very recent past when producing forecasts. Figlewski (1997)[1] shows that this might fail to produce an accurate forecast and instead advocate the use of a longer horizon. With this in mind I have chosen the in-sample data to consist of five years of daily observations 1990-01-01 to 1994-12-31 and the out-of-sample forecasted period 1995-01-01 to 2014-02-07.

Before testing, the behavior of the raw data is analyzed and presented in the descriptive statistics (Section 5). No filtering of outliers have been performed. The reason for not filtering the outliers is the assumption that the outliers represent extreme events like war, natural disasters, and financial crisis. Filtering the outliers might give better forecasts for tranquil periods should such events no longer occur. Including the outliers will lead to the opportunity of capturing such events. I believe it is naive to assume such events will stop to occur when looking at our history which is filled with them.

The raw data is tested for normality according to the earlier presented (Section 4.1) Jarque-Bera test. The results of the test combined with the skewness and excess kurtosis shown in the descriptive statistics (Section 5) of all the series is used to justify the use of different distributions for forecasting with ARCH-family models.

The data is imported to R where all the results is generated manly by the use of the *rugarch* package. The code is available in the Appendix.

6.2 Forecasting procedure

The volatility in the out-of-sample period is forecasted through the in-sample data using a rolling window which is adjusted to forecast the amount of data points in the next month. This results in a forecast which produce as many data points for each of the months in the forecast horizon as there is observations in the out-of-sample data, which is more accurate than just using a 21-day window as an approximation. This implies that the first forecasted month uses the entire in-sample data. For the next month the oldest month of the in-sample data is excluded and the realized daily values of the first forecast period is used in-sample to produce the next forecast. This is repeated throughout the out-of-sample period giving the conditional variance for the forecast. The monthly forecast of the standard deviation is than produced by,

$$\hat{\sigma}_m = \sqrt{\sum_{i=1}^N \hat{\sigma}_i^2}, \quad (17)$$

where the monthly standard deviation $\hat{\sigma}_m$ is calculated by the square root of the sum of squared daily variance $\hat{\sigma}_i^2$ in month m .

The proxy used for monthly realized volatility is computed by finding the daily standard deviation using the realized returns in each month and then applying the square root of time rule to get the monthly standard deviation using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^N (r_i - \bar{r})^2}, \quad (18)$$

where N is the number of days in month m .

6.3 Model specifications and tests

6.3.1 Naïve models

I use tree models that is considered naive: RW, SMA, and EWMA which are earlier presented (Section 3). The reason that the models are considered naive is that the model bases forecasts of future volatility on past realizations, with different weighting schemes. In the paper by Poon and Granger (2003)[4] the naive model fall under the category of historical volatility models. For the RW model the realized volatility is estimated using a one month look back SMA. When it comes to the SMA model it is a question of how many periods the forecast is based on. Having too few data points contains to little information, while including too many data points runs the risk of adding obsolete observations that have no added explanatory power with regards to the volatility in recent time. I will forecast on the bases of 1-,6-, and 12-months where the 1 month is referred to as the RW model. When forecasting with the EWMA-model the selection of the smoothing parameter value (λ) is an empirical issue. I have chosen to follow RiskMetrics™ and thus setting $\lambda = 0.97$.

6.3.2 In-sample tests

It is important for sake of an ARMA model that the data is stationary for it to avoid giving misleading statistics. The ADF test describes earlier (Section 4.2) is used to test the in sample absolute returns

for unit roots. There are several other test for stationarity that could be applied but the use of many tests is more relevant for data with few observations as that is the main weakness of the tests.

Engle's ARCH LM test is run on the residuals from an ARMA(1,1)-model to test for presence of volatility clustering in the in-sample data. Rejecting the null implies that there is significant heteroskedasticity in the data, which in turn justifies producing forecasts by applying the more complex ARCH-family models to the data.

The sign bias test by Engle and Ng's (1993)[20] is than applied based on fitting a GARCH(1,1)-model to the in-sample data. The tests is a joint sign bias tests for asymmetries. The standard GARCH model is symmetric and is unable to capture asymmetric effects like the leverage effect which is commonly found in asset return series. Significant presence of such asymmetry is cause for forecasting using the asymmetric models in the ARCH-family.

6.3.3 Regression based forecast models

When is comes to regression based models used for forecasting the ARMA(1,1) is commonly used and recommended over higher order ARMA models that might be a better fit for in-sample data in some cases, but tend to be unstable when used for forecasting. The ARMA(1,1) is the one chosen in this thesis and applied to the squared returns which is a proxy for realizes daily volatility. One alternative approach could be to find the optimal model for the in-sample data with the intention of finding the model which gives the best forecast when the optimal model specification is known but this suffers form look ahead bias, and is not used in this thesis.

In recent times the ARCH-family models have become more popular than the ARMA model for volatility forecasting and a wide array of them have been produced. The most common one used for forecasting is no doubt the GARCH(1,1)-model and is the only symmetric ARCH-family model that I will apply for forecasting. The asymmetric model specifications used is EGARCH(1,1), GJR-GARCH(1,1) and t-GARCH(1,1). All the ARCH-family model will be run using the normal, skewed Student-t, and skewed Generalized Error Distribution. The reason for trying multiple distributions is that the descriptive statistics and JB test show that the excess skewness and kurtosis found in the return series suggests that the normal distribution might not produce the best forecasts.

6.4 Problems with ARCH-family models

The ARCH-type models require a large amount of data to give robust estimations. As previously mentioned models with many parameters might fit in-sample data better, but often fall apart when used for out-of-sample, thus failing to produce an accurate forecast. To produce an accurate forecast the model has to be sufficiently stable and continue to hold over time Figlewski (1997)[1].

For some of the models coefficients might be non-positive or sum to values greater than one. Should a model have coefficients which sum to be larger than one it will experience long run instability. If the coefficients are negative or do not sum to one, the maximum might lie outside of the theoretically accepted region Figlewski (1997)[1]. Verbeek (2008)[18] points out that a parameter value that sums to more than one gives a non-stationary process, but is a typical finding in empirical studies.

6.5 Evaluation with loss functions

The evaluation models chosen need to be robust against the presence of noise. The impact from extreme outcomes may have a dominating influence on the forecast evaluation. According to Patton (2011)[22], a robust loss function is not only a function that is robust to noise in the proxy Huber (1981)[23] but also to an expected loss ranking of between two volatility forecasts is the same. Results produced by Patton (2011)[22] indicate that the only evaluation model that is robust is the Mean Squared Error (MSE). The preferred forecasting models in this thesis will be selected by MSE. The inputs are given by the realized volatility proxy at time t and the forecasted variance from the different models at time t . They are summed up for every month in the out-of-sample period giving an average for the loss which can be compared between the different models.

Villhelmsson (2006)[15] opted to use the mean absolute error (MAE) instead of MSE, due to the fact that MSE is a loss function sensitive to outliers. MAE does not square the errors so larger errors does not dominate the evaluation to the extent they do for the MSE. The MAE will be presented to show that the choice of evaluation model can influence the ranking of the models significantly.

MSE is defined by,

$$MSE = \frac{1}{n} \sum_{t=1}^n (\hat{\sigma}_t - \sigma_t)^2, \quad (19)$$

where n is the number of observations and $\hat{\sigma}_t - \sigma_t$ is the deviation of the estimated - and the proxy realized standard deviation.

MAE is defined by,

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{\sigma}_t - \sigma_t|, \quad (20)$$

where n is the number of observations and $\hat{\sigma}_t - \sigma_t$ is the deviation of the estimated - and the proxy realized standard deviation.

7 Empirical results and analysis

7.1 In-sample test results

Table 5.1 shows the results of the augmented Dickey-Fuller test that is applied to the in-sample absolute returns. For all the series the null is rejected on a 1% level suggesting that all the series are stationary

Table 5.1: ADF test for unit roots

ADF	S&P-500	NASDAQ Composite	DJIA
Statistic	-8.6851	-8.6158	-8.65
P-value	<0.01	<0.01	<0.01
	CBOE Interest Rate	LBMA Gold	USD/GBP
Statistic	-7.8495	-6.5575	-7.9207
P-value	<0.01	<0.01	<0.01

Table 5.2 shows the results of Engle’s ARCH LM test described in 2.4.3 at lag 5 for all the series. The null is rejected at a 1% level for all the series. This is evidence that all the series have squared residuals for previous lags that are correlated with the squared residuals at time t . This indicates that there is heteroskedasticity in the data, and motivates the use of models that does not assume a constant variance, as they might provide a better forecast.

Table 5.2: Engle’s ARCH LM test at lag 5

ARCH LM	S&P-500	NASDAQ Composite	DJIA
Statistic	57.23	84.05	34.71
P-value	<0.01	<0.01	<0.01
	CBOE Interest Rate	LBMA Gold	USD/GBP
Statistic	17.10	15.792	39.69
P-value	<0.01	<0.01	<0.01

In Table 5.3 the results from the sign bias test for asymmetry in volatility. Two of the series the NASDAQ Composite index and the USD/GBP have asymmetry that is significant on a 1% level. DJIA have a negative sign bias that is significant at a 10% level while the S&P-500 index have a joint effect that is significant at a 15% level. The CBOE Interest Rate and LBMA Gold series does not have any significant in-sample asymmetry at levels 15% or lower. This result is ambiguous as some series display clear signs of leverage effects in the standardized residuals while other series have no clear significant asymmetry. Still there might be asymmetry in the out-of-sample data even if the in-sample data show no significant sign of it. The fact that the leverage effect is commonly found in financial time series gives reason to apply asymmetric model to all the series even though the in-sample results are ambiguous.

Table 5.3: Sign Bias Test of Engle and Ng (1993) t-value with p-values in the brackets

Sign Bias Test	S&P500	NASDAQ Composite	DJIA
Sign Bias	1.22(0.222)	2.998(<0.01)	0.602(0.547)
Negative Sign Bias	0.569(0.569)	0.239(0.811)	1.915(0.056)
Positive Sign Bias	0.082(0.934)	0.425(0.664)	0.653(0.514)
Joint Effect	5.482(0.140)	18.80(<0.01)	5.200(0.158)
	CBOE Interest Rate	LBMA Gold	USD/GBP
Sign Bias	1.367(0.172)	0.100(0.921)	2.453(0.014)
Negative Sign Bias	1.422(0.155)	0.768(0.443)	0.128(0.898)
Positive Sign Bias	0.330(0.741)	0.487(0.626)	0.647(0.518)
Joint Effect	3.036(0.386)	1.278(0.735)	13.20(<0.01)

7.2 In-sample parameter estimates

There are in total 5 regression based models used for producing the results of each of the 6 series 4 of them using 3 distributions which gives a total of 78 different in-sample parameter estimates. I will limit this Section to contain only the in-sample parameter estimates of the most accurate models in terms of MSE for the out-of-sample period, with the best performing distribution. This totals to five models as EWMA where the preferred model for the USD/GBP series. The notation follows the one used in Section 3.

7.2.1 S&P500 GJR-GARCH(1,1)-skew.Student-t

Table 5.4 shows the parameter estimates of the preferred model for the S&P500 series in terms of the MSE loss function. The ω is zero and α close to zero and the P-value suggests that they are not significantly different from zero. The rest of the parameters are significant at a 1% level. I see that $\alpha + b + \gamma > 1$ which according to Verbeek (2008)[18] gives a non-stationary process which is not uncommon in empirical work but might result in getting a sample outside the accepted region when maximizing. The persistence in this case being larger than one means that the volatility persists and grows. The significant γ parameter suggests that the levered effect is present in the in-sample data, even if the sign bias test only suggests that the joint effect is significant given a 15% level.

Table 5.4: In-sample parameter estimates for S&P500 GJR-GARCH(1,1)-skew.Student-t model

GJR-GARCH(1,1)-skew.Student-t	Estimate	P-value
ω	0.000000	0.691004
α	0.000027	0.978187
β	0.984377	0.000000
γ	0.029456	0.000046

7.2.2 NASDAQ Composite GJR-GARCH(1,1)-norm

Table 5.5 shows the parameter estimates of the preferred model for the NASDAQ Composite series in terms of the MSE loss function. All the parameters are significant at a 5% level. The parameters $\alpha + b + \gamma < 1$ the sum being lower than 1 but relatively high indicates a relatively high persistence and slow decay of the volatility shocks. The significant γ parameter suggests that the levered effect is

present in the in-sample data, which is in line with the in-sample sign bias test for the series.

Table 5.5: In-sample parameter estimates for NASDAQ Composite GJR-GARCH(1,1)-norm model

GJR-GARCH(1,1)-norm	Estimate	P-value
ω	0.000019	0.001677
α	0.059612	0.036893
β	0.552284	0.000002
γ	0.231844	0.003250

7.2.3 DJIA GJR-GARCH(1,1)-skew.Student-t

Table 5.6 shows the parameter estimates of the preferred model for the DJIA series in terms of the MSE loss function. The ω is zero and α close to zero and the P-value suggests that they not significantly different from zero. The rest of the parameters are significant at a 5% level. I see that $\alpha + b + \gamma > 1$ which according to Verbeek (2008)[18] gives a non-stationary process which is not uncommon in empirical work but might result in getting a sample outside the accepted region when maximizing. The persistence in this case being slightly larger than one means that the volatility persist and grows. The significant γ parameter suggests that the levered effect is present in the in-sample data, which is in line with the in-sample sign bias test for the series.

Table 5.6: In-sample parameter estimates for DJIA GJR-GARCH(1,1)-skew.Student-t model

GJR-GARCH(1,1)-skew.student-t	Estimate	P-value
ω	0.000000	0.731614
α	0.001047	0.963420
β	0.982396	0.000000
γ	0.028907	0.012486

7.2.4 CBOE Interest Rate EGARCH(1,1)-skew.Student-t

Table 5.7 shows the parameter estimates of the preferred model for the CBOE Interest Rate series in terms of the MSE loss function. The ω and α is close to zero and the P-value suggests that they

not significantly different from zero. The rest of the parameters are significant at a 5% level. The persistence of the EGARCH model is determined by the beta parameter alone which is close to 1, which indicates high persistence and a slow decay of the volatility shocks. The significant γ parameter suggests that the levered effect is present in the in-sample data, even if the in-sample sign bias test does not show significant asymmetry in the data.

Table 5.7: In-sample parameter estimates for CBOE Interest Rate EGARCH(1,1)-skew.Student-t model

EGARCH(1,1)-skew.Student-t	Estimate	P-value
ω	-0.068147	0.385855
α	-0.014669	0.232351
β	0.992875	0.000000
γ	0.066451	0.002524

7.2.5 LBMA Gold EGARCH(1,1)-norm

Table 5.8 shows the parameter estimates of the preferred model for the LBMA Gold series in terms of the MSE loss function. The ω is close to zero and the P-value suggests that it is not significantly different from zero. The rest of the parameters are significant at a 1% level. The persistence of the EGARCH model is determined by the beta parameter alone which is close to 1, which indicates high persistence and a slow decay of the volatility shocks. The significant γ parameter suggests that the levered effect is present in the in-sample data, even if the in-sample sign bias test does not show significant asymmetry in the data.

Table 5.8: In-sample parameter estimates for LBMA Gold EGARCH(1,1)-norm model

EGARCH(1,1)-skew.Student-t	Estimate	P-value
ω	-0.009504	0.663625
α	0.048995	0.000001
β	0.998449	0.000000
γ	0.086810	0.000000

7.3 Out-of-sample comparison

In this Section, I present the out-of-sample predictive power of the forecasting models described in Section 2.3. The main results of model evaluation in tables ranked by MSE and MAE. For every series a figure showing the graph of the preferred models forecast versus the realized volatility is provided. As stated in Section 6.5, a robust loss function is not only a function that is robust to noise in the proxy Huber, (1981)[23] but also to an expected loss ranking of between two volatility forecasts is the same Patton (2011)[22]. Patton's results indicate that only MSE fulfills his criterion and is the loss function I will use to choose the most accurate forecasting model. The raw data have not been filtered for outliers and thus I expect there to be several large residuals that MSE are sensitive towards and a case could certainly be made for choosing the most accurate model based on MAE which is more robust against outliers. It will in any case be interesting to see the deviation in ranking between the loss functions.

7.3.1 Evaluation models for S&P500 index

Table 5.9 shows that the model with the best out-of-sample volatility forecast for the S&P500 series is the GJR-GARCH(1,1) under the skewed Student-t distribution. This is not to surprising with regards to distribution as the descriptive data show in Section 3 that the series have excess skewness and kurtosis and the Jarque-Bera test rejected the null for normality on a 1% level. The sign bias test done on the in-sample data did however not show a significant asymmetry unless the level where set to 15% but the fact that a asymmetric model is the most accurate suggest that the levered effect is present in the data. There where high significant arch effects in the data according the the ARCH LM test and it is reasonable that the ARCH-family models dominate the ranking for both MSE and MAE. The only regression based models that failed to beat all the naive models where the ARMA(1,1), EGARCH(1,1)-norm, EGARCH(1,1)-skew.GED w.r.t MSE and ARMA(1,1) w.r.t MAE. MAE also favors the tGARCH model over the GJR-GARCH model, shown by the tGARCH(1,1) with a skewed GED having the lowest MAE.

Table 5.9: Evaluation of predictive power of the forecast models for the S&P500 index series.

S&P500	MSE*	Rank	MAE	Rank
RW	0.041705	11	0.013991	14
SMA-6 months	0.060146	16	0.015810	16
SMA-12 months	0.073863	17	0.017760	17
EWMA	0.042749	14	0.013665	13
ARMA(1,1)	0.052120	15	0.014712	15
GARCH(1,1)-norm	0.038017	7	0.013391	10
GARCH(1,1)-skew.Student-t	0.038196	8	0.013589	12
GARCH(1,1)-skew.GED	0.038248	9	0.013541	11
EGARCH(1,1)-norm	0.041919	12	0.012978	9
EGARCH(1,1)-skew.Student-t	0.040983	10	0.012848	4
EGARCH(1,1)-skew.GED	0.041982	13	0.012940	8
tGARCH(1,1)-norm	0.036775	6	0.012789	3
tGARCH(1,1)-skew.Student-t	0.035631	4	0.012612	2
tGARCH(1,1)-skew.GED	0.035975	5	<i>0.012608</i>	1
GJR-GARCH(1,1)-norm	0.035331	3	0.012874	6
GJR-GARCH(1,1)-skew.Student-t	<i>0.033764</i>	1	0.012851	5
GJR-GARCH(1,1)-skew.GED	0.034139	2	0.012936	7

*MSE is multiplied by 100

Figure 1: Histogram of the MSE sorted by rank

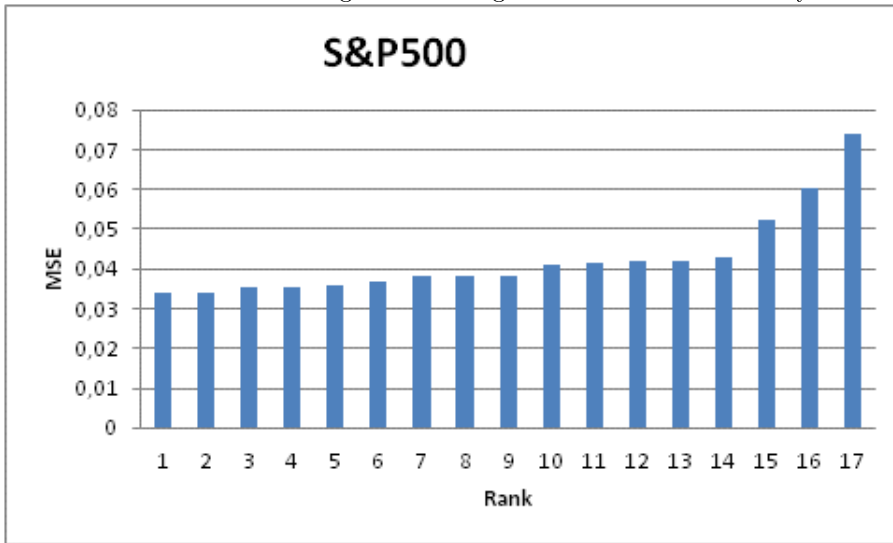


Figure 2: GJR-GARCH(1,1)-skew.Student-t forecasted std.dev vs proxy realized std.dev for S&P500

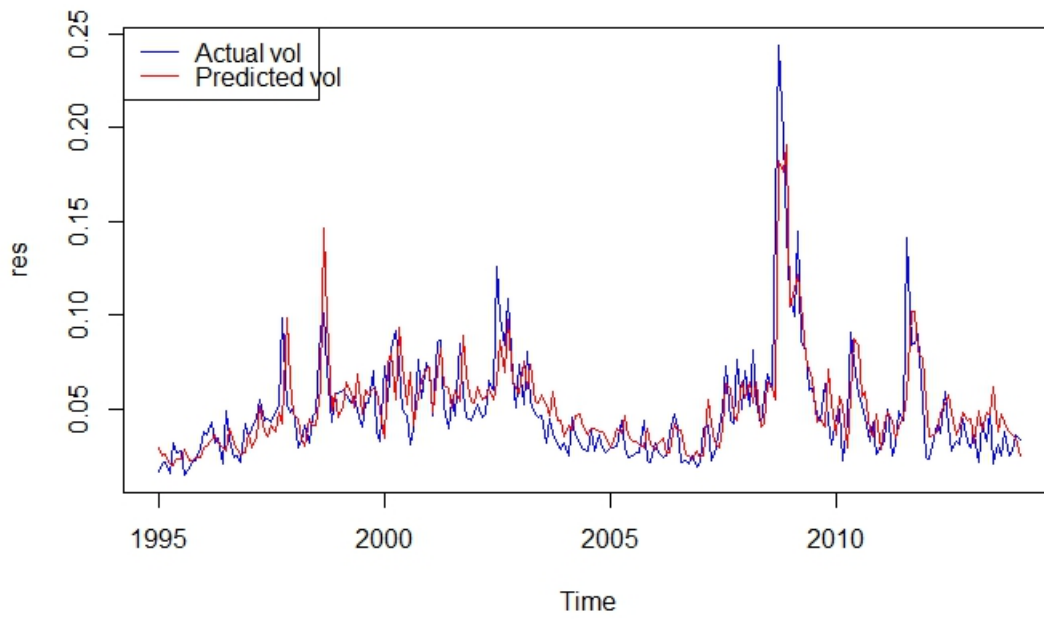


Figure 2 shows the predicted volatility vs the proxy realized volatility for the out-of-sample period. The volatility clustering is evident just by looking at the graph, one can see that periods with high volatility tend to persist followed by periods with relative tranquility. The model seems to fit reasonably well in the tranquil periods while there are some large deviations in periods with high volatility most pronounced during the financial crisis 2007-2008 where the model estimate is off by about 0.05 at a forecasted standard deviation of approximately 0.19 while the proxy realized volatility is at almost 0.25. The large deviation during this period and a few other periods with relative high volatility will be punished greatly by the MSE loss function. It is no surprise after looking at the result that MAE gives different ranking as there might very well be a model that predicts the volatility a bit better on average but slightly worse during periods with high volatility.

7.3.2 Evaluation models for NASDAQ Composite index

Table 5.10 shows that the model that best predicts the out-of-sample volatility for the NASDAQ Composite index series is the GJR-GARCH(1,1) under the normal distribution. This is not expected with regard to the distribution as the descriptive data show in Section 3 that the series has excess skewness and kurtosis and the Jarque-Bera test rejected the null for normality on a 1% level. The sign bias test done on the in-sample data shows a significant asymmetry at the 1% level, thus it is no surprise that an asymmetric model is the most accurate given the significant in-sample levered effect present in the data. The ARCH LM test showed significant arch effects in the in-sample data, and it is reasonable that the ARCH-family models outperform the ARMA(1,1) in terms of MSE. The only regression-based models that failed to beat all the naive models were the ARMA(1,1) and EGARCH(1,1)-norm w.r.t MSE. In the case of MAE the story is quite different as the EWMA was the second best only beaten by EGARCH(1,1) with a skewed Student-t distribution. This shows how one can very easily have a methodology and a time series where the conclusion that the naive model outperforms the more complex regression-based models holds, and this is quite frequently the case in studies. This is also the case for the USD/GDP series as shown under. As shown by the massive study of 93 papers by Poon and Granger (2003)[4] the ARCH-family models are about equal to the historical volatility models which contain among others the models referred to in this thesis as the naive models when it came to giving the best forecasts.

Table 5.10: Evaluation of predictive power of the forecast models for the NASDAQ Composite index series.

NASDAQ Composite	MSE*	Rank	MAE	Rank
RW	0.065848	14	0.017643	14
SMA-6 months	0.082203	15	0.018748	15
SMA-12 months	0.097158	17	0.021382	17
EWMA	0.061012	12	0.016409	2
ARMA(1,1)	0.085893	16	0.019357	16
GARCH(1,1)-norm	0.056858	6	0.016653	6
GARCH(1,1)-skew.Student-t	0.056420	4	0.016762	10
GARCH(1,1)-skew.GED	0.056500	5	0.016751	9
EGARCH(1,1)-norm	0.063464	13	0.016879	11
EGARCH(1,1)-skew.Student -t	0.058488	9	0.016390	1
EGARCH(1,1)-skew.GED	0.059245	10	0.016442	3
tGARCH(1,1)-norm	0.059597	11	0.016693	8
tGARCH(1,1)-skew.Student-t	0.057351	7	0.016621	5
tGARCH(1,1)-skew.GED	0.057795	8	0.016619	4
GJR-GARCH(1,1)-norm	0.056006	1	0.016675	7
GJR-GARCH(1,1)-skew.Student-t	0.056282	3	0.017103	13
GJR-GARCH(1,1)-skew.GED	0.056218	2	0.017102	12

*MSE is multiplied by 100

Figure 3: Histogram of the MSE sorted by rank

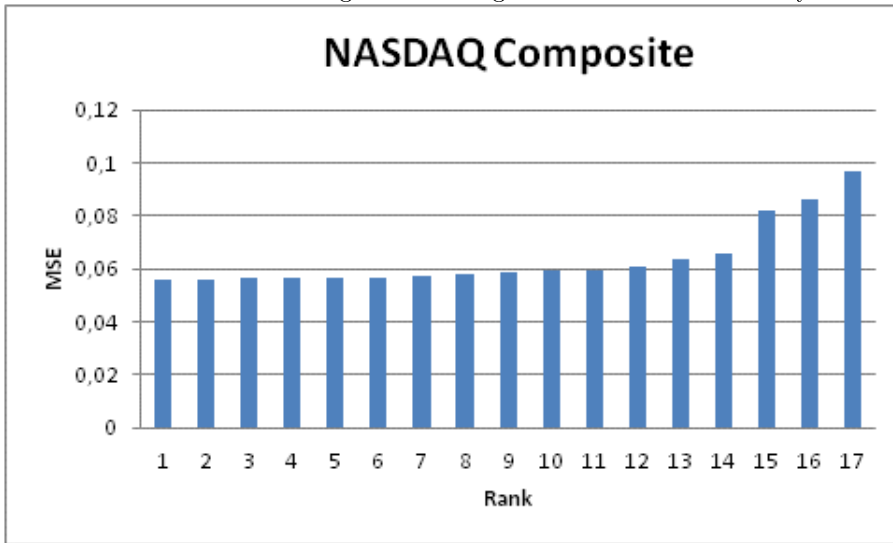


Figure 4: GJR-GARCH(1,1)-norm forecasted std.dev vs proxy realized std.dev for NASDAQ Composite

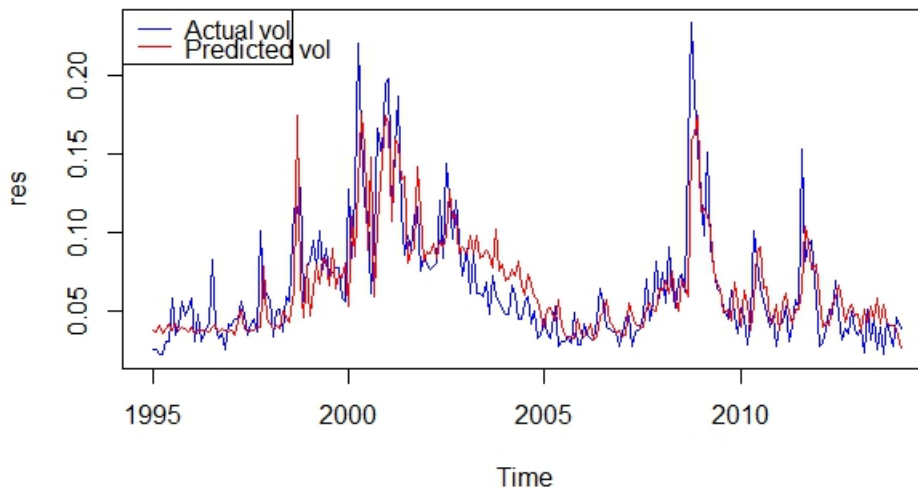


Figure 4 shows the predicted volatility vs the proxy realized volatility for the out-of-sample period. The volatility clustering is evident just by looking at the graph one can see that periods with high volatility tend to persist followed by periods with relative tranquility. The model leaves something to be desired in both the tranquil and the more turbulent periods. The collapse of the dot-com bubble 1999-2001 is easy to recognize in the graph, as well as the 2007-2008 financial crisis by the massive spikes in volatility. Similar to the S&P500 series the model performs rather poorly in the high volatility periods leaving large residuals which will have a heavy influence on the MSE. Also for this series the different scoring among the MSE and MAE is not surprising as this series seems to have even more extreme events than the S&P500 series.

7.3.3 Evaluation models for DJIA index

Table 5.11 shows that the model that best predicts the out-of-sample volatility for the DJIA series is the GJR-GARCH(1,1) under the skewed Student-t distribution. The descriptive data shown in Section 3 that the series has excess skewness and kurtosis and the Jarque-Bera test rejected the null for normality on a 1% level so one would expect the distributions that allow for more kurtosis and skewness to outperform the normal distribution. The sign bias test done on the in-sample data did show a significant negative sign bias at the level 10%, and it is reasonable that an asymmetric model is the most accurate given a significant levered effect is present in the data. There were significant ARCH effects in the data according to the ARCH LM test which is consistent with the ARCH-family models outperforming the ARMA(1,1) model in terms of both MSE and MAE. The only regression-based models that failed to beat all the naive models were the ARMA(1,1) w.r.t MSE, and ARMA(1,1) as well as the GARCH(1,1) with a skewed GED w.r.t MAE. MAE also favors the tGARCH model over the GJR-GARCH model, shown by the tGARCH(1,1) with a skewed GED having the lowest MAE.

Table 5.11: Evaluation of predictive power of the forecast models for the DJIA index series.

DJIA	MSE*	Rank	MAE	Rank
RW	0.039424	13	0.013584	14
SMA-6 months	0.054270	16	0.015472	16
SMA-12 months	0.065622	17	0.017081	17
EWMA	0.040017	14	0.013438	12
ARMA(1,1)	0.047652	15	0.013836	15
GARCH(1,1)-norm	0.035321	9	0.012945	10
GARCH(1,1)-skew.Student-t	0.035389	12	0.013165	11
GARCH(1,1)-skew.GED	0.035337	10	0.013478	13
EGARCH(1,1)-norm	0.035785	11	0.012220	6
EGARCH(1,1)-skew.Student-t	0.034431	7	0.011983	3
EGARCH(1,1)-skew.GED	0.035305	8	0.012058	4
tGARCH(1,1)-norm	0.032214	6	0.012112	5
tGARCH(1,1)-skew.Student-t	0.031381	3	0.011891	2
tGARCH(1,1)-skew.GED	0.031589	4	0.011881	1
GJR-GARCH(1,1)-norm	0.031794	5	0.012263	9
GJR-GARCH(1,1)-skew.Student-t	0.030668	1	0.012242	8
GJR-GARCH(1,1)-skew.GED	0.030955	2	0.012238	7

*MSE is multiplied by 100

Figure 5: Histogram of the MSE sorted by rank

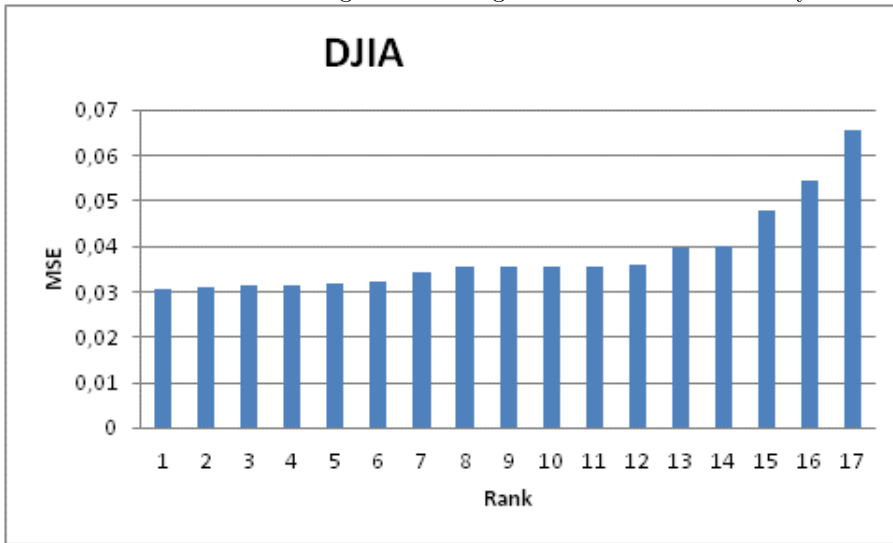


Figure 6: GJR-GARCH(1,1)-skew.Student-t forecasted std.dev vs proxy realized std.dev for DJIA

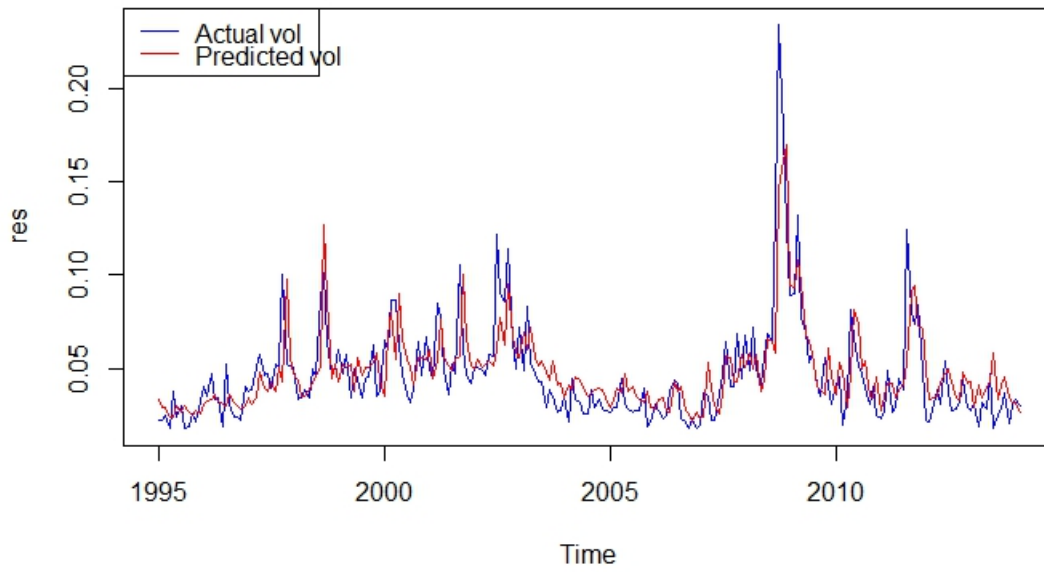


Figure 6 shows that the DJIA are very similar to the S&P500 in terms of volatility during the out-of-sample period. It is not unreasonable that the same model specification predicts the out-of-sample volatility best for both series. The model seem to fit reasonably well in the tranquil periods while there is some large deviations in periods with high volatility most pronounced during the financial crisis 2007-2008. The large residuals created by extreme events will influence MSE more than MAE so the difference in ranking the models is to be expected.

7.3.4 Evaluation models for CBOE Interest Rate

The CBOE Interest Rate is based on 10 times the yield-to-maturity on the most recently auctioned 10-year Treasury note. Table 5.12 shows that the model that best predict the out-of-sample volatility for the CBOE Interest Rate series is the EGARCH(1,1) under the skewed Student-t distribution. The descriptive data show in Section 3 that the series have excess skewness and kurtosis and the Jarque-Bera test rejected the null for normality on a 1% level so one would expect the distributions that allow for more kurtosis and skewness to outperform the normal distribution. The sign bias test done on the in-sample data did however not show a significant asymmetry, but the fact that a asymmetric model is the most accurate suggest that the levered effect is present in the data. There where significant arch effects in the data according the the ARCH LM test which is consistent with the ARCH-family models outperforming the ARMA(1,1) in terms for both MSE and MAE. Only the ARMA(1,1) failed to beat the naive models among the regression based models and got the lowest rank w.r.t MSE and MAE. In terms of MAE the GJR-GARCH(1,1) with a skewed GED where the most accurate model.

Table 5.12: Evaluation of predictive power of the forecast models for the CBOE Interest Rate series.

CBOE Interest Rate	MSE*	Rank	MAE	Rank
RW	0.040573	14	0.014352	14
SMA-6 months	0.060462	15	0.018080	16
SMA-12 months	0.065560	16	0.018026	15
EWMA	0.038607	13	0.014015	13
ARMA(1,1)	0.068767	17	0.018493	17
GARCH(1,1)-norm	0.034128	10	0.013224	11
GARCH(1,1)-skew.Student-t	0.034402	12	0.013331	12
GARCH(1,1)-skew.GED	0.034212	11	0.013219	10
EGARCH(1,1)-norm	0.032962	9	0.013216	9
EGARCH(1,1)-skew.Student-t	<i>0.032446</i>	1	0.013081	5
EGARCH(1,1)-skew.GED	0.032558	3	0.013063	4
tGARCH(1,1)-norm	0.032847	8	0.013201	8
tGARCH(1,1)-skew.Student-t	0.032774	6	0.013161	7
tGARCH(1,1)-skew.GED	0.032743	5	0.013085	6
GJR-GARCH(1,1)-norm	0.032827	7	0.012839	2
GJR-GARCH(1,1)-skew.Student-t	0.032675	4	0.012877	3
GJR-GARCH(1,1)-skew.GED	0.032552	2	<i>0.012804</i>	1

*MSE is multiplied by 100

Figure 7: Histogram of the MSE sorted by rank

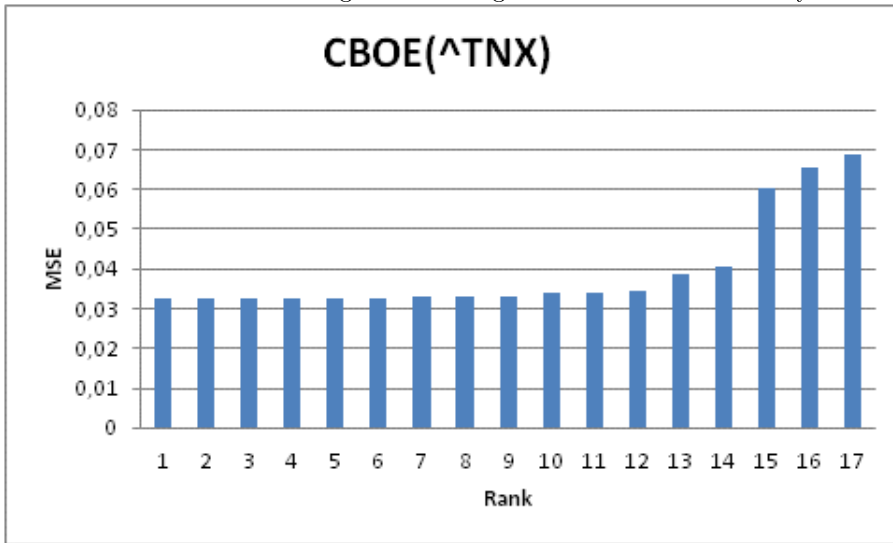


Figure 8: EGARCH(1,1)-skew.Student-t forecasted std.dev vs proxy realized std.dev for CBOE Interest Rate

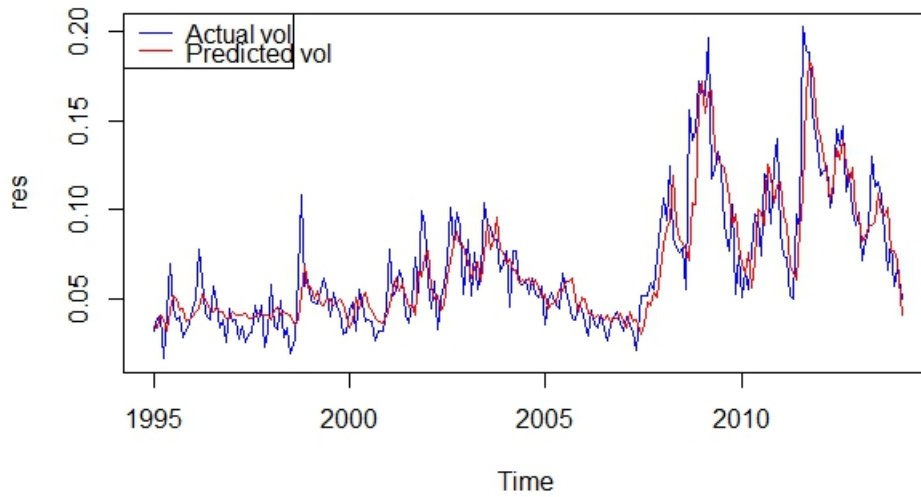


Figure 8 shows the predicted volatility vs the proxy realized volatility for the out-of-sample period. Also for this series the volatility clustering is easy to recognize with the volatility relative calm until the financial crisis starts in 2007-2008. This looks a lot different than the stock index series with more persistent volatility in the period after the crash in 2007-2008. The reason for this might be the uncertainty with regard to the interest rate in the USA during this period. The predicted volatility seem to fit reasonably well for this series not leaving too many large residuals. The prediction in the period 1995-2000 would seem to be the weakest part. With a lower amount of large residuals the similarity in the ranking of the models between MSE and MAE compared to the stock index series is reasonable.

7.3.5 Evaluation models for USD/GBP

Table 5.13 shows that the model that best predict the out-of-sample volatility for the USD/GBP series is the EWMA model for both MSE and MAE. The second best model where the GJR-GARCH(1,1) with a normal distribution for both loss functions. The implementation of other distributions had a clear negative effect on the predictive power even if the descriptive data showed excess kurtosis and skewness and the Jarque-Bera test rejected the null at a 1% level. The effect of the distributions were dramatic to the point where the software failed to produce results in the case of tGARCH(1,1) with a skewed GED. There are some large differences between this and the other series which might cause the instability, one of them is as pointed out in the descriptive statistics is that the standard deviation of the returns is less than half of the other series. The series also have a positive skewness where most of the other series have negative skewness. The exchange rate is recorded for every day in the month not only on trading days like the other series this leads to a larger n-step ahead prediction for every month which might be a reason for the unstable behavior under distributions other than the normal distribution.

Table 5.13: Evaluation of predictive power of the forecast models for the USD/GBP series.

USD/GBP	MSE*	Rank	MAE**	Rank
RW	0.04393	4	0.04675	4
SMA-6 months	0.05204	7	0.04868	6
SMA-12 months	0.06113	9	0.05358	7
EWMA	<i>0.03870</i>	1	<i>0.04270</i>	1
ARMA(1,1)	0.09045	11	0.06627	11
GARCH(1,1)-norm	0.04044	3	0.04474	3
GARCH(1,1)-skew.Student-t	0.06003	8	0.06007	8
GARCH(1,1)-skew.GED	1.15835	14	0.13459	14
EGARCH(1,1)-norm	0.04544	6	0.04786	5
EGARCH(1,1)-skew.Student-t	0.10091	12	0.07545	12
EGARCH(1,1)-skew.GED	100.674	16	1.20295	16
tGARCH(1,1)-norm	0.04501	5	0.06626	10
tGARCH(1,1)-skew.Student-t	0.31981	13	0.07886	13
tGARCH(1,1)-skew.GED	NA	17	NA	17
GJR-GARCH(1,1)-norm	0.04006	2	0.04363	2
GJR-GARCH(1,1)-skew.Student-t	0.06591	10	0.06222	9
GJR-GARCH(1,1)-skew.GED	13.6575	15	0.17104	15

*MSE is multiplied by 1000, **MAE is multiplied by 10

Figure 9: Histogram of the MSE sorted by rank

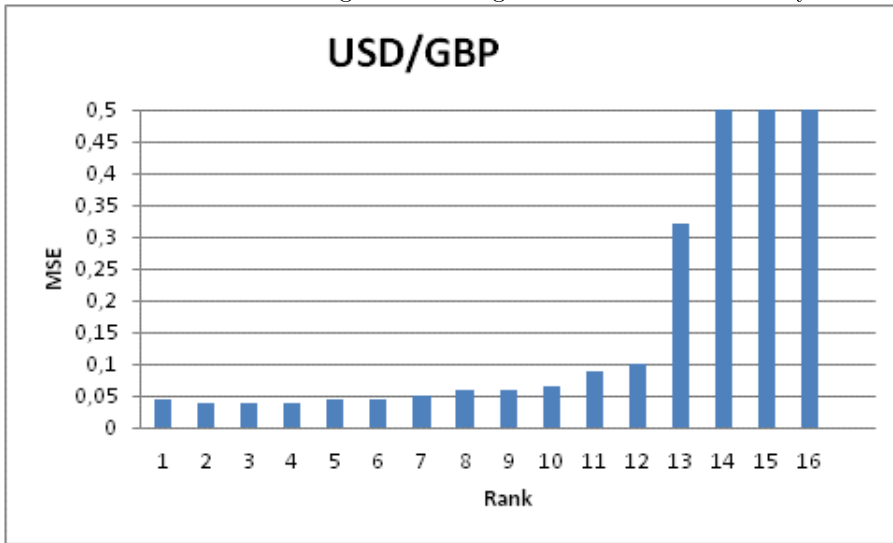


Figure 10: EWMA forecasted std.dev vs proxy realized std.dev for USD/GBP

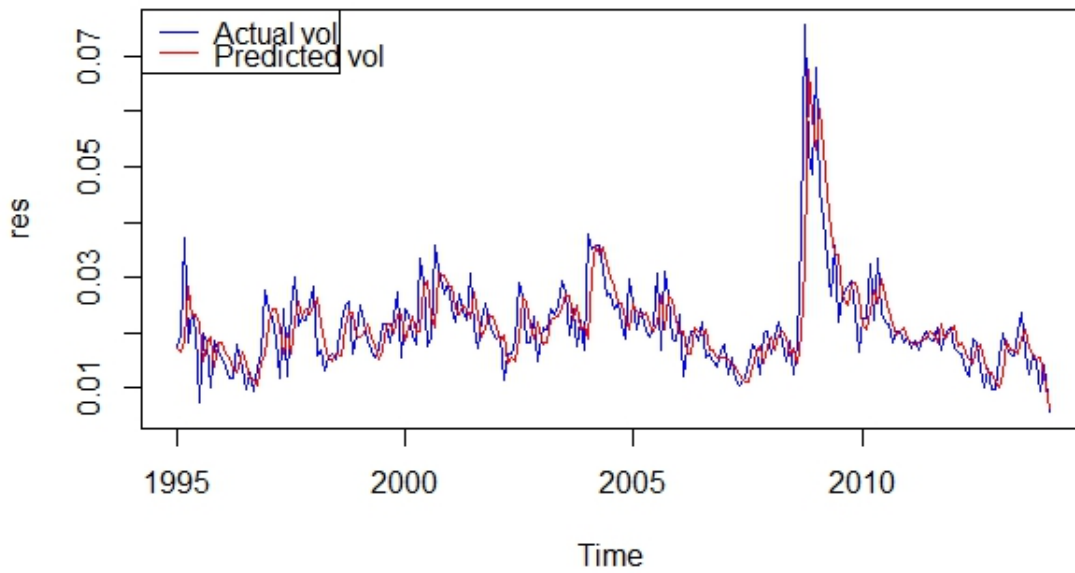


Figure 10 shows the proxy realized volatility and the predicted volatility of the EWMA over the period. The model estimates the volatility quite well for the most part and the series seem to have a volatility ranging from 0.01-0.03 during most of the period only interrupted by a large spike peaking at above 0.07 in volatility during the financial crisis 2008-2009 and seem to have stabilized after that. The series have a lot lower volatility in general than all the other series that peak at 0.2+ during the financial crisis.

7.3.6 Evaluation models for LBMA Gold

For the last series I take a look at the gold returns during the out-of-sample period. Table 5.14 shows that the model which best predicts the out-of-sample volatility for the LBMA Gold series is the EGARCH(1,1) with the normal distribution. The descriptive data show in Section 3 that the series have excess skewness and kurtosis and the Jarque-Bera test rejected the null for normality on a 1% level so one would expect the distributions that allow for more kurtosis and skewness to outperform the normal distribution. When looking at the other ARCH family models the skewed GED improved the accuracy for all of them. The sign bias test done on the in-sample data did however not show a significant asymmetry, but the fact that an asymmetric model is the most accurate suggests that the levered effect is present in the data. There were significant ARCH effects in the data according to the ARCH LM test which is consistent with the ARCH-family models outperforming the ARMA(1,1) in terms of both MSE and MAE. Only the EWMA among the naive models managed to beat several regression based models getting rank 7 for MSE and 5 in terms of MAE. The most accurate model given the MAE loss function was the EGARCH(1,1) with a skewed GED.

Table 5.14: Evaluation of predictive power of the forecast models for the LBMA Gold series.

LBMA Gold	MSE*	Rank	MAE	Rank
RW	0.034340	15	0.012150	14
SMA-6 months	0.034297	14	0.013079	15
SMA-12 months	0.042021	17	0.014815	17
EWMA	0.029720	7	0.011615	5
ARMA(1,1)	0.040559	16	0.013479	16
GARCH(1,1)-norm	0.030807	9	0.012013	11
GARCH(1,1)-skew.Student-t	0.031451	10	0.011961	10
GARCH(1,1)-skew.GED	0.030622	8	0.011771	7
EGARCH(1,1)-norm	<i>0.027768</i>	1	0.011337	3
EGARCH(1,1)-skew.Student-t	0.028404	3	0.011332	2
EGARCH(1,1)-skew.GED	0.027931	2	<i>0.011037</i>	1
tGARCH(1,1)-norm	0.029547	6	0.011853	8
tGARCH(1,1)-skew.Student-t	0.029315	5	0.011646	6
tGARCH(1,1)-skew.GED	0.028580	4	0.011356	4
GJR-GARCH(1,1)-norm	0.031694	12	0.012084	13
GJR-GARCH(1,1)-skew.Student-t	0.032209	13	0.012072	11
GJR-GARCH(1,1)-skew.GED	0.031651	11	0.011935	9

*MSE is multiplied by 100

Figure 11: Histogram of the MSE sorted by rank

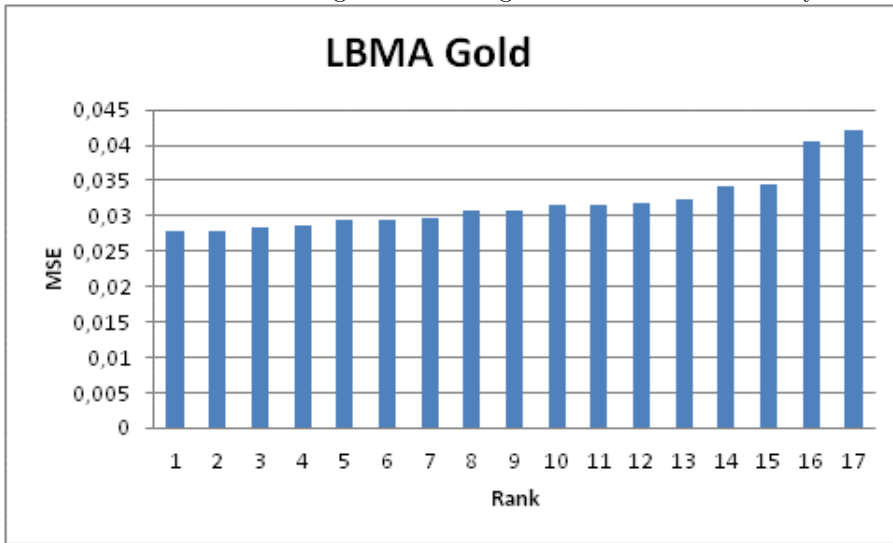


Figure 12: EGARCH(1,1)-norm forecasted std.dev vs proxy realized std.dev for LBMA Gold

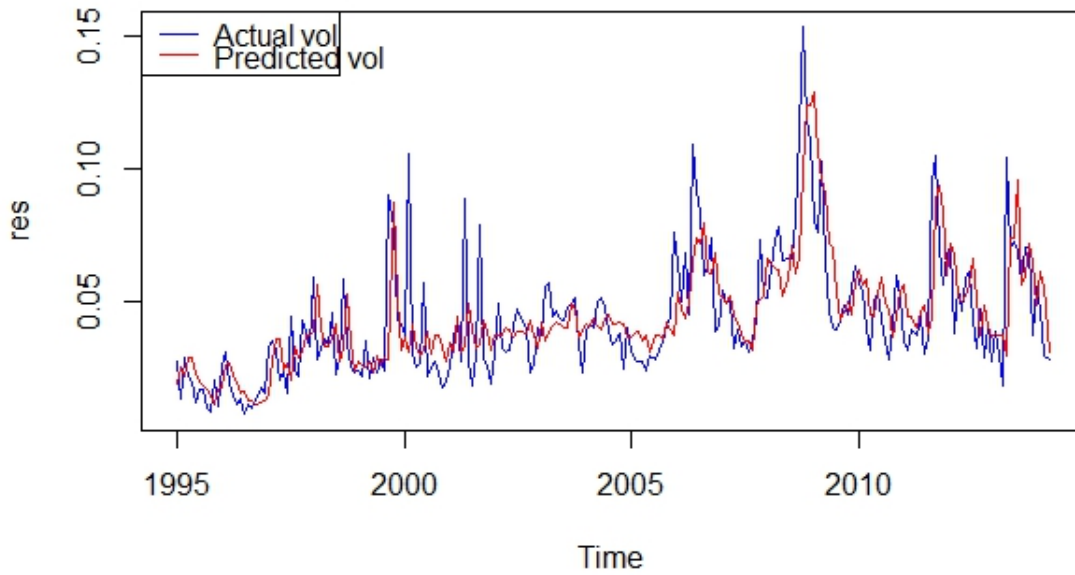


Figure 12 shows that the model have a fairly close fit in the period 1995-2000 but there are several short bursts in volatility in the early 2000 that leaves large residuals. Also for gold returns the highest volatility during the out of sample period comes at the start of the financial crisis 2007-2008 with a standard deviation reaching 0,15. From 2010 until the end of the period the model does a relatively good job predicting the volatility. The ranking of models in terms of the loss functions are fairly even despite the set of lager residuals which MSE is sensitive for.

8 Discussion

The result is consistent with the previous research presented by Poon and Granger (2003)[4] that generally the GARCH models outperform the ARMA model and the asymmetric GARCH models outperform the symmetric GARCH models. Also the ambiguous results when it comes to finding a superior model shown in previous research is reflected as there are tree different models that provide the most accurate forecast among the six time series. When it comes to the distributions used the skewed Student-t improved forecast accuracy of the preferred model in tree of the series but the gain where very small. However when looking at all the models where several distributions was used, and not only the best models both the skewed Student-t and the GED improved the forecast accuracy for the majority of the models, but the gain is still very small in most cases.

The RW model were only able to beat a few of the GARCH-family models for the USD/GBP and the S&P500 series. In the case of the USD/GBP series the GARCH-family models were unstable and produced significantly worse predictions than they did for the other series. The RW model did however outperform the 6 and 12 month SMA model as well as the ARMA(1,1) model consistently. The 6 month SMA gave better prediction than the 12 month SMA for all the series implying that increasing the period used to predict the volatility decrease the forecast accuracy for this model. The EWMA model were the best of the historical volatility models it was preferred for the USD/GBP series and only slightly worse than the best asymmetric GARCH-models that were preferred for the other series. The difference in complexity between an EWMA model and the asymmetric GARCH models combined with the small gain they give in terms of forecast accuracy makes it questionable whether it is worth it to implement the much more complex models for the marginal gain in forecasting accuracy.

As I have studied the subject of volatility forecasting and worked with this thesis there are several

questions that became apparent, which I do not address but could be explore in future research. It would be interesting to see how the preferred models in this thesis hold up when compared to implied volatility which is assumed to contain not only historic information, but also the markets expectation of future volatility. Expanding on the amount of models used might lead to finding models that provide better forecasting accuracy. There are several GARCH-family models both symmetric and asymmetric which have not been applied to the data like IGARCH and asymmetric power GARCH just to mention a few. I am also curious about how the models would preform when extending the forecasting horizon from one month to for instance half a year or a year. One interesting observation is that the GARCH-family models used became unstable when the skewed Student-t and the skewed GED distribution were used for the USD/GBP series. A closer diagnostic of this would be interesting as there is no difference in the procedure applied to this series than used on the other series, which did not suffer from this problem.

9 Conclusion

Most previous research show that data from financial time series have excess kurtosis and skewness. This have been accounted for by using skewed Student-t and skewed GED, in addition to the normal distribution. Student-t and GED allow for more observations in the tails. In-sample test suggests on a 1% level that all the series are stationary, and that there are volatility clustering. The test for asymmetry show ambiguous results with NASDAQ Composite and USD/GBP having significant in-sample asymmetry on a 1% level while S&P500 and DJIA have significant asymmetry on a 10% level. CBOE Interest Rate and LBMA Gold did not show any significant in-sample asymmetry. The volatility is tend to be persistent, meaning that the lags of previous volatility tend to die away slowly. For S&P500 and DJIA the volatility persistent in the in-sample data is slightly above 1, which according to Verbeek (2008)[18] gives a non-stationary process which is not uncommon in empirical work but might result in getting a sample outside the accepted region when maximizing. The models used in this thesis can be grouped into historical volatility models containing: RW, SMA and EWMA and regression based models containing: ARMA(1,1), GARCH(1,1), EGARCH(1,1) GJR-GARCH(1,1) and TGARCH(1,1). To evaluate the models with regards to forecasting performance I have used two loss functions MSE and MAE. The preferred models have been selected on the basis of the MSE loss function that is robust against noise while the MAE that is robust against outliers is produced to show that the model selection is sensitive to choice of loss functions. It is important that a good proxy for

realized volatility is calculated. This have been done using the daily data in each month which is easy to attain compared to intraday data that might give a even more accurate proxy.

Poon and Granger (2003)[4] summarized the findings of 93 papers on volatility forecasting showed that historical volatility models beat the regression based models in about 50% of the cases. In this case only the USD/GBP series preferred a historical volatility model namely the EWMA model over the GARCH-family models. However in all the series the ARMA(1,1) got outperformed by one or more historical volatility models. For the stock index series S&P500, NASDAQ Composite and DJIA it was the GJR-GARCH(1,1) that gave the most accurate forecast. For S&P500 and DJIA the GJR-GARCH(1,1) with the skewed Student-t distribution did best while for NASDAQ Composite the normal distribution preformed best despite the Jarque-Bera test rejected normal distributed returns on a 1% level. When it comes to the CBOE Interest Rate and LBMA Gold series EGARCH(1,1) was the most accurate model. The normal distribution was the most accurate for LBMA Gold while the skewed Student-t provided the best forecast for the CBOE Interest Rate series. With the exception of the USD/GBP series where there were stability problems when producing forecast with regression based models the GARCH-family models out-preformed the historical volatility models. The EWMA model was the only model that could compete with the GARCH-family models. The results when looked at in the histograms show that there is very small differences in predictive power between the top 3-5 models depending on the series. There is not much lost by choosing a less complex model for instance in the NASDAQ Composite series the EWMA model got rank 12 and is still very close to the best model in terms of MSE.

Acknowledgments

I would like to express my very great appreciation to Professor Valeriy I Zakamulin for excellent guidance accumulating in my ability to write this thesis. I would also like to tank my fellow finance students for providing a grate work environment and support during the master program. A special tanks to Vegard Hernæs, Anders Braathen, Øystein Dulsrud and Jacob Ask for countless hours productive group work through the studies.

References

- [1] S. Figlewski, "Forecasting Volatility." *Financial Markets, Institutions, and Instruments*. Boston MA: Blackwell Publishers, 1997, vol. 6, no. 1, (monograph).
- [2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
- [3] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, pp. 77–91, 1952.
- [4] S.-H. Poon and C. Granger, "Forecasting volatility in financial markets: A review," *Journal of Economic Literature*, vol. 41, no. 2, pp. 478–539, 2003.
- [5] S.-H. Poon, *A Practical Guide to Forecasting Financial Market Volatility*. Wiley, 2005.
- [6] G. Box and G. Jenkins, *Time series analysis: Forecasting and control*, 2nd ed. Holden-Day, 1976.
- [7] R. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom in ation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [8] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [9] A. Christie, "The stochastic behaviour of common stock variances: Value, leverage, and interest rate effect," *Journal of Financial Economics*, vol. 10, pp. 407–432, 1982.
- [10] D. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.
- [11] Z. J. M, "Threshold heteroskedastic models," 1991, institut National de la Statistique et des Etudes Economiques, Paris.
- [12] R. J. L.R. Glosten and D. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, vol. 48, pp. 1779–1801, 1993.
- [13] C. Brooks, *Introductory Econometrics for Finance*, 2nd ed. Cambridge University Press, 2008.
- [14] K. Dowd, *Measuring market risk*, 2nd ed. Wiley, 2005.
- [15] A. Wilhelmsson, "Garh forecasting performance under different distribution assumptions," *Journal of Forecasting*, vol. 25, pp. 561–578, 2006.

- [16] A. Ghalanos, *rugarch: Univariate GARCH models.*, 2013, r package version 1.2-7.
- [17] F. J. Fabozzi, Ed., *Handbook of Finance.* Wiley, 2008.
- [18] M. Verbeek, *A guide to modern econometrics.* John Willy and Sons, 2008.
- [19] W. A. Dickey, D. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *Journal of the American Statistical Association*, vol. 74, pp. 427–431, 1979.
- [20] R. Engle and V. Ng, “Measuring and testing the impact of news on volatility,” *Journal of Finance*, vol. 48, no. 5, pp. 1749–1778, 1993.
- [21] C. Jarque and A. Bera, “A test for normality of observations and regression residuals,” *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.
- [22] A. J. Patton, “Volatility forecast comparison using imperfect volatility proxies,” *Journal of Econometrics*, vol. 160, no. 1, pp. 246–256, 2011.
- [23] P. Huber, *Robust Statistics*, 6th ed. Wiley, 1981.
- [24] A. Trapletti and K. Hornik, *tseries: Time Series Analysis and Computational Finance*, 2013, r package version 0.10-32. [Online]. Available: <http://CRAN.R-project.org/package=tseries>
- [25] J. Ryan and J. Ulrich, *xts: eXtensible Time Series*, 2013, r package version 0.9-7. [Online]. Available: <http://CRAN.R-project.org/package=xts>
- [26] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2012, r package version 1.6-1. [Online]. Available: <http://CRAN.R-project.org/package=e1071>

Appendix

Credit to the creators of the packages used `tseries` by Trapletti and Hornik (2013) [24], `rugarch` by Ghalanos (2013) [16], `xts` by Ryan and Ulrich (2013) [25], `e1071` by Meyer et al. (2012) [26], and a special tanks to Valeriy I Zakamulin for providing some of his personal functions.

```
1 rm(list=ls(all=TRUE))
2 library(tseries)
3 library(rugarch)
4 library(xts)
5 library(e1071)
6 source("Functions.R")
7
8 # Read the price data from a file
9
10 data <- read.table("LBMA_Gold_Price.csv", header=TRUE, sep=",")
11 dates <- as.Date(data$Date, "%Y-%m-%d")
12 dataxts <- as.xts(data$Close, order.by=dates)
13
14 # DEFINE THE PARAMETERS
15
16 # define the starting point for the first portfolio return
17 year.start <- 1995
18 month.start <- 1
19
20 # The length of the rolling window, in the number of months
21 # Set to 1,6 and 12 when computing SMA
22
23 nLookback <- 5*12
24
25 # Compute returns
26
```

```

27 Close <- coredata(dataxts)
28 n <- length(Close)
29 rt <- diff(Close)/Close[1:(n-1)]
30 rt <- rt[-1]
31
32 #Daily standard deviations
33
34 sigmaD <- abs(rt)
35
36 # Descriptive statistics
37
38 summary(rt)
39 sd(rt)
40 skewness(rt)
41 kurtosis(rt)
42 jarque.bera.test(rt)
43
44 # In-sample tests
45
46 #Number of in-sample observations for the series
47 #1246 NASDAQ 1824 USD/GBP
48 #1254 LBMA Gold 1264 S&P500
49 #1250 CBOE Interest Rate 1264 DJIA
50
51 sigmaDis <- sigmaD[1:1254]
52 adf.test(sigmaDis)
53
54 arimaspec <- arfimaspec(mean.model = list(armaOrder = c(1, 1)))
55 show(arimaspec)
56
57 arimaF <- arfimafit(arimaspec, rt, out.sample = 4810)

```

```

58 arimaF # object containing Engle's ARCH LM test
59
60 Gspec <- ugarchspec(variance.model = list(model = "sGARCH",
61     garchOrder = c(1, 1), submodel = NULL)
62     ,mean.model = list(armaOrder = c(0,0), include.mean = T),
63     distribution.model = "norm")
64 Gspec
65
66 garchF <- ugarchfit(Gspec, rt, out.sample = 4810)
67 garchF # object containing Sign bias test
68
69
70 # EVALUATION OF THE OUT-OF-SAMPLE FORECAST
71
72 # Find the start and end date
73
74 start <- as.yearmon(paste(year.start, month.start), "%Y_%m")
75 end <- as.yearmon(dates[nobs])
76
77 # compute the number of months
78
79 nMonths <- numMonthsBetween(start, end)
80
81 # reserve the place to hold predicted and actual standard deviations
82
83 std.predict <- rep(0, nMonths)
84 std.actual <- rep(0, nMonths)
85
86 year <- year.start
87 month <- month.start
88

```

```

89 #RW model
90
91
92 for (i in 1:nMonths) {
93
94 # select returns from the start to the end of IS period
95 ret.past <- selectRange(ret, dates, year, month, nLookback)
96
97 # prepare next iteration
98
99 lst <- futYearMon(year, month, 1)
100 year <- lst$year; month <- lst$month
101
102 # select returns for the next month
103
104 ret.future <- selectRange(ret, dates, year, month, 1)
105
106 # compute the actual std and convert to monthly values
107
108 std <- sd(ret.future)
109 nDays <- length(ret.future)
110 std.actual[i] <- std*sqrt(nDays)
111 }
112
113 # A lag of k=1 applied to the actual std to get the naive prediction
114 # Done by removing the first month in actual and last month in the predicted
115
116 std.predict <- std.actual
117 std.predict <- std.predict[-231]
118 std.actual <- std.actual[-1]
119

```

```

120 #SMA model
121
122 for (i in 1:nMonths) {
123
124 # select returns from the start to the end of IS period
125
126 ret.past <- selectRange(ret, dates, year, month, nLookback)
127
128 # prepare next iteration
129
130 lst <- futYearMon(year, month, 1)
131 year <- lst$year; month <- lst$month
132
133 # select returns for the next month
134
135 ret.future <- selectRange(ret, dates, year, month, 1)
136
137 # compute the actual std and convert to monthly values
138
139 std <- sd(ret.future)
140 nDays <- length(ret.future)
141 std.actual[i] <- std*sqrt(nDays)
142
143 # compute the forecasted volatility and converting to monthly volatility
144
145 n.days.past <- length(ret.past)
146 std.predict[i] <- sqrt((sum(ret.past^2)/n.days.past))*sqrt(nDays)
147 }
148
149
150 #EWMA model

```

```

151
152 for (i in 1:nMonths) {
153
154 # select returns from the start to the end of IS period
155
156 ret.past <- selectRange(ret, dates, year, month, nLookback)
157
158 # prepare next iteration
159
160 lst <- futYearMon(year, month, 1)
161 year <- lst$year
162 month <- lst$month
163
164 # select returns for the next month
165
166 ret.future <- selectRange(ret, dates, year, month, 1)
167
168 # compute the actual std and convert to monthly values
169
170 std <- sd(ret.future)
171 nDays <- length(ret.future)
172 std.actual[i] <- std*sqrt(nDays)
173 k <- length(ret.past)
174 var.ewma <- covEWMA(as.data.frame(ret.past), lambda=0.97)
175 var <- var.ewma[k]
176 std.predict[i] <- sqrt(var*nDays)
177 }
178
179
180 # ARMA model
181

```

```

182 # Define ARIMA spec
183
184 arimaspec <- arfimaspec(mean.model = list(armaOrder = c(1, 1)))
185
186 for (i in 1:nMonths) {
187
188 # select returns from the start to the end of IS period
189
190 sig.past <- selectRange(sigmaD, dates, year, month, nLookback)
191
192 # prepare next iteration
193
194 lst <- futYearMon(year, month, 1)
195 year <- lst$year; month <- lst$month
196
197 # select returns for the next month
198
199 ret.future <- selectRange(ret, dates, year, month, 1)
200
201 # compute the actual std and convert to monthly values
202
203 std <- sd(ret.future)
204 nDays <- length(ret.future)
205 std.actual[i] <- std*sqrt(nDays)
206
207 # fit the ARIMA model
208
209 fit.arima = arfimafit(spec = arimaspec, data = sig.past, solver = 'hybrid')
210
211 # forecast for the next month, day by day
212 for.arima <- arfimaforecast(fit.arima, n.ahead = nDays)

```

```

213
214 # monthly variance is the sum of squared daily standard deviations
215
216 sigmapred <- fitted(for.arima)
217 std.predict[i] <- sqrt(sum(sigmapred^2))
218 }
219
220
221 #ARCH-family models
222
223 # Define the GARCH model
224
225 uspec <- ugarchspec(variance.model = list(model = "eGARCH",
226     garchOrder = c(1, 1), submodel = NULL)
227     ,mean.model = list(armaOrder = c(0, 0),
228     include.mean = FALSE), distribution.model = "norm")
229
230 for (i in 1:nMonths) {
231
232 # select returns from the start to the end of IS period
233
234 ret.past <- selectRange(ret, dates, year, month, nLookback)
235
236 # prepare next iteration
237
238 lst <- futYearMon(year, month, 1)
239 year <- lst$year; month <- lst$month
240
241 # select returns for the next month
242
243 ret.future <- selectRange(ret, dates, year, month, 1)

```



```

244
245 # compute the actual std and convert to monthly values
246
247 std <- sd(ret.future)
248 nDays <- length(ret.future)
249 std.actual[i] <- std*sqrt(nDays)
250
251 # fit the GARCH model
252
253 fit.garch = ugarchfit(spec = uspec, data = ret.past, solver = 'hybrid')
254
255 # forecast for the next month, day by day
256
257 for.garch <- ugarchforecast(fit.garch, n.ahead = nDays)
258
259 # monthly variance is the sum of squared daily standard deviations
260
261 std.predict[i] <- sqrt(sum(sigma(for.garch)^2)) }
262
263
264 # computation of the
265 #Mean Squared Forecasting Error
266 #Mean absolute forecasting Error
267
268 MSFE <- mean((residuals)^2)
269
270 MSFE MAFE <- mean(abs(residuals)) MAFE

```

- S&P500 <http://finance.yahoo.com/q/hp?s=%5EGSPC&a=00&b=1&c=1995&d=01&e=7&f=2014&g=d>
- NASDAQ <http://finance.yahoo.com/q/hp?s=%5EIXIC&a=00&b=1&c=1995&d=01&e=7&f=2014&g=d>
- DJIA <http://finance.yahoo.com/q/hp?s=%5EDJI&a=00&b=1&c=1995&d=01&e=7&f=2014&g=d>
- CBOE Interest Rate <http://finance.yahoo.com/q/hp?s=%5ETNX&a=00&b=1&c=1995&d=01&e=7&f=2014&g=d>
- USD/GBP <http://www.oanda.com/currency/historical-rates/>
- LBMA Gold http://www.quandl.com/OFDP/GOLD_2-LBMA-Gold-Price-London-Fixings-P-M